



**I
N
A
O
E**

Aprendizaje de Representaciones de Datos de EEG sobre Múltiples Sujetos usando Modelos basados en Aprendizaje Profundo

Magdiel Jiménez Guarneros, María del Pilar Gómez Gil

Reporte técnico No. CCC-17-002
15 de Junio de 2017

© **Coordinación de Ciencias Computacionales
INAOE**

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.



Aprendizaje de Representaciones de Datos de EEG sobre Múltiples Sujetos usando Modelos basados en Aprendizaje Profundo

Magdiel Jiménez Guarneros, María del Pilar Gómez Gil

Coordinación de Ciencias Computacionales, Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro #1, Santa María Tonantzintla, Puebla, México, CP. 72840.
{*magdiel.jg, pgomez*}@*inaoep.mx*

Resumen

Determinar los niveles de carga cognitiva a través de las señales de EEG es una tarea que puede ser vista como un problema de clasificación, donde la correcta diferenciación entre los niveles de carga está fuertemente sujeta a una apropiada extracción de características, tanto temporales como espectrales. En este sentido, el aprendizaje de representaciones es la línea de investigación dedicada al estudio y el desarrollo de modelos capaces de extraer, de manera automática, características más útiles para la clasificación. Entre los diferentes paradigmas existentes, el aprendizaje profundo muestra aprender estructuras jerárquicas y permite obtener representaciones más abstractas, mejorando los resultados de clasificación. Recientemente, diversos trabajos basados en aprendizaje profundo han dedicado esfuerzos para aprender representaciones de datos de EEG. Sin embargo, un inconveniente de la mayor parte de los trabajos reportados, consiste en extraer características para el entrenamiento de clasificadores en sujetos específicos. Este enfoque implica, además de un consumo de tiempo, un costo en los recursos del etiquetado de los datos y el entrenamiento de nuevos modelos para cada nuevo sujeto. Para abordar esta problemática, esta propuesta de investigación plantea desarrollar soluciones para aprender representaciones de características de señales de EEG sobre múltiples sujetos. Las características aprendidas serán utilizadas para construir modelos de clasificación y de adaptación de dominio, a fin de categorizar actividades cerebrales de sujetos desconocidos en el estudio de la carga cognitiva. Como resultado preliminar, se propone una primera arquitectura de red profunda para el aprendizaje de representaciones en espacio, tiempo y frecuencia de las señales de EEG.

Palabras clave - Aprendizaje de representaciones, Aprendizaje profundo, Modelos generativos, Modelos discriminativos, Aprendizaje basado en transferencia, Clasificación, EEG, Carga cognitiva.

Contenido

1.	Introducción	1
2.	Conceptos básicos	3
2.1.	Electroencefalograma	3
2.2.	Carga cognitiva.....	4
2.3.	Medición de la carga cognitiva usando EEG	5
2.3.1.	Procesamiento de señales en tiempo-frecuencia	5
2.4.	Aprendizaje de representaciones	6
2.4.1.	Aprendizaje profundo	6
2.5.	Medidas de evaluación	12
3.	Trabajo relacionado.....	14
3.1.	Aprendizaje de representaciones de datos de EEG	14
3.1.1.	Aprendizaje profundo	14
4.	Propuesta de investigación	17
4.1.	Problema a resolver	17
4.2.	Motivación y justificación.....	18
4.3.	Preguntas de investigación	19
4.4.	Hipótesis.....	20
4.5.	Objetivo general	20
4.6.	Objetivos específicos.....	20
4.7.	Contribuciones.....	21
5.	Metodología	21
5.1.	Cronograma de actividades	29
5.2.	Alcances y limitaciones.....	30
5.3.	Plan de publicaciones	30
6.	Resultados preliminares	30
6.1.	Arquitectura de red neuronal profunda para el aprendizaje de características en espacio, tiempo y frecuencia, de señales de EEG	30
6.1.1.	Descripción del conjunto de datos de carga cognitiva	31
6.1.2.	Pre-procesamiento: extracción de imágenes de datos de EEG.....	31
6.1.3.	CGRUD: arquitectura de red profunda propuesta.....	32
6.1.4.	Validación de CGRUD: métodos de comparación	34
6.1.5.	Diseño del experimento	35

6.1.6. Resultados	35
6.1.7. Discusión.....	37
7. Conclusiones	38
8. Referencias.....	38

Lista de Tablas

Tabla 2-1. Bandas de frecuencia y regiones de activación asociadas con los estados mentales [17].	5
Tabla 2-2. Matriz de confusión para una clasificación de dos clases.....	13
Tabla 5-1. Cronograma de actividades de las tareas a realizar por bimestres. Las celdas coloreadas en azul indican que la tarea ha sido realizada, mientras que las celdas en color gris indican las actividades por realizar.	29
Tabla 6-1. Conjunto de datos de carga cognitiva. La tabla muestra el total de ensayos por sujeto y niveles de carga cognitiva (clases) inducidos.....	31
Tabla 6-2. Resultados de clasificación ($\mu \pm \sigma$) en exactitud de los métodos comparados. Los resultados más altos por columna se indican en negrita.....	36
Tabla 6-3. Resultados de clasificación ($\mu \pm \sigma$) en <i>F-measure</i> de los métodos comparados. Los resultados más altos por columna se indican en negrita.....	36
Tabla 6-4. Resultados de clasificación ($\mu \pm \sigma$) en AUC de los métodos comparados. Los resultados más altos por columna se indican en negrita.....	37

Lista de Figuras

Figura 2.1. Sistema internacional 10-20 para la colocación de electrodos sobre el cuero cabelludo [46].	4
Figura 2.2. Diagrama de bloque de aprendizaje residual [55]......	8
Figura 2.3 Diferentes configuraciones de aprendizaje de transferencia [43].	12
Figura 3.1 Modelo propuesto en [32] para el aprendizaje de características espaciales, temporales y espectrales, de señales de EGG sobre múltiples sujetos.	16
Figura 5.1. Diagrama de flujo de la metodología propuesta.	22
Figura 5.2. Modelo generativo esperado para aprender características robustas en espacio, tiempo y frecuencia.	23
Figura 6.1. Arquitectura CGRUD.	33

Lista de Acrónimos

EEG	Electroencefalograma
DL	<i>Deep Learning</i> (Aprendizaje profundo)
FT	<i>Fourier Transform</i> (Transformada de Fourier)
STFT	<i>Short-Time Fourier Transform</i> (Transformada de Fourier de tiempo corto)
CNN	<i>Convolutional Neural Network</i> (Red Neuronal de Convolución)
RNN	<i>Recurrent Neural Network</i> (Red Neuronal Recurrente)
GRU	<i>Gated Recurrent Unit</i> (Unidad Recurrente con compuerta)
AE	<i>Autoencoder</i> (Autocodificador)
RBM	<i>Restricted Boltzman Machines</i> (Máquinas de Boltzman Restringidas)
DBN	<i>Deep Belief Network</i> (Redes de creencias profundas)
GAN	<i>Generative Adversarial Networks</i> (Redes Adversarias Generativas)
VAE	<i>Variational Autoencoders</i> (Autocodificadores variacionales)
PCA	<i>Principal Component Analysis</i> (Análisis de Componentes principales)

1. Introducción

En la actualidad diferentes técnicas en neurociencia han permitido el estudio no invasivo del funcionamiento cerebral, tales como las imágenes de resonancia magnética (IRM) o las mediciones obtenidas de la actividad eléctrica del cerebro a través de Electroencefalogramas (EEG) [1]. Esta última técnica se caracteriza por poseer una alta resolución temporal, por lo que su uso es de gran relevancia para el estudio de la dinámica de los distintos procesos neuronales. Algunos de estos estudios incluyen a aquellos relacionados con la epilepsia [2, 3, 4], los estados afectivos [5, 6], los estados de sueño y alerta [7, 8, 9], la carga cognitiva [10] y el diseño de interfaces cerebro computadora [11, 12].

Recientemente, la supervisión de los estados mentales, tal como el de la carga cognitiva, ha recibido mucha atención en diversos campos [13, 14, 15, 16]. La carga cognitiva es considerada como la carga de los recursos cognitivos de la mente, generada por la demanda cognitiva en una tarea dada [17, 18]. El ser humano es propenso a cometer errores y sufre una pérdida en su eficiencia cuando la tarea cognitiva impuesta es muy alta [17]. Por estos motivos, la necesidad de medir la carga cognitiva de manera objetiva toma mayor relevancia, principalmente en dominios donde se realizan tareas críticas, tal como en conducción de vehículos [13], o no tan críticas como en el contexto de psicología educativa [18]. Conocer el nivel de carga permitiría diseñar estrategias que mejoren la productividad de un individuo para una tarea asignada [17, 15].

El estudio de las señales de EEG para medir el nivel de carga cognitiva puede ser visto como un problema de clasificación, donde el objetivo consiste en discriminar entre los diferentes niveles de carga cognitiva. Para la clasificación de estos niveles, la extracción de características es una parte importante, por lo que es esencial desarrollar modelos que consideren la alta correlación espacial y temporal, además de la alta dimensionalidad y ruido inherente a las señales de EEG.

Diferentes investigaciones se han reportado para capturar las diferentes respuestas cerebrales en los diferentes estudios de carga cognitiva inducidos, para observar los cambios en los potenciales eléctricos durante el proceso cognitivo [10, 15, 19, 20]. Investigaciones sugieren que los cambios de potencia específicos, en ciertas bandas del espectro de frecuencias, muestran ser características robustas para la clasificación, pero dependen de la ubicación cortical, el tiempo y la naturaleza de la tarea cognitiva [10, 21]. Por lo tanto, diferentes características temporales y espectrales son a menudo empleadas como características representativas del EEG, las cuales varían con respecto a la carga cognitiva.

Actualmente, los métodos empleados para describir la función cerebral en el estudio de la carga cognitiva, presentan una precisión limitada debido a las propiedades discriminativas de las características extraídas [8]. Este proceso de extracción es una labor intensiva e implica un consumo de tiempo, pues es usual que sea utilizado el conocimiento del dominio de los datos de un experto, para definir manualmente las características a utilizarse. Además, cuando los tipos de datos o tareas de clasificación cambian, el proceso necesita ser repetido [22, 23]. En este sentido, el aprendizaje de representaciones es la línea de investigación que se dedica al estudio y desarrollo de métodos capaces de extraer, de manera automática, un conjunto de características útiles para la clasificación, a fin de representar estructuras jerárquicas ricas en espacio y tiempo de señales de EEG, con una menor intervención humana [24].

Diversos trabajos de aprendizaje de representaciones han sido propuestos para el análisis de señales de EEG bajo diferentes paradigmas, tales como aprendizaje de variedades [25], aprendizaje de diccionarios [3] o técnicas probabilistas [26]. Por otra parte, un paradigma de gran interés en la actualidad que forma parte del aprendizaje de representaciones es el aprendizaje profundo (DL por sus siglas del inglés *Deep Learning*). DL forma una familia de métodos de aprendizaje que ha mostrado aprender mejores representaciones de características que otros métodos, al usar una

arquitectura jerárquica organizada en múltiples niveles y aprender características más abstractas en niveles superiores [27, 28, 29].

Los trabajos reportados en la literatura, basados en aprendizaje profundo para datos de EEG, se pueden agrupar principalmente en dos categorías:

Modelos discriminativos: Este tipo de modelos determinan la probabilidad condicional $p(\mathbf{y}|\mathbf{x})$ de las clases \mathbf{y} dado un conjunto de ejemplos de entrada \mathbf{x} [30, 31]. Los trabajos basados en modelos discriminativos utilizan un aprendizaje supervisado, y están basados en redes de convolución [32, 33], recurrentes [34], y autocodificadores apilados [35, 36, 37].

Modelos generativos: son modelos probabilistas que intentan determinar la probabilidad conjunta de los datos de entrada, $p(\mathbf{x})$ [24]. Este tipo de modelos están basados en aprendizaje no supervisado y reciben un ejemplo de entrada x para codificar una representación latente h . Esta representación h preserva los diferentes factores de variación de los datos y se utiliza para reconstruir a x en la capa de salida [38]. Una ventaja de modelos generativos consiste en compartir fortalezas estadísticas que son de utilidad en el desarrollo de modelos semi-supervisados, multi-tareas o de adaptación de dominio [24]. Los trabajos actualmente reportados para el análisis de EEG están basados en máquinas restringidas de *Boltzman* y redes de creencias profundas [9, 6, 39].

Un inconveniente de la mayoría de los modelos reportados es su énfasis en el aprendizaje de representaciones para entrenar clasificadores en sujetos específicos. En la práctica, este enfoque implica, además de un consumo de tiempo, un costo en los recursos del etiquetado de los datos y el entrenamiento de nuevos modelos para cada nuevo sujeto [40]. Algunos estudios incluso sugieren que podría ser necesario construir nuevos clasificadores cada día para un mismo sujeto [41]. Para abordar esta problemática, una solución podría ser extraer representaciones de características sobre múltiples sujetos para construir un modelo de clasificación, tal que permita clasificar a sujetos desconocidos; sin embargo, este proceso de extracción representa un reto pues las distribuciones de probabilidad de los datos de EEG varían a través de los diferentes sujetos [40, 42].

En la literatura, algunos esfuerzos han sido reportados para aprender una representación de características sobre múltiples sujetos [32, 8]. El trabajo reportado en [32], basado en un modelo discriminativo, permite aprender una representación en espacio, tiempo y frecuencia usando una red convolucional-recurrente; mientras que en [8], basado en un modelo generativo, utiliza una modificación de una red de creencias profundas para extraer una representación en el dominio temporal. En general en el procesamiento de señales de EEG, la ventaja de modelos discriminativos consiste en la flexibilidad de procesar los datos, conservando su estructura inherente en espacio, tiempo y frecuencia. Sin embargo, este tipo de modelos no permiten capturar los diferentes factores de variación, lo cual podría ayudar a modelar las diferencias existentes entre individuos; además, requieren una gran cantidad de datos etiquetados para mostrar su mejor rendimiento [43]. Por otra parte, los modelos generativos tienen la ventaja de capturar los diferentes factores de variación y modelar sus relaciones [24]; ciertamente, el aprendizaje de estos modelos se realiza sobre datos no etiquetados, e incluso existen estudios sobre la capacidad de aprendizaje de datos con una diferente distribución de probabilidad [44, 45]. No obstante, los trabajos de modelos generativos actuales sólo permiten procesar los datos en forma de vectores de características, omitiendo la estructura inherente en espacio, tiempo y frecuencia. Cabe recordar que para discriminar entre los distintos niveles de carga cognitiva, se han utilizado características basadas en los cambios en el potencial eléctrico de determinadas bandas de frecuencia [10]. Estos cambios dependen de la ubicación cortical y de la información temporal de las señales de EEG. Por lo tanto, para abordar la variabilidad entre las señales de distintos individuos, un análisis espacial, temporal y espectral, resulta necesario.

Por los motivos anteriores, esta investigación está enfocada en proponer un marco de trabajo, tal que permita proveer de mejores modelos de aprendizaje de representaciones sobre múltiples sujetos que los trabajos actuales. El marco de trabajo estará basado en modelos generativos, con los cuales se pretende modelar la variabilidad existente en los datos de EEG ante las diferencias individuales, integrando componentes de modelos discriminativos que permiten conservar la estructura inherente en espacio, tiempo y frecuencia. Para incrementar el desempeño de aprendizaje de nuestros modelos propuestos, se integrará un modelo de aprendizaje basado en transferencia, incorporando conocimiento de otros dominios de EEG con una diferente distribución de probabilidad. Las características aprendidas de los modelos propuestos serán utilizadas para construir modelos de clasificación y de adaptación de dominio, con el propósito de clasificar actividades cerebrales de sujetos desconocidos en el estudio de la carga cognitiva.

Como resultado preliminar, se presenta una primera arquitectura basada en redes neuronales profundas para el aprendizaje de características de datos de EEG en espacio, tiempo y frecuencia. Esta arquitectura recibe como representación inicial una secuencia de imágenes de datos de EEG, la cual se obtiene a partir de dividir cada segmento de tiempo en una serie de ventanas; cada ventana forma una imagen al proyectar las coordenadas espaciales de los múltiples electrodos del EEG con su respectiva información de frecuencia. El propósito de la tarea consiste en determinar los cuatro niveles de carga cognitiva inducidos a un grupo de sujetos. Para la evaluación de la arquitectura propuesta se utilizó un enfoque de aprendizaje supervisado.

Este documento está organizado de la siguiente manera: en la sección 2 se introducen los conceptos básicos referentes a EEG y los diferentes modelos de aprendizaje de representaciones basados en aprendizaje profundo. En la sección 3, se analizan los trabajos más relevantes relacionados con esta investigación. En la sección 4 se presenta el problema, las preguntas de investigación, la hipótesis y los objetivos; la sección 5 describe la metodología y el diseño experimental para dar cumplimiento a los objetivos planteados. Finalmente, en las secciones 6 y 7 se muestran los resultados preliminares y las conclusiones.

2. Conceptos básicos

2.1. Electroencefalograma

Una señal de Electroencefalograma (EEG) es una medida de las corrientes eléctricas que fluyen durante las excitaciones sinápticas de las dendritas de muchas neuronas en la corteza cerebral. El EEG permite registrar los potenciales eléctricos producidos por las diferentes actividades cerebrales mediante C_T electrodos (o canales) colocados sobre la superficie del cuero cabelludo [46].

Los sistemas de posicionamiento de electrodos estándar permiten obtener registros más comparables dentro y entre sujetos [47]. El sistema de posicionamiento establecido por la Federación Internacional de Sociedades de Electroencefalografía y Neurofisiología Clínica es el “10-20”, mostrado en la Figura 2.1. La ventaja de este sistema de posicionamiento sobre otros es que las colocaciones son calculadas como porcentajes de distancias estándar entre puntos preestablecidos: los electrodos son ubicados a lo largo de la línea media en 10, 20, 20, 20, 20 y 10 %, del total de la distancia de los huesos de la nariz y el frontal (Nasión), a la parte más sobresaliente del occipital (Inión).

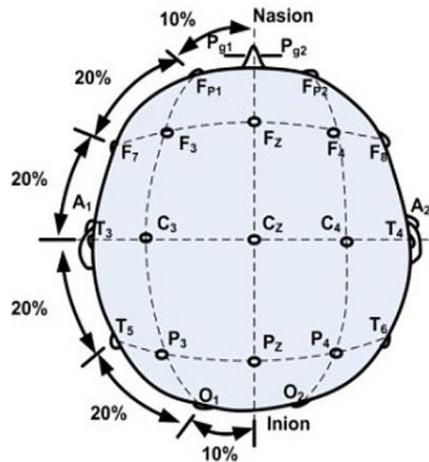


Figura 2.1. Sistema internacional 10-20 para la colocación de electrodos sobre el cuero cabelludo [47].

Las señales de EEG se caracterizan por ser no estacionarias, es decir, mantienen una estadística variable en el tiempo [47]. El EEG aproxima una señal continua $x_c(t)$, que denota la evolución temporal t de los voltajes de cada electrodo c , al medir dicha señal en puntos de tiempo discretos a lo largo de un intervalo Δt [1], esto es:

$$x_c[n] = x_c(n\Delta t), \quad n = 1, 2, 3 \dots \quad (2.1)$$

donde $x_c[n]$ es el registro discreto de la señal en el tiempo $x_c(t)$ para intervalos de tiempo $t = \Delta, 2\Delta t, \dots, n\Delta t$.

El EEG es sensible a ruido, provocado por la actividad que no se origina en el cerebro [1]. Este ruido, formado por elementos denominados “artefactos”, puede ser de tipo externo, tal como el ambiente, la interferencia electromagnética o la unión inapropiada de los electrodos; así como fisiológicos, tal como el parpadeo de los ojos, la masticación, el movimiento muscular o latidos del corazón. Además de que las señales de EEG poseen una señal a ruido baja, éstas tienen una naturaleza irregular y nunca se replican exactamente.

2.2. Carga cognitiva

La carga cognitiva es la cantidad total del esfuerzo mental utilizado en la memoria de trabajo de un individuo [48, 49]. Cuando se trata con información nueva, la memoria de trabajo funciona como un almacén temporal y permite procesar la información necesaria para ejecutar tareas cognitivas complejas [18]. La memoria de trabajo tiene una capacidad limitada, pues únicamente 7 ± 2 elementos de información pueden ser almacenados en esta memoria de trabajo; incluso este número disminuye cuando la información tiene que ser no sólo recordada sino también procesada [48]. Por lo tanto, una demanda cognitiva más allá de este límite resulta en una sobrecarga y en un decremento en el rendimiento de los sujetos para una tarea asignada.

En la actualidad existe un gran interés en el uso de EEG como herramienta de medición fisiológica de la carga cognitiva, debido a la necesidad de identificar de manera objetiva los recursos mentales utilizados para una tarea cognitiva en particular [17, 15]. La alta resolución temporal del EEG permite medir los cambios en la actividad eléctrica en la escala de milisegundos, lo que lleva a entender la demanda cognitiva que una tarea ha impuesto, en un paso dado, a los limitados recursos mentales de un sujeto.

2.3. Medición de la carga cognitiva usando EEG

La actividad neurológica en el cerebro produce un rango de ondas por segundo (frecuencias) que son característicamente diferentes para diversos niveles de actividad cerebral [47]. En adultos saludables por ejemplo, las amplitudes y frecuencias de tales señales cambian de un estado a otro, dependiendo si el sujeto se encuentra en estado de vigilia o de sueño [1]. Además, las características de las ondas cambian también con la edad.

La actividad rítmica del cerebro se ha dividido según las bandas de frecuencia que ocupa, entre otras se identifican las bandas: *delta* (0.1-3.9 Hz), *theta* (4.0-7.9 Hz), *alfa* (8.0-12.9 Hz), *beta* (13-29.9 Hz) y *gamma* (30-60 Hz) [47]. En la Tabla 2-1 se describen las actividades mentales asociadas y las regiones de activación con estas bandas de frecuencia.

Tabla 2-1. Bandas de frecuencia y regiones de activación asociadas con los estados mentales [17].

Banda de Frecuencia	Estado mental asociado
<i>Delta</i> (0.1 - 3.9 Hz)	En sueño profundo y en tareas de atención continua, existe un incremento en la potencia espectral de esta banda; sobresale en el lóbulo temporal.
<i>Theta</i> (4.0 - 7.9 Hz)	El poder espectral de esta banda se incrementa con el aumento de la carga mental; sobresale en el lóbulo frontal. La banda está relacionada con somnolencia, asignación de recursos y excitación.
<i>Alfa</i> (8.0 – 12.9 Hz)	El poder espectral de esta banda aumenta con el incremento de la carga mental; sobresale en el lóbulo temporal y occipital, además del lóbulo frontal durante tareas complejas. La banda está relacionada con el estado de relajación y reflectante.
<i>Beta</i> (13 Hz – 29.9 Hz)	La banda es utilizada ampliamente para la medición de carga de trabajo mental; sobresale en el lóbulo temporal y occipital. La banda indica un estado de alerta, excitación, frustración y trabajo.
<i>Gamma</i> (30 Hz – 60 Hz)	La banda está relacionada con el procesamiento de información sensorial; sobresale en la línea media central del cerebro.

Algunos estudios de carga cognitiva sugieren que los cambios en amplitud y frecuencia dependen de la ubicación cortical, el tiempo y la naturaleza de la tarea cognitiva dada [50]. Por este motivo, en la siguiente sección se describe uno de los métodos más utilizados para el procesamiento de señales en tiempo y frecuencia.

2.3.1. Procesamiento de señales en tiempo-frecuencia

La transformada de Fourier, *Fourier Transform* (FT), es un modelo que permite describir a una señal en términos de sus componentes de frecuencia. La FT de una señal discreta $x[n]$ para las muestras $n = 1, 2, \dots, N$ en una ventana arbitraria está dada por [51]:

$$FT[\omega] = \sum_{n=1}^N x[n]e^{-i\omega n}, \omega = \frac{2\pi m}{N}, m = 0, 1, \dots, N - 1 \quad (2.2)$$

donde N es el número de muestras de la ventana que se analizarán, $x[n]$ indica la muestra tomada en el instante n , y m es el índice de frecuencia cuyo valor se quiere obtener. La expresión exponencial representa el número complejo en coordenadas polares $re^{j\theta}$, donde r es el módulo y argumento; dado que, según la fórmula de Euler:

$$e^{-i\omega n} = \cos(\omega n) - i \sin(\omega n) \quad (2.3)$$

Las funciones de base sinusoidal $e^{-i\omega n}$ son capaces de aislar la actividad en diferentes frecuencias ω , medida en radianes. El valor de la FT en cada ω representa la contribución relativa de eventos que ocurren en esa frecuencia para $x[n]$. La FT es definida para $0 \leq \omega = \frac{2\pi m}{N} < 2\pi$ con $m = 0, 1, \dots, N - 1$.

La FT se utiliza para evaluar suficientes componentes de frecuencia, tal que permita reconstruir el segmento finito que se analiza. Sin embargo, cuando una señal es no estacionaria, la FT no puede obtener información del instante en que aparecen y desaparecen los componentes de frecuencia. Esta falta de localización temporal se debe a que las funciones base senos y cosenos son periódicas. Para solucionar este problema, una ventana en movimiento $g[n - a]$ es aplicada sobre la señal $x[n]$, variante conocida como *Short-Time Fourier Transform* (STFT) [52], la cual está definida como:

$$STFT[\omega, a] = \sum_{n=-\infty}^{\infty} x[n]g[n - a]e^{-i\omega n} \quad (2.4)$$

donde $g[n - a]$ corresponde a la ventana de análisis con un número de muestras de tamaño $n - a$. De esta manera, la STFT forma una representación en tiempo-frecuencia de la señal. La STFT tiene una resolución fija para la ventana $g[n - a]$; por lo que un tamaño angosto de ventana provee de una alta resolución temporal y una baja resolución en frecuencia. Por otra parte, un tamaño ancho de ventana provee una alta resolución en frecuencia pero no así temporal. Ambos casos son conocidos como banda amplia y angosta [47].

2.4. Aprendizaje de representaciones

El aprendizaje de representaciones está formado por un conjunto de métodos que permiten a un algoritmo de aprendizaje de máquina ser alimentado con datos crudos, y descubrir automáticamente las representaciones necesarias para la detección o la clasificación [27]. Cuando se diseñan características o algoritmos de aprendizaje de características, se busca separar los factores de variación que explican los datos observados [30, 24]. Bajo el contexto de aprendizaje de representaciones, la frase *separar factores de variación* hace referencia a separar las fuentes de influencia o las características relevantes en la entrada, las cuales pueden ser pensadas como conceptos o abstracciones, que ayudan a dar sentido a la variabilidad encontrada en los datos [43]. Por ejemplo, en el reconocimiento del habla, estos factores de variación incluyen edad, acento, sexo, entre otros.

2.4.1. Aprendizaje profundo

El aprendizaje profundo es un tipo de aprendizaje de representaciones que está formado por la composición de múltiples transformaciones no lineales, usualmente redes neuronales, con el objetivo de producir representaciones más abstractas y ultimadamente más útiles para la clasificación [24].

Un modelo matemático de una red neuronal aproxima una función f^* [53], el cual puede estar asociado a la composición de diversas funciones. Por ejemplo, si se tienen las funciones $f^{(1)}$, $f^{(2)}$ y $f^{(3)}$ conectadas en cadena, éstas forman a $f^*(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$. Así, cada función $f^{(i)}$ es denominada capa y el número total de capas define la profundidad del modelo [29, 43]. Esta terminología es utilizada en redes neuronales profundas y en general en aprendizaje profundo.

En las siguientes secciones se describen los conceptos relacionados con algunos de los diferentes enfoques basados en aprendizaje profundo.

2.4.1.1. Modelos discriminativos

Los modelos discriminativos, basados en aprendizaje supervisado, determinan la probabilidad condicional de las etiquetas \mathbf{y} dado un conjunto de ejemplos de entrada \mathbf{x} , $p(\mathbf{y}|\mathbf{x})$. Los modelos de este tipo, involucrados en la presente investigación, se definen brevemente a continuación.

2.4.1.1.1 Redes neuronales de convolución

La red neuronal de convolución (CNN de las siglas del inglés *Convolutional Neural Network*), es una variante de una red de perceptrones multicapa, diseñada para procesar datos que tienen una estructura topológica en forma de cuadrícula [54, 55]. La primera capa oculta de una CNN es equivalente a una operación de convolución, la cual produce un mapa de características al aplicar repetidamente una función o *kernel* (o filtros) a través de sub-regiones (campos receptivos) sobre un ejemplo de entrada x . Este mapa se obtiene al convolucionar x con un *kernel* lineal, añadiendo un sesgo y aplicando posteriormente una función de activación ϕ :

$$h_{ij}^k = \phi(W_{ij}^k * x_{ij} + b_k) \quad (2.5)$$

donde k denota el mapa de características de salida en una capa dada, cuyos filtros son determinados por los pesos W^k y el sesgo b_k . W_{ij}^k es la matriz de pesos que conecta a cada pixel del k -ésimo mapa de características en la capa m , con el pixel localizado en las coordenadas (i, j) del l -ésimo mapa de características en la capa $m - 1$. El operador de convolución $*$, aplicado en señales discretas de dos dimensiones se define como:

$$f[m, n] * g[m, n] = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} f[u, v]g[m - u, n - v] \quad (2.6)$$

Cada *kernel* es replicado a través de un campo receptivo, de manera que estas unidades replicadas comparten los mismos pesos y sesgos, para formar un mapa de características [54].

Una segunda capa, denominada *pooling*, consiste en una forma de sub-muestreo no lineal, lo que permite ser vista como un elemento de disminución de la resolución de los mapas de características. La capa de *pooling* reemplaza la salida de una capa de convolución con un resumen estadístico. Uno de los métodos más populares es *max-pooling*, el cual divide una entrada en un conjunto de rectángulos no superpuestos, tal que para cada sub-región analizada produce el máximo valor. Esta herramienta es útil debido a que reduce los cálculos para capas de niveles más altos y convierte a la representación invariante a translaciones de x . Finalmente, es usual que en los últimos niveles de una red de convolución se apilen capas multiconectadas para la discriminación de características aprendidas en capas de niveles inferiores.

2.4.1.1.1 Bloque de aprendizaje residual

El modelo de bloque residual presentado en [56], propone la idea de dividir y contribuir a la propagación de un gradiente en múltiples rutas [57]. Este bloque está compuesto de un módulo de residuo f_i y una conexión de salto que omite f_i (ver Figura 2.2). Generalmente el módulo f_i se compone de múltiples capas de convolución. La entrada al bloque i -ésimo está dada por y_{i-1} , mientras la salida está definida por:

$$y_i \equiv f_i(y_{i-1}) + y_{i-1} \quad (2.7)$$

donde $f_i(y)$ usualmente es una secuencia de capas de convolución, que contiene una normalización por lotes [58] y funciones de activación basadas en unidades lineales rectificadas (ReLU) [59].

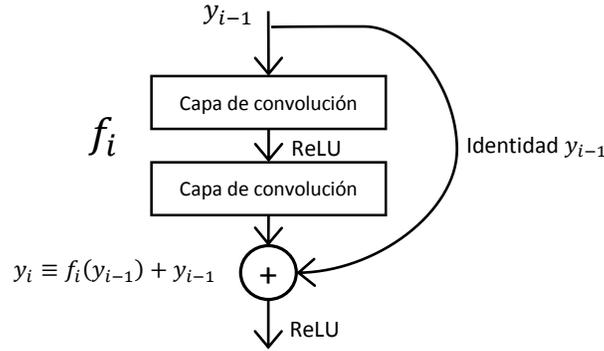


Figura 2.2. Diagrama de bloque de aprendizaje residual [56].

2.4.1.1.2 Redes neuronales recurrentes

Un tipo de red neuronal recurrente, referida simplemente como *Recurrent Neural Network* (RNN) en [60], es una extensión de una red de propagación hacia adelante, la cual es capaz de manejar secuencias de entrada de tamaño variable en su entrada. La RNN tiene un estado oculto recurrente h_t cuya activación en cada instante de tiempo depende de instantes de tiempo previos:

$$h_t = \begin{cases} 0, & t = 0 \\ \phi(h_{t-1}, x_t), & \text{en otro caso} \end{cases} \quad (2.8)$$

donde ϕ es una función de activación, x_t es una entrada en el tiempo t y h_{t-1} la salida del estado oculto anterior. La expresión 2.8 puede ser reescrita como:

$$h_t = \phi(Wx_t + Uh_{t-1}) \quad (2.9)$$

W y U corresponden a los pesos del ejemplo de entrada x_t y del estado oculto anterior h_{t-1} .

Un inconveniente que presenta una RNN es la dificultad de capturar *dependencias a largo plazo* [61], es decir, aprender información relevante de múltiples pasos en el tiempo de una secuencia de datos. La razón de esto es porque los gradientes tienden a crecer o disminuir en cada instante de tiempo, lo cual provoca que desaparezcan o crezcan excesivamente en varios pasos de tiempo. Una solución para superar este problema consiste en usar una función de activación más sofisticada, la cual está compuesta por varias unidades que controlan el flujo de la información.

2.4.1.1.2.1 GRU

Una *Gated Recurrent Unit* (GRU) [62] es una función de activación que está compuesta por dos compuertas, una de reinicio r y otra de actualización z para modificar la información almacenada en un estado oculto h_t . La activación de cada compuerta depende del estado previo oculto h_{t-1} y de la entrada actual x_t . El estado h_t está determinado por una combinación lineal entre una activación previa h_{t-1} y una activación candidata \tilde{h}_t :

$$h_t^j = (1 - z_t^j)h_{t-1}^j + z_t^j\tilde{h}_t^j, \quad (2.10)$$

donde j es el número de unidad (neurona) y z_t^j es la compuerta de actualización que decide el contenido a actualizar:

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j, \quad (2.11)$$

siendo W_z y U_z los pesos de la entrada x en el tiempo t y el estado previo h_{t-1} , respectivamente; mientras que σ es la función de activación sigmoideal. La activación candidata se define como:

$$\tilde{h}_t^j = \tanh(W_h x_t + r_t(U_h h_{t-1}))^j \quad (2.12)$$

r_t representa un conjunto de compuertas de reinicio; W_h y U_h son nuevamente los pesos del ejemplo de entrada en el instante de tiempo t y el estado oculto previo h_{t-1} , con una función de activación basada en tangente hiperbólica. La compuerta r_t se define como:

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j \quad (2.13)$$

siendo σ una función de activación sigmoideal, con los pesos W_r y U_r del ejemplo de entrada x en el instante de tiempo t y en el estado previo h_{t-1} .

2.4.1.1.3 Autocodificadores

Un autocodificador [63] intenta aprender una función $\hat{x} = h_W(x) \approx x$, tal que la salida \hat{x} se aproxime a una entrada x , y se define como:

$$h_W(x) = \phi(Wx + b) \quad (2.14)$$

donde W y b representan los pesos y sesgos respectivamente; ϕ es una función de activación. El proceso de aprendizaje de un autocodificador está basado en aprendizaje no supervisado. En la literatura, los autocodificadores han sido utilizados comúnmente para inicializar los pesos de cada nivel oculto en una red profunda multi-conectada.

2.4.1.2. Modelos generativos profundos

Una manera de aprender una representación \mathbf{h} consiste en modelar la distribución de probabilidad conjunta de los datos de entrada \mathbf{x} , $p(\mathbf{x})$ [30]. Si esta representación \mathbf{h} posee muchos de los factores subyacentes de \mathbf{x} , y las etiquetas de clase \mathbf{y} se encuentran entre los factores más relacionados, entonces es posible determinar \mathbf{y} a partir de \mathbf{h} . Por lo tanto, \mathbf{y} es uno de los factores de \mathbf{x} , y \mathbf{h} representa todos esos factores [24]. Este proceso se puede representar como:

$$p(\mathbf{h}, \mathbf{x}) = p(\mathbf{x}|\mathbf{h})p(\mathbf{h}) \quad (2.15)$$

Como consecuencia los datos tienen una probabilidad marginal, definida como:

$$p(\mathbf{x}) = \mathbb{E}_{\mathbf{h}} p(\mathbf{x}|\mathbf{h}) \quad (2.16)$$

donde $\mathbb{E}_{\mathbf{h}}$ es el valor esperado por marginalizar \mathbf{h} .

Este tipo de modelos son conocidos como generativos y su propósito es descubrir la estructura de los datos con \mathbf{h} como variables latentes estocásticas, tal que explique las variaciones observadas en \mathbf{x} .

2.4.1.2.1 Redes adversarias generativas (GAN)

Las redes adversarias generativas, en inglés *Generative Adversarial Network* (GAN) [64], son modelos generativos basados en teoría de juegos. La idea de GAN consiste en entrenar una red generativa $G(\mathbf{z}; \theta_g)$ con parámetros θ_g para aprender una distribución de probabilidad p_G sobre el conjunto de entrada \mathbf{x} , transformando variables de entrada de ruido \mathbf{z} de una distribución $p_z(\mathbf{z})$ a ejemplos similares a los de los datos originales \mathbf{x} , mediante $\tilde{\mathbf{x}} = G(\mathbf{z}; \theta_g)$. Para este propósito, una GAN entrena una red discriminativa $D(\mathbf{x}; \theta_d)$ con parámetros θ_d para distinguir entre los datos reales \mathbf{x} y los suministrados por $G(\mathbf{z}; \theta_g)$. La red generativa $G(\mathbf{z}; \theta_g)$ se entrena para que $D(\mathbf{x}; \theta_d)$ acepte sus salidas como datos reales. De esta manera, $G(\mathbf{z}; \theta_g)$ intenta aproximar de mejor manera la distribución de probabilidad $p(\mathbf{x})$. Esta idea utiliza un juego de suma cero para el aprendizaje de $D(\mathbf{x}; \theta_d)$ y $G(\mathbf{z}; \theta_g)$. Las funciones $v(\theta_g, \theta_d)$ y $-v(\theta_g, \theta_d)$ corresponden a la recompensa del discriminador y generador. Cada jugador intenta maximizar su propia recompensa mediante la siguiente expresión:

$$g^* = \arg \min_G \max_D v(G, D) \quad (2.17)$$

donde v está definida como:

$$v(\theta^{(g)}, \theta^{(d)}) = \mathbb{E}_{x \sim p_{data}(x)} \log D(x) + \mathbb{E}_{z \sim p_z(z)} \log(1 - D(G(z))) \quad (2.18)$$

siendo $\mathbb{E}_{x \sim p_{data}}$ el valor esperado del ejemplo x que sigue la distribución de probabilidad sobre el conjunto de los datos reales \mathbf{x} , y $\mathbb{E}_{z \sim p_z(z)}$ el valor esperado de la entrada de ruido z que sigue la distribución de probabilidad del modelo generativo $p_z(\mathbf{z})$.

2.4.1.2.2 Autocodificadores variacionales

Este tipo de modelos, en inglés *Variational Autoencoders* (VAE) [65], intentan modelar la probabilidad conjunta de los datos de entrada $p(\mathbf{x})$, generando una representación basada en un conjunto de variables latentes estocásticas \mathbf{h} . Así, la probabilidad de los datos de entrada \mathbf{x} está dada por:

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{h})p(\mathbf{h})d\mathbf{z} \quad (2.19)$$

Los VAE utilizan una red neuronal para hacer inferencia aproximada, introduciendo un conjunto de variables latentes aleatorias \mathbf{h} para capturar las variaciones existentes en las variables observadas \mathbf{x} . La distribución de probabilidad conjunta se define como:

$$p(\mathbf{x}, \mathbf{h}) = p(\mathbf{x}|\mathbf{h})p(\mathbf{h}) \quad (2.20)$$

La probabilidad *a priori* $p(\mathbf{h})$ usualmente está definida como una distribución de probabilidad Gaussiana, mientras que la probabilidad condicional $p(\mathbf{x}|\mathbf{h})$ es un modelo de observación, cuyos

parámetros se calculan mediante una función paramétrica de \mathbf{h} . En VAE se utiliza una función altamente no lineal, usualmente una red neuronal, para parametrizar $p(\mathbf{x}|\mathbf{h})$. Para calcular la probabilidad *a posteriori* $p(\mathbf{h}|\mathbf{x})$, VAE emplea una aproximación variacional $q(\mathbf{h}|\mathbf{x})$ de ésta, tal que habilita el uso del límite variacional más bajo:

$$\log p(\mathbf{x}) \geq -\text{KL}(q(\mathbf{h}|\mathbf{x})||p(\mathbf{h})) + \mathbb{E}_{q(\mathbf{h}|\mathbf{x})} \log p(\mathbf{x}|\mathbf{h}) \quad (2.21)$$

donde $\text{KL}(q(\mathbf{h}|\mathbf{x})||p(\mathbf{h}))$ es la divergencia Kullback-Leibler [66] entre dos distribuciones de probabilidad Q y P . Esta divergencia determina la similitud entre dos distribuciones de probabilidad.

El modelo generativo $p(\mathbf{x}|\mathbf{h})$ y el modelo de inferencia $q(\mathbf{h}|\mathbf{x})$ se entrenan conjuntamente al maximizar el límite variacional más bajo con respecto a sus parámetros. La inferencia de este modelo se realiza mediante el algoritmo de gradiente descendente.

2.4.1.3. Regularización

Según [53], regularización es cualquier modificación que se hace a cualquier algoritmo de aprendizaje, que está intentando reducir el error de generalización, pero no su error de entrenamiento. Las técnicas más usuales de regularización en aprendizaje profundo son:

- **Aumento de datos** [67]: el aumento del número de ejemplos en el conjunto de entrenamiento mediante transformaciones de los datos de un dominio específico.
- **Weight decay** [68]: este enfoque se basa en limitar la capacidad de los modelos por añadir una penalización Ω a una función objetivo J . Este enfoque evita que los valores de los pesos crezcan demasiado a menos de que sea necesario.
- **Dropout** [69]: esta técnica previene el sobre-ajuste y proporciona una manera de promediar el resultado de muchas arquitecturas de redes profundas. La idea de *Dropout* consiste en retirar un conjunto de neuronas de la red, de manera temporal, junto a todas sus conexiones entrantes y salientes, formando una arquitectura con número menor de unidades y parámetros. La elección de las unidades se realiza de manera aleatoria.

2.4.1.4. Aprendizaje de transferencia

Dado un dominio fuente D_s y una tarea de aprendizaje T_s , un dominio objetivo D_t y una tarea de aprendizaje T_t . El aprendizaje de transferencia propone ayudar a mejorar el aprendizaje de una función de clasificación objetivo $f_t(\cdot)$ en D_t usando el conocimiento en D_s y T_s , donde $D_s \neq D_t$, o $T_s \neq T_t$ [44]. $D = \{\mathbf{x}, p(\mathbf{x})\}$ es un dominio que consiste de un conjunto de ejemplos de entrada \mathbf{x} con su respectiva distribución de probabilidad $p(\mathbf{x})$; $T = \{\mathbf{y}, f(\cdot)\}$ es una tarea de aprendizaje compuesta de una función de clasificación $f(\cdot)$ con sus respectivas etiquetas \mathbf{y} .

Una taxonomía de los principales enfoques de aprendizaje de transferencia se muestra en la Figura 2.3 [44]. Los enfoques están organizados de acuerdo a la disponibilidad de las etiquetas en el dominio fuente o en el destino, y se encuentran agrupados en aprendizaje de transferencia inductivo, transductivo y no supervisado. En un enfoque inductivo las etiquetas de los datos están disponibles únicamente en el dominio objetivo, mientras que en un enfoque transductivo las etiquetas están disponibles en el dominio fuente. Finalmente, cuando no se dispone de etiquetas en ninguno de los dominios, éste se denomina aprendizaje de transferencia no supervisado. En la presente investigación se abordarán los enfoques aprendizaje autodidacta y adaptación de dominio, los cuales se describen brevemente en seguida.

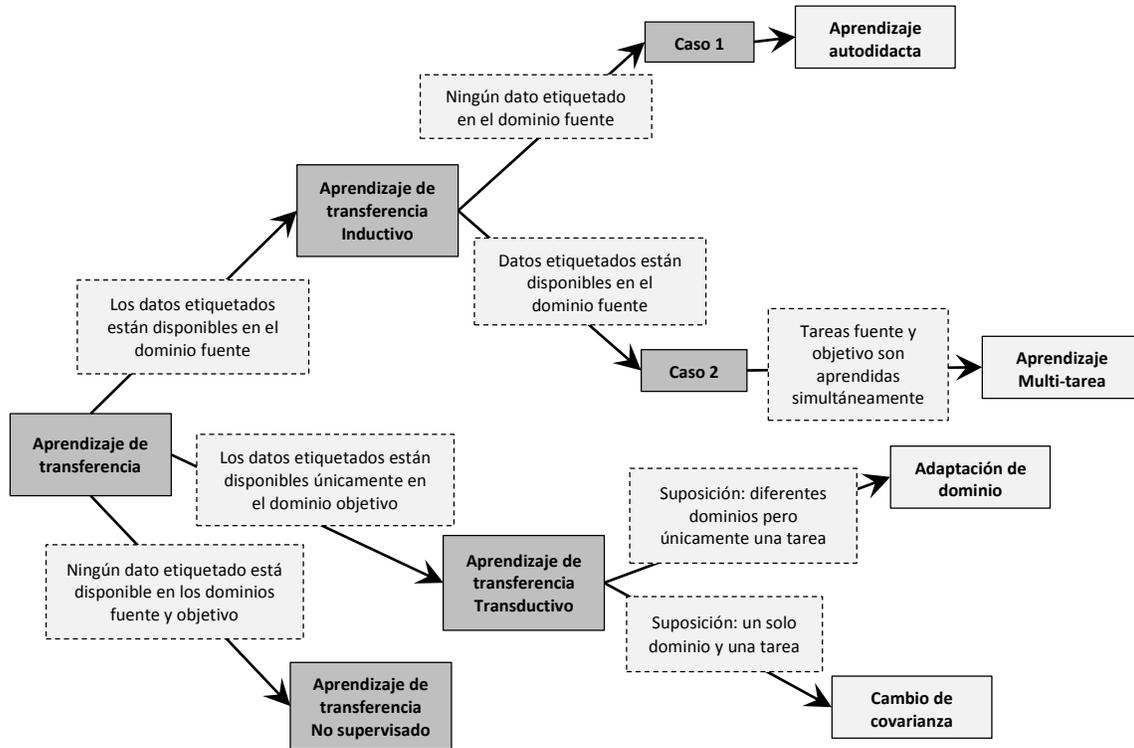


Figura 2.3 Diferentes configuraciones de aprendizaje de transferencia [44].

2.4.1.4.1 Aprendizaje autodidacta

Sea $\mathbf{x}_l = ((x_l^1, y^1), \dots, (x_l^m, y^m))$ un conjunto de entrenamiento etiquetado con m ejemplos, independientes e idénticamente distribuidos, de alguna distribución D , donde $\mathbf{x}_l \in \mathbb{R}^n$ y $\mathbf{y} = \{1, \dots, C\}$ es el conjunto de etiquetas de clase; sea $\mathbf{x}_u = x_u^1, \dots, x_u^k \in \mathbb{R}^n$ un conjunto de k ejemplos no etiquetados, producidos de una distribución de probabilidad diferente a la de \mathbf{x}_l . De acuerdo a Raina et al. (2007) [45], dados los conjuntos \mathbf{x}_l y \mathbf{x}_u , un algoritmo de aprendizaje autodidacta produce una función $f : \mathbb{R}^n \rightarrow \{1, \dots, C\}$ que intenta imitar la relación entrada-etiqueta para ambos conjuntos; esta función f es evaluada bajo la misma distribución de probabilidad que los datos etiquetados.

2.4.1.4.2 Adaptación de dominio

Dado un dominio fuente D_S y su respectiva tarea de aprendizaje T_S , un dominio objetivo D_T y su tarea de aprendizaje T_T , la adaptación de dominio propone mejorar el aprendizaje de una función de clasificación $f_T(\cdot)$ en D_T usando el conocimiento de D_S y T_S , donde $D_T \neq D_S$ y $T_S = T_T$ [44]. Adicionalmente, algunos datos no etiquetados del dominio objetivo deben estar disponibles en la etapa de entrenamiento.

2.5. Medidas de evaluación

El criterio utilizado para evaluar un método de aprendizaje de representaciones es un factor importante para conocer con exactitud su rendimiento. En esta investigación los métodos propuestos se evaluarán con respecto a su desempeño en tareas de clasificación. La Tabla 2-2 muestra la definición de una matriz de confusión, para el caso más básico de clasificación de dos clases.

Tabla 2-2. Matriz de confusión para una clasificación de dos clases.

	Positiva Real	Negativa Real
Positiva predicha	Verdaderos positivos (VP)	Falsos Positivos (FP)
Negativa predicha	Falsos Negativos (FN)	Verdaderos Negativos (VN)

En la literatura se reportan diversas medidas para la evaluación de la clasificación, entre éstas una de las más utilizadas es la razón de exactitud [70]:

$$\text{Exactitud} = \frac{VP+VN}{VP+FN+FP+VN} \quad (2.22)$$

Un problema de esta medida es que no considera los resultados por clase, por lo que en un problema con clases desbalanceadas existe una tendencia a favorecer la clase mayoritaria. Por otro lado, una medida que permite medir el desempeño en la clasificación en problemas con clases desbalanceadas es *F-measure* [71], misma que se expresa como:

$$F - \text{measure} = \frac{2 * \text{Sensitividad} * \text{Precisión}}{\text{Precisión} + \text{Sensitividad}} \quad (2.23)$$

donde precisión y sensibilidad están definidos como:

$$\text{Precisión} = \frac{VP}{VP+FP} \quad (2.24)$$

$$\text{Sensitividad} = \frac{VP}{VP+VN} \quad (2.25)$$

La métrica *F-measure* es una medida que puede ser vista como el compromiso entre precisión y sensibilidad. El valor de esta métrica se encuentra en el intervalo de entre cero y uno, donde un valor más alto indica un mejor desempeño del clasificador y es producido cuando los valores de precisión y sensibilidad son altos. Asumiendo un sistema de recuperación de información, la precisión expresa la fracción de objetos que son relevantes, mientras que sensibilidad es la fracción de objetos relevantes que son recuperados.

Otra medida muy utilizada para la evaluación de la clasificación es una gráfica basada en *Receiver Operating Characteristic* (ROC) [72]. Este tipo de gráficas muestran que un clasificador no puede incrementar su valor de verdaderos positivos sin aumentar los falsos positivos. Una medida basada en gráficos ROC es calcular el área bajo la curva (AUC) [73], la cual se define como:

$$AUC = \frac{(1 + \text{Especificidad}) * \text{Sensitividad} + \text{Especificidad} * (1 - \text{Sensitividad})}{2} \quad (2.26)$$

donde:

$$\text{Especificidad} = \frac{VN}{FP+VN} \quad (2.27)$$

y sensibilidad se definió anteriormente.

AUC puede verse como la probabilidad que un clasificador obtenga un ejemplo de la clase positiva, elegido aleatoriamente, más que un ejemplo de la clase negativa. Los puntajes de AUC están en el rango entre [0,1], donde el valor más alto es el mejor.

3. Trabajo relacionado

En esta sección se muestra el análisis de los principales trabajos que abordan el problema de aprender una representación de características de datos de EEG sobre múltiples sujetos en espacio, tiempo y frecuencia. El análisis se enfoca principalmente en los trabajos de aprendizaje profundo.

3.1. Aprendizaje de representaciones de datos de EEG

Las investigaciones existentes para describir una función cerebral han provisto de un número de soluciones que utilizan la información de señales de EEG en el dominio temporal [74, 25, 75] o en un dominio transferido como el espectral [76]. De esta manera, diferentes características temporales y espectrales son usualmente utilizadas como características representativas para entrenar un clasificador [77, 78]. Sin embargo, a pesar de estos avances, una clasificación precisa y robusta de las actividades neuronales es todavía un reto, pues esta precisión es dependiente de las propiedades discriminativas de las características extraídas. Como se mencionó, la línea de investigación de aprendizaje de representaciones busca superar esta limitante, por extraer de manera automática un conjunto de características útiles para la clasificación y hacerse menos dependiente de una ingeniería humana.

En la literatura es posible encontrar diversos trabajos para aprender características de datos de EEG. Bajo este esquema, algunos trabajos se basan principalmente en aprendizaje de variedades [3, 79, 80], aprendizaje de diccionarios [81, 82, 3, 83, 84], e incluso aquellos basados en modelos probabilistas [85, 26, 86, 22]. Recientemente, existe un interés por los métodos de aprendizaje profundo, pues muestran aprender estructuras jerárquicas y permiten obtener representaciones abstractas, mejorando la eficacia en la clasificación [32, 8]. Además que este tipo de modelos, al estar basados en redes neuronales, surgen como algoritmos para el análisis de series de tiempo no lineales y no estacionarias, los cuales podrían ser de utilidad para la detección de eventos relacionados con la actividad cerebral bajo estudio. En la siguiente sección se describen los trabajos más relevantes para datos de EEG basados en aprendizaje profundo.

3.1.1. Aprendizaje profundo

Estos modelos están categorizados en dos grupos: discriminativos y generativos. A continuación se describen los principales trabajos reportados para cada grupo. La idea de este estudio consiste en identificar fortalezas, así como los trabajos más cercanos a nuestra investigación. Finalmente, se da una discusión general de los trabajos reportados.

3.1.1.1. Modelos discriminativos

Las principales características de los trabajos basados en modelos discriminativos se describen a continuación.

Redes de convolución: los trabajos basados en redes de convolución proveen soluciones que procesan los datos de EEG en el dominio temporal [87, 88, 8, 33], espectral [2, 89, 90] y temporal-espectral [2, 79]. Para estos dos últimos tipos de procesamiento, se emplean las transformadas de *Fourier* o *Wavelets* en una primera etapa, filtrando las señales en diferentes bandas de frecuencia de acuerdo a cada actividad neuronal bajo estudio. La ventaja de las redes de convolución consiste en la capacidad de aprender un conjunto de características conservando la estructura de los datos (ej. tiempo y frecuencia), por lo que diferentes arquitecturas de esta red se han reportado en la literatura [33, 91, 92].

Redes recurrentes: este tipo de redes poseen la capacidad de aprender dependencias temporales a partir de una secuencia. Por lo tanto, los trabajos reportados en la literatura realizan su análisis sobre datos temporales [93] o temporales-espectrales mediante alguna transformación previa [34].

Autocodificadores: los trabajos basados en autocodificadores apilados reportan una extracción de características principalmente en el dominio espectral [35, 11, 94, 36, 95, 37]. En trabajos como en [35], se reporta el uso adicional de métodos como análisis de componentes principales y cambio de covarianza para reducir la alta dimensionalidad y la varianza en el tiempo respectivamente. Otros trabajos emplean un autocodificador de un solo nivel para generar un conjunto de funciones base y entrenar así un diccionario [3]. Específicamente, de un segmento de tiempo se obtiene una representación de características de la cual se forma una palabra, así una señal continua en el tiempo está descrita por una secuencia de palabras. Por otra parte, los autocodificadores se han empleado también para generar una representación de cada segmento de tiempo y entrenar posteriormente un clasificador con esta nueva representación [80, 96]. Bajo este esquema, un caso de interés surge en el uso de redes recurrentes como modelos autocodificadores para la extracción de características temporales [97].

Un inconveniente de los trabajos reportados basados en redes de convolución, recurrentes y autocodificadores, es su esfuerzo en desarrollar modelos para sujetos específicos. Además que los trabajos actuales no reportan incorporar la información espacial. A continuación se describen los trabajos reportados basados en una combinación de arquitecturas de redes profundas.

Arquitecturas híbridas: estos trabajos combinan diferentes redes profundas con el propósito de aprovechar sus fortalezas [98, 32]. Los autores en [32] proponen transformar cada segmento de tiempo en una secuencia de imágenes (ver Figura 3.1). Esta secuencia representa la información en espacio, tiempo y frecuencia de las señales de EEG. Las señales se filtran en las bandas de frecuencia *theta*, *alpha* y *beta*, luego de aplicar la Transformada de *Fourier* de tiempo corto. Una imagen contiene la información espectral y espacial, mientras que la secuencia de imágenes provee de la información temporal. Esta secuencia de imágenes se utiliza para entrenar una arquitectura de red profunda, la cual consiste de un ensamble de siete arquitecturas de convolución que aprenden características espaciales-espectrales en una primera etapa, mientras que en una segunda etapa se incorpora una red recurrente para aprender características temporales. Para el mejor de nuestro conocimiento, este es el único trabajo basado en modelos discriminativos que pretende aprender una representación sobre múltiples sujetos.

En [98] se emplea una estrategia basada en *Cellular Neural Network* y una red recurrente básica bidireccional (BRNN) para aprender características espacio-temporales basado en imágenes de EEG. No obstante, el trabajo no incorpora información espectral, ni pretende aprender características sobre múltiples sujetos.

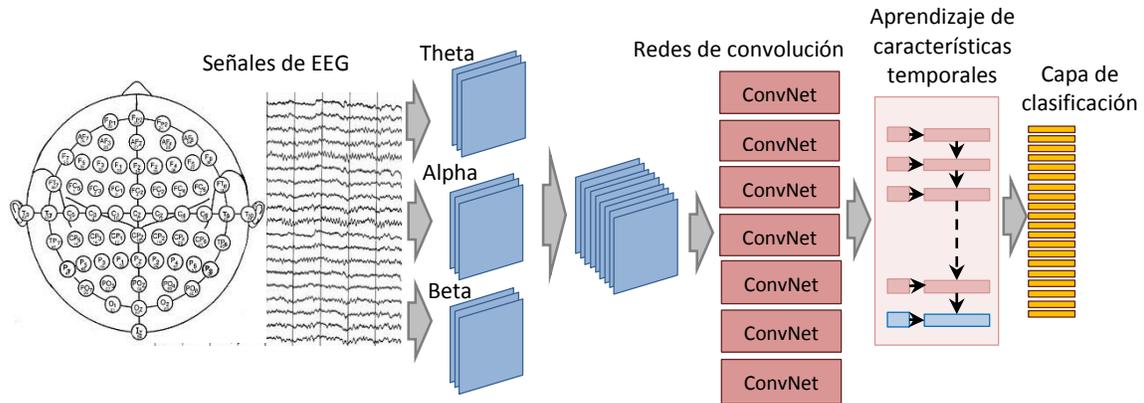


Figura 3.1 Modelo propuesto en [32] para el aprendizaje de características espaciales, temporales y espectrales, de señales de EEG sobre múltiples sujetos.

3.1.1.2. Modelos generativos

Los trabajos reportados para este tipo de modelos están basados en Máquinas restringidas de *Boltzman* y Redes de creencias profundas. Estos modelos se han utilizado principalmente para aprender características en el dominio de frecuencia debido a que no capturan dependencias temporales [99, 100, 12, 101, 102, 103, 104, 105, 106, 107]. No obstante, se han reportado algunos trabajos que combinan redes de creencias profundas con redes recurrentes o Modelos Ocultos de Markov para solventar esta limitante [100, 22, 108]. Aunque en [109, 39, 110, 111, 112] se muestran esfuerzos para aprender características de datos crudos, estos trabajos no superan a los resultados alcanzados por los trabajos que utilizan adicionalmente alguna técnica de procesamiento de señales [113, 114, 22].

Entre algunos otros trabajos relevantes, se encuentra el uso Máquinas de *Boltzman* para aprender una representación de características de manera no supervisada; luego, las características aprendidas son utilizadas para entrenar un clasificador [9, 6, 39]. Así también, es posible encontrar soluciones basadas en métodos semi-supervisados [115], ante la disponibilidad de pocos ejemplos etiquetados en el entrenamiento.

En el contexto de modelos generativos, solo un trabajo reporta aprender una representación sobre datos de múltiples sujetos, mostrando un rendimiento superior con respecto a una arquitectura basada en redes de convolución [8]. Sin embargo, este trabajo no toma en consideración la estructura inherente de los datos en espacio, tiempo y frecuencia.

3.1.1.3. Discusión

Analizando los trabajos reportados y considerando sus características, se puede resaltar lo siguiente:

- Los trabajos basados en redes neuronales profundas (modelos discriminativos) permiten procesar la estructura inherente en un tensor de datos de EEG en espacio, tiempo y frecuencia, durante el proceso de aprendizaje [2, 32]. Por otro lado, el procesamiento de los trabajos basados en modelos generativos se enfocan sólo en vectores de características [100, 12, 102]. Además, una gran parte de los trabajos en ambos tipos de modelos se encaminan a aprender

representaciones de actividades neuronales focalizadas. Para nuestro caso de estudio de carga cognitiva, resulta adecuado involucrar la información espacial, pues se ha mostrado evidencia que esta información permite una mejora en la precisión de la clasificación [50]. Aunque diferentes esfuerzos de modelos discriminativos y generativos se han reportado en la literatura para aprender una representación de datos crudos, una combinación con técnicas de procesamiento de señales muestran potencializar la eficacia de los modelos de aprendizaje profundo [113, 114, 22].

- La mayor parte de trabajos reportados se enfocan en aprender características de un sujeto específico. En lo que respecta a modelos discriminativos, el único trabajo que aprende una representación sobre múltiples sujetos en espacio, tiempo y frecuencia, es el basado en una red recurrente-convolucional [32]. De la misma manera, en el caso de modelos generativos, únicamente el trabajo reportado en [8] extrae una representación sobre múltiples sujetos mediante una red de creencias profundas, presentando mejores resultados que una arquitectura basada en una red de convolución. Sin embargo, el modelo propuesto se basa en aprender características en el dominio temporal, omitiendo la información espectral y espacial.
- En general, los trabajos actualmente reportados para el aprendizaje de características sobre múltiples sujetos parten del supuesto que los datos provienen de una misma distribución de probabilidad [32]. Esto provoca que las características aprendidas sean sub-óptimas y obtengan por ende, una menor eficacia en la clasificación.
- Los modelos generativos han sido poco explorados en el análisis de EEG, pues los trabajos actuales se basan únicamente en Máquinas de *Boltzmann* y redes de creencias profundas. La ventaja de modelos generativos se muestra principalmente en el uso de aprendizaje no supervisado, permitiendo re-utilizar las características aprendidas de manera no supervisada para entrenar un clasificador [9, 6, 39]. Así también, este tipo de modelos ha permitido el desarrollo métodos semi-supervisados para abordar el problema de un número reducido de ejemplos etiquetados [115]. Aunque los modelos basados en autocodificadores permiten desarrollar este mismo tipo de trabajos [80, 96], los modelos generativos aventajan a autocodificadores en modelar los diferentes factores de variación en los datos y sus relaciones. Estas ventajas motivan a seguir explorando otros enfoques basados en modelos generativos.

4. Propuesta de investigación

En esta sección se presenta el problema a resolver, la motivación y la justificación a la solución propuesta, así como las preguntas de investigación, la hipótesis y los objetivos planteados en esta propuesta de investigación.

4.1. Problema a resolver

Como se comentó en la sección 3, diferentes métodos de aprendizaje de representaciones han sido aplicados para extraer y representar características de señales de EEG [100, 12, 32, 33, 3]. Hasta ahora, la mayoría de los estudios se enfatizan en la extracción de características para sujetos específicos, sin considerar las diferencias individuales entre los múltiples sujetos. Entre los principales problemas se encuentran los siguientes:

- **Variación espacial:** la mayor parte de los métodos actuales no toman en consideración las diferencias de la anatomía cortical y la localización de los electrodos. En ciertos dominios, tales como el de carga cognitiva, es preferible considerar este tipo de información, pues la actividad cerebral puede manifestarse en diferentes áreas corticales.
- **Varianza en el tiempo y dependencias temporales:** las señales de EEG tienen una dependencia explícita con respecto al tiempo. Esto es, para una entrada en un instante de tiempo t , un modelo puede clasificar un valor $y(t)$, pero para una entrada idéntica en un tiempo $t + \Delta t$, ésta podría ser asociada con un resultado diferente [47].
- **Factores subyacentes de variación:** la mayoría de los trabajos actuales no se enfocan en capturar los factores ocultos de variación asociados con los datos y su relación entre estos [24].
- **Supuesto de una distribución idéntica e independiente:** los métodos existentes se basan en el supuesto estadístico básico que los datos son muestreados de una distribución idéntica e independiente [42, 40]. Esto genera una representación de características sub-óptima, provocando que su rendimiento en tareas de clasificación sea bajo, cuando se aplica en otros conjuntos con diferentes distribuciones y desconocidas para el modelo.
- **Número reducido de ejemplos etiquetados para entrenar:** reunir datos etiquetados en dominios de EEG y para sujetos específicos, es una barrera en el análisis y clasificación de este tipo de señales. La mayoría de estudios actuales están basados en un paradigma de aprendizaje supervisado, requiriendo así una gran cantidad de datos etiquetados para un desempeño exitoso [43].

El problema a atacar en esta investigación se enfoca en el *aprendizaje de representaciones* de series de tiempo, obtenidas de datos de EEG y para casos de estudio de la carga cognitiva, tal que las características aprendidas sean más robustas ante las diferencias individuales entre los múltiples sujetos. Una mayor robustez en la representación aprendida implica una mejora en la eficacia de la clasificación de datos de EEG con respecto a los trabajos reportados en el estado del arte. Por robustez nos referimos a producir representaciones que permitan construir clasificadores que no sean afectados por variaciones respecto a las hipótesis de los modelos.

4.2. Motivación y justificación

El problema que se plantea aquí es considerado de suma importancia en varios campos de la investigación, tanto en ciencias de la computación como en el área médica [39], la psicológica [18], entre otras [17]. En el caso particular de estudio como la supervisión de estados mentales a través de la carga cognitiva, ha tomado una especial atención en áreas de aplicación donde se requiere tomar una decisión crítica, tal como en conducción de vehículos [8], en control de tráfico aéreo [116], o no tan críticas, como en el contexto de aprendizaje [18, 20, 16]. El ser humano está expuesto a cometer errores y sufre una pérdida de eficiencia cuando una tarea cognitiva es muy alta [17]. Entonces, es necesario contar con modelos de apoyo a individuos, que permitan medir la carga cognitiva de manera objetiva, pues detectar este nivel de carga permitiría el diseño de estrategias que mejoren su productividad [15, 18, 17].

Diversas soluciones se han propuesto para clasificar las diferentes funciones cerebrales como la carga cognitiva [117, 15]. A pesar del éxito mostrado por estas soluciones, su eficacia está restringida por las propiedades de las características extraídas, pues se requiere tratar con un conocimiento

extenso en el dominio y existe la posibilidad de no utilizar eficazmente toda la información en los datos o utilizarla de forma ineficaz [30]. Diversos esfuerzos se han reportado para aprender representaciones de características de datos de EEG para superar estas limitantes. Entre los diferentes paradigmas existentes, el aprendizaje profundo ha tomado gran relevancia por su capacidad de aprender capas jerárquicas, permitiendo obtener representaciones más abstractas [27]. Estos modelos, basados usualmente en redes neuronales, permiten el análisis de series de tiempo no lineales y no estacionarias, por lo que podrían ser de utilidad para la extracción de eventos relacionados con las actividades neuronales.

Bajo el contexto de aprendizaje profundo, una gran cantidad de esfuerzos se han reportado en dirección de aprender mejores representaciones [100, 12, 33, 3]. Sin embargo, es usual que los trabajos reportados se enfoquen en modelos de aprendizaje de representaciones para entrenar clasificadores en sujetos específicos. En aplicaciones reales, las distribuciones de probabilidad de los datos varían a través de sujetos y entre múltiples sesiones dentro del mismo sujeto [42, 40], lo cual podría requerir del entrenamiento de nuevos clasificadores para cada sujeto e incluso para cada día [41]. En la práctica, esto resulta prohibitivo, debido al consumo de tiempo y el costo en los recursos del etiquetado de los datos y el entrenamiento de nuevos modelos para cada nuevo sujeto. Además, sólo unos pocos de los trabajos reportados llevan a cabo un análisis en espacio, tiempo y frecuencia. Este tipo de análisis es de nuestro interés, pues las actividades cerebrales de nuestro caso de estudio por lo general existen en determinadas ubicaciones espaciales y, en particular, bandas de frecuencia [10]. Por los motivos antes descritos, es necesario de proveer de modelos que permitan aprender características de señales EEG sobre múltiples sujetos, tal que obtengan una mejor eficacia en la clasificación de sujetos desconocidos que los trabajos actuales.

El desarrollo de un modelo de aprendizaje de representaciones toma gran relevancia, pues permite compartir fortalezas estadísticas que son de utilidad en modelos de aprendizaje semi-supervisados, multi-tarea, adaptación de dominio y aprendizaje autodidacta [24]. Si bien, aunque nuestra investigación está acotada para casos de estudio de las señales de EEG para determinar los niveles de carga cognitiva, ésta podría tener una amplia aplicación en otros campos, tales como en el diagnóstico médico, la clasificación de estados de sueño, la detección de emociones, en el diseño de aplicaciones interfaces cerebro-computadora.

4.3. Preguntas de investigación

Considerando el problema de investigación descrito en la sección 4.1, surge la siguiente pregunta de investigación:

- ¿Cómo se pueden aprender representaciones de características más robustas en espacio, tiempo y frecuencia sobre múltiples sujetos, tal que permita mejorar la eficacia en la clasificación de señales de EEG que los métodos reportados en el estado del arte?

A partir de esta pregunta de investigación se plantean las siguientes preguntas secundarias:

- **Q1:** ¿Cómo se puede aprender una representación, tal que capture la información espacial, temporal y de frecuencia, acorde a las propiedades inherentes de las señales de EEG?
- **Q2:** ¿De qué manera se puede mejorar la representación aprendida para modelar la variabilidad asociada a los datos de EEG entre múltiples sujetos?
- **Q3:** ¿De qué manera se puede incorporar información relevante obtenida de otros dominios que analizan señales de EEG, tal que esta información relevante permita mejorar el aprendizaje de representaciones robustas?

4.4. Hipótesis

En esta propuesta de investigación se plantea la siguiente hipótesis:

Un marco de trabajo, basado en redes neuronales profundas, modelos generativos profundos y aprendizaje de transferencia, puede aprender una representación más robusta en espacio, tiempo y frecuencia sobre múltiples sujetos, tal que la representación aprendida obtiene una mejor eficacia en la clasificación de datos de EEG que los trabajos del estado del arte.

En el contexto de esta investigación, la capacidad de representación del modelo se validará en términos de su capacidad de clasificar en casos de estudio de carga cognitiva, la cual será evaluada en términos de su eficacia, utilizando las siguientes métricas: Exactitud, *F-measure* y Área bajo la curva.

4.5. Objetivo general

El objetivo principal de esta investigación es el desarrollar un *marco de trabajo* que permita aprender representaciones más robustas en espacio, tiempo y frecuencia de datos de EEG sobre múltiples sujetos, comparado con los trabajos reportados actualmente en el estado del arte. Una representación más robusta permitirá obtener una *mejor eficacia en la clasificación*.

4.6. Objetivos específicos

El objetivo general de esta investigación será alcanzado a través del cumplimiento de los siguientes objetivos específicos:

- **O1:** Proponer una arquitectura de red neuronal profunda que permita aprender una representación con la información espacial, temporal y espectral, de las señales de EEG.

Esta arquitectura será utilizada para extraer una representación de características basada en una mejora estructural de arquitecturas de redes profundas. La evaluación en esta etapa consistirá en un enfoque de aprendizaje supervisado.

- **O2:** Proponer un modelo generativo profundo para modelar probabilísticamente la variabilidad en espacio, tiempo y frecuencia, de las señales de EEG sobre múltiples sujetos, tal que integre la arquitectura propuesta en el objetivo uno.

Este modelo será utilizado para extraer los diferentes factores de variación en los datos de EEG, así como modelar las relaciones entre estos factores. Los factores poco informativos serán descartados, con el propósito de proveer características más robustas para la clasificación. En esta etapa se abordará también el problema del supuesto que los datos no provienen de una distribución idéntica e independiente. La evaluación consistirá en un enfoque de aprendizaje semi-supervisado; enfoque que permitirá incorporar datos etiquetados y no etiquetados al aprendizaje del modelo propuesto.

- **O3:** Extender el modelo generativo profundo a un paradigma de aprendizaje autodidacta.

Este enfoque, basado en aprendizaje de transferencia y a diferencia de un enfoque semi-supervisado, parte del supuesto que los datos no etiquetados provienen de una distribución

de probabilidad diferente a la de los datos etiquetados disponibles. El modelo será utilizado para mejorar el rendimiento del modelo generativo propuesto en el segundo objetivo, usando datos no etiquetados con diferentes distribuciones de probabilidad.

4.7. Contribuciones

Como resultado de este trabajo, se espera contribuir al conocimiento a través del diseño de:

1. Una nueva arquitectura de red neuronal profunda que permite aprender una representación con la información espacial, temporal y espectral, de las señales de EEG.
2. Un modelo generativo profundo novedoso que permite modelar la variabilidad entre los múltiples sujetos en espacio, tiempo y frecuencia de las señales de EEG.
3. Un modelo generativo profundo basado en aprendizaje autodidacta que permite incorporar datos no etiquetados provenientes de diferentes distribuciones de probabilidad.

5. Metodología

En la Figura 5.1 se muestra el diagrama de la metodología a seguir, la cual contiene un total de 11 actividades. El modelo esperado luego de seguir la metodología propuesta se muestra en la Figura 5.2. Las actividades (1) y (3) en estos diagramas conforman la parte de pre-procesamiento de las señales de EEG. Posteriormente, en las actividades (2) y (4) se plantea abordar el primer objetivo específico, proponiendo una arquitectura de red neuronal profunda para el aprendizaje de representaciones en espacio, tiempo y frecuencia.

En las actividades (5) y (6) se prevé estudiar los métodos y técnicas necesarias para proponer un modelo generativo profundo, tal que permita modelar la variabilidad entre los sujetos. Los modelos a explorar se describen en los pasos (7) y (8), los cuales son Autocodificadores Variacionales (VAEs) y Redes Adversarias Generativas (GANs). Estos dos pasos consideran incorporar la estructura de red neuronal profunda propuesta en el paso (4). El modelo generativo propuesto consistirá en un aprendizaje semi-supervisado, el cual permite hacerse menos dependiente de datos etiquetados, incorporando información no etiquetada. Este último modelo daría cumplimiento al segundo objetivo específico.

En las actividades (9) y (10) se tiene contemplado extender el modelo generativo propuesto en (7)-(8) a un aprendizaje autodidacta. Una suposición de métodos semi-supervisados consiste en que los datos no etiquetados provienen de la misma distribución de probabilidad que los datos etiquetados. En un aprendizaje autodidacta los datos no etiquetados pueden provenir de cualquier distribución de probabilidad. Finalmente, la actividad (11) está asociada con la aplicación del modelo, el diseño experimental y la evaluación.

A continuación se describen a detalle las 11 actividades relacionadas con la metodología propuesta:

1. Análisis y estudio de técnicas de procesamiento de señales para describir a los datos de EEG, de manera inicial, en un tensor de datos con información en espacio, tiempo y frecuencia, tal que se facilite la aplicación de técnicas de aprendizaje de representaciones.
 - a) Los métodos para la representación en tiempo-frecuencia, tal como *Short-Time Fourier Transform* (STFT) y *Wavelets*, serán estudiados y analizados.
 - b) Estudio y análisis de métodos existentes que incorporan la información de las localizaciones espaciales de los múltiples electrodos del EEG.

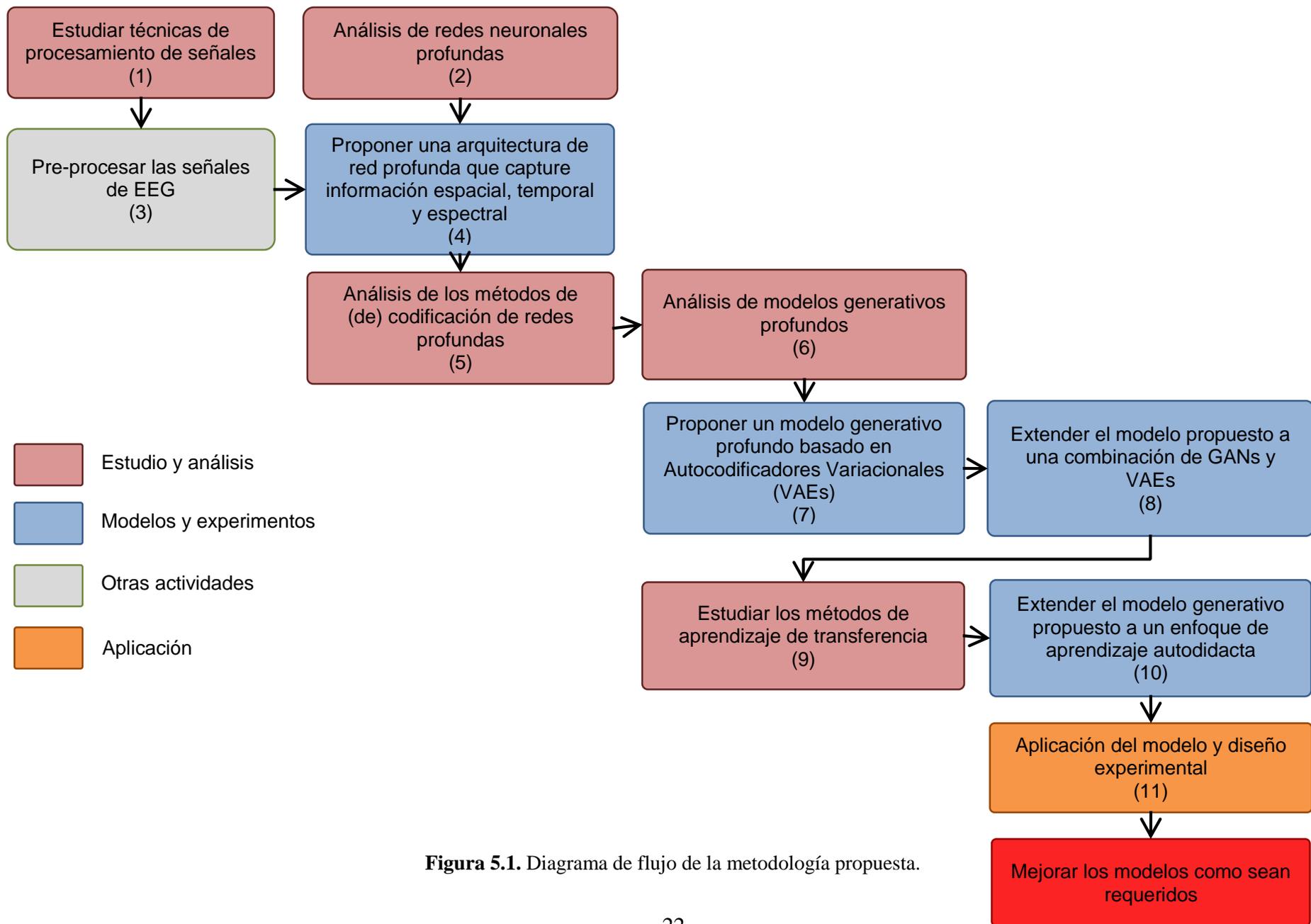


Figura 5.1. Diagrama de flujo de la metodología propuesta.

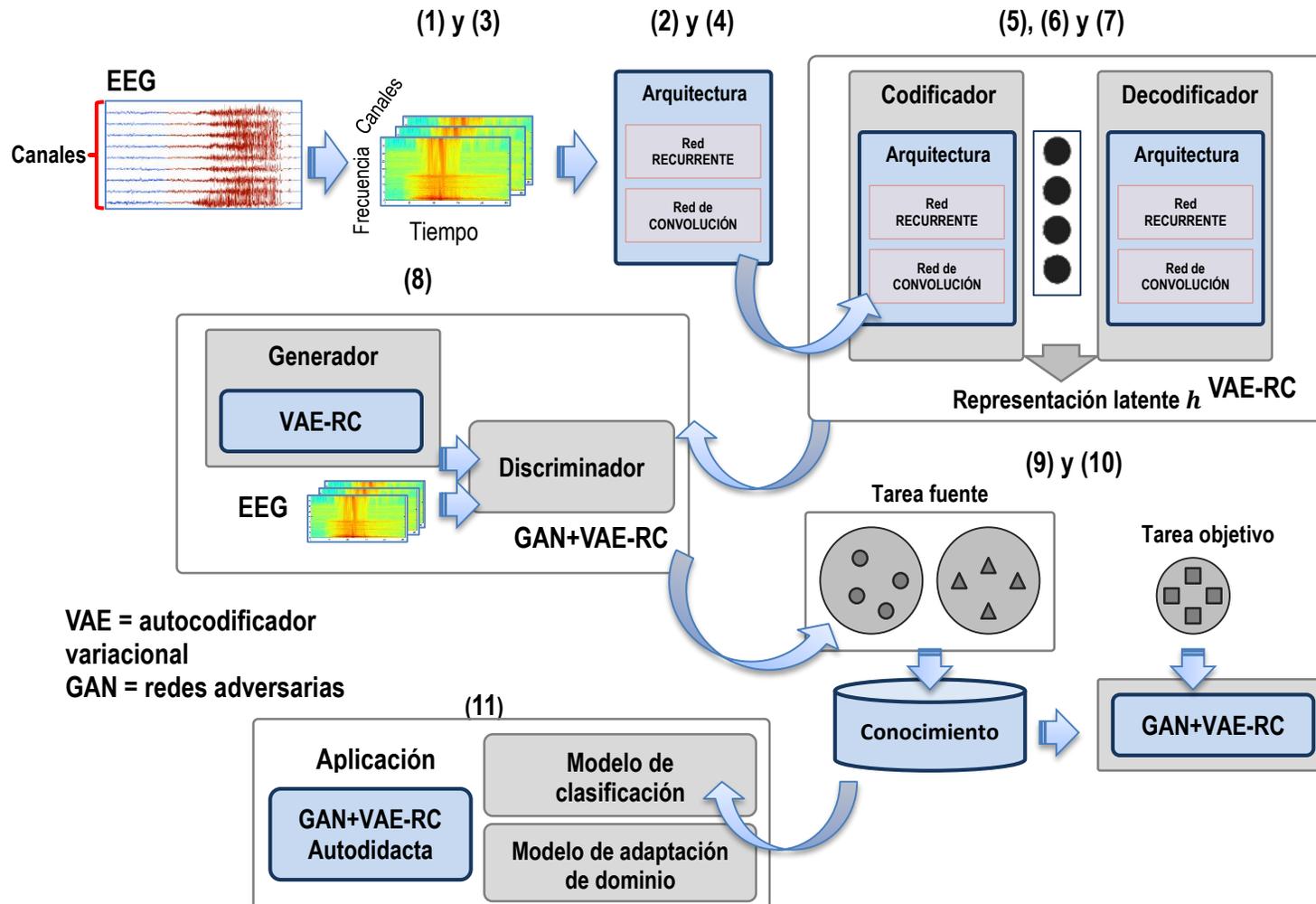


Figura 5.2. Modelo esperado para aprender características robustas en espacio, tiempo y frecuencia.

2. Análisis y estudio de los distintos modelos de redes neuronales profundas, disponibles actualmente para el aprendizaje de representaciones en señales de EEG.
 - a) En esta etapa se estudiarán las diferentes arquitecturas de redes neuronales profundas de convolución, recurrentes y multi-conectadas.
 - b) Análisis y estudio de algoritmos de entrenamiento y regularización.

3. Pre-procesamiento de las señales de EEG para obtener una representación inicial basada en un tensor de datos con la información en espacio, tiempo y frecuencia.
 - a) Las técnicas seleccionadas del estudio realizado en el paso (1) serán aplicadas.

4. Propuesta de una arquitectura basada en redes neuronales profundas para aprender una representación de características en espacio, tiempo y frecuencia, de las señales de EEG.
 - a) Proponer una arquitectura basada en una red de convolución, para el aprendizaje de características en espacio y frecuencia.
 - b) Integrar a la arquitectura propuesta en (4.a), una red recurrente para el aprendizaje de características temporales. Modificar la arquitectura como sea requerida.
 - c) Evaluar la arquitectura propuesta usando un enfoque de aprendizaje supervisado.

5. Análisis y estudio de métodos de codificación/decodificación de redes neuronales que servirán para el diseño y desarrollo de modelos generativos.
 - a) Esta etapa se centrará en el estudio de las diferentes arquitecturas de codificación/decodificación de redes recurrentes, convolución, multi-conectadas y multi-tareas reportadas en la literatura.

6. Análisis y estudio de los diferentes modelos generativos profundos reportados en la literatura que permitan modelar de manera probabilista la variabilidad en espacio, tiempo y frecuencia, de las señales de EEG.
 - a) El estudio se enfocará en dos tipos de modelos generativos: 1) Autocodificadores Variacionales (VAEs) y 2) Redes Adversarias Generativas (GANs).

7. Propuesta de modelo generativo basado en VAEs.
 - a) La arquitectura del paso (4) servirá como base de una arquitectura codificadora, mientras que la arquitectura decodificadora será desarrollada con base el estudio del paso (5). El resultado esperado es un autocodificador variacional recurrente-convolucional (VAE-RC) basado en aprendizaje no supervisado, el cual permitirá identificar los factores de variación en los datos y modelar sus relaciones. Los factores irrelevantes serán descartados para proveer de una mayor robustez a las características aprendidas.
 - b) Extensión y evaluación del modelo VAE-RC a un enfoque semi-supervisado.

8. Extensión del modelo generativo VAE-RC en combinación con GANs.
 - a) El modelo VAE-RC del paso (7) será utilizado junto a un entrenamiento por adversarios.
 - b) Extensión y evaluación del modelo propuesto basado en un enfoque semi-supervisado.

9. Análisis y estudio de los métodos de aprendizaje de transferencia reportados en la literatura para la incorporación de datos de EEG con diferente distribución de probabilidad.
 - a) El marco de trabajo que se explorará está basado en aprendizaje autodidacta, el cual es un tipo de aprendizaje de transferencia de datos no etiquetados.

10. Extensión del modelo generativo (VAE-RC+GAN) de los pasos (6)-(8) a un modelo basado en aprendizaje autodidacta.
 - a) Un modelo será propuesto con base en el análisis del paso (9) y el mejor modelo generativo (VAE-RC+GAN) resultante de los pasos (6)-(8).

11. Aplicación y diseño experimental para la validación del modelo generativo autodidacta propuesto.
 - a) El modelo generativo autodidacta será utilizado bajo dos escenarios de clasificación para una evaluación inter-sujetos:
 - i) Modelo de clasificación genérico.
 - ii) Modelo de adaptación de dominio.

 - b) Obtención de datos de EEG:
 - i) Conjuntos de datos relacionados con carga cognitiva:
 - (1) Bashivan et al. (2014) [10]. Conjunto de datos con un total de 2,670 ejemplos de 3.5 segundos cada uno, extraídos de 13 sujetos. El experimento aplicado a cada sujeto consistió en mostrarle un arreglo de caracteres, seguido de un carácter de prueba. El sujeto debía decidir si el carácter de prueba se encontraba en el arreglo inicial. Los datos fueron extraídos usando un gorro de 64 electrodos a una frecuencia de muestreo de 500 Hz. En total se indujeron cuatro niveles de carga cognitiva (clases), variando el número de caracteres en el arreglo en 2, 4, 6 y 8. La tarea de clasificación consiste en diferenciar entre los cuatro niveles de carga a través de las señales de EEG.

 - (2) Onton et al. (2005) [118]. Conjunto de datos extraído de 23 sujetos, con un total de 100 a 150 ejemplos por sujeto. El experimento consistió en mostrar a cada sujeto una secuencia de 8 letras consonantes. Entre éstas, 3, 5 y 7 (coloreadas en negro) fueron memorizadas, mientras el resto 5, 3 y 1 (coloreadas en verde), eran ignoradas; la duración del estímulo fue de 1.2 segundos. En seguida, una cruz fue mostrada en pantalla durante un periodo de mantenimiento de 2 a 4 segundos. Finalmente, a cada sujeto se le mostró una letra de prueba, de la cual debía decidir si ésta se encontraba o no en la secuencia de caracteres original. Los datos fueron extraídos de un gorro de 71 electrodos a una frecuencia de muestreo de 250 Hz. En total se indujeron tres niveles de carga cognitiva, basados en el número de letras consonantes a memorizar (3, 5 y 7).

 - (3) Estudio y análisis sobre la posibilidad de recolectar datos de carga cognitiva para generar una base de datos propia.

- ii) Obtener otras bases de datos de otros dominios que apoyan al aprendizaje basado en transferencia. Algunos datos iniciales a considerar son: (1) Conjunto de tareas de atención visual¹, (2) Conjunto de datos de interfaces humano-computadora^{2 3}.
- c) Diseño experimental.
- i) Unidades experimentales:
 - (1) U1: Conjunto de datos Bashivan et al. (2014).
 - (2) U2: Conjunto de datos de Onton et al. (2005).
 - ii) Posibles fuentes de sesgo: librerías de aprendizaje profundo, equipo de cómputo, el responsable desarrollador del modelo.
 - iii) Experimento **E1**.
 - (1) Pregunta **Q1**.
 - (2) Hipótesis de investigación: La mejora estructural de una *arquitectura de red profunda*, basada en CNN, RNN y *Fully connected* (FNN), puede aprender una representación de características sobre múltiples sujetos, acorde a las propiedades en espacio, tiempo y frecuencia de las señales de EEG, tal que la representación aprendida mejora la eficacia en la clasificación que los trabajos en el estado del arte.
 - (3) Racional: se espera una arquitectura de red profunda, basada en una mejora estructural, para aprender una representación en espacio, tiempo y frecuencia.
 - (4) Objetivo experimental. Validar la arquitectura propuesta con respecto a los resultados alcanzados por la mejor arquitectura de Bashivan et al. (2016) [32].
 - La arquitectura propuesta será evaluada en tareas de clasificación usando un enfoque de aprendizaje supervisado.
 - (5) Tipos de validación:
 - (a) Interna: *Leave One Subject Out Cross Validation*.
 - (b) Concurrente: Comparación con el trabajo de Bashivan et al. (2016) [32].
 - (6) Variables independientes:
 - (a) F1: Arquitectura propuesta (Nivel 1).
 - (b) F2: Arquitectura Bashivan et al. (2016) (Nivel 1).
 - (7) Variable dependiente:
 - (a) C: Representación aprendida.
 - (8) Tratamiento sobre unidades experimentales: F1-U1, F2-U1, F1-U2, F2-U2.
 - i) Experimento **E2**.
 - (1) Pregunta **Q2**.
 - (2) Hipótesis de investigación: Un *modelo generativo profundo*, basado en VAEs y GANs, permite modelar la variabilidad entre sujetos en espacio, tiempo y frecuencia, de los datos de EEG, tal que la representación aprendida permite mejorar la eficacia en la clasificación que trabajos en el estado del arte.
 - (3) Racional: se espera un modelo generativo que permita capturar los diferentes factores de variación y modele las relaciones entre estos factores entre los múltiples sujetos. Los factores poco informativos serán descartados para proveer de una mayor robustez a las características aprendidas.

¹ <http://headit.ucsd.edu/studies/2747adbc-385a-11e3-b29d-0050563f2612/description>

² <https://www.physionet.org/physiobank/database/erpbc1/>

³ <https://www.physionet.org/physiobank/database/eegmmidb/>

- (4) Objetivo experimental. Validar el modelo generativo propuesto con respecto a los resultados alcanzados por los trabajos en el estado del arte.
 - El modelo propuesto será evaluado en tareas de clasificación usando un enfoque de aprendizaje semi-supervisado.
- (5) Tipos de validación:
 - (a) Interna: *Leave One Subject Out Cross Validation*.
 - (b) Concurrente: para el mejor de nuestro conocimiento, no existe un modelo generativo que dé cumplimiento a nuestra hipótesis de investigación. Por tal motivo, en esta etapa se hará una comparación con el trabajo de Chung et al. (2015) [119], el cual propone un modelo generativo para el procesamiento de datos secuenciales. No obstante, la principal desventaja de este trabajo es el de procesar una secuencia a la vez y no partir del supuesto que se clasificarán datos de una distribución de probabilidad diferente. Como se mencionó, en el caso de nuestro trabajo, se pretende procesar múltiples secuencias simultáneamente, obtenidas de n electrodos, modelando la variabilidad sobre múltiples sujetos en espacio, tiempo y frecuencia.
- (6) Variables independientes:
 - (a) F1: Modelo generativo propuesto (Nivel 1).
 - (b) F2: Modelo generativo de Chung et al. (2015) (Nivel 1).
- (7) Variable dependiente:
 - (a) C: Representación aprendida.
- (8) Tratamiento de sobre unidades experimentales: F1-U1, F2-U1, F1-U2, F2-U2.

ii) Experimento **E3**.

- (1) Pregunta **Q3**.
- (2) Hipótesis de investigación: Un modelo generativo, basado en *aprendizaje autodidacta*, permite mejorar el aprendizaje de representaciones robustas de datos de EEG sobre múltiples sujetos en espacio, tiempo y frecuencia, tal que la representación aprendida mejora la eficacia en la clasificación que trabajos en el estado del arte.
- (3) Racional: se espera mejorar el aprendizaje de representaciones robustas en espacio, tiempo y frecuencia sobre múltiples sujetos, incorporando información de datos de EEG con diferentes distribuciones de probabilidad.
- (4) Objetivo experimental 1. Validar la estrategia de aprendizaje autodidacta, incorporada al modelo generativo, con respecto a otras estrategias reportadas en el estado del arte.
 - (a) Validación usando un modelo clasificador:
 - (i) Interna: *Leave One Subject Out Cross Validation*.
 - (ii) Concurrente: comparación con Kingma et al. (2014) [120].
 - (b) Variables independientes:
 - 1. F1: Estrategia autodidacta propuesta.
 - 2. F2: Estrategia propuesta por Kingma et al. (2014).
 - (c) Variable dependiente:
 - (i) C: Representación aprendida.
 - (d) Tratamiento de sobre unidades experimentales: F1-U1, F2-U1, F1-U2, F2-U2.

- (5) Objetivo experimental 2. Validar el modelo de aprendizaje autodidacta con respecto a otros trabajos reportados en el estado del arte.
- Este experimento busca dar aplicación al modelo generativo autodidacta, el cual será evaluado bajo dos contextos: (1) un modelo de clasificación y (2) un modelo de adaptación de dominio.
- (a) Validación usando un modelo clasificador:
- (i) Interna: *Leave One Subject Out Cross Validation*.
 - (ii) Concurrente: Comparación con el modelo generativo propuesto en nuestro segundo objetivo de investigación.
 - (iii) Variables independientes:
 1. F1: Modelo generativo autodidacta propuesto.
 2. F2: Modelo generativo propuesto en nuestro segundo objetivo de investigación.
 - (iv) Variable dependiente:
 1. C: Representación aprendida.
 - (v) Tratamiento de sobre unidades experimentales: F1-U1, F2-U1, F1-U2, F2-U2.
- (b) Validación usando un modelo de adaptación de dominio:
- (i) Interna: *Leave One Subject Out Cross Validation*.
 - (ii) Concurrente: comparación con el trabajo de Purushotham et al. (2017) [121]. Este trabajo es una extensión de Chung et al. (2015) para un modelo de adaptación de dominio. El trabajo de Purushotham et al. (2017) tiene las mismas desventajas mencionadas en nuestro segundo experimento (E2).
 - (iii) Variables independientes:
 1. F1: Modelo generativo autodidacta propuesto usando adaptación de dominio.
 2. F2: Modelo propuesto por Purushotham et al. (2017).
 - (iv) Variable dependiente:
 1. C: Representación aprendida.
 - (v) Tratamiento de sobre unidades experimentales: F1-U1, F2-U1, F1-U2, F2-U2.
- iii) Para nuestro diseño experimental se determinará el número de muestras y el cálculo del poder estadístico.
- iv) Las métricas de evaluación para los tres experimentos son: Exactitud, *F-measure* y Área bajo la curva.

5.1. Cronograma de actividades

El cronograma de actividades propuesto se muestra a continuación. Las tareas corresponden a los puntos principales de la metodología propuesta.

Tabla 5-1. Cronograma de actividades de las tareas a realizar por bimestres. Las celdas coloreadas en azul indican que la tarea ha sido realizada, mientras que las celdas en color gris indican las actividades por realizar.

Actividad	2016						2017						2018						2019					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Estudio crítico del estado del arte	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Elaboración de propuesta de investigación	█	█	█	█	█	█	█	█																
Defensa de la propuesta de investigación							█																	
Preparación de reporte técnico.									█	█														
(1) Estudio de técnicas de procesamiento de señales				█																				
(2) Análisis de redes neuronales profundas				█	█	█																		
(3) Pre-procesamiento de las señales de EEG				█																				
(4) Proponer una arquitectura de red neuronal profunda que capture información en espacio, tiempo y frecuencia.					█	█	█	█																
(5) Análisis de los métodos de codificación/decodificación de redes profundas							█																	
(6) Análisis de modelos generativos profundos.							█		█	█														
(7) Proponer un modelo generativo basado en autocodificadores variacionales (VAEs)									█	█	█	█												
(8) Extender el modelo generativo propuesto a una combinación de VAEs y GANs.													█	█	█	█	█							
(9) Estudiar métodos de aprendizaje de transferencia.															█	█	█							
(10) Extender el modelo generativo propuesto a un enfoque de aprendizaje autodidacta																█	█	█	█					
(11) Aplicación del modelo							█																	
Redacción de artículos para conferencias.									█						█						█			
Redacción de artículos para revistas indizadas.															█	█	█							
Escritura de tesis doctoral.																					█	█		
Revisión por parte del comité doctoral.																						█	█	
Realizar correcciones de tesis.																							█	
Defensa final de tesis de doctorado.																							█	

5.2. Alcances y limitaciones

La presente propuesta de investigación se centra en proponer un marco de trabajo para aprender características robustas de datos de EEG sobre múltiples sujetos. La evaluación de las características aprendidas se encuentran bajo el contexto de determinar los niveles de carga cognitiva en sujetos para una tarea asignada. Dos aplicaciones serán desarrolladas usando el modelo de aprendizaje de representaciones final, un modelo de clasificación y un modelo de adaptación de dominio.

Como se mencionó previamente, el modelo propuesto pretende mejorar la eficacia en la clasificación de señales de EEG. Por tal motivo, este tipo de modelos están pensados para sistemas de asistencia y no como parte de soluciones de sistemas interpretables.

5.3. Plan de publicaciones

- Arquitectura de red profunda que permite aprender representaciones de datos de EEG en espacio, tiempo y frecuencia. Conferencia: *2018 IEEE World Congress on Computational Intelligence*, fecha límite 15 diciembre, 2017.
- Modelo generativo profundo que permite modelar la variabilidad entre sujetos en espacio, tiempo y frecuencia de las señales de EEG. Revista: *Neurocomputing* – Elsevier.
- Modelos de aplicación del modelo generativo profundo basado en aprendizaje autodidacta que permite incorporar datos no etiquetados provenientes de diferentes distribuciones de probabilidad. Revista: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

6. Resultados preliminares

En esta sección se presentan los resultados preliminares alcanzados a la fecha, los cuales se enfocan en aprender una representación en espacio, tiempo y frecuencia, de las señales de EEG. Específicamente, se presenta una evaluación inter-sujetos de una primera arquitectura de red neuronal profunda sobre un conjunto de datos para determinar el nivel de carga cognitiva, utilizando un esquema de aprendizaje supervisado. Estos resultados forman parte del primer objetivo específico de la presente propuesta de investigación y los cuatro primeros pasos de la metodología.

6.1. Arquitectura de red neuronal profunda para el aprendizaje de características en espacio, tiempo y frecuencia, de señales de EEG

Como se mencionó en la sección 3, existen diversos trabajos basados en redes neuronales profundas para el aprendizaje de representaciones en señales de EEG. Particularmente, hay una especial atención en aquellos trabajos que utilizan una arquitectura compuesta de múltiples bloques basados en diferentes tipos de redes neuronales. Bajo esta idea, se propone una primera arquitectura (ver Figura 6.1), la cual consiste de una combinación de bloques compuestos por CNN, GRU, y FNN, que son una red de convolución, una red recurrente y una red multi-conectada, respectivamente. La arquitectura propuesta recibe como entrada un tensor de datos de EEG, basado en una secuencia de imágenes con la información espacial-espectral-temporal.

6.1.1. Descripción del conjunto de datos de carga cognitiva

Quince estudiantes (7 hombres) de la Universidad de Memphis participaron en un estudio de la carga cognitiva [10]. Las edades de los participantes se encontraban entre 24 y 33 años (28 ± 3 años), todos diestros; ninguno reportó antecedentes de trastornos visuales o neuro-psiquiátricos. Cada participante completó 240 ensayos de una tarea visual, la cual consistió en memorizar un conjunto de caracteres en inglés y responder si un carácter de prueba, aleatoriamente presentado, estaba dentro del conjunto memorizado o no. Específicamente, en 0.5 segundos se le muestra un arreglo de caracteres a cada participante, luego al término de este tiempo se le muestra una pantalla con una cruz durante 3 segundos; seguido, un carácter de prueba es mostrado para que en un plazo máximo de 5 segundos el participante decida si este carácter se encuentra dentro del arreglo de caracteres. El número de caracteres presentados durante cada ensayo de la tarea fue variado entre 2, 4, 6, y 8 elementos, con el propósito de inducir cuatro niveles de carga cognitiva.

El registro de EEG para el experimento fue realizado usando un gorro de 64 electrodos, colocados en localizaciones estándar 10-10. Los datos fueron digitalizados a una frecuencia de muestreo de 500 Hz. Las porciones de EEG con artefactos oculares fueron descartadas usando análisis de componentes principales. Los datos registrados fueron sub-muestreados a 250 Hz para su análisis. De acuerdo al estudio, los datos de dos de los participantes fueron excluidos del conjunto de datos por exceso de ruido y artefactos. En el conjunto de datos final se incluyeron únicamente los ensayos que fueron respondidos correctamente, por lo que en total se disponen de 2,670 ensayos de 3.5 segundos para 13 participantes, con cuatro niveles de carga cognitiva (ver Tabla 6-1).

Tabla 6-1. Conjunto de datos de carga cognitiva. La tabla muestra el total de ensayos por sujeto y niveles de carga cognitiva (clases) inducidos.

Sujetos	Niveles de carga cognitiva (clases)				Total
	1	2	3	4	
S1	57	48	47	33	185
S2	64	54	49	45	212
S3	55	54	47	43	199
S4	60	57	45	39	201
S5	59	56	45	36	196
S6	58	54	47	42	201
S7	56	53	42	42	193
S8	58	58	50	36	202
S9	57	58	51	44	210
S10	59	58	57	51	225
S11	57	58	53	49	217
S12	58	56	52	43	209
S13	60	59	54	47	220

6.1.2. Pre-procesamiento: extracción de imágenes de datos de EEG

El pre-procesamiento utilizado en esta etapa es el reportado en [32]. Cada ensayo de 3.5 segundos se divide en ventanas de 0.5 segundos, aplicando para cada ventana y cada electrodo la transformada rápida de Fourier, con el propósito de calcular el poder espectral promedio en las bandas de frecuencia

alpha, *beta* y *theta*. Estas bandas fueron seleccionadas debido a evidencias de su rol en varias funciones cognitivas (ver sección 2.1) [21, 50].

Los electrodos del EEG distribuidos sobre el cuero cabelludo en un espacio de tres dimensiones forman mapas de activación con la información de frecuencia. Estos mapas de activación son transformados a imágenes de dos dimensiones usando la proyección equidistante Acimutal, en inglés *Azimuthal Equidistant Projection* (AEP) [122]. Las proyecciones producidas por AEP forman un plano tangente a cualquier polo, el ecuador o cualquier punto intermedio. Las distancias del centro de la proyección a cualquier punto son preservados. En el caso específico del EEG, la forma del gorro usado en la cabeza de un ser humano puede ser aproximada por una esfera y el mismo método podría utilizarse para calcular la proyección de las posiciones de los electrodos en una superficie 2D, el cual es tangente a un punto central superior de la cabeza. La imagen resultante representa la distribución espacial de las actividades cerebrales sobre la corteza. Posteriormente, se aplicó *Clough-Tocher scheme* [123] para interpolar las mediciones del poder espectral sobre el cuero cabelludo y estimar los valores de los electrodos en una malla de 32×32 . Este procedimiento es aplicado para cada banda de frecuencia, produciendo tres mapas de actividad topográficos para cada ventana de 0.5 segundos. Estos tres mapas son agrupados para formar una imagen con tres canales (colores). Así, de un ensayo de 3.5 segundos se forma una secuencia de siete imágenes de 32×32 con tres canales.

6.1.3. CGRUD: arquitectura de red profunda propuesta

La arquitectura *Convolutional-Gated Recurrent Unit-Deep* (CGRUD) utiliza una red de convolución para aprender una representación en espacio y frecuencia, una red recurrente para aprender características temporales y una red multi-conectada para discriminar las características aprendidas por niveles inferiores. El diseño de la arquitectura propuesta se muestra en la Figura 6.1.

CGRUD procesa la información en espacio-frecuencia de cada imagen de EEG utilizando una arquitectura de red de convolución en tres dimensiones. Las dos primeras capas de convolución emplean una función de activación basada en tangente hiperbólica. Esta función fue seleccionada debido a que en estudios previos [124], ha mostrado buenos resultados para el modelado de señales de EEG. La segunda capa, a diferencia de la primera, incrementa el número de pasos de convolución con el propósito de reducir el tamaño de los mapas de características de salida. En la salida de cada capa de convolución se usa una normalización por lotes, la cual agrega un método de regularización implícito más [58]. Por último, se utiliza un bloque residual con la idea de dividir y contribuir a la propagación del gradiente en múltiples rutas [57]. La arquitectura de convolución propuesta se asemeja a un ensamble de redes de convolución para el procesamiento simultáneo en espacio-frecuencia de cada imagen dentro de la secuencia de imágenes.

Para incorporar un aprendizaje de características temporales sobre la secuencia de imágenes, se apila una arquitectura de red recurrente basada en unidades GRU. La selección de este tipo de unidades recurrentes sobre otras como LSTM, se deriva del estudio presentado en [125], donde se menciona que no se tiene claro un dominio para el cual las unidades LSTM aventajen a las unidades GRU. Sin embargo, las unidades GRU se caracterizan por obtener un mejor rendimiento cuando se dispone de pocos datos para entrenar.

Finalmente, las características aprendidas por la capa recurrente son alimentadas a dos capas multi-conectadas. Como se describe en [126], estas capas son apropiadas en un orden más alto de una arquitectura, pues producen una representación de características de más alto nivel, lo que permite que las clases sean más fáciles de discriminar.

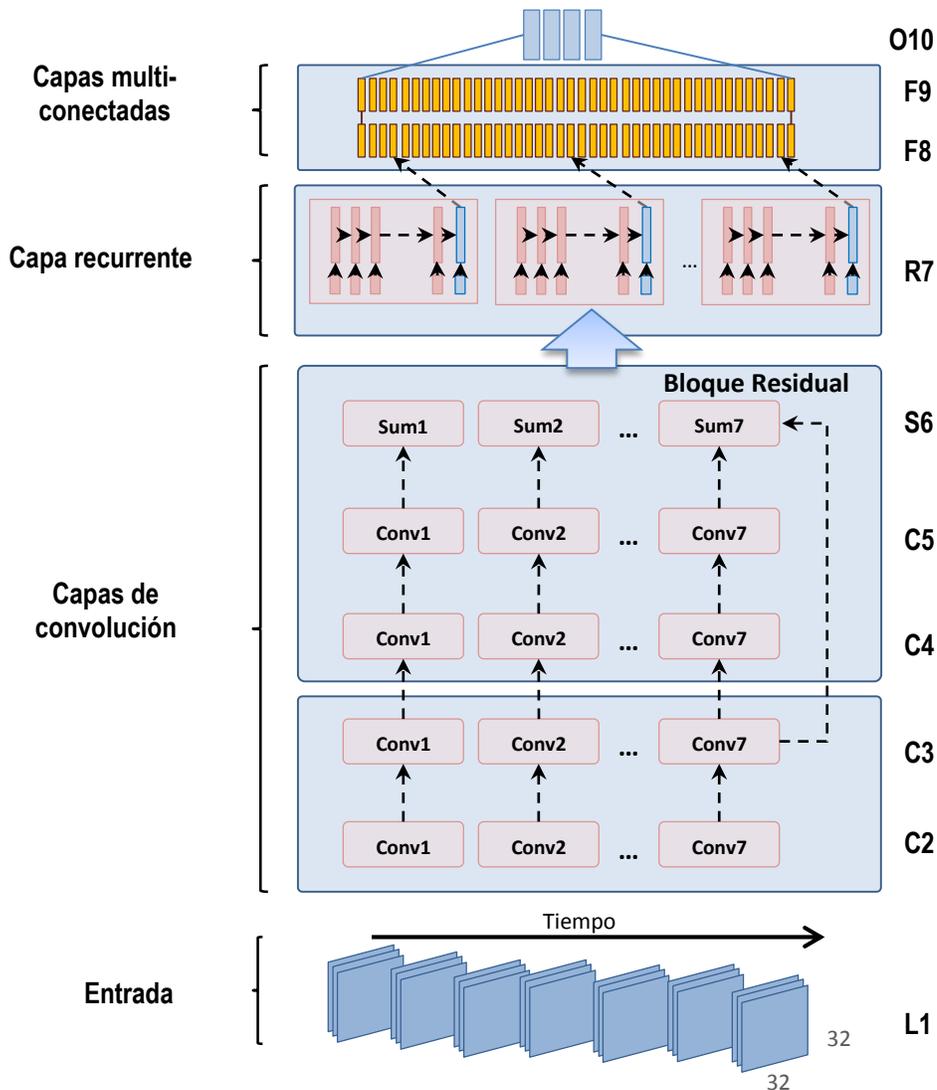


Figura 6.1. Arquitectura CGRUD.

A continuación se describe con detalle cada uno de los componentes de CGRUD:

- 1) L1: capa de entrada. La arquitectura recibe como entrada un conjunto de ejemplos, donde cada ejemplo es un tensor de datos con una secuencia de imágenes de la forma “No. de imágenes en la secuencia” × “No. canales de la imagen” × “ancho de imagen” × “alto de imagen”.
- 2) Arquitectura de red de convolución en 3-dimensiones:
 - a. C2: capa de convolución. Esta capa propone filtrar la señal en espacio y frecuencia usando 16 filtros, dejando sin modificar la información temporal. El tamaño de cada filtro de convolución es de $1 \times 3 \times 3$ con pasos de 1, utilizando como función de activación a la tangente hiperbólica. En la salida de esta capa se utiliza una normalización por lotes, definida en [58]. Los pesos son inicializados usando una distribución uniforme.

- b. C3: capa de convolución. Se utiliza una capa similar a la anterior con la única diferencia que el tamaño de los pasos de convolución es puesto a dos. El propósito de este cambio consiste en reducir el tamaño de los mapas de características de salida.
- c. B1: Bloque de residuo. Este bloque es el propuesto en [56] y está formado por un módulo de residuo f_i , además de un salto de conexión por pasar f_i (ver Figura 2.2 de la sección 2). El módulo de residuo utilizado consiste en:
 - i. C4: capa de convolución: Esta capa filtra la señal en el dominio del espacio por usar 16 filtros con un tamaño de $1 \times 3 \times 3$, usando una unidad de lineal rectificadas (ReLU). En la salida de esta capa se utiliza una normalización por lotes. Los pesos son inicializados usando una distribución normal.
 - ii. C5: capa de convolución: La misma configuración que la capa previa con la única diferencia que no se utiliza una función de activación en la salida de cada filtro.
 - iii. S6: Combinación de capas: La salida de la capa C3 es sumada a la salida de la capa C5. Posteriormente, una función de activación ReLU es aplicada.
- 3) R7: capa recurrente. Esta capa contiene 128 unidades GRU para capturar la información temporal. Un regularizador *Dropout* con una probabilidad de 0.5 se utiliza en la salida de esta capa.
- 4) F8: capa multi-conectada. Cada neurona en F8 es conectada a cada neurona de R7. La capa está compuesta de 100 neuronas con funciones de activación basadas en ReLU y una función de inicialización de pesos basada en una distribución uniforme. Similar a la capa anterior, se utiliza el regularizador *Dropout* para evitar el sobreajuste.
- 5) F9: capa multi-conectada. Cada neurona en F9 es conectada a cada neurona en F8. El mismo número de neuronas y el mismo regularizador utilizados en F8 se emplean en esta capa.
- 6) O10: clasificador. La salida contiene un número de neuronas basado en el número de clases del conjunto de datos (4 clases). Se utiliza entropía cruzada como función de costo.

6.1.4. Validación de CGRUD: métodos de comparación

Para validar CGRUD se utilizó una validación concurrente, realizando una comparación con la mejor arquitectura propuesta por Bashivan et al. (2016) [32]. Esta arquitectura aprende una representación en espacio, tiempo y frecuencia, y se describe a continuación:

CNN+MIX: en una primera etapa se propone utilizar un ensamble de arquitecturas de redes de convolución para el aprendizaje de características en espacio-frecuencia. Cada arquitectura sigue el diseño de la red VGG [127], utilizada en la competencia de clasificación de imágenes “*Imagenet*”. Posteriormente, los autores apilan sobre el ensamble de redes de convolución, una red de convolución en una dimensión y una capa basada en unidades recurrentes LSTM, para incorporar características temporales.

De manera adicional, los siguientes métodos de línea base fueron utilizados para comparar la red propuesta:

1. **PCA+FFN.** La secuencia de imágenes se convierte a un vector de características. Luego, PCA es utilizado para reducir la dimensionalidad del vector. Para nuestro caso en específico,

se utilizaron 50 componentes. Este proceso es repetido para cada ejemplo de entrenamiento. Finalmente, el nuevo conjunto de datos es utilizado para entrenar una red de multiconectada *Feed Forward Network* (FFN) con una sola capa oculta compuesta de 128 unidades.

2. **Random Forest.** De manera similar a PCA+FNN, cada secuencia de imágenes es transformada a un vector de características. Luego, el clasificador *Random Forest* es entrenado con 3,000 árboles.

6.1.5. Diseño del experimento

La hipótesis experimental es: *una mejora en el diseño estructural de una arquitectura de red profunda que incorpora información espacial, temporal y espectral sobre múltiples sujetos, permite mejorar la eficacia de la clasificación que los trabajos reportados en el estado del arte.* El objetivo del experimento consiste en: *validar la arquitectura CGRUD con los resultados alcanzados por la mejor arquitectura reportada en Bashivan et al. (2016) y dos métodos de línea base.*

La validación interna a utilizar está basada en una replicación cruzada entre sujetos: *Leave Subject-Out Cross Validation* (LSOCV). Aquí, los datos de un sujeto son usados como conjunto de prueba mientras el resto de sujetos son utilizados como conjunto de entrenamiento. Adicionalmente, el tamaño del conjunto de entrenamiento es redimensionado con el propósito de extraer de éste, un conjunto de validación con un número de ensayos igual al del conjunto de prueba. En este último conjunto se distribuyeron aleatoriamente los ensayos con una misma proporción de las clases (4 niveles de carga cognitiva). Este procedimiento es repetido para todos los sujetos.

El entrenamiento de CGRUD, CNN+MIX y FNN, se realizó usando una razón de aprendizaje de 0.001, un tamaño de lote de 20, el algoritmo de entrenamiento *Adam* y una función de costo basada en entropía cruzada. El algoritmo *Adam* se seleccionó para nuestros experimentos, pues de acuerdo a [128] ha mostrado converger mejor que otros algoritmos reportados para el entrenamiento de redes neuronales profundas. Las arquitecturas son entrenadas hasta alcanzar un máximo de 30 épocas, seleccionando aquel modelo de red con el mayor valor de exactitud obtenido sobre el conjunto de validación. Los resultados alcanzados por cada uno de los modelos comparados fueron evaluados en términos de Exactitud, *F-measure* y Área bajo la curva ROC (AUC).

6.1.6. Resultados

Los resultados de clasificación obtenidos para las tres medidas de evaluación utilizadas en nuestro diseño experimental se muestran en las Tablas 6-2, 6-3 y 6-4. Los valores representan la media±la desviación estándar utilizando LSOCV. En las Tablas 6-2 y 6-3 se observa que la arquitectura CGRUD obtiene los mejores resultados de clasificación en Exactitud y *F-measure* promedio, con valores de 0.9071 ± 0.124 y 0.8963 ± 0.148 respectivamente. Estos resultados son alcanzados con una media ROC de área bajo la curva de 0.9729.

Tabla 6-2. Resultados de clasificación ($\mu \pm \sigma$) en exactitud de los métodos comparados. Los resultados más altos por columna se indican en negrita.

Sujetos	Exactitud			
	PCA+FNN	Random Forest	CNN+MIX (Bashivan et al. 2016)	CGRUD (modelo propuesto)
S1	0.5567	0.6594	0.6162	0.7892
S2	0.6509	0.7594	0.7453	0.7264
S3	0.8492	0.9597	0.9347	0.9146
S4	0.9303	0.9950	0.9950	1.0000
S5	0.9693	1.0000	1.0000	0.9898
S6	0.9601	0.9850	0.9900	1.0000
S7	0.9015	0.9792	0.9948	1.0000
S8	0.9950	0.9900	1.0000	1.0000
S9	0.9714	0.9809	0.9910	0.9905
S10	0.9200	0.9822	0.9911	1.0000
S11	0.8110	0.9585	0.9860	0.9862
S12	0.6650	0.6794	0.7460	0.7273
S13	0.4045	0.4863	0.4636	0.6682
Prom.	0.8142±0.172	0.8781±0.164	0.8811±0.172	0.9071±0.124

Tabla 6-3. Resultados de clasificación ($\mu \pm \sigma$) en F-measure de los métodos comparados. Los resultados más altos por columna se indican en negrita.

Sujetos	F-Measure			
	PCA+FNN	Random Forest	CNN+MIX (Bashivan et al. 2016)	CGRUD (modelo propuesto)
S1	0.5099	0.5995	0.5588	0.7833
S2	0.6266	0.7205	0.7092	0.6907
S3	0.8474	0.9597	0.9348	0.9139
S4	0.9275	0.9950	0.9950	1.0000
S5	0.9692	1.0000	1.0000	0.9897
S6	0.9601	0.9850	0.9900	1.0000
S7	0.9021	0.9792	0.9948	1.0000
S8	0.9950	0.9900	1.0000	1.0000
S9	0.9714	0.9810	0.9905	0.9905
S10	0.9185	0.9822	0.9911	1.0000
S11	0.8090	0.9587	0.9862	0.9862
S12	0.6343	0.6437	0.6973	0.7090
S13	0.3511	0.4153	0.3781	0.5889
Prom.	0.8017±0.206	0.8623±0.197	0.8635±0.208	0.8963±0.148

Tabla 6-4. Resultados de clasificación ($\mu \pm \sigma$) en AUC de los métodos comparados. Los resultados más altos por columna se indican en negrita.

Sujetos	AUC			
	PCA+FNN	Random Forest	CNN+MIX (Bashivan et al. 2016)	CGRUD (modelo propuesto)
S1	0.8210	0.8725	0.8620	0.8661
S2	0.9088	0.9353	0.9861	0.9530
S3	0.9546	0.9983	0.9978	0.9937
S4	0.9886	1.0000	1.0000	1.0000
S5	0.9951	1.0000	1.0000	1.0000
S6	0.9860	0.9850	1.0000	1.0000
S7	0.9763	0.9992	1.0000	1.0000
S8	0.9972	1.0000	1.0000	1.0000
S9	0.9977	0.9997	0.9999	1.0000
S10	0.9817	0.9994	1.0000	1.0000
S11	0.9411	0.9987	1.0000	0.9997
S12	0.9371	0.9337	0.9939	0.9716
S13	0.8135	0.7685	0.8626	0.8634
Prom.	0.9461±0.063	0.9608±0.069	0.9771±0.051	0.9729±0.050

Un análisis estadístico basado en la prueba de Friedman con un nivel de significancia de 0.05, mostró que no existe diferencia estadística significativa entre CGRUD y CNN+MIX para las medidas de Exactitud y *F-Measure*; encontrando sólo diferencia estadística significativa entre CGRUD y PCA+FNN. Sin embargo, la desviación estándar obtenida por CGRUD, tal y como se muestra en las Tablas 6-2 y 6-3, se reduce en comparación con los demás métodos. Cabe recordar que la desviación estándar es un índice de precisión, lo cual evidencia la factibilidad de un incremento en la eficacia de la clasificación mediante una mejora estructural de una red profunda.

6.1.7. Discusión

El objetivo experimental planteado consistió en validar que un mejor diseño de una arquitectura de red profunda en espacio, tiempo y frecuencia de datos de EEG sobre múltiples sujetos, puede mejorar la eficacia de clasificación de los diferentes niveles de carga cognitiva. Nuestros resultados muestran que CGRUD supera en clasificación promedio y una menor desviación estándar a los resultados alcanzados por CNN+MIX; arquitectura con el mismo propósito. CNN+MIX reporta en la literatura una exactitud de 0.9110 ± 0.148 para una evaluación basada en sujetos cruzados. Sin embargo, nuestro estudio experimental muestra que CNN+MIX obtiene una exactitud promedio de 0.8811 ± 0.172 ; un valor inferior al resultado alcanzado por CGRUD que fue de 0.9071 ± 0.124 . El resultado de CNN+MIX podría deberse a una inicialización diferente de los pesos en la red y el uso de estrategias adicionales de entrenamiento, tal como interrupción anticipada. Cabe señalar que para nuestra comparación se utilizó la arquitectura implementada por el autor, así como la misma configuración experimental reportada.

El éxito de nuestra arquitectura para la discriminación de señales de EEG en el estudio de la carga cognitiva, depende de utilizar desarrollos en aprendizaje profundo, originalmente diseñados para otros dominios de aplicación. Primero, con la motivación de incrementar la profundidad de la red neuronal para aprender características más abstractas, se incorporó un bloque residual para los niveles de redes de convolución, además de dos capas multi-conectadas en niveles superiores de la

arquitectura. Un bloque residual descompone el gradiente en dos términos aditivos: un sub-bloque compuesto de capas de convolución (C4 y C5), que propaga la información del gradiente a través de éste, y un salto de conexión que permite propagar directamente el gradiente a capas inferiores [57]. Este bloque puede proporcionar mayor flexibilidad en el número de niveles de la arquitectura propuesta, ante las diferencias existentes entre los datos de los múltiples sujetos. Para el mejor de nuestro conocimiento, esta es la primera vez que se muestra evidencia de la utilidad de este tipo de bloques en el procesamiento de señales de EEG. Por otra parte, para evitar el posible sobreajuste de la red, se adoptan estrategias de inicialización de los pesos, una normalización por lotes, así como el uso del regularizador *dropout* en diferentes niveles de la arquitectura. Cabe mencionar que unidades GRU son utilizadas en la capa recurrente, las cuales utilizan un número menor de parámetros que otras unidades como LSTM [119].

7. Conclusiones

La presente propuesta de investigación se enfoca en el aprendizaje de representaciones de características más robustas para datos de EEG sobre múltiples sujetos, derivado de la variabilidad existente entre los diferentes sujetos en espacio, tiempo y frecuencia, con el propósito de mejorar la eficacia en tareas de clasificación que los trabajos actuales. La evaluación de estos trabajos es realizada en el contexto de carga cognitiva.

Como un resultado preliminar se propuso una primera arquitectura basada en redes neuronales profundas para aprender de datos de EEG sobre múltiples sujetos, una representación en espacio, tiempo y frecuencia. Esta arquitectura, denominada CGRUD, está integrada por la composición de una red de convolución, una red recurrente, y una red multi-conectada, además de la incorporación de diferentes desarrollos en aprendizaje profundo de otros dominios, tal que mejoran el diseño estructural. Los resultados muestran la factibilidad de CGRUD para producir una mejora en la eficacia de la clasificación, al generar una representación que supera en términos de clasificación promedio y una menor desviación estándar, para una evaluación inter-sujetos, a los resultados alcanzados por CNN+MIX y dos métodos de línea base.

Con base en los resultados preliminares presentados, esta investigación se muestra como adecuada y la propuesta muestra ser factible al abordar uno de los objetivos específicos.

8. Referencias

- [1] A. Varsavsky, I. Mareels y M. Cook, *Epileptic Seizures and the EEG: Measurement, Models, Detection and Prediction*, CRC Press, 2016.
- [2] P. Mirowski, D. Madhavan, Y. LeCun y R. Kuzniecky, «Classification of patterns of EEG\ synchronization for seizure prediction,» *Clinical Neurophysiology*, vol. 120, pp. 1927-1940, 2009.
- [3] G. Xun, X. Jia y A. Zhang, «Detecting epileptic seizures with electroencephalogram via a context-learning model,» *BMC Medical Informatics and Decision Making*, vol. 16, pp. 97-109, 2016.
- [4] C. A. Teixeira, B. Direito, M. Bandarabadi, M. L. V. Quyen, M. Valderrama, B. Schelter, A. Schulze-Bonhage, V. Navarro, F. Sales y A. Dourado, «Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients,» *Computer Methods and Programs in Biomedicine*, vol. 114, pp. 324-336, 2014.

- [5] W.-L. Zheng y B.-L. Lu, «Personalizing EEG-Based Affective Models with Transfer Learning,» de *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, {IJCAI} 2016, New York, NY, USA, 9-15 July 2016*, 2016.
- [6] W. Liu, W.-L. Zheng y B.-L. Lu, «Emotion Recognition Using Multimodal Deep Learning,» de *Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16--21, 2016, Proceedings, Part II*, A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee y D. Liu, Edits., Ch, a: Springer International Publishing, 2016, pp. 521-529.
- [7] J. Zhang, Y. Wu, J. Bai y F. Chen, «Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers,» *Transactions of the Institute of Measurement and Control*, p. 0142331215587568, 2015.
- [8] M. Hajinorozi, Z. Mao, T.-P. Jung, C.-T. Lin y Y. Huang, «EEG-based prediction of driver's cognitive performance by deep convolutional neural network,» *Signal Processing: Image Communication*, pp. - , 2016.
- [9] P. P. San, S. H. Ling, R. Chai, Y. Tran, A. Craig y H. T. Nguyen, «EEG-based Driver Fatigue Detection using Hybrid Deep Generic Model,» de *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2016.
- [10] P. Bashivan, G. M. Bidelman y M. Yeasin, «Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity,» *European Journal of Neuroscience*, vol. 40, pp. 3774-3784, 2014.
- [11] J. Li y A. Cichocki, «Deep Learning of Multifractal Attributes from Motor Imagery Induced EEG,» de *Neural Information Processing: 21st International Conference, ICONIP 2014, Kuching, Malaysia, November 3-6, 2014. Proceedings, Part I*, C. K. Loo, K. S. Yap, K. W. Wong, A. Teoh y K. Huang, Edits., Ch, a: Springer International Publishing, 2014, pp. 503-510.
- [12] N. Lu, T. Li, X. Ren y H. Miao, «A Deep Learning Scheme for Motor Imagery Classification based on Restricted Boltzmann Machines,» *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, pp. 1-1, 2016.
- [13] B. T. Jap, S. Lal y P. Fischer, «Comparing combinations of EEG\ activity in train drivers during monotonous driving,» *Expert Systems with Applications*, vol. 38, pp. 996-1003, 2011.
- [14] S. K. L. Lal, A. Craig, P. Boord, L. Kirkup y H. Nguyen, «Development of an algorithm for an EEG-based driver fatigue countermeasure,» *Journal of Safety Research*, vol. 34, pp. 321-328, 2003.
- [15] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet y P. L. Craven, «EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks,» *Aviation, space, and environmental medicine*, vol. 78, pp. B231--B244, 2007.
- [16] P. Zarjam, J. Epps, F. Chen y N. H. Lovell, «Classification of Working Memory Load Using Wavelet Complexity Features of EEG Signals,» de *Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part II*, T. Huang, Z. Zeng, C. Li y C. S. Leung, Edits., Berlin, Heidelber: Springer Berlin Heidelberg, 2012, pp. 692-699.
- [17] N. Kumar y J. Kumar, «Measurement of Cognitive Load in HCI Systems Using EEG Power Spectrum: An Experimental Study,» *Procedia Computer Science*, vol. 84, pp. 70-78, 2016.
- [18] P. Antonenko, F. Paas, R. Grabner y T. van Gog, «Using Electroencephalography to Measure Cognitive Load,» *Educational Psychology Review*, vol. 22, pp. 425-438, 2010.

- [19] A. F. Rabbi, A. Zony, P. de Leon y R. Fazel-Rezai, «Mental workload and task engagement evaluation based on changes in electroencephalogram,» *Biomedical Engineering Letters*, vol. 2, pp. 139-146, 2012.
- [20] P. Zarjam, J. Epps, F. Chen y N. H. Lovell, «Estimating cognitive workload using wavelet entropy-based features during an arithmetic task,» *Computers in Biology and Medicine*, vol. 43, pp. 2186-2195, 2013.
- [21] C. Rottschy, R. Langner, I. Dogan, K. Reetz, A. R. Laird, J. B. Schulz, P. T. Fox y S. B. Eickhoff, «Modelling neural correlates of working memory: A coordinate-based meta-analysis,» *NeuroImage*, vol. 60, pp. 830-846, 2012.
- [22] M. Långkvist, L. Karlsson y A. Loutfi, «Sleep stage classification using unsupervised feature learning,» *Advances in Artificial Neural Systems*, vol. 2012, p. 5, 2012.
- [23] G. Ruffini, D. Ibañez, M. Castellano, S. Dunne y A. Soria-Frisch, «EEG-driven RNN Classification for Prognosis of Neurodegeneration in At-Risk Patients,» de *Artificial Neural Networks and Machine Learning -- ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings, Part I*, E. P. A. Villa, P. Masulli y J. A. Pons Rivero, Edits., Ch. a: Springer International Publishing, 2016, pp. 306-313.
- [24] Y. Bengio, A. Courville y P. Vincent, «Representation Learning: A Review and New Perspectives,» *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, pp. 1798-1828, #aug# 2013.
- [25] K. Mahajan, M. R. Vargantwar y S. M. Rajput, «Article: Classification of EEG using PCA, ICA and Neural Network,» *IJCA Proceedings on International Conference in Computational Intelligence (ICCIA 2012)*, vol. ICCIA, pp. 1-4, March 2012.
- [26] H. Zhang, Y. Liu, J. Liang, J. Cao y L. Zhang, «Gaussian mixture modeling in stroke patients' rehabilitation EEG data analysis,» de *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.
- [27] Y. LeCun, Y. Bengio y G. Hinton, «Deep learning,» *Nature*, vol. 521, pp. 436-444, 2015.
- [28] Z. Mao, V. Lawhern, L. M. Meriño, K. Ball, L. Deng, B. J. Lance, K. Robbins y Y. Huang, «Classification of non-time-locked rapid serial visual presentation events for brain-computer interaction using deep learning,» de *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit International Conference on*, 2014.
- [29] J. Schmidhuber, «Deep learning in neural networks: An overview,» *Neural Networks*, vol. 61, pp. 85-117, 2015.
- [30] Y. Bengio, «Deep Learning of Representations for Unsupervised and Transfer Learning,» de *Unsupervised and Transfer Learning - Workshop held at {ICML} 2011, Bellevue, Washington, USA, July 2, 2011*, 2012.
- [31] L. Deng y N. Jaitly, «Deep discriminative and generative models for pattern recognition,» *USENIX--Advanced Computing Systems Association*, 2015.
- [32] P. Bashivan, I. Rish, M. Yeasin y N. Codella, «Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks,» *International Conference in Learning Representations 2016*, 2016.
- [33] R. Manor y A. B. Geva, «Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI,» *Frontiers in Computational Neuroscience*, vol. 9, p. 146, 2015.
- [34] M. Soleymani, S. Asghari-Esfeden, M. Pantic y Y. Fu, «Continuous emotion detection using EEG signals and facial expressions,» de *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

- [35] S. Jirayucharoensak, S. Pan-Ngum y P. Israsena, «EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation,» *The Scientific World Journal*, vol. 2014, 2014.
- [36] Y. Qi, Y. Wang, J. Zhang, J. Zhu y X. Zheng, «Robust deep network with maximum correntropy criterion for seizure detection,» *BioMed research international*, vol. 2014, 2014.
- [37] O. Tsinalis, P. M. Matthews y Y. Guo, «Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders,» *Annals of Biomedical Engineering*, vol. 44, pp. 1587-1597, 2016.
- [38] M. Bianchini, M. Maggini y L. C. Jain, *Handbook on Neural Information Processing*, Springer Publishing Company, Incorporated, 2013.
- [39] Y. Zhao y L. He, «Deep Learning in the EEG Diagnosis of Alzheimer's Disease,» de *Computer Vision - ACCV 2014 Workshops: Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part I*, C. V. Jawahar y S. Shan, Edits., Ch. a: Springer International Publishing, 2015, pp. 340-353.
- [40] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf y M. Grosse-Wentrup, «Transfer Learning in Brain-Computer Interfaces,» *IEEE Computational Intelligence Magazine*, vol. 11, pp. 20-31, Feb 2016.
- [41] G. F. Wilson, C. A. Russell, J. W. Monnin, J. R. Estep y J. C. Christensen, «How Does Day-to-Day Variability in Psychophysiological Data Affect Classifier Accuracy?,» *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, pp. 264-268, 2010.
- [42] K. Muandet, D. Balduzzi y B. Schölkopf, «Domain Generalization via Invariant Feature Representation,» de *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, Atlanta, 2013.
- [43] I. Goodfellow, Y. Bengio y A. Courville, «Deep Learning,» 2016.
- [44] S. J. Pan y Q. Yang, «A Survey on Transfer Learning,» *IEEE Trans. on Knowl. and Data Eng.*, vol. 22, pp. 1345-1359, #oct# 2010.
- [45] R. Raina, A. Battle, H. Lee, B. Packer y A. Y. Ng, «Self-taught Learning: Transfer Learning from Unlabeled Data,» de *Proceedings of the 24th International Conference on Machine Learning*, New York, NY, USA, 2007.
- [46] K. Mahajan, M. R. Vargantwar y S. M. Rajput, «Classification of EEG using PCA, ICA and Neural Network,» *IJCA Proceedings on International Conference in Computational Intelligence (ICCI 2012)*, vol. ICCIA, pp. 1-4, March 2012.
- [47] S. Sanei y J. A. Chambers, *EEG Signal Processing*, Wiley, 2008.
- [48] J. Sweller, «Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load,» *Educational Psychology Review*, vol. 22, pp. 123-138, 2010.
- [49] F. Paas, J. E. Tuovinen, H. Tabbers y P. W. M. Van Gerven, «Cognitive load measurement as a means to advance cognitive load theory,» *Educational psychologist*, vol. 38, pp. 63-71, 2003.
- [50] P. Bashivan, G. M. Bidelman y M. Yeasin, «Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity,» *European Journal of Neuroscience*, vol. 40, pp. 3774-3784, 2014.
- [51] R. C. Singleton, «On Computing the Fast Fourier Transform,» *Commun. ACM*, vol. 10, pp. 647-654, #oct# 1967.
- [52] D. P. Subha, P. K. Joseph, R. Acharya U y C. M. Lim, «EEG Signal Analysis: A Survey,» *Journal of Medical Systems*, vol. 34, pp. 195-212, 2010.
- [53] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

- [54] Y. Lecun, L. Bottou, Y. Bengio y P. Haffner, «Gradient-based learning applied to document recognition,» *Proceedings of the IEEE*, vol. 86, pp. 2278-2324, Nov 1998.
- [55] A. Krizhevsky, I. Sutskever y G. E. Hinton, «ImageNet Classification with Deep Convolutional Neural Networks,» de *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou y K. Q. Weinberger, Edits., Curran Associates, Inc., 2012, pp. 1097-1105.
- [56] K. He, X. Zhang, S. Ren y J. Sun, «Deep Residual Learning for Image Recognition,» de *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [57] A. Veit, M. J. Wilber y S. Belongie, «Residual Networks Behave Like Ensembles of Relatively Shallow Networks,» de *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon y R. Garnett, Edits., Curran Associates, Inc., 2016, pp. 550-558.
- [58] S. Ioffe y C. Szegedy, «Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,» de *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, 2015.
- [59] X. Glorot, A. Bordes y Y. Bengio, «Deep Sparse Rectifier Neural Networks,» de *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, 2011.
- [60] Y. Bengio, N. Boulanger-Lewandowski y R. Pascanu, «Advances in optimizing recurrent networks,» de *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.
- [61] Y. Bengio, P. Simard y P. Frasconi, «Learning Long-term Dependencies with Gradient Descent is Difficult,» *Trans. Neur. Netw.*, vol. 5, pp. 157-166, #mar# 1994.
- [62] K. Cho, B. van Merriënboer, D. Bahdanau y Y. Bengio, «On the Properties of Neural Machine Translation: Encoder-Decoder Approaches,» *CoRR*, vol. abs/1409.1259, 2014.
- [63] G. E. Hinton y R. R. Salakhutdinov, «Reducing the dimensionality of data with neural networks,» *Science*, vol. 313, pp. 504-507, #Jul# 2006.
- [64] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville y Y. Bengio, «Generative Adversarial Nets,» de *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence y K. Q. Weinberger, Edits., Curran Associates, Inc., 2014, pp. 2672-2680.
- [65] D. P. Kingma y M. Welling, «Auto-Encoding Variational Bayes,» *International Conference in Learning Representations 2014*, 2014.
- [66] S. Kullback, *Information Theory And Statistics*, Dover Pubns, 1997.
- [67] X. Cui, V. Goel y B. Kingsbury, «Data Augmentation for Deep Neural Network Acoustic Modeling,» *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 23, pp. 1469-1477, #sep# 2015.
- [68] A. Krogh y J. A. Hertz, «A Simple Weight Decay Can Improve Generalization,» de *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 4*, 1992.
- [69] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, «Dropout: A Simple Way to Prevent Neural Networks from Overfitting,» *J. Mach. Learn. Res.*, vol. 15, pp. 1929-1958, #jan# 2014.
- [70] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.
- [71] R. A. Baeza-Yates y B. Ribeiro-Neto, *Modern Information Retrieval*, Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1999.

- [72] A. P. Bradley, «The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms,» *Pattern Recogn.*, vol. 30, pp. 1145-1159, #jul# 1997.
- [73] J. Huang y C. X. Ling, «Using AUC and accuracy in evaluating learning algorithms,» *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, pp. 299-310, March 2005.
- [74] X. Yu, P. Chum y K.-B. Sim, «Analysis the effect of PCA\ for feature reduction in non-stationary EEG\ based motor imagery of BCI\ system,» *Optik - International Journal for Light and Electron Optics* , vol. 125, pp. 1498-1502, 2014.
- [75] C. J. James y D. Lowe, «Extracting Multisource Brain Activity from a Single Electromagnetic Channel,» *Artif. Intell. Med.*, vol. 28, pp. 89-104, #may# 2003.
- [76] K. R. Rao, D. N. Kim y J.-J. Hwang, *Fast Fourier Transform - Algorithms and Applications*, 1st ed., Springer Publishing Company, Incorporated, 2010.
- [77] B.-h. Yang, G.-z. Yan, R.-g. Yan y T. Wu, «Feature extraction for EEG-based brain-computer interfaces by wavelet packet best basis decomposition,» *Journal of Neural Engineering*, vol. 3, p. 251, 2006.
- [78] M. Vetterli y C. Herley, «Wavelets and filter banks: theory and design,» *IEEE Transactions on Signal Processing*, vol. 40, pp. 2207-2232, Sep 1992.
- [79] X. Li, X. Jia, G. Xun y A. Zhang, «Improving EEG feature learning via synchronized facial video,» de *Big Data (Big Data)*, 2015 *IEEE International Conference on*, 2015.
- [80] B. Yan, Y. Wang, Y. Li, Y. Gong, L. Guan y S. Yu, «An EEG signal classification method based on sparse auto-encoders and support vector machine,» de *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, 2016.
- [81] Q. Barthélemy, C. Gouy-Pailler, Y. Isaac, A. Souloumiac, A. Larue y J. I. Mars, «Multivariate temporal dictionary learning for EEG,» *Journal of Neuroscience Methods* , vol. 215, pp. 19-28, 2013.
- [82] W. Zhou, Y. Yang y Z. Yu, «Discriminative dictionary learning for EEG signal classification in Brain-computer interface,» de *2012 12th International Conference on Control Automation Robotics Vision (ICARCV)*, 2012.
- [83] A. J. Brockmeier y J. C. Principe, «Learning Recurrent Waveforms Within EEGs,» *{IEEE} Trans. Biomed. Engineering*, vol. 63, pp. 43-54, 2016.
- [84] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot y P. Vandergheynst, «Shift-invariant dictionary learning for sparse representations: Extending K-SVD,» de *2008 16th European Signal Processing Conference*, 2008.
- [85] G. J. McLachlan y D. Peel, *Finite mixture models*, New, Yor: J. Wiley & Sons, 2000.
- [86] R. Salakhutdinov, A. Mnih y G. Hinton, «Restricted Boltzmann Machines for Collaborative Filtering,» de *Proceedings of the 24th International Conference on Machine Learning*, New York, NY, USA, 2007.
- [87] H. Cecotti y A. Graeser, «Convolutional Neural Network with embedded Fourier Transform for EEG classification,» de *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008.
- [88] H. Cecotti y A. Graser, «Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 433-445, March 2011.
- [89] S. Sakhavi, C. Guan y S. Yan, «Parallel convolutional-linear neural network for motor imagery classification,» de *Signal Processing Conference (EUSIPCO), 2015 23rd European*, 2015.
- [90] B. Yang, K. Duan y T. Zhang, «Removal of EOG\ artifacts from EEG\ using a cascade of sparse autoencoder and recursive least squares adaptive filter,» *Neurocomputing* , pp. - , 2016.

- [91] Z. Tang, C. Li y S. Sun, «Single-trial EEG\ classification of motor imagery using deep convolutional neural networks,» *Optik - International Journal for Light and Electron Optics* , vol. 130, pp. 11-18, 2017.
- [92] A. Antoniadou, L. Spyrou, C. C. Took y S. Sanei, «Deep learning for epileptic intracranial EEG data,» de *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2016.
- [93] P. R. Davidson, R. D. Jones y M. T. R. Peiris, «EEG-Based Lapse Detection With High Temporal Resolution,» *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 832-839, May 2007.
- [94] Z. Wang, S. Lyu, G. Schalk y Q. Ji, «Deep feature learning using target priors with applications in ECoG signal decoding for BCI,» 2013.
- [95] A. Supratak, L. Li y Y. Guo, «Feature extraction with stacked autoencoders for epileptic seizure detection,» de *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.
- [96] W. Gao, J.-a. Guan, J. Gao y D. Zhou, «Multi-ganglion ANN\ based feature learning with application to P300-BCI signal classification,» *Biomedical Signal Processing and Control* , vol. 18, pp. 127-137, 2015.
- [97] N. Zhang, W.-L. Zheng, W. Liu y B.-L. Lu, «Continuous Vigilance Estimation Using LSTM Neural Networks,» de *Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16--21, 2016, Proceedings, Part II*, A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee y D. Liu, Edits., Ch. a: Springer International Publishing, 2016, pp. 530-537.
- [98] L. Vidyaratne, A. Glandon, M. Alam y K. M. Iftekharuddin, «Deep recurrent neural network for seizure detection,» de *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016.
- [99] D. Tan, R. Zhao, J. Sun y W. Qin, «Sleep spindle detection using deep learning: A validation study based on crowdsourcing,» de *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
- [100] W. L. Zheng, J. Y. Zhu, Y. Peng y B. L. Lu, «EEG-based emotion classification using deep belief networks,» de *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014.
- [101] Z. V. Freudenburg, N. F. Ramsey, M. Wronkiewicz, W. D. Smart, R. Pless y E. C. Leuthardt, «Real-time naive learning of neural correlates in ECoG Electrophysiology,» *International Journal of Machine Learning and Computing*, vol. 1, p. 269, 2011.
- [102] W. L. Zheng y B. L. Lu, «Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks,» *IEEE Transactions on Autonomous Mental Development*, vol. 7, pp. 162-175, Sept 2015.
- [103] W. L. Zheng, H. T. Guo y B. L. Lu, «Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network,» de *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015.
- [104] F. Li, G. Zhang, W. Wang, R. Xu, T. Schnell, J. Wen, F. McKenzie y J. Li, «Deep Models for Engagement Assessment With Scarce Label Information,» *IEEE Transactions on Human-Machine Systems*, vol. PP, pp. 1-8, 2016.
- [105] T. Ma, H. Li, H. Yang, X. Lv, P. Li, T. Liu, D. Yao y P. Xu, «The extraction of motion-onset VEP\ BCI\ features based on deep learning and compressed sensing,» *Journal of Neuroscience Methods* , vol. 275, pp. 80-92, 2017.

- [106] N. Thammasan, K. i. Fukui y M. Numao, «Application of deep belief networks in eeg-based dynamic music-emotion recognition,» de *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016.
- [107] K. Li, X. Li, Y. Zhang y A. Zhang, «Affective state recognition from EEG with deep belief networks,» de *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*, 2013.
- [108] Y. Ren y Y. Wu, «Convolutional deep belief networks for feature extraction of EEG signal,» de *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014.
- [109] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco y B. Litt, «Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement,» *Journal of Neural Engineering*, vol. 8, p. 036015, 2011.
- [110] D. Wang y Y. Shang, «Modeling physiological data with deep belief networks,» *International journal of information and education technology (IJJET)*, vol. 3, p. 505, 2013.
- [111] M. H. Cilasan y H. Yalçın, «A deep learning approach to EEG based epilepsy seizure determination,» de *2016 24th Signal Processing and Communication Application Conference (SIU)*, 2016.
- [112] M. Långkvist, L. Karlsson y A. Loutfi, «A review of unsupervised feature learning and deep learning for time-series modeling,» *Pattern Recognition Letters*, vol. 42, p. 11?24, 2014.
- [113] J. T. Turner, A. Page, T. Mohsenin y T. Oates, «Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection,» de *2014 AAAI Spring Symposium Series*, 2014.
- [114] S. Zhao y F. Rudzicz, «Classifying phonological categories in imagined and articulated speech,» de *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.
- [115] X. Jia, K. Li, X. Li y A. Zhang, «A Novel Semi-Supervised Deep Learning Framework for Affective State Recognition on EEG Signals,» de *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on*, 2014.
- [116] G. Shou y L. Ding, «Frontal theta EEG dynamics in a real-world air traffic control task,» de *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.
- [117] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston y C. T. Silva, «A User Study of Visualization Effectiveness Using EEG and Cognitive Load,» de *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, Chichester, 2011.
- [118] J. Onton, A. Delorme y S. Makeig, «Frontal midline EEG\ dynamics during working memory,» *NeuroImage* , vol. 27, pp. 341-356, 2005.
- [119] J. Chung, K. Kastner, L. Dinh, K. Goel, A. C. Courville y Y. Bengio, «A Recurrent Latent Variable Model for Sequential Data,» de *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama y R. Garnett, Edits., Curran Associates, Inc., 2015, pp. 2962-2970.
- [120] D. P. Kingma, S. Mohamed, D. J. Rezende y M. Welling, «Semi-supervised Learning with Deep Generative Models,» de *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence y K. Weinberger, Edits., Curran Associates, Inc., 2014, pp. 3581-3589.
- [121] S. Purushotham, W. Carvalho, T. Nilanon y Y. Liu, «VARIATIONAL RECURRENT ADVERSARIAL DEEP DOMAIN ADAPTATION,» *International Conference in Learning Representations 2016*, 2017.

- [122] J. P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections*, University of Chicago Press, 1997.
- [123] P. Alfeld, «A trivariate clough—tocher scheme for tetrahedral data,» *Computer Aided Geometric Design*, vol. 1, pp. 169-181, 1984.
- [124] P. W. Mirowski, Y. LeCun, D. Madhavan y R. Kuzniecky, «Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG,» de *2008 IEEE Workshop on Machine Learning for Signal Processing*, 2008.
- [125] R. Józefowicz, W. Zaremba y I. Sutskever, «An Empirical Exploration of Recurrent Network Architectures,» de *Proceedings of the 32nd International Conference on Machine Learning, {ICML} 2015, Lille, France, 6-11 July 2015*, 2015.
- [126] A. r. Mohamed, G. Hinton y G. Penn, «Understanding how Deep Belief Networks perform acoustic modelling,» de *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.
- [127] K. Simonyan y A. Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition,» *International Conference in Learning Representations 2016*, 2016.
- [128] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization,» *CoRR*, published in *International Conference on Learning Representations 2015*, vol. abs/1412.6980, 2014.