



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

ANÁLISIS DE COMPONENTES INDEPENDIENTES Y DESCOMPOSICIÓN DE  
MODO EMPÍRICO PARA LA CLASIFICACIÓN DE TAREAS DE IMAGINERÍA  
MOTORA EN INTERFAZ CEREBRO COMPUTADOR

TESIS PARA OPTAR AL GRADO DE DOCTOR EN INGENIERÍA ELÉCTRICA

CLAUDIO TAPIA MALEBRÁN

PROFESOR GUÍA

DR. JAVIER RUIZ DEL SOLAR SAN MARTÍN

MIEMBROS DE LA COMISIÓN

DIEGO COSMELLI SÁNCHEZ

MATÍAS ZAÑARTU SALAS

JORGE SILVA SÁNCHEZ

SANTIAGO DE CHILE

2018

RESUMEN DE TESIS PARA OPTAR  
AL GRADO DE DOCTOR EN INGENIERIA ELÉCTRICA

POR: CLAUDIO TAPIA MALEBRÁN

FECHA: 05 DE ENERO DE 2018

PROF. GUÍA: JAVIER RUIZ DEL SOLAR

## **Análisis de Componentes Independientes y Descomposición de Modo Empírico para la Clasificación de Tareas de Imaginería Motora en Interfaz Cerebro Computador**

La interfaz cerebro computador permite la interacción entre las señales extraídas del cerebro de una persona y su entorno, lo que se logra mediante el procesamiento de estas señales en un computador. Sin embargo, aún no existe consenso respecto a los métodos de análisis y procesamiento de estas señales para una correcta clasificación en tareas de imaginería motora. La presente Tesis propone un nuevo método para optimizar la clasificación de tareas de imaginería motora mediante el uso automatizado de análisis de componentes independientes (ICA) y descomposición de modo empírico (EMD) en la fase de procesamiento de señales de electroencefalografía. La señales utilizadas corresponden a la base de datos pública Physionet. Con el fin de evaluar la eficacia de la clasificación se comparan cuatro clasificadores estadísticos, Máquina de soporte vectorial con optimización mínima secuencial (SMO), Bayes, Dagging y Logistic. Éstos fueron seleccionados considerando las propiedades de las señales de EEG.

Las tareas de imaginería motora a clasificar fueron: imaginar abrir y cerrar el puño izquierdo, imaginar abrir y cerrar el puño derecho e imaginar reposo de extremidades. Se compararon los resultados de la clasificación para las 3 tareas de imaginería motora usando los cuatro clasificadores con y sin el uso del método propuesto en 102 sujetos de la base de datos. La metodología propuesta demostró que el uso automatizado de ICA y EMD en secuencia es eficaz para mejorar la clasificación de tareas de imaginería motora. Los mejores resultados se obtuvieron con el clasificador Logistic, con un porcentaje de 98,8% de aciertos en la clasificación, seguido por SMO con 97,7%, Dagging con 93,9% y finalmente Bayes con 58,8%. Además se demostró que existe diferencia estadísticamente significativa al comparar los resultados entre el método propuesto y el método tradicional en los cuatro clasificadores usados.

A mis padres  
A mi familia

# Agradecimientos

Quiero agradecer a quienes son el soporte y la fuerza para avanzar, mis hijos.

Toda mi gratitud a quienes les resté tiempo para emprender esta tarea.

Muchas gracias a mi querido maestro Javier Ruiz del Solar, quien me enseñó un camino nuevo y me permitió descubrir la belleza de la Ingeniería. Gracias por confiar en mi, por ser un guía, apoyarme y reprenderme cuando fue necesario. Gracias Profesor, sin usted esto no sería posible.

Gracias a mis compañeros, con quienes compartí hermosos momentos, cada instante aportando un grano de arena a mi desarrollo como persona. Gracias por ayudarme a ser una persona comprometida con una sociedad más justa.

Gracias a los académicos y administrativos del Departamento de Ingeniería Eléctrica de la Universidad de Chile, aquí descubrí a personas maravillosas. Siempre me sentí acogido e incluso protegido. Aquí reí, sufrí, crecí, aquí viví plenamente.

Gracias a los profesores que revisaron mi Tesis, sus aportes han sido invaluable en este proceso. Gracias por su extraordinaria calidad humana, más allá de lo académico, siempre encontré a personas dispuestas a ayudarme.

Especial gracias a Patricio, Loreto, Wilmita, Mauricio, Fernando, Marcelo, Leo, Omar, Teresita. Con todos aprendí mucho, más allá de lo que se ve.

Gracias a Dios.

# Tabla de contenido

<b>1. Introducción</b>	<b>2</b>
1.1. Antecedentes Generales . . . . .	2
1.1.1. Electroencefalografía . . . . .	3
1.1.2. Potenciales Relacionados con Evento . . . . .	6
1.1.3. Imaginería Motora . . . . .	6
1.2. Definición del Problema a Abordar . . . . .	7
1.3. Objetivos . . . . .	9
1.3.1. Objetivo General . . . . .	9
1.3.2. Objetivos Específicos . . . . .	9
1.4. Hipótesis . . . . .	9
1.5. Estructura de la Tesis . . . . .	10
<b>2. Estado del Arte</b>	<b>11</b>
2.1. Modelos y Arquitectura de la Interfaz Cerebro Computador . . . . .	11
2.2. Preprocesamiento de la Señal . . . . .	14
2.2.1. Filtros Espaciales . . . . .	15
2.2.1.1. Filtros de Patrones Espaciales Comunes . . . . .	15
2.2.1.2. Referencia Común Promedio . . . . .	17
2.2.1.3. Filtro Laplaciano . . . . .	17
2.2.2. Análisis de Componentes Principales . . . . .	18
2.2.3. Análisis de Componentes Independientes . . . . .	19
2.2.4. Comparación de Métodos . . . . .	21
2.3. Procesamiento . . . . .	22
2.3.1. Enfoque Autoregresivo . . . . .	23
2.3.2. Transformada de Wavelet . . . . .	24
2.3.3. Densidad Espectral de Potencia . . . . .	25
2.3.4. Resumen . . . . .	26
2.4. Clasificación en Interfaz Cerebro Computador . . . . .	26
2.4.1. Análisis Discriminante lineal . . . . .	27
2.4.2. Máquina de Soporte Vectorial usando Optimización Mínima Secuencial	30
2.4.3. Clasificador Dagging . . . . .	34
2.4.4. Clasificador Logístico . . . . .	35
2.4.5. Redes Neuronales . . . . .	36
2.4.6. Clasificador de Bayes . . . . .	36
2.4.7. Clasificador Basado en Vecino más Cercano . . . . .	38
2.5. Clasificación Usando Bases de Datos públicas . . . . .	41

<b>3. Metodología Propuesta</b>	<b>43</b>
3.1. Metodología Experimental . . . . .	43
3.2. Procesamiento de la Señal de EEG con Método Propuesto . . . . .	45
3.2.1. Preprocesamiento . . . . .	46
3.2.2. Procesamiento ICA y EMD . . . . .	47
3.3. Extracción de Características . . . . .	51
3.4. Clasificación . . . . .	52
<b>4. Resultados</b>	<b>54</b>
<b>5. Conclusión y Discusión</b>	<b>57</b>
<b>6. Contribución Científica y Trabajo Futuro</b>	<b>61</b>
6.1. Optimización de la Clasificación de Tareas de Imaginería Motora . . . . .	61
6.2. Automatización en la Eliminación de Componentes Independientes . . . . .	62
6.3. Enfoque No Lineal y No Estacionario en el Procesamiento de las Señales de EEG . . . . .	62
6.4. Clasificación de Tres Clases . . . . .	62
6.5. Limitaciones y Trabajo Futuro . . . . .	63
6.6. Trabajo Relacionado-Estudio de Casos . . . . .	64
<b>Bibliografía</b>	<b>68</b>

# Índice de figuras

1.1.	Localización y etiquetado de los electrodos según el Sistema Internacional 10-20 extendido. . . . .	4
1.2.	Bandas de frecuencia para un registro de electroencefalografía. . . . .	5
2.1.	Arquitectura tradicional de un sistema BCI. Notar que el módulo de procesamiento tradicionalmente considera las etapas de extracción de características y clasificación. . . . .	13
2.2.	BCI invasivo para el control de un brazo robótico en primates (modificado de [30]). . . . .	14
2.3.	Hiperplano de separación, los vectores de soporte se definen con puntos encerrados en un círculo (modificado de [27]). . . . .	31
2.4.	Clasificador por distancia mínima. La muestra se clasifica en la clase para la cual la distancia a su representante es mínima, en este caso se clasifica en la clase $c_k$ . . . . .	39
2.5.	Clasificación por la regla del vecino más cercano. La muestra $x$ se clasifica en la clase para la que la distancia a uno de sus puntos es mínima. En este ejemplo el prototipo más cercano pertenece a la clase $c_k$ . . . . .	40
3.1.	Secuencia de los 14 ensayos experimentales utilizados en la base de datos Physionet. . . . .	44
3.2.	Diagrama de bloque del método empleado. La etapa de preprocesamiento incluyó z-score y blanqueamiento de los datos. . . . .	46
3.3.	Método para determinar IMFs. Modificado de [12]. . . . .	50
3.4.	Diagrama de bloque que resume la metodología de preprocesamiento y procesamiento de la señal de EEG. . . . .	51
4.1.	Comparación de gráfico de cajas respecto al resultado de la clasificación para los cuatro clasificadores usando el método propuesto versus el método tradicional. Cada caja muestra la distribución de los resultados entre el percentil 25 y el 75. La línea roja muestra la mediana y el rombo la media. . . . .	56
6.1.	Sistema de registro en condiciones de funcionamiento habitual del Laboratorio de Visión Computacional de la Universidad de Chile. El usuario está expuesto a múltiples estímulos externos diferentes a las tareas de imaginación motora. . . . .	65
6.2.	Izquierda: Panel de control utilizado en el paradigma experimental; Derecha: Sujeto utilizando el sistema de registro Emotiv Eloc. . . . .	66
6.3.	Metodología de Registro . . . . .	66

6.4. Resultados preliminares de un estudio de casos aplicando procesamiento con ICA y EMD para la clasificación de tareas de imaginería motora. . . . .	68
---	----



# Índice de tablas

2.1. Comparación de métodos de preprocesamiento de señales de EEG. Resumen de acuerdo a lo presentado en [81, 65, 107]. . . . .	22
2.2. Resumen de métodos de extracción de características de acuerdo al mecanismo neural involucrado (Modificado de [20]). . . . .	26
4.1. Comparación de los resultados de la clasificación para 3 tareas de imaginería motora usando cuatro clasificadores con y sin el uso del método propuesto. . . . .	55
4.2. Evaluación del test de normalidad Shapiro Wilk para la distribución de los resultados de 102 sujetos evaluados con y sin el método propuesto. . . . .	55
4.3. Evaluación de significancia estadística en la clasificación de tres tareas de imaginería motora al utilizar procesamiento con ICA y EMD. . . . .	56
5.1. Clasificación de tareas de imaginería motora usando bases de datos públicas. Revisión bibliográfica. . . . .	59
6.1. Comparación de resultados de clasificación para 3 tareas de imaginería motora usando cuatro clasificadores con y sin el uso del método propuesto en 8 sujetos voluntarios (resultado preliminar). . . . .	67

# Capítulo 1

## Introducción

### 1.1. Antecedentes Generales

La interfaz cerebro computador, BCI (por su sigla en inglés “Brain Computer Interface”), es un sistema que permite la comunicación entre las señales extraídas del cerebro de un sujeto (habitualmente a través de electroencefalografía) y un computador. Este sistema puede ser usado para controlar dispositivos, tales como el cursor en una pantalla de computador o una silla de ruedas. Originalmente su objetivo fue permitir a las personas con deterioro neuro-motor interactuar con su entorno, mediante un sistema de comunicación y control que no dependiera de las relaciones neuromotoras normales, disminuyendo de este modo la discapacidad de una persona. Sin embargo, en la actualidad sus aplicaciones son múltiples y van desde video-juegos a usos militares [139]. La tecnología de BCI ha ganado terreno en los últimos diez años, principalmente debido a los múltiples avances en las estrategias usadas para el procesamiento de las señales, y sobre todo por la amplia gama de posibilidades que otorga su uso, más allá de aplicaciones médicas o de rehabilitación. El principal objetivo de la BCI es decodificar la actividad cerebral de una persona, y usar la representación de estas señales para comunicarse con un dispositivo externo. Por ejemplo, desde el punto de vista clínico se han descrito aplicaciones en pacientes con la condición de *locked-in* o enclaustramiento, estado caracterizado por la ausencia de actividad motora y de lenguaje, pero con funciones cognitivas completas [9]. También se ha descrito su uso en pacientes con desordenes degenerativos como esclerosis lateral amiotrófica, esclerosis múltiple o distrofia muscular y en pacientes con traumatismo raquímedular [152]. Aún cuando este sistema no permite la reparación de la vía afectada, sí puede ayudar a restaurar la función perdida, evitar deterioros físicos y psicológicos, y favorecer la interacción del sujeto con su entorno.

El método de adquisición más usado en aplicaciones de BCI, debido a su simplicidad y costo, es la electroencefalografía (EEG). Sin embargo, también se han usado otros métodos no invasivos como la magnetoencefalografía (MEG), la tomografía por emisión de positrones (PET), la imagen por resonancia magnética funcional (fMRI) y la imagen óptica cercana al infrarrojo (fNRI). Por otro lado, también existen métodos invasivos; éstos implican el uso de electrodos en contacto directo con la corteza cerebral, lo que aumenta la posibilidad de daño y expone al paciente a los riesgos propios de una cirugía. Los métodos invasivos pueden ser aplicados tanto de manera experimental como clínico. En el caso de uso experimental, se ha utilizado para el estudio de BCI tanto en humanos como en poblaciones no humanas; mientras que el uso clínico está dado principalmente por aplicaciones con fines diagnósticos, por ejemplo: la detección de focos epilépticos.

### 1.1.1. Electroencefalografía

La actividad cerebral, generada por la activación de la población neuronal, produce señales eléctricas que pueden ser detectadas en la superficie externa del cráneo (cuero cabelludo), desde la corteza cerebral o por registros en el interior del cerebro. Desde esta perspectiva, los sistemas de adquisición de estas señales pueden ser invasivos o no invasivos. Los sistemas invasivos, que utilizan registros a partir de electrodos implantados en la corteza cerebral, implican mayor riesgo clínico, y requieren mayor complejidad en la técnica de implantación. Por este motivo su uso se ha limitado principalmente al estudio en primates no humanos. En la actualidad se ha privilegiado el uso de sistemas de registro no invasivos, particularmente la EEG. Ésta corresponde a la medición de las fluctuaciones de voltaje resultante de los flujos de corrientes iónicas en las neuronas de la corteza cerebral. La información obtenida mediante estos potenciales eléctricos puede ser usada tanto en clínica como en investigación. En este ámbito, la tecnología de BCI ha intentado utilizar estos registros de manera que puedan permitir la comunicación con un dispositivo externo.

La señal de EEG corresponde a la suma espacial de corrientes inducidas por potenciales post-sinápticos sincronizados que ocurren en grupos de neuronas, estimándose que para un electrodo de aproximadamente 10 mm de diámetro cerca de 250.000 neuronas deberían descargar de manera sincrónica para generar actividad visible en el registro de EEG [28]. Bajo esta premisa, es lícito pensar que bajas frecuencias corresponden a un alto grado de sincronía, mientras que altas frecuencias equivalen a uno menor. En un sujeto adulto la amplitud de la actividad eléctrica del EEG se encuentra entre 10 y 100  $\mu\text{V}$ , comúnmente en 50  $\mu\text{V}$  [28]. Sin embargo, algunos potenciales de interés, como los potenciales evocados sensoriales<sup>1</sup> (SEP), presentan una amplitud menor (aproximadamente 5  $\mu\text{V}$ ) y sólo pueden ser visualizados promediando múltiples ventanas [45]. Las señales de EEG siempre representan la diferencia de potencial entre dos electrodos, uno activo (correspondiente a la célula excitable) y otro de referencia (correspondiente a una zona eléctricamente neutra), por lo tanto la calidad del registro dependerá de la interfaz electrodo-piel. Por otra parte, con el fin de establecer un lenguaje común en los registros de EEG, y así comparar hallazgos y resultados entre diferentes mediciones, se cuenta con el sistema internacional de localización de electrodos [117, 71] (ver figura 1.1).

---

<sup>1</sup>Los SEP corresponden a señales corticales registradas luego de la estimulación sensorial, ya sea visual, auditiva o somatosensorial.



y en algunos casos con la habilidad del cerebro para integrar varios aspectos de un estímulo en un todo coherente.

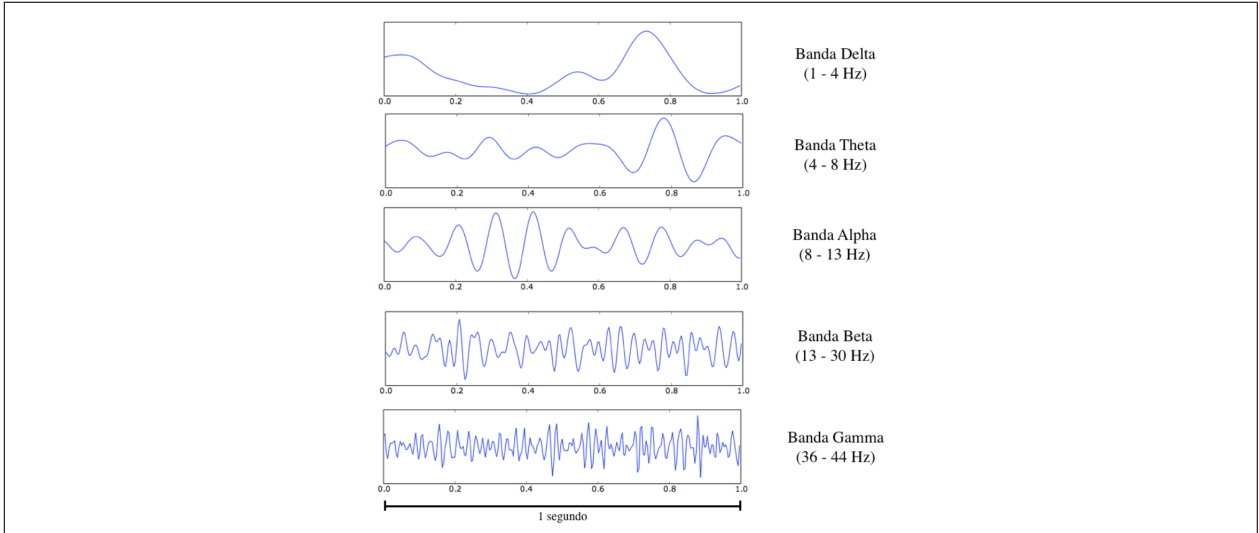


Figura 1.2: Bandas de frecuencia para un registro de electroencefalografía.

Por otra parte, en el mismo espectro de la banda alpha, se encuentra la banda mu (8-12Hz), onda espontánea del EEG, asociada con actividades motoras, que se observa en registros sobre el cortex motor. Esta señal disminuye (o se desincroniza) con el movimiento o la intención de movimiento sobre la corteza sensoriomotora, mostrando que el movimiento y la imaginación de movimiento están asociados con la activación de las mismas regiones corticales [11]. De acuerdo a los estudios de [99], las diferencias entre la imaginación de movimiento/reposo son considerablemente más pequeñas que las diferencias entre movimiento real/reposo, este aspecto es fundamental en la implementación de una interfaz cerebro computador, pues la señal a decodificar corresponde a imaginación de movimiento, por esta razón cobra especial importancia el o los mecanismos implementados para diferenciar la actividad mental asociada a movimiento y a reposo.

El fenómeno de supresión o disminución de la banda de frecuencia se denomina desincronización relacionada con evento “ERD” (Event Related desynchronization), y puede ser observado en respuesta a diversos tipos de tareas. El ERD refleja el procesamiento activo de información de procesos cerebrales excitatorios, esto es, en tareas motoras podría reflejar una excitabilidad de la corteza motora primaria contralateral a la extremidad en movimiento [18]. Las bandas mu junto con las bandas alpha corresponden a las señales más usadas en BCI, debido a que las personas pueden aprender a cambiar la amplitud haciendo un esfuerzo mental apropiado [110]. Esto se puede lograr recordando una imagen (altamente estimulante) o aumentando los niveles de concentración [99, 110]. En otras palabras, las personas pueden aprender a controlar las bandas de frecuencia entre los 8 y 12 Hz.

Los ritmos sensorio motores han sido ampliamente usados para establecer comandos en BCI. Cambios en la amplitud de los ritmos sensorio motores, conocidos como desincronización/sincronización (ERD/ERS), pueden ser detectados mediante EEG cuando la actividad cerebral está relacionada a cualquier tarea motora, aun cuando no exista movimiento. Sin embargo, el uso de estos ritmos en ambientes reales es aún limitado.

### 1.1.2. Potenciales Relacionados con Evento

En el estudio de las señales de EEG, y particularmente en el uso de interfaz cerebro computador, cobran vital importancia los potenciales eléctricos, ya sea de tipo espontáneo o relacionado a eventos. El primero se presenta en el EEG sin relación a eventos particulares más allá del estado del sujeto, el segundo considera diferentes tipos de actividades normalmente relacionadas a eventos (ya sea internos o externos) y que incluye a su vez actividad evocada (Evoked Potentials) y actividad propiamente tal relacionada a los eventos (Event Related Potentials).

Los potenciales evocados normalmente se asocian a aquellos tempranamente evocados por un estímulo, generalmente sensorial, y que está muy precisamente relacionado temporalmente con el estímulo que lo genera (por ejemplo la actividad P1/N1 asociada a la presentación de estímulos visuales de alto contraste). Los potenciales relacionados a evento no necesariamente están determinados temporalmente de manera fija por estímulo y dependen en gran medida del estado del sujeto (por ejemplo P300 o N400).

Los relacionados con evento (ERP) reflejan la actividad cerebral de poblaciones de neuronas activadas en sincronía que ocurre en preparación a eventos discretos (como el caso de potenciales preparatorios), sean internos o externos al sujeto; o bien se pueden presentar como respuesta a eventos (como el potencial P300, N2pc o P2). Son considerados manifestaciones neurológicas de funciones psicológicas específicas que reflejan procesos cognitivos, perceptuales o motores [28]. Corresponden a potenciales eléctricos de unos pocos microvolts, por lo tanto es extremadamente importante separar los componentes que constituyen la señal biológica de los que son ruido. Actualmente el ERP es uno de los métodos más ampliamente usado en neurociencia cognitiva para estudiar las relaciones fisiológicas de la actividad sensorial, perceptual y cognitiva.

### 1.1.3. Imaginería Motora

La imaginería es definida como una experiencia que se asemeja a la experiencia perceptual, pero que ocurre en ausencia del estímulo apropiado para que ésta sea relevante [136]. En general, se refiere a un cúmulo de habilidades que incluyen, entre otras: la imaginería visual, imaginería cinestésica, imaginería de imágenes en movimiento, o combinaciones de diferentes modalidades. En este marco, la imaginería motora representa el resultado del acceso consciente al contenido de intención de movimiento, el cual usualmente se realiza de manera no consciente durante la preparación del movimiento [93]. Se define como el acto de imaginar repetidamente un movimiento sin ejecutar el movimiento y con la intención primaria de adquirir y optimizar habilidades motoras [124].

La imaginería motora es una herramienta cognitiva utilizada estratégicamente para aprender y optimizar tareas específicas de movimiento [124]. Desde el punto de vista práctico se ha observado que el entrenamiento con imaginería en atletas, músicos y pacientes con enfermedad cerebro vascular se asocia a mejoras en la activación electromiográfica de los músculos involucrados cuando se comparan con controles sin imaginería [93]. Todo esto ha llevado a considerar que, desde una perspectiva funcional, las tareas mentales de imaginería sean fenómenos neurológicos que pueden ser altamente aprovechados en el diseño de un BCI.

## 1.2. Definición del Problema a Abordar

A pesar que la BCI se ha utilizado con múltiples fines, aún no existe claridad respecto a su implementación y a los métodos de procesamiento y análisis que se deberían emplear para un uso óptimo. Brevemente, debido a la pobre resolución espacial que poseen los métodos de registro no invasivos, como la electroencefalografía, algunos investigadores se han enfocado en los registros a nivel cortical. Sin embargo, éstos aun se encuentran en fase experimental debido al riesgo que implica su uso en humanos. Por esta razón, la mayoría de los estudios en BCI se han realizado con registros de superficie. No obstante, tampoco hay claridad respecto al número de electrodos a utilizar ni su localización. Otro problema frecuentemente observado es la falta de homogeneidad en los métodos utilizados para extracción de características y posterior procesamiento de las señales. Evidentemente, resulta altamente complejo y computacionalmente costoso procesar un alto número de datos contenidos en las señales cerebrales cuando se intenta establecer asociaciones entre éstas y diferentes tareas motoras. Otro problema importante corresponde al elevado costo de los sistemas implementados, situación que ha confinado su uso a centros especializados capaces de invertir grandes cantidades de dinero en investigación. Finalmente, la gran dificultad que implica el largo entrenamiento y aprendizaje del usuario para lograr el movimiento deseado ha llevado a que la interfaz cerebro computador esté confinada a pruebas sencillas y controladas, la mayoría de las veces al interior de laboratorios de investigación, impidiendo su uso en condiciones no controladas, propias de la vida diaria de las personas que podrían requerir su uso.

En un reciente estudio [97] se compara el rendimiento de dos sistemas de registro, uno profesional (Biosemi Active 2 EEG headset) y otro de bajo costo (Emotiv EPOC) en la clasificación de imaginería motora en la base de datos de la BCI Competition III, dataset V (3 sujetos). Por un lado, los autores observaron que el sistema profesional tuvo un mejor rendimiento (aproximadamente un 28 % mejor) cuando se usó 32 canales en comparación al mismo registro utilizando sólo la posición de los electrodos del Emotiv EPOC (14 canales). Sin embargo al comparar el registro de Biosemi sólo usando la posición de los canales de Emotiv EPOC en 3 sujetos versus el registro utilizando el sistema Emotiv EPOC propiamente tal en 2 sujetos, se observó un rendimiento 14,04 % mejor para este último. A pesar de este mejor resultado, sólo se obtuvo un rendimiento de 62,61 % en la clasificación de tres clases. Los autores de este estudio consideran que Emotiv EPOC puede ser una herramienta interesante para el desarrollo de aplicaciones que utilicen imaginería motora. Por otro lado, también señalan: «*Otras áreas de interés incluyen la prueba de las metodologías de preprocesamiento y clasificación, ya que cualquier mejora obtenida mediante el uso de registros de bajo costo razonablemente se trasladará a los sistemas médicos*» [97].

Considerando lo anterior, la presente Tesis apunta a una de las etapas críticas en el manejo de señales de EEG, el «PROCESAMIENTO DE LAS SEÑALES ELECTROENCEFALOGRÁFICAS».

Aun cuando se han descrito múltiples técnicas de análisis de datos en el uso de BCI, haciendo énfasis en la definición de atributos o características a procesar o en la determinación del mejor clasificador, no se ha considerado la etapa de procesamiento como un hito fundamental en el tratamiento de las señales. Por lo tanto se plantea como pregunta de investigación: ¿EL USO DE PROCESAMIENTO DE LA SEÑAL DE EEG CON ANÁLISIS DE COMPONENTES INDEPENDIENTES Y DESCOMPOSICIÓN DE MODO EMPÍRICO PUEDE AFECTAR EL RENDIMIENTO DE UN CLASIFICADOR DE TAREAS MENTALES?

En este documento se propone un nuevo método de procesamiento de señales de EEG,

basado en el uso automatizado y en cadena de Análisis de Componentes Independientes (ICA) y Descomposición de Modo Empírico (EMD) para mejorar el rendimiento de la clasificación de tareas de imaginación motora. Este método será validado mediante su uso en una base de datos pública, de esta manera se podrá comparar con el rendimiento entregado por otros métodos usados en la actualidad.

Particularmente el uso de ICA ha sido estudiado de manera frecuente en el análisis de señales de EEG, sin embargo un factor determinante que limita su aplicación en BCI es la dependencia del operador para eliminar componentes independientes, siendo esta eliminación establecida mediante inspección visual [72, 114], situación evidentemente incompatible con la implementación de un sistema automatizado de clasificación de tareas de imaginación motora en BCI. Por otra parte, el uso de EMD se ha descrito para clasificar tareas de imaginación motora en BCI, principalmente como estrategia en la extracción de características. En [38] se usó EMD espacio-temporal multivariado para determinar tareas unitarias de imaginación motora en 11 sujetos pertenecientes a la base de datos pública Physionet. En este estudio se usaron 5 IMFs como características y máquina de soporte vectorial como clasificador, obteniendo una sensibilidad de 75,7% y especificidad de 71,5%. En [134] se usaron las cuatro primeras funciones de modo intrínseco, generadas con EMD, como características para clasificar 2 tareas de imaginación motora en 5 sujetos pertenecientes a la base de datos BCI Competition III, dataset IVa. Aquí se usó como clasificador máquina de soporte vectorial de mínimos cuadrados, obteniendo un 97,56% de éxito en la clasificación. En [153] se usó EMD como estrategia en la selección de características para clasificar cinco estados de epilepsia en la base de datos pública Epilepsy Data de la Universidad de Bonn, aquí los IMFs fueron seleccionados de acuerdo al nivel de significancia de acuerdo a la prueba t-test, usando redes neuronales artificiales con algoritmo Back Propagation se observó un 87,3% de éxito en la clasificación. En [19] se usó EMD multivariado en combinación con transformada de Fourier en la extracción de características para la clasificación de 2 tareas de imaginación motora en la base de datos BCI Competition II dataset III, aquí se usó como clasificador el método de k vecinos más cercanos, obteniendo un 90,71% de éxito en la clasificación. En [46] se usó una combinación de EMD multivariado y reconstrucción espacio-fase para la construcción del vector de características en tareas cognitivas no motoras en 7 sujetos obtenidos de la base de datos Brain-Computer Interface de la Universidad Colorado State<sup>2</sup>. Aquí se usó como clasificador máquina de soporte vectorial de mínimos cuadrados, obteniendo un 83,33% de éxito en la clasificación.

Además se ha usado como herramienta para mejorar la localización de la información de frecuencia en EEG [111, 52]. Sólo en un estudio se usó como filtro de señal de EEG para reducir el ruido en una tarea de imaginación motora de fuerza de apriete de mano [150].

Sin embargo no se ha estudiado el efecto del uso automatizado y en cadena de ICA y EMD en el procesamiento de la señal de EEG previo a la extracción de características con el fin de mejorar los indicadores de clasificación de tareas de imaginación motora.

---

<sup>2</sup>Esta base de datos se encuentra disponible en el repositorio Brain-Computer Interfaces Laboratory de Colorado State University. Sin embargo los datos fueron registrados originalmente por Zak Keirn en Purdue University.



## 1.3. Objetivos

### 1.3.1. Objetivo General

Optimizar la clasificación de tareas de imaginación motora mediante el uso automatizado de ICA y EMD en la fase de procesamiento de señales de EEG obtenidas de una base de datos pública.

### 1.3.2. Objetivos Específicos

- Implementar un sistema automatizado de Análisis de Componentes Independientes y Descomposición de Modo Empírico en el procesamiento de señales de EEG.
- Validar el método propuesto usando una base de datos pública de EEG.
- Comparar la clasificación para tres clases entre las señales procesadas con ICA y EMD versus las señales que no fueron procesadas con este método.
- Comparar el rendimiento de los clasificadores Dagging, Máquina de soporte vectorial con optimización mínima secuencial, Bayes y Logistic en la clasificación de tres clases de las señales procesadas con ICA y EMD y las señales que no fueron procesadas con este método.

## 1.4. Hipótesis

«El procesamiento automatizado de señales de EEG con ICA y EMD mejora el porcentaje de éxito en la clasificación de las tareas de imaginación motora».

Se plantea la presente hipótesis considerando las características de los métodos propuestos. Particularmente ICA permitiría encontrar una representación lineal de múltiples fuentes sin poseer información de las señales originales ni de las ponderaciones de la mezcla, lo que facilitaría la selección de los componentes de interés de la señal de EEG. Por otra parte, se considera el uso de EMD, ya que resultaría útil para descomponer datos no estacionarios y no lineales, características propias de la señal de EEG.

La hipótesis será evaluada en la base de datos pública *Physionet*<sup>3</sup> (ver capítulo 3), la cual ha sido utilizada en diversas publicaciones, permitiendo de este modo comparar el rendimiento de la metodología propuesta con el estado del arte en cuanto a clasificación de tareas de imaginación motora.

Además, con el fin de evaluar la eficacia de la clasificación, se compararán diferentes clasificadores estadísticos. Éstos fueron seleccionados considerando las propiedades de las señales de EEG (no estacionariedad, no linealidad, variabilidad en el tiempo).

Se hace hincapié en que la hipótesis busca probar que el procesamiento con ICA y EMD mejora el porcentaje de éxito en la clasificación y no pretende evaluar un método de selección de características particular. Para tal efecto se comparará indicadores estadísticos de la clasificación obtenida con el método de procesamiento propuesto versus los resultados obtenidos sin este método (el cual llamaremos «Método Tradicional»).

---

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

## 1.5. Estructura de la Tesis

En el capítulo 2 se describe el estado del arte en cuanto a BCI, haciendo énfasis en los modelos y arquitectura de la BCI, métodos de procesamiento de señales de EEG, extracción y selección de características, y los métodos de clasificación usados. En el capítulo 3 se presenta la metodología propuesta en esta Tesis, destacando la descripción del método de procesamiento de la señal de EEG con Análisis de Componentes Independientes y Descomposición de Modo Empírico. Además se describen las características de la base de datos usada, el método de extracción de características y la clasificación propiamente tal. En el capítulo 4 se presentan los resultados obtenidos para la clasificación de tres clases con y sin el método propuesto. Además se comparan con los resultados de otras investigaciones que utilizaron base de datos públicas. En el capítulo 5 se presentan las conclusiones y discusión de la presente investigación.

# Capítulo 2

## Estado del Arte

### 2.1. Modelos y Arquitectura de la Interfaz Cerebro Computador

La BCI permite la comunicación en tiempo real entre el cerebro y algún dispositivo de salida, como un computador [155]. En un BCI la actividad cerebral debe ser registrada, digitalizada y clasificada de manera tal que esta información pueda generar una señal de salida, la cual será controlada por la actividad cerebral. Inicialmente el objetivo de una BCI fue permitir a un individuo, con discapacidad motora severa, tener control efectivo sobre algún dispositivo, como un computador, sintetizador de voz, prótesis u otro [20], aunque también fue usada con el fin de restaurar funciones cognitivas [155]. Actualmente los alcances de sus aplicaciones parecen inconmensurables y van más allá de los usos en el campo médico. En particular, el propósito de la BCI es aumentar o mejorar algunas capacidades humanas permitiendo la interacción con un computador a través de una modulación consciente o espontánea de las oscilaciones cerebrales después de un período de entrenamiento [48]. De este modo, la BCI se convierte en un sistema cerebro-actuado que permite canales alternativos de comunicación y control [48]. Un sistema de BCI es capaz de detectar ciertas características de la actividad cerebral (habitualmente asociada al movimiento) que tienen relación con la intención de controlar un movimiento [20]. Para reconocer esta actividad cerebral se emplean múltiples algoritmos del ámbito del procesamiento de señales, algunos serán discutidos en los Capítulos 2.2 y 2.3. La investigación acerca de sistemas BCI se inició en 1970 en la Universidad de California, Los Angeles. Inicialmente se concentró en el desarrollo de aplicaciones clínicas, particularmente en pacientes con alteraciones del sistema nervioso. Sin embargo, el desarrollo de nuevas tecnologías ha incrementado su espectro de aplicaciones y las áreas de investigación, de esta manera sus campos de acción han crecido considerablemente. De este modo, también se ha usado como alternativa de comunicación en individuos sanos [48]. Las modalidades de adquisición usadas para BCI, ya sea en humanos o primates no humanos incluyen EEG adquiridas desde electrodos localizados en el cuero cabelludo y registros específicos de neuronas a través de micro electrodos implantados. La técnica con EEG se considera un método seguro, además de bajo costo comparado con los métodos invasivos, pero tiene baja resolución espacial y pobre relación señal-ruido [155], provocado por la atenuación que generan los huesos del cráneo y por la contaminación de la señal debido a los músculos vecinos. Por el contrario el registro específico de unidades neuronales presenta una mejor resolución espacial, permitiendo además el uso de más canales independientes. Estos

permiten el registro de múltiples neuronas en un mismo tiempo otorgando la posibilidad de activar una prótesis mediante movimiento multiplanares [135]. Kennedy et al. describieron un método de implante de electrodos cónicos que son fijados con un sustrato que contiene factor de crecimiento nervioso. Este sistema requiere indemnidad del tejido neural de manera de permitir el crecimiento dentro del cono, proceso que ocurre en un período de varias semanas. Esta modalidad permite obtener registros por períodos prolongados [74, 77, 75, 73, 76]. Sin embargo, este tipo de registro requiere una gran proximidad con el tejido neuronal (alrededor de 100 micrones), por lo que no es recomendable para estudio en humanos por el alto riesgo clínico que esto implica, además de la falta de efecto en el tiempo debido a la respuesta cicatrizal en la zona de registro y del alto costo de implementación [147, 49]. Wolpaw y McFarland demostraron que métodos de registro no invasivos en humanos, a través de electrodos en el cuero cabelludo, permiten lograr control multidimensional del cursor de un computador, similar a los obtenidos con registros invasivos en monos [148]. Esto sugiere que personas con desordenes motores podrían controlar movimientos complejos sin necesidad del uso de registros con electrodos implantados. Además, los trabajos de Wolpaw [147, 148] y los de Pfurtscheller [116] muestran que personas sanas y con parálisis pueden realizar control voluntario de los ritmos sensorio-motores en la corteza motora a través de la imaginación de movimientos. Por otro lado, también existe evidencia que personas con deterioro motor pueden aprender a manipular potenciales eléctrico cerebrales, como los ritmos mu o beta [42, 129, 148]. Pfurtscheller localizó los ritmos sensorio motores en regiones cerebrales siguiendo la organización homuncular del cerebro. La imaginación del movimiento de la mano anuló los ritmos sensorio motores sobre la región de la mano, y el movimiento imaginario o real de las extremidades inferiores bloqueó los ritmos sensorio motores en el sulcus interhemisférico [115]. La mayoría de los estudios de BCI con EEG como sistema de registro han utilizado las bandas de frecuencias conocidas como ritmos mu y beta asociados a la corteza motora. Se piensa que los ritmos mu y beta son el producto de circuitos tálamo corticales que muestran atenuación de la amplitud durante la activación cortical. Esta supresión se conoce como Even Related Descynchronization (ERD) y puede ser inducida por movimiento real o imaginario. Los ritmos mu y beta están separados respecto a la secuencia de activación y a la distribución topográfica [133], pero tienden a mostrar supresión bilateral difusa (contra-lateral dominante) con la actividad motora. Adicionalmente, bandas de frecuencias más altas, conocidas como ritmos gamma, también han sido investigadas. La bandas gamma ( $>30$  Hz) están a menudo relacionadas con un incremento en la amplitud (Event Related Synchronization, ERS), la cual se asocia a una activación cortical. También se ha postulado que estarían vinculadas con programación motora, atención e integración sensorio motora [69]. Sin embargo estas altas frecuencias de oscilaciones no han sido utilizadas en sistemas BCI.

En términos generales un sistema de BCI está formado por cuatro módulos, el de adquisición de señales, el de preprocesamiento, el de procesamiento y el de actuación. El módulo de adquisición de señales está formado por el sistema de registro, habitualmente electrodos de superficie, considera además la amplificación y digitalización de la señal de EEG. El módulo de preprocesamiento incluye técnicas que buscan mejorar la calidad de la señal previo al procesamiento, entre éstas se cuentan el re-muestreo, normalización, referenciación y uso de filtros. El módulo de procesamiento tradicionalmente incluye dos etapas, extracción de características y clasificación. La etapa de extracción de características implica el uso de diversos métodos para filtrar y extraer información relevante (o características) a partir de los datos registrados. Esta información se puede obtener por ejemplo de los ritmos sensorio-motores,

potenciales evocados como el P300 u otro. La etapa de clasificación permite identificar las tareas de imaginación realizadas por el usuario del BCI mediante la detección de un patrón de actividad conocido previamente, para ello utiliza las características extraídas de la señal de EEG. El módulo de actuación desarrolla la acción correspondiente, de esta manera se logra el control del dispositivo deseado, por ejemplo: silla de ruedas, cursor del computador, neuroprótesis, etc. La figura 2.1 muestra un diagrama de bloques con estos módulos básicos.

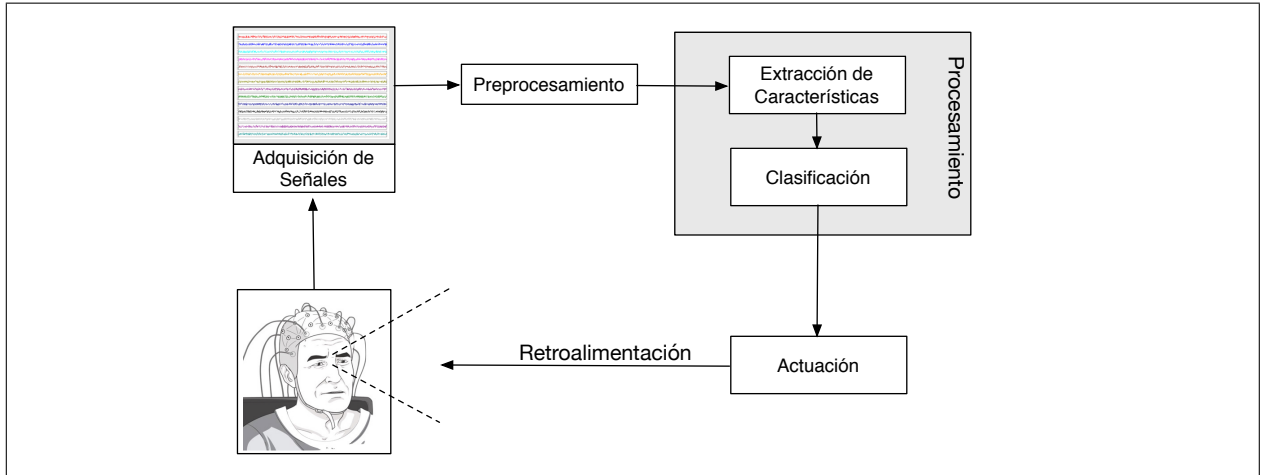


Figura 2.1: Arquitectura tradicional de un sistema BCI. Notar que el módulo de procesamiento tradicionalmente considera las etapas de extracción de características y clasificación.

Como se mencionó anteriormente, la mayoría de las tecnologías de BCI se basa en el uso de EEG como sistema de registro. Sin embargo, éste presenta desventajas como la elevada impedancia de entrada, la mala relación señal-ruido o variabilidad en el posicionamiento de los electrodos [48] y la baja resolución espacial [155]. A pesar de ello, el EEG se utiliza en más del 80% de los trabajos publicados en BCI [9]. A grandes rasgos, existen dos tipos de extracción de señales desde el cerebro, BCI invasivo y BCI no invasivo. En el BCI invasivo se utilizan arreglos de electrodos implantados directamente en el cerebro, aquí se registra la actividad de cientos de neuronas, obteniéndose una alta calidad de señal. Esta tecnología ha sido explorada principalmente en primates no humanos.

Los experimentos realizados en monos rhesus por [30] permitieron registrar la actividad de amplias poblaciones de neuronas cerebrales para generar movimiento en un brazo robótico. Los monos fueron entrenados para alcanzar y tomar objetos en una pantalla de computador mediante la manipulación de un joystick, mientras el correspondiente movimiento de un brazo robótico ocurría de manera oculta. Posteriormente los monos aprendieron a controlar el brazo robótico al ver sus movimientos [30]. La figura 2.2 muestra la metodología empleada.

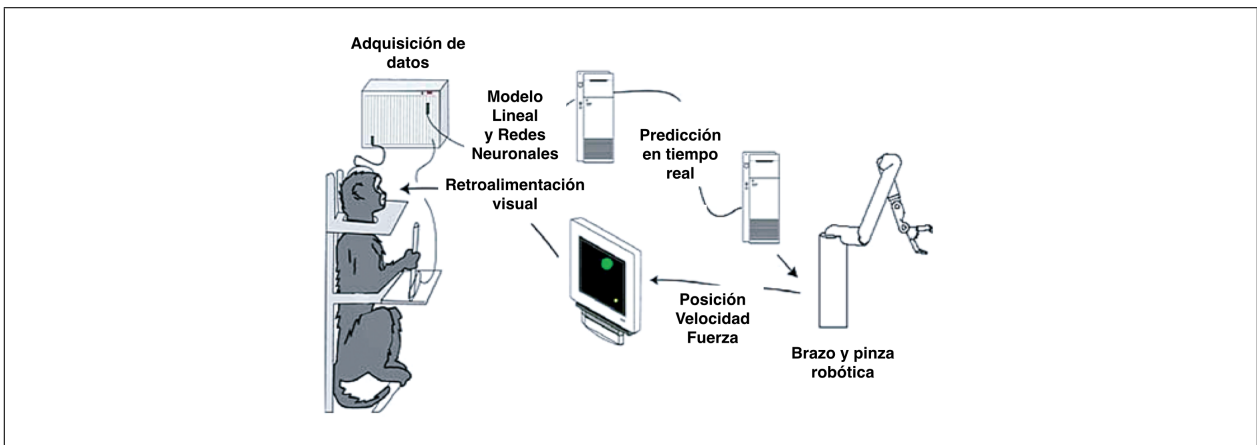


Figura 2.2: BCI invasivo para el control de un brazo robótico en primates (modificado de [30]).

El BCI no invasivo corresponde al método de registro indirecto, donde los sensores no están en contacto directo con el tejido nervioso. El mecanismo más estudiado es la EEG, aquí los electrodos de registro son de superficie y se posicionan en el cuero cabelludo, registrando la actividad de millones de neuronas. A pesar de las desventajas del uso de la EEG, ésta presenta buena resolución temporal, es fácil de usar, es de menor costo que un sistema invasivo y se reduce el riesgo para el sujeto examinado [28].

## 2.2. Preprocesamiento de la Señal

La naturaleza no estacionaria de la señal de EEG [106], la no linealidad, la variabilidad temporal y la presencia de ruido contaminante, sumado al hecho que las diversas conductas y estados mentales cambian las propiedades estadísticas de ésta, hacen necesario un enfoque diferente (no lineal, variante en el tiempo y no estacionario) en el procesamiento de las señales de EEG.

La variabilidad temporal en las tareas de imaginación motora se debe a la cambiante dinámica cerebral y dificultad para desarrollar tareas mentales por periodos prolongados de tiempo [106]. Esto obliga a obtener datos relevantes en periodos cortos de registro y transforma en esencial la etapa de preprocesamiento de la señal de EEG, ya sea para eliminar artefactos o factores contaminantes en el dominio del tiempo, de la frecuencia o de ambos [103].

El advenimiento de nuevas tecnologías ha permitido el desarrollo de sistemas de adquisición inalámbricos, con electrodos secos y de bajo costo, los que tienen un gran potencial para el uso en la vida diaria, en la industria o el entretenimiento. Sin embargo, estas tecnologías han traído nuevos desafíos en cuanto al preprocesamiento de la señal de EEG.

Considerando el objetivo general de este documento, así como la hipótesis planteada, se hace necesario describir algunos métodos utilizados para preparar la señal previo a la fase de selección/extracción de características y de clasificación.

En [20] se revisaron 96 estudios de BCI que utilizaron técnicas de preprocesamiento para mejorar la señal antes de extraer características, de ellos 32 % usan filtros Laplacianos, 22 % usan Análisis de Componentes Principales o Análisis de Componentes Independientes y 11 % usan técnicas de referencia común.

A continuación se detallan los métodos más frecuentemente usados en la etapa de procesamiento de señales de EEG de acuerdo a [20].

## 2.2.1. Filtros Espaciales

El propósito de un filtro espacial es reducir el efecto de la distorsión espacial en la señal de EEG original. Esta distorsión espacial ocurre como efecto de la separación que existe entre el sensor y las fuentes de la señal de EEG en el cerebro. Como se mencionó en el capítulo 1.1.1, la señal de EEG corresponde a la suma espacial de corrientes inducidas por potenciales post-sinápticos sincronizados que ocurren en grupos de neuronas, éstos potenciales eléctricos corticales son conducidos a través del tejido circundante (volumen de conducción), lo que tiene como consecuencia una pobre resolución espacial en la señal de EEG [24, 151].

### 2.2.1.1. Filtros de Patrones Espaciales Comunes

En estricto rigor, los filtros de patrones espaciales comunes (CSP, por sus sigla en inglés) corresponden a la proyección ortogonal de la señal de EEG desde varios electrodos por medio del discriminante entre dos clases. Si  $\Sigma^{(+)} \in \mathbb{R}^{C \times C}$  y  $\Sigma^{(-)} \in \mathbb{R}^{C \times C}$  corresponden a las matrices de covarianza de la señal de EEG (filtrada pasa banda) en dos condiciones, ejemplo: imaginación mano derecha (+) e imaginación mano izquierda (-).

$$\Sigma^{(C)} = \frac{1}{|I_C|} \sum_{i \in I_C} x_i x_i^T, (C \in [+,-]) \quad (2.1)$$

donde  $I_C (C \in [+,-])$  es el set de índices correspondientes a los ensayos que pertenecen a la condición o clase  $C$ ,  $|I|$  denota el tamaño del set  $I$ ,  $\mathbf{x} \in \mathbb{R}^{Ch \times S}$  es un segmento de señal de EEG correspondiente a un ensayo de imaginación de movimiento,  $Ch$  es el número de canales y  $S$  es el número de muestras. Es necesario considerar que la señal está centrada y escalada (z-score). La matriz promedio de covarianza espacial normalizada  $\Sigma^{(+)}$  y  $\Sigma^{(-)}$  es obtenida mediante el promedio de todas la matrices de covarianza de cada grupo de registros para cada clase. La matriz de covarianza espacial compuesta está dada por:

$$\bar{\Sigma} = \Sigma^{(+)} + \Sigma^{(-)} \quad (2.2)$$

esta se puede factorizar como

$$\bar{\Sigma} = \Sigma^{(+)} + \Sigma^{(-)} = U \Lambda U^T \quad (2.3)$$

donde  $U$  son los vectores propios correspondientes a matriz diagonal de valores propios  $\Lambda$ . Los valores propios se consideran ordenados en orden descendente. Luego, mediante blanqueamiento (ver Capítulo 3.2.1)

$$P = \Lambda^{-\left(\frac{1}{2}\right)} U^T \quad (2.4)$$

se transforma las matrices de covarianza promedio como

$$\hat{\Sigma}^{(+)} = P \Sigma^{(+)} P^T \quad (2.5)$$

$$\hat{\Sigma}^{(-)} = P \Sigma^{(-)} P^T \quad (2.6)$$

notar que  $\hat{\Sigma}^{(+)} + \hat{\Sigma}^{(-)} = \mathbf{I}$ , matriz de identidad. Además  $\hat{\Sigma}^{(+)} + \hat{\Sigma}^{(-)}$  comparten la misma matriz de vectores propios. Por lo tanto, cualquier matriz ortonormal  $\mathbf{B}$  satisface  $\mathbf{B}^T(\hat{\Sigma}^{(+)} + \hat{\Sigma}^{(-)})\mathbf{B} = \mathbf{I}$ , finalmente

$$\hat{\Sigma}^{(+)} = \mathbf{B}\Lambda^{(+)}\mathbf{B}^T \quad (2.7)$$

$$\hat{\Sigma}^{(-)} = \mathbf{B}\Lambda^{(-)}\mathbf{B}^T \quad (2.8)$$

donde  $\mathbf{B}$  es un conjunto de vectores propios y  $\Lambda$  es la matriz diagonal de los valores propios. Esto implica que los valores propios de  $\Lambda^{(+)}$  se ordenan en orden descendente y al mismo tiempo los valores propios de  $\Lambda^{(-)}$  se clasifican en orden ascendente. Por lo tanto el autovector asociado con el autovalor más grande tiene la varianza más grande para la clase (+) y tiene la menor varianza para la clase (-) y viceversa. Por lo tanto, si se proyectan señales EEG blanqueadas sobre el primer y último vector propios de  $\mathbf{B}$ , se obtendrá la información discriminatoria óptima para el proceso de clasificación. El mapeo de cada ensayo se realiza usando la matriz de proyección

$$\mathbf{w} = P^T\mathbf{B} \quad (2.9)$$

Entonces  $\mathbf{w}$  corresponde al filtro espacial que maximiza la varianza para una clase y minimiza la varianza para la otra. Con la matriz de proyección  $\mathbf{w}$  la señal de EEG original puede ser transformada en componentes no correlacionados

$$s(t) = \mathbf{w}^T x(t) = (s_1(t), \dots, s_d(t))^T \quad (2.10)$$

Aquí el primer filtro CSP entrega la máxima varianza de la clase (+) y el último filtro CSP entrega la máxima varianza de la clase (-). La dimensión  $d$  de los filtros CSP estará dada por el número de los primeros y últimos  $m$  filtros seleccionados.

Finalmente el vector de característica estará dado por

$$f_i = \log \left( \frac{\text{var}[s_i(t)]}{\sum_{i=1}^d \text{var}[s_i(t)]} \right) \quad (2.11)$$

En resumen, este método detecta patrones en señales de EEG mediante la construcción de filtros espaciales que maximizan la varianza de una tarea y simultáneamente minimizan la varianza de la otra [81]. Una ventaja de este método es que no requiere una selección a priori de una banda de frecuencia específica. Sin embargo, tiene la desventaja que necesita ser usado con múltiples electrodos y su desempeño puede ser afectado al cambiar de posición los electrodos.

El principal problema de este método es su sensibilidad a los artefactos en las señales de EEG, ya que las matrices de covarianza son usadas como base para el cálculo de estos filtros. Por lo tanto, un registro contaminado con artefactos puede causar cambios extremos en los filtros. Considerando que los filtros espaciales comunes detectan patrones espaciales en la EEG, cualquier cambio en la posición de los electrodos puede hacer inútiles las mejoras en la precisión de la clasificación obtenida por este método. Por lo tanto, este método requiere posiciones de electrodos casi idénticas para todos los ensayos y sesiones, lo que puede ser difícil de lograr [151, 24, 112, 158, 107, 57, 132, 144].



### 2.2.1.2. Referencia Común Promedio

La referencia común promedio (CAR, por su sigla en inglés) corresponde a un filtro que permite la sustracción de la actividad común de EEG, dejando sólo la actividad de cada electrodo específico de la EEG. El potencial de cada electrodo después del filtrado podría ser calculado como:

$$x_t^{CAR}(t) = x_i(t) - \frac{1}{C} \sum_{j=1}^C x_j(t) \quad (2.12)$$

donde  $x_t^{CAR}(t)$  es la salida filtrada del  $i$ -ésimo electrodo,  $x_i(t)$  es el potencial entre el  $j$ -ésimo electrodo y la referencia,  $C$  es el número total de electrodos en el sistema de registro<sup>1</sup>.

En [151] se demostró que el filtro espacial de referencia común promedio entregó mejores resultados al compararlo con filtros Laplacianos y filtros Espaciales Comunes.

### 2.2.1.3. Filtro Laplaciano

Este filtro es usado para mejorar la relación señal-ruido y como un identificador de la fuente en una señal de EEG. Este filtro resuelve el problema del electrodo de referencia y es robusto contra artefactos generados en zonas no cubiertas por los electrodos. Además se considera un método con alta resolución espacial. Sin embargo es demasiado sensible a la elección de parámetros durante la interpolación de la función *spline*<sup>2</sup> [65].

El filtro corresponde a una medida de la segunda derivada espacial de una señal. Habitualmente se aplica a una señal que previamente ha sido suavizada con alguna aproximación Gaussiana para reducir su sensibilidad al ruido [151]. El filtro Laplaciano se puede describir como<sup>3</sup>:

$$x_i^{LAP}(t) = x_i(t) - \sum_{j \in S_i} \omega_{ij} x_j(t) \quad (2.13)$$

$$\omega_{ij} = \frac{1/d_{ij}}{\sum_{j \in S_i} 1/d_{ij}} \quad (2.14)$$

donde  $x_i^{LAP}(t)$  es la señal filtrada del electrodo  $i$ ,  $x_i(t)$  es el potencial del electrodo  $i$  comparado al electrodo de referencia,  $\omega_{ij}$  es la constante peso calculada por la ecuación 2.14, donde  $d_{ij}$  es la distancia Euclidiana desde el electrodo  $i$  al electrodo  $j$ ,  $S_i$  es el set de electrodos vecinos al electrodo  $i$ .

---

<sup>1</sup>CAR corresponde a la abreviación de Common Average Reference.

<sup>2</sup>El término *spline* hace referencia a una amplia clase de funciones que son utilizadas en aplicaciones que requieren la interpolación de datos, o un suavizado de curvas.

<sup>3</sup>LAP: Laplaciano.

## 2.2.2. Análisis de Componentes Principales

El Análisis de Componentes principales (PCA, acrónimo en inglés de Principal Components Analysis) consiste en encontrar transformaciones ortogonales de las variables originales para conseguir un nuevo conjunto de variables no correlacionadas, denominadas Componentes Principales, que se obtienen en orden decreciente de importancia. Los componentes son combinaciones lineales de las variables originales y se espera que sólo unos pocos (los primeros) recojan la mayor parte de la variabilidad de los datos, obteniéndose una reducción de la dimensión en los mismos. Luego, el propósito fundamental de la técnica consiste en la reducción de la dimensión de los datos con el fin de simplificar el problema en estudio. Desde el punto de vista gráfico (bidimensional), el subespacio quedará definido mediante un sistema de referencia, determinado por vectores perpendiculares, el primero corresponde a la dirección en la que más varían los datos, el segundo, perpendicular al primero, recogiendo la mayor parte de la variabilidad restante. Los vectores del sistema de referencia definen nuevas variables, que son combinaciones lineales de las variables de partida y se denominan componentes principales. De esta forma, podemos reducir la dimensión seleccionando solamente a los primeros componentes. La reducción de la dimensión se deriva del hecho que las variables están relacionadas entre sí y, por tanto, tienen información común. De esta manera, la información común a todas ellas se extrae en los componentes principales. En muchas ocasiones es difícil encontrar el significado de los componentes, como variables compuestas, por lo que el uso principal de la técnica es la reducción de la dimensión como paso previo a la aplicación de otros análisis posteriores, por ejemplo, Análisis de Componentes Independientes.

Hay dos enfoques ampliamente aceptados para realizar PCA: (1) La descomposición de valores propios (eigenvalue) de la matriz de covarianza (o de la matriz de correlación); (2) La descomposición de valor singular (SVD) de la matriz de datos centrada [105].

En el presente estudio sólo se describe el método basado en la maximización de la varianza a partir de la descomposición de valores propios de la matriz de covarianza.

Dadas  $\mathbf{x}$  variables originales  $(x_1, x_2, \dots, x_p)$  se quiere calcular, a partir de ellas, un nuevo conjunto de variables  $y_1, y_2, \dots, y_p$ , no correlacionadas entre sí, cuyas varianzas vayan decreciendo progresivamente, donde cada  $y_j$  ( $j = 1, \dots, p$ ) es una combinación lineal de las variables originales  $(x_1, x_2, \dots, x_p)$ , es decir

$$y_j = a_{1j}x_1 + a_{2j}x_2 + \dots + a_{pj}x_p = \mathbf{a}'_j \mathbf{x} \quad (2.15)$$

siendo  $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{pj})$  un vector de constantes, y

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$$

si se quiere maximizar la varianza, se pueden aumentar los coeficientes  $a_{ij}$ . Por lo cual, para mantener la ortogonalidad de la transformación se impone que el módulo del vector  $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{pj})$  sea 1, es decir  $\mathbf{a}'_j \mathbf{a}_j = 1$ . Luego, el primer componente se calcula eligiendo  $\mathbf{a}_1$  de modo que  $y_1$  tenga la mayor varianza posible. El segundo componente principal se calcula obteniendo  $\mathbf{a}_2$  de modo que la variable  $y_2$  no esté correlacionada con  $y_1$ . Del mismo modo se eligen  $y_3 \dots y_p$ , de manera que las variables obtenidas vayan teniendo cada vez menor varianza.

Para el caso de  $\mathbf{a}_1$ , se tiene que

$$Var(y_1) = Var(\mathbf{a}'_1 \mathbf{x}) = \mathbf{a}'_1 \Sigma \mathbf{a}_1 \quad (2.16)$$

donde  $\Sigma$  corresponde a la matriz de covarianza<sup>4</sup>. Por lo tanto el problema consiste en maximizar la función  $\mathbf{a}'_1 \Sigma \mathbf{a}_1$  sujeta a la restricción  $\mathbf{a}'_1 \mathbf{a}_1 = 1$ . luego, basado en el método de los multiplicadores de Lagrange se busca maximizar una función de varias variables sujeta a restricciones. Si el vector desconocido que permite la combinación lineal óptima es  $\mathbf{a}_1$ , se tiene que la nueva función es  $L(\mathbf{a}_1) = \mathbf{a}'_1 \Sigma \mathbf{a}_1 - \lambda(\mathbf{a}'_1 \mathbf{a}_1 - 1)$ , para buscar el máximo se deriva e iguala a cero, obteniendo  $(\Sigma - \lambda I)\mathbf{a}_1 = 0$ . Para que el sistema tenga una solución distinta de 0 la matriz  $(\Sigma - \lambda I)$  tiene que ser singular. Esto implica que el determinante debe ser igual a cero:  $|\Sigma - \lambda I| = 0$ . De lo anterior se tiene que  $\lambda$  es un autovalor de  $\Sigma$ , la matriz de covarianza, la cual es de orden  $p$  y se define como positiva. Por lo tanto tendrá  $p$  autovalores distintos  $\{\lambda_1, \lambda_2, \dots, \lambda_p\}$  tales que  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ . Luego, para maximizar la varianza de  $y_1$  se tiene que tomar el mayor autovalor,  $\lambda_1$ , y el correspondiente autovector  $\mathbf{a}_1$ . Este último es un vector que otorga la combinación de las variables originales que tiene mayor varianza, esto es, si  $\mathbf{a}'_1 = (a_{11}, a_{12}, \dots, a_{1p})$ , entonces  $y_1 = \mathbf{a}'_1 \mathbf{x} = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$ .

### 2.2.3. Análisis de Componentes Independientes

El análisis de componentes independientes (ICA, acrónimo en inglés de Independent Components Analysis) es una técnica iterativa que tiene por objetivo encontrar una representación lineal de múltiples fuentes sin poseer información de las señales originales ni de las ponderaciones de la mezcla [121]. Se basa en el problema de la separación ciega de fuentes, donde se considera que las señales recibidas por los sensores son mezclas que proceden de varias fuentes estadísticamente independientes entre sí. El objetivo aquí será tomar estas mezclas y obtener a partir de ellas las señales originales puras [8]. Es decir, identificar las señales independientes y los artefactos asociados a la señal original, permitiendo separar el ruido de la señal. El siguiente ejemplo es mencionado comúnmente en la literatura [8, 104, 64] para explicar el uso de ICA: Si se imagina a  $n$  personas en un salón hablando simultáneamente y se tienen diferentes micrófonos ubicados en distintas locaciones. Los micrófonos entregarán señales temporales  $x_1(t), x_2(t), \dots, x_n(t)$ , las cuales serán sumas ponderadas de las señales de voz, denotadas por  $s_1, s_2, \dots, s_n$ . Como se expresan en la ecuación lineal:

$$\begin{aligned} x_1(t) &= a_{11}s_1 + a_{12}s_2 + \dots + a_{1n}s_n \\ x_2(t) &= a_{21}s_1 + a_{22}s_2 + \dots + a_{2n}s_n \\ &\vdots \\ x_n(t) &= a_{n1}s_1 + a_{n2}s_2 + \dots + a_{nn}s_n \end{aligned} \quad (2.17)$$

---

4

$$\begin{aligned} \mathbf{a}'_1 \Sigma \mathbf{a}_1 &= \mathbf{a}'_1 E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'] \mathbf{a}_1 \\ E[\mathbf{a}'_1 (\mathbf{x} - \mu)(\mathbf{x} - \mu)' \mathbf{a}_1] &= E[\mathbf{a}'_1 (\mathbf{x} - \mu)]^2 = Var(\mathbf{a}'_1 \mathbf{x}) \end{aligned}$$

donde  $E(\Phi)$  corresponde a la esperanza matemática de la distribución de probabilidad de una variable aleatoria  $\Phi$  y  $\mu = E(\Phi)$ , esto implica la propiedad  $E(\Phi - \mu) = 0$ , y  $Var(\Phi) = E[(\Phi - \mu)^2]$

donde  $a$  corresponde a parámetros que dependen de la distancia de los micrófonos a las personas. De este modo el problema consiste en estimar las señales de voz originales  $s_1(t), s_2(t), \dots, s_n(t)$  a partir de los registros  $x_1(t), x_2(t), \dots, x_n(t)$ . Este caso es conocido como el problema del cóctel (cocktail party problem). Evidentemente si el parámetro  $a$  es conocido, fácilmente se podría resolver estas ecuaciones a través de métodos tradicionales. Sin embargo, si  $a$  es desconocido la solución se vuelve difícil. Una aproximación a esta solución se puede lograr asumiendo que cada señal original es estadísticamente independiente de otra en cada instante de tiempo. Este supuesto es cierto en muchos casos y en otros no necesariamente debe serlo. Sin embargo, ICA sólo es efectivo si se asume como condición que las señales de entrada son estadísticamente independientes entre sí.

Para efectos del análisis es conveniente usar notación matricial y no suma de variables. Aquí se define  $\mathbf{x}$  como el vector, cuyos elementos son las mezclas  $x_1, \dots, x_n$  y  $\mathbf{s}$  el vector con elementos  $s_1, \dots, s_n$ . Además se define la matriz  $\mathbb{A}$  cuyos elementos son  $a_{ij}$ . De este modo el modelo de mezcla anterior se redefine como

$$\mathbf{x} = \mathbb{A}\mathbf{s} \quad (2.18)$$

El objetivo consiste en recuperar el vector de fuentes  $\mathbf{s}$ , disponiendo únicamente de un vector  $\mathbf{x}$  de  $n$  observaciones, para ello se debe estimar la matriz  $\mathbb{A}$  calculando su inversa, la cual denominaremos  $\mathbb{W}$ , de esta manera se puede obtener el componente independiente por

$$\mathbf{s} = \mathbb{W}\mathbf{x} \quad (2.19)$$

es decir, se busca calcular una matriz inversa  $\mathbb{W}$  con algunos cambios en escala y permutaciones, de modo que se puedan recuperar las señales originales de forma independiente.

Por simplicidad se asumirá que los componentes independientes tienen distribución idéntica, por tanto para estimar uno de los componentes independientes consideramos una combinación lineal de  $x_i$  (ver ecuación 2.18), ésta la denotaremos  $y = \mathbf{w}^T \mathbf{x} = \sum_i w_i x_i$ , donde  $\mathbf{w}$  es un vector a ser determinado. Si  $\mathbf{w}$  fuera una de las filas de la matriz inversa de  $\mathbb{A}$  esta combinación lineal sería igual a uno de los componentes independientes. Ahora bien, no es posible determinar  $\mathbf{w}$  tal que sea igual a una fila de la matriz inversa de  $\mathbb{A}$ , pero en la práctica es posible encontrar un estimador que da una buena aproximación, para ello estableceremos una nueva definición  $\mathbf{z} = \mathbb{A}^T \mathbf{w}$ . Luego, de acuerdo a lo antes descrito se tiene

$$y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbb{A}\mathbf{s} = \mathbf{z}^T \mathbf{s} \quad (2.20)$$

Por lo tanto  $y$  es una combinación lineal de  $s_i$ , con pesos dados por  $z_i$ . Como una suma de dos variables aleatorias independientes es más gaussiana que las variables originales<sup>5</sup>,  $\mathbf{z}^T \mathbf{s}$  es más gaussiana que cualquier  $s_i$ , y se vuelve menos gaussiano cuando se iguala a un  $s_i$ , aquí sólo uno de los elementos  $z_i$  de  $\mathbf{z}$  es diferente de cero [31]. Por lo tanto se debe elegir un vector  $\mathbf{w}$  que maximice la no gaussianidad de  $\mathbf{w}^T \mathbf{x}$ . Este vector debería corresponder a un  $\mathbf{z}$

---

<sup>5</sup>Según el Teorema Central del Límite, la distribución de la suma de variables aleatorias independientes tiende a ser Gaussiana. Por lo tanto la suma de dos variables aleatorias independientes tendrá una distribución más cercana a la Gaussiana que cualquiera de las dos variables aleatorias originales.

que sólo tenga un componente diferente de cero, esto significa que  $\mathbf{w}^T \mathbf{x} = \mathbf{z}^T \mathbf{s}$  es igual a uno de los componentes independientes.

En términos generales el éxito del algoritmo ICA depende que se cumplan las siguientes condiciones [41]:

1. Las fuentes deben ser estadísticamente independientes<sup>6</sup> (componentes independientes).
2. Las fuentes deben tener una distribución no Gaussiana (como máximo sólo una de ellas puede tener distribución Gaussiana).
3. La mezcla debe ser una combinación lineal de fuentes independientes.
4. No hay retraso desde las fuentes a los sensores
5. Debe existir el mismo número de fuentes y de sensores

ICA se ha usado tradicionalmente como herramienta de procesamiento, previo a la extracción de características, con el fin de remover artefactos. Sin embargo, aún no existe consenso respecto a los criterios para seleccionar o eliminar los componentes independientes, utilizándose en áreas biológicas criterios dependientes del operador, basados en inspección visual de los estos componentes [72, 114]. Situación que, evidentemente, se transforma en una limitación al impedir que la selección de componentes independientes, la posterior eliminación de los componentes seleccionados y finalmente la reconstrucción de la señal sea realizada de manera automatizada para su uso en BCI.

Por otra parte, aunque ha sido considerado un método robusto para eliminar artefactos, también se ha señalado que la supresión de artefactos podría afectar la densidad espectral de potencia [107, 140].

Es importante destacar que ICA también ha mostrado buenos resultados para eliminar señales contaminantes de fuentes no deseadas en el EEG, como señales de origen muscular o cardíaco [146].

## 2.2.4. Comparación de Métodos

Como se mencionó anteriormente el preprocesamiento de la señal de EEG juega un importante rol previo a la siguiente etapa en una estructura de BCI, esto es, la extracción de características y la clasificación dependerán de la eficiencia del método empleado en el preprocesamiento de la señal.

La mayoría de los métodos se enfocan en filtros espaciales (Laplaciano, PCA, ICA, CSP) y en filtros temporales (Fourier, métodos autoregresivos). Sin embargo no hay un método estándar para ser aplicado [65].

En [81, 65, 107] se realizaron revisiones de las diferentes etapas que componen la arquitectura de un BCI. En la tabla 2.1 se presenta un resumen de estos estudios con las ventajas y desventajas de los métodos citados durante la etapa de preprocesamiento en BCI.

---

<sup>6</sup>En teoría de probabilidades, se dice que dos sucesos aleatorios son independientes entre sí cuando la probabilidad de cada uno de ellos no está influida porque el otro suceso ocurra o no, es decir, cuando ambos sucesos no están relacionados.

Tabla 2.1: Comparación de métodos de preprocesamiento de señales de EEG. Resumen de acuerdo a lo presentado en [81, 65, 107].

Comparación de Métodos de Preprocesamiento de EEG		
Método	Ventajas	Desventajas
CSP	Trabaja bien en imaginería motora / No requiere selección específica de frecuencias / Algunas ubicaciones de electrodos ofrecen información más discriminatoria para algunas actividades específicas del cerebro que otras	Requiere múltiples electrodos / Su desempeño es afectado por la resolución espacial / Cambios en la posición de los electrodos puede afectar la posterior clasificación.
PCA	Reduce la dimensión de los datos / Permite hacer blanqueamiento (cap. 3.2.1)	No funciona bien como filtro / Requiere que los artefactos no estén correlacionados con la señal de EEG
CAR	Excelente método para referenciar / Mejora la relación Señal/Ruido	Necesita cubrir grandes áreas de Registro / La densidad de muestra finita causa problemas al calcular el promedios
LAP	Robusto contra artefactos generados en zonas no cubiertas por los electrodos / Movimientos oculares pueden ser eliminados durante la adquisición de la señal.	Sensible a los artefactos / Sensible a la elección de parámetros de la función <i>spline</i>
ICA	Buen desempeño para registros con múltiples electrodos / Robusto para la eliminación de artefactos	Requiere un alto costo computacional / Asume independencia estadística de las fuentes / Puede afectar el espectro de potencia / Eficacia depende del método de selección de componentes independientes

## 2.3. Procesamiento

Si se considera un BCI como un sistema de reconocimiento de patrones que clasifica clases (diferentes tareas mentales) de acuerdo a las características de estos patrones, resulta fundamental extraer información relevante que caracterice dichos patrones. Éstas se denominan *características* o *atributos* y reflejan similitudes con una determinada clase, así como diferencias con el resto de las otras. De este modo, las características se miden o se derivan de las propiedades de las señales que contienen la información discriminativa necesaria para distinguir sus diferentes tipos.

De acuerdo a lo presentado en la figura 2.1, el módulo de procesamiento considera habitualmente las etapas de *Extracción de Características y Clasificación*. En este capítulo sólo se describirán los métodos de extracción de características, mientras que la etapa de clasificación se abordará en el Capítulo 2.4.

Las múltiples características se pueden extraer de varios canales y de varios segmentos de tiempo antes de concatenarse en un solo vector de característica. Una de las principales dificultades en el diseño de BCI es la elección de características relevantes a partir de la gran cantidad de características posibles.

Dos métodos ampliamente usados en la etapa de procesamiento corresponden a PCA e ICA (previamente descritos). Particularmente PCA, además de ser utilizado durante el preprocesamiento de la señales de EEG también es usado para la extracción de características. Su uso está orientado principalmente para proyectar los datos o reducir dimensionalidad, esto último permite seleccionar atributos o características. Sin embargo, este método presenta pobres resultados ya que asume linealidad y continuidad de la señal de EEG, la cual no cumple estas condiciones [65].

En [151] se evalúa el efecto de PCA para reducción de características en señales no estacionarias de EEG basadas en imaginación motora, demostrando que el uso de PCA no ayuda a mejorar el desempeño de los clasificadores, pero sí mantiene el nivel de la clasificación a pesar de disminuir la dimensionalidad de los datos.

En [131] se usa PCA como herramienta para disminuir la dimensionalidad de los datos evitando información redundante, considerando que los cinco primeros valores propios contienen más del 99 % de la información. Estos autores [131] observaron que el uso de 64 canales de registro y el uso de sólo los canales C3, C4 y Cz producía prácticamente el mismo resultado en la clasificación, mientras que el uso de PCA entregó peores resultados.

En [58] se plantea que el uso de PCA muestra malos resultados debido a la dificultad para determinar buenas características en el dominio espacial (selección de canales), temporal o de frecuencia debido a la baja relación señal-ruido.

En [51] se comparan tres métodos de selección de características basados en reducción de dimensionalidad: PCA, Locality Preserving Projections y Análisis Discriminante Local de Fisher (LFDA), encontrando el mejor desempeño con LFDA y el peor con PCA.

De acuerdo a [107], PCA no siempre garantiza una buena clasificación cuando éste ha sido usado para reducir dimensionalidad, ya que los mejores componentes discriminantes pueden no estar incluidos en los componentes principales, de este modo no se garantiza que las características seleccionadas sean óptimas para clasificación. Por otro lado también se destaca que su uso para detectar artefactos requiere que éstos no estén correlacionados con la señal de EEG.

A continuación se presentan los métodos más usados en la etapa de extracción de características.

### 2.3.1. Enfoque Autoregresivo

El enfoque autoregresivo (AR) modela la señal de EEG como una señal aleatoria de salida de un filtro lineal invariable en el tiempo, donde la entrada es ruido blanco con media cero y una varianza constante. El objetivo de este método es obtener los coeficientes de filtro (parámetros), los cuales serán usados como características de la señal para la clasificación<sup>7</sup>.

---

<sup>7</sup>Existen varios métodos para calcular los coeficientes de filtro o parámetros, como: ecuaciones de Yule-Walker, Burg, covarianza y algoritmos forward-backward [70, 107].

Este método ha sido usado para modelar señales de EEG mediante la representación de cada canal (electrodo) como una combinación lineal de la señal en puntos previos; es decir, la variable de salida depende linealmente de sus propios valores anteriores:

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad (2.21)$$

donde  $\varphi_1, \dots, \varphi_p$  son los parámetros del modelo,  $c$  es una constante y  $\varepsilon_t$  es ruido blanco.

Los coeficientes o parámetros resultantes pueden ser usados para estimar la densidad espectral de potencia de la señal de EEG como:

$$X(\omega) = \frac{\sigma_Z^2}{|1 - \sum_{k=1}^p \varphi_k e^{-i2\pi f k}|^2} \quad (2.22)$$

donde  $\sigma_Z^2$  es la varianza del ruido,  $\varphi_k$  es el número de coeficientes estimados y  $p$  es el orden del modelo AR. Si el orden es demasiado bajo, el resultado no será representativo de la señal, ya que el espectro será demasiado suave, mientras que si el orden es demasiado alto, el espectro puede mostrar espigas espúreas [107].

El AR se prefiere a la transformada de Fourier, pues tiene mejor resolución para segmentos cortos de tiempo. Sin embargo, su desempeño es pobre cuando la señal es no estacionaria.

Su principal desventaja es la dificultad para la determinación de sus parámetros.

### 2.3.2. Transformada de Wavelet

Es una técnica extensamente utilizada para extraer información de señales continuas. La transformada de Wavelet permite transformar la señal al dominio tiempo-frecuencia, relacionando el contenido espectral con el momento de su ocurrencia.

Este método es útil para señales no estacionarias debido a la flexibilidad de representación en dominio de tiempo y frecuencia. Las ventajas de este método dicen relación con la capacidad de analizar la señal en el dominio del tiempo y frecuencia a través de ventanas de tamaño variable, permitiendo diferentes resoluciones para diferentes frecuencias. Esto es, en segmentos de señal en los que interese información precisa de baja frecuencia, basta con tomar una ventana muy ancha; al contrario cuando requiramos más información de alta frecuencia, se deben tomar ventanas más angostas.

La transformada continua de wavelet se define como la convolución de la señal de EEG  $x(t)$  con la función wavelet  $\psi_{s,\tau}(t)$

$$w(s, \tau) = \int_{-\infty}^{\infty} x(t) \psi_{s,\tau}^*(t) dt \quad (2.23)$$

donde  $w(s, \tau)$  es el coeficiente de wavelet que corresponde a la frecuencia asociada con la escala  $s$  y el tiempo  $\tau$  de la función  $\psi_{s,\tau}(t)$ , y el símbolo «\*» representa el complejo conjugado. La función wavelet  $\psi_{s,\tau}(t)$  es generada a partir de la traslación y cambio de escala de una función llamada *wavelet madre*  $\psi(t)$

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t - \tau}{s} \right) \quad (2.24)$$



donde  $s$  es el factor de escala y  $\tau$  es el factor de traslación.

Una wavelet madre puede tomar muchas formas, pero siempre satisface la condición

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (2.25)$$

En [47] el espacio de características se definió con coeficientes de una transformada discreta de wavelet para el reconocimiento de patrones y clasificación de potenciales de EEG asociados a movimiento.

Múltiples wavelets madres han sido descritas para aplicaciones en BCI. En [25] se plantea que la wavelet sombrero mexicano presenta mejor rendimiento en el dominio del tiempo y es útil para localizar potenciales relacionados con evento. En [126] se establece que la wavelet Morlet tiene mejor rendimiento en el dominio de la frecuencia y es útil para identificar actividad en la banda gamma. En [98] se usa wavelet bi-scale para diseñar un BCI basado en detección de imaginación motora en el rango de frecuencia de 1-4 Hz. En [60] se describe el uso de wavelet Daubechies para clasificar potenciales corticales lentos<sup>8</sup>.

### 2.3.3. Densidad Espectral de Potencia

Los métodos de extracción de características basados en frecuencia han sido ampliamente usados debido a su fácil aplicación, costo computacional y directa interpretación de los resultados. De acuerdo a [20], hasta 2007 alrededor de un tercio de los diseños de BCI usaban características en el dominio de la frecuencia, particularmente el 41 % de los BCI basados en ritmos sensoriomotores usan la densidad espectral de potencia (PSD, por sus siglas en inglés) en la definición de características y el 64 % de los BCI basados en potenciales evocados visuales también lo hacían.

Mediante el uso de Transformada Rápida de Fourier se pueden extraer características transformando las señales desde el dominio del tiempo al de la frecuencia. Funciona bien para señales estacionarias y en sistemas lineales. La transformada de Fourier sólo permite el análisis en el dominio de la frecuencia (a diferencia de la transformada de wavelet), entregando información del contenido de frecuencia, pero no entrega información de cuándo ocurren estas frecuencias.

Una forma de analizar una señal no estacionaria es realizar un análisis espectral dependiente del tiempo. Es decir, la señal es dividida en una secuencia de segmentos de tiempo en los cuales la señal puede ser considerada como cuasi-estacionaria y la Transformada de Fourier puede ser aplicada a cada segmento local de la señal. Este enfoque se asocia a una relación costo beneficio, ya que mientras más pequeña sea la ventana más alta será la resolución temporal, pero más baja la resolución en frecuencia y viceversa. Por lo tanto, una importante limitación de este método es no poder trabajar en sistemas no lineales y no estacionarios.

La unidad de PSD es energía por ancho de frecuencia y se puede obtener energía en de un rango de frecuencia específico mediante la integración de PSD dentro de ese rango de frecuencia.

---

<sup>8</sup>Los potenciales corticales lentos son cambios de voltaje en el EEG que tienen una duración de uno a varios segundos y están asociados a cambios en el nivel de actividad cortical.

### 2.3.4. Resumen

La tabla 2.2 resume los diferentes métodos de extracción de características en los diseños de BCI, considerando el mecanismo neural usado en el registro de EEG y los métodos comúnmente asociados para la extracción de características.

Tabla 2.2: Resumen de métodos de extracción de características de acuerdo al mecanismo neural involucrado (Modificado de [20]).

Mecanismo Neural y Método de Extracción de Características Asociado	
Mecanismo Neural	Método de Extracción de Características
Actividad Sensorio-motora	Parámetros espectrales
	Modelado paramétrico (Autoregresivo, adaptativo autoregresivo, adaptativo con entradas exógenas)
	Representación tiempo-frecuencia
	Comparación basada en correlación cruzada
	Parámetros basados en la envolvente y en correlación cruzada
P300	Representación tiempo-frecuencia
	Amplitud de potenciales corticales lentos
	Mezcla de filtros
	Análisis discriminante, correlación cruzada
	Cálculo de área
Potenciales evocados visuales	Diferencias en bandas de potencia
	Parámetros espectrales
	Correlación cruzada
Respuesta a tareas mentales	Modelado paramétrico (Autoregresivo, adaptativo autoregresivo)
	Parámetros temporales
	Parámetros espectrales
	Vectores propios de matriz de covarianza (o correlación)

## 2.4. Clasificación en Interfaz Cerebro Computador

La etapa de clasificación forma parte del módulo de procesamiento y es fundamental en la arquitectura de un sistema de BCI. Aquí se busca clasificar las señales de EEG tomando en cuenta los vectores de características, de ahí lo determinante de las etapas anteriores. Particularmente en este documento la etapa de clasificación busca reconocer la intención del usuario (imaginería motora).

Tradicionalmente los algoritmos de clasificación han sido entrenados mediante aprendizaje supervisado, etiquetando la base de datos de entrenamiento y de prueba, con el objetivo de detectar patrones de actividad cerebral. Sin embargo, considerando la naturaleza no estacionaria de las señales cerebrales, este enfoque podría no ser del todo apropiado. Por un lado los patrones observados durante el período de muestreo para la calibración puede ser diferente a los registrados durante las sesiones de prueba. Por otra parte, el entrenamiento mental progresivo de los usuarios o incluso cambios en la concentración, la atención o la motivación pueden afectar las señales cerebrales. La adaptación a señales no estacionarias

es particularmente necesaria en los BCIs no invasivos, de ahí la propuesta presentada en la hipótesis de este estudio (ver capítulo 1.4) basado en el uso del método de *descomposición de modo empírico*.

A continuación se presentan los métodos de clasificación más utilizados en aplicaciones de BCI:

### 2.4.1. Análisis Discriminante lineal

Aunque PCA encuentra componentes que son útiles para representar datos determinados, no hay evidencia para asumir que estos componentes serán útiles para discriminar entre diferentes clases [44].

El análisis discriminante lineal (LDA) o análisis discriminante lineal de Fisher es un clasificador muy simple de usar y tiene muy bajo costo computacional. LDA es muy común en aplicaciones de BCI principalmente debido a su rápida respuesta y relativamente buenos resultados. Sin embargo para distribuciones no Gaussianas puede no mantener la estructura compleja de los datos [65, 16].

En [51] se evalúa una variante denominada *Análisis Discriminante local de Fisher (LFDA)* al compararla con PCA y *Proyección Lineal Local* para la clasificación de dos tareas de imaginación motora (mano izquierda vs mano derecha) en cinco sujetos, encontrando el mejor resultado en LFDA con un promedio de 77,3 % de éxito en la clasificación. Sin embargo LFDA no presenta buenos resultados cuando el número de características es demasiado grande [61]. En [106] se propone un método modificado, Stacked Regularized Linear Discriminant Analysis (SRLDA), el cual mostró ser efectivo en la clasificación binaria y multiclase (4 tareas).

Para un problema de dos clases, LDA asume que las dos clases son linealmente separables. Brevemente, si se considera el problema de proyectar datos desde  $d$  dimensiones, una simple solución para reducir su dimensionalidad sería proyectarlos en una línea. Evidentemente, aún cuando las muestras formen grupos compactos y bien separado en el espacio de  $d$  dimensiones, la proyección en una línea arbitraria producirá una mezcla de las muestras y por lo tanto dificultará la clasificación de ellas. Sin embargo, al mover la línea se podría encontrar una orientación para la cual las muestras proyectadas estén bien separadas. Éste es exactamente el objetivo del análisis discriminante clásico.

Si se tiene un conjunto  $d - dimensional$  de  $n$  muestras  $\mathbf{x}_1, \dots, \mathbf{x}_n$ ,  $n_1$  pertenecen a la clase  $D_1$  etiquetado  $\omega_1$  y  $n_2$  pertenecen a la clase  $D_2$  etiquetado  $\omega_2$ . La combinación lineal de los componentes de  $\mathbf{x}$  permite obtener el producto escalar  $y = \mathbf{w}^T \mathbf{x}$  y un correspondiente conjunto de  $n$  muestras  $y_1, \dots, y_n$  distribuidas en los subconjuntos  $Y_1$  y  $Y_2$  [44].

Si  $\|\mathbf{w}\| = 1$ , cada  $y_i$  es la proyección del correspondiente  $\mathbf{x}_i$  sobre una línea en la dirección de  $\mathbf{w}$ . La magnitud de  $\mathbf{w}$  no es relevante, ya que sólo escala a  $y$ . La dirección de  $\mathbf{w}$  sí es importante, ya que determinará la dirección de la línea donde deberán caer las muestras  $\omega_1$  y  $\omega_2$ . Si la muestra original tiene una distribución altamente superpuesta entre las clases, se vuelve fundamental encontrar la mejor dirección para un  $\mathbf{w}$  que permita separar las clases. Una medida de la separación entre los puntos proyectados es la diferencia de las medias de las muestras. Si  $\mathbf{m}_i$  es la media de la muestra  $d - dimensional$  dada por

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x} \quad (2.26)$$

entonces la media de la muestra para los puntos proyectados está dada por

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y \quad (2.27)$$

$$= \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{m}_i \quad (2.28)$$

que corresponde a la proyección de  $\mathbf{m}_i$ .

Se deduce, por lo tanto, que la distancia entre las medias proyectadas es

$$|\tilde{m}_1 - \tilde{m}_2| = |\mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)| \quad (2.29)$$

esta diferencia puede ser tan grande como se desee, simplemente escalando por  $\mathbf{w}$ . Para obtener una buena separación de los datos proyectados se requiere que la diferencia entre las medias sea grande en relación con alguna medida de las desviaciones estándar para cada clase. Luego, se define la distribución de las muestras proyectadas y etiquetada como  $\omega_i$  por

$$\tilde{s}_i = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 \quad (2.30)$$

Por lo tanto,  $(1/n)(\tilde{s}_1^2 - \tilde{s}_2^2)$  es una estimación de la varianza de los datos agrupados, y  $\tilde{s}_1^2 - \tilde{s}_2^2$  se denomina *dispersión total intra clase* de las muestras proyectadas. El discriminante lineal de Fisher emplea la función lineal  $\mathbf{w}^T \mathbf{x}$  para la cual se puede definir una función objetivo  $J(\mathbf{w})$  máxima e independiente de  $\|\mathbf{w}\|$ . La maximización de esta función conduce a la mejor separación entre los dos conjuntos proyectados. Sin embargo, también se requiere un criterio de umbral para tener un buen clasificador.

Para obtener  $J(\cdot)$  como una función explícita de  $\mathbf{w}$  se definen las matrices de dispersión  $\mathbf{S}_i$  y  $\mathbf{S}_w$  por

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \quad (2.31)$$

$$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2 \quad (2.32)$$

donde  $\mathbf{S}_w$  se denomina *matriz intraclase* luego, se puede escribir

$$\tilde{s}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_i)^2 \quad (2.33)$$

$$= \sum_{\mathbf{x} \in D_i} \mathbf{w}^T (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \mathbf{w} \quad (2.34)$$

$$= \mathbf{w}^T \mathbf{S}_i \mathbf{w} \quad (2.35)$$

por lo tanto la suma (matriz intraclase) es

$$\tilde{s}_1 + \tilde{s}_2 = \mathbf{w}^T \mathbf{S}_w \mathbf{w} \quad (2.36)$$

de manera similar, la separación de las medias proyectadas obedece a

$$(\tilde{m}_1 - \tilde{m}_2)^2 = (\mathbf{w}^T \mathbf{m}_1 - \mathbf{w}^T \mathbf{m}_2)^2 \quad (2.37)$$

$$= \mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} \quad (2.38)$$

$$= \mathbf{w}^T \mathbf{S}_B \mathbf{w} \quad (2.39)$$

donde  $\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$  se denomina *matriz interclase*.

En términos de  $\mathbf{S}_B$  y de  $\mathbf{S}_w$  la función  $J(\cdot)$  o criterio de Fisher se puede escribir

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \quad (2.40)$$

esta expresión también se conoce como *cuociente generalizado de Rayleigh*.

De aquí, un vector  $\mathbf{w}$  que maximiza  $J(\cdot)$  debe satisfacer  $\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w}$  para una constante  $\lambda$ , lo que corresponde a un problema generalizado de valores propios.

Si  $\mathbf{S}_w$  es invertible (no singular) se obtiene un problema convencional de valores propios de la forma  $\mathbf{S}_w^{-1} \mathbf{S}_B \mathbf{w} = \lambda \mathbf{w}$ . En este caso no es necesario resolver para vectores y valores propios de  $\mathbf{S}_w^{-1} \mathbf{S}_B$  ya que  $\mathbf{S}_B \mathbf{w}$  siempre está en la dirección  $\mathbf{m}_1 - \mathbf{m}_2$ . Debido a que el factor de escala para  $\mathbf{w}$  es inmaterial<sup>9</sup>, la solución para  $\mathbf{w}$  que optimice  $J(\cdot)$  es

$$\mathbf{w} = \mathbf{S}_w^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \quad (2.41)$$

De esta manera se ha obtenido  $\mathbf{w}$  para el discriminante lineal de Fisher, es decir, la función lineal que produce la razón máxima en la dispersión entre clases y la dispersión dentro de la clase. Ahora se debe encontrar el umbral, aquel punto a lo largo del subespacio unidimensional que separa los puntos proyectados.

Por lo tanto LDA define una función de discriminación lineal que representa un hiperplano en el espacio de características para distinguir las clases. La clase a la que pertenece el vector de características dependerá del lado del plano donde se encuentre el vector, si el problema fuese de un número mayor de clases, se deberían utilizar varios hiperplanos.

El plano de separación puede ser representado como  $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$ , donde  $\mathbf{w}$  es el vector de peso,  $\mathbf{x}$  es el vector de características de entrada y  $w_0$  es un umbral. El vector de características de entrada es asignado a una u otra clase dependiendo del signo de  $g(\mathbf{x})$ .  $\mathbf{x}$  es una matriz conteniendo  $n$  vectores de características  $x_1, x_2, \dots, x_n \in \mathbb{R}^d$ . Aquí  $\mathbf{w}$  puede ser calculado como:

$$\mathbf{w} = \Sigma_c^{-1} (\mu_2 - \mu_1) \quad (2.42)$$

donde  $\mu_i$  es la media estimada para la clase  $i$  y  $\Sigma_c = \frac{1}{2}(\Sigma_1 + \Sigma_2)$  es la matriz de covarianza común. Los estimadores de la matriz de covarianza son calculados como

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (2.43)$$

y evidentemente

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.44)$$

La matriz de covarianza (ecuación 2.43) puede verse afectada en casos donde la dimensionalidad de las características es demasiado alta comparada al número de ensayos disponibles.

---

<sup>9</sup>La escala aplicada a los vectores propios originales es inmaterial cuando el proceso de convertirlos a forma ortonormal anula cualquier factor de escala.

Finalmente, si se utilizan las medias de la muestra y la matriz de covarianza de la muestra para estimar  $\mu_i$  y  $\Sigma$ , se obtendrá un vector en la misma dirección que  $\mathbf{w}$  en la ecuación 2.42 que maximiza  $J(\cdot)$ . Por lo tanto, para el caso normal de la igualdad de covarianza, la regla de decisión óptima es simplemente decidir  $\omega_1$  si el discriminante lineal de Fisher excede algún umbral, y decidir  $\omega_2$  en caso contrario.

## 2.4.2. Máquina de Soporte Vectorial usando Optimización Mínima Secuencial

Máquina de soporte vectorial (SVM del inglés Support Vector Machine) es un clasificador usado ampliamente en sistemas de BCI ya que no requiere un gran conjunto de entrenamiento para obtener buenos resultados. Funciona de manera similar a los clasificadores LDA, construye un hiperplano o conjunto de hiperplanos, con el fin de separar los vectores de características en varias clases. Sin embargo, a diferencia de la LDA, SVM selecciona los hiperplanos que maximizan los márgenes, es decir, la distancia entre las muestras de entrenamiento más cercanas y los hiperplanos.

El objetivo de SVM es mapear los datos en un espacio de alta dimensión y encontrar un hiperplano de separación con el margen máximo de acuerdo con el *teorema de Cover*<sup>10</sup> sobre la separabilidad de los patrones. Se ha implementado en clasificaciones binarias y multiclase [130, 107, 20, 21].

SVM induce separación lineal, ya sea en el espacio original de los ejemplos de entrada, si éstos son separables o cuasi-separables, o en un espacio transformado (espacio de características) si las entradas no son separables linealmente en el espacio original. Como se verá más adelante, la búsqueda del hiperplano de separación en estos espacios transformados, normalmente de muy alta dimensión, se hará de forma implícita utilizando las denominadas funciones kernel. De este modo también será posible crear un SVM no lineal. El kernel que se utiliza normalmente en el campo BCI es la función de base gaussiana o radial (RBF).

En estricto rigor, se busca representar patrones en una alta dimensión (típicamente mucho más alta que el espacio de características original). Con un apropiado mapeo no lineal a una suficientemente alta dimensión los datos de dos clases siempre pueden ser separables por un hiperplano [44].

La optimización mínima secuencial (SMO del inglés Sequential Minimal Optimization) es un algoritmo para entrenar SVM. Este método consiste en dividir el problema cuadrático<sup>11</sup> en una serie de problemas cuadráticos de menor dimensión los cuales pueden ser resueltos de manera analítica.

---

<sup>10</sup>El Teorema de Cover establece que dado un conjunto de datos de entrenamiento que no es linealmente separable, éste se puede transformar con alta probabilidad en un conjunto de entrenamiento que sí es linealmente separable proyectándolo en un espacio de mayor dimensión a través de alguna transformación no lineal [107].

<sup>11</sup>Programación cuadrática es un tipo especial de problema de optimización, específicamente el problema de optimizar (minimizar o maximizar) una función cuadrática de varias variables sujetas a restricciones lineales sobre estas variables. Es un tipo particular de programación no lineal, donde el objetivo es encontrar un vector n-dimensional  $\mathbf{x}$

$$f(x) = \frac{1}{2}\mathbf{x}^T B\mathbf{x} - \mathbf{x}^T \mathbf{b}, \quad \text{sujeto a } A\mathbf{x} \leq \mathbf{b}$$

La notación  $A\mathbf{x} \leq \mathbf{b}$  significa que cada entrada del vector  $A\mathbf{x}$  es menor o igual que la entrada correspondiente del vector  $\mathbf{b}$ .

Brevemente, SVM mapea los puntos de entrada en un espacio de características de una dimensión mayor, encontrando un hiperplano que separa y maximiza el margen entre las clases. La solución del hiperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada que son llamados vectores de soporte.

Los vectores de soporte corresponden a las muestras de entrenamiento que definen el hiperplano de separación óptimo. Aquí los puntos  $\mathbf{x}$  que se encuentran en el hiperplano satisfacen la condición  $\mathbf{w} \cdot \mathbf{x} + b = 0$ , donde  $\mathbf{w}$  es un vector normal al hiperplano,  $|b| / \|\mathbf{w}\|$  es la distancia perpendicular desde el hiperplano al origen, y  $\|\mathbf{w}\|$  es la norma euclidiana de  $\mathbf{w}$ . El margen se define por  $d_+$  y  $d_-$  como la distancia más corta desde el hiperplano de separación al vector positivo (negativo) más cercano (figura 2.3).

Para el caso de separación lineal el SVM busca el hiperplano de separación con el mayor margen. Por lo tanto, si se tienen los datos de entrenamiento  $\{\mathbf{x}_i, y_i\}$ ; con  $i = 1, \dots, l$ ;  $y_i \in \{-1, 1\}$ ; y  $\mathbf{x}_i \in \mathbb{R}^d$ , se sabe que

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 \quad \text{for } y_i = +1 \quad (2.45)$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 \quad \text{for } y_i = -1 \quad (2.46)$$

lo que puede ser generalizado

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall i \quad (2.47)$$

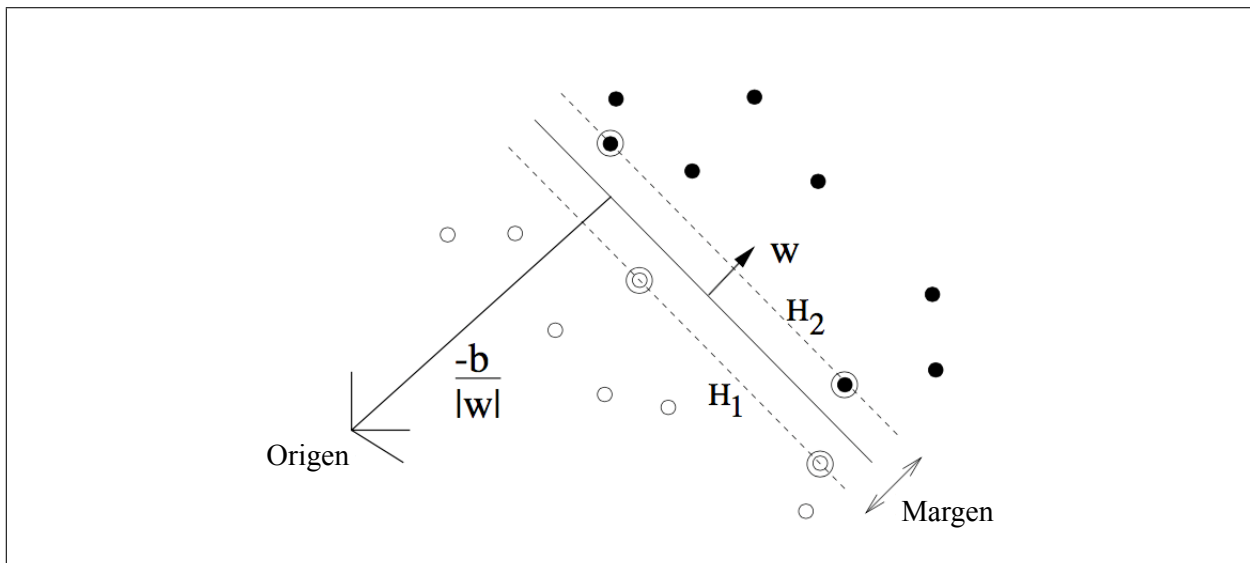


Figura 2.3: Hiperplano de separación, los vectores de soporte se definen con puntos encerrados en un círculo (modificado de [27]).

Maximizar el margen es un problema de programación cuadrática (QP) y puede ser resuelto por su problema dual<sup>12</sup> introduciendo multiplicadores de Lagrange. Hay dos razones

<sup>12</sup>Asociado a cada problema lineal existe otro problema de programación lineal denominado problema dual (PD), que posee importantes propiedades y relaciones con respecto al problema lineal original, problema que para diferenciar del dual se denomina entonces como problema primal (PP).

para hacerlo: La primera es que las restricciones en la ecuación 2.47 serán reemplazadas por restricciones sobre los propios multiplicadores de Lagrange, que serán mucho más fáciles de manejar. La segunda es que en esta reformulación del problema, los datos de entrenamiento sólo aparecerán (en los algoritmos de entrenamiento y prueba reales) en forma de producto punto entre vectores. Esta es una propiedad crucial que permite generalizar el procedimiento al caso no lineal.

Al introducir multiplicadores de Lagrange<sup>13</sup>  $\alpha_i = 1, \dots, l$  se tiene

$$L_P = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^l \alpha_i \quad (2.48)$$

donde  $L_P$  es el Lagrangiano Primal. Como se mencionó anteriormente, la maximización del margen se puede resolver mediante su problema dual<sup>14</sup>, esto es: Maximizar  $L_P$  sujeto a la restricción que el gradiente de  $L_P$  con respecto a  $\mathbf{w}$  y  $b$  sea cero, y sujeto también a la restricción que  $\alpha_i \geq 0$ .

De acuerdo a la maximización de  $L_P$  se tiene que:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \quad (2.49)$$

$$\sum \alpha_i y_i = 0 \quad (2.50)$$

Considerando la restricción dual, a partir de la ecuación 2.48 se tiene

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (2.51)$$

De las ecuaciones 2.48 y 2.51 se tiene que la solución se obtendrá por la minimización de  $L_P$  o por la maximización de  $L_D$ . Se hace notar que hay un multiplicador de Lagrange  $\alpha_i$  para cada elemento del vector de entrenamiento, en el caso que  $\alpha_i > 0$  estos elementos serán llamados vectores de soporte.

En el caso de datos que tienden a ser no separables es necesario introducir una variable de costo para evitar que la función objetivo (ecuación 2.51) crezca. Esta variable modifica las condiciones de las ecuaciones 2.45 y 2.46 de la forma

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 - \xi_i \quad \text{for } y_i = +1 \quad (2.52)$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 + \xi_i \quad \text{for } y_i = -1 \quad (2.53)$$

$$\xi_i \geq 0 \quad \forall i \quad (2.54)$$

<sup>13</sup>La regla es que para las restricciones de la forma  $C_i \geq 0$ , las ecuaciones de restricción se multiplican por multiplicadores de Lagrange positivos y se restan de la función objetivo, para formar el lagrangiano. Para las restricciones de igualdad, los multiplicadores de Lagrange no tienen restricciones.

<sup>14</sup>Se denomina Problema dual de Wolfe a la propiedad que establece que el máximo de  $L_P$ , sujeto a la restricción que el gradiente de  $L_P$  con respecto a  $\mathbf{w}$  y  $b$  sea cero, y sujeto también a la restricción que  $\alpha_i \geq 0$ , se produce con los mismos valores de  $\mathbf{w}$ ,  $b$  y  $\alpha$  que el mínimo de  $L_P$  sujeto a que las derivadas de  $L_P$  respecto a todos los  $\alpha_i$  sean cero y que  $\alpha_i \geq 0$ .



Por lo tanto,  $\sum_i \xi_i$  es el límite superior en el número de errores de entrenamiento. Una forma de asignar un costo a los errores es cambiar la función objetivo a ser minimizada de  $\frac{\|\mathbf{w}\|^2}{2}$  a  $\frac{\|\mathbf{w}\|^2}{2} + C(\sum_i \xi_i)$ , donde  $C$  es un parámetro arbitrario (un valor alto de  $C$  corresponde a asignar una penalización mayor a los errores). Ahora la ecuación 2.51 queda sujeta a  $0 \leq \alpha_i \leq C$  y  $\sum_i \alpha_i y_i = 0$ . Nuevamente la solución es similar a la ecuación 2.49.

El caso de un problema no separable linealmente puede ser resuelto mediante el uso de una función kernel (coloquialmente llamado «kernel trick»). Notar que en la ecuación 2.51 los datos se presentan como producto punto  $x_i \cdot x_j$ , si mapeamos los datos en un espacio euclidiano  $\mathcal{H}$  usando un mapeo  $\Phi$ , tal que  $\Phi : \mathbb{R}^d \mapsto \mathcal{H}$ , el algoritmo de entrenamiento dependerá del producto punto en  $\mathcal{H}$ . Esto es, si se tiene funciones de la forma  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ , sólo se necesitaría usar una función kernel  $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$  en el algoritmo de entrenamiento, sin necesidad de conocer  $\Phi$ . Un ejemplo de esta función es el kernel gaussiano  $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2}$ .

Finalmente, se necesita que  $\mathbf{w}$  exista en el espacio  $\mathcal{H}$ , por lo que la SVM es usada para calcular el producto punto de un punto determinado  $\mathbf{x}$  con  $\mathbf{w}$ , o más específicamente la función signo del plano de separación

$$f(x) = \text{sign} \left( \sum_{i=1}^{N_s} \alpha_i y_i \Phi(\mathbf{S}_i) \cdot \Phi(\mathbf{x}) + b \right) = \text{sign} \left( \sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{S}_i, \mathbf{x}) + b \right) \quad (2.55)$$

donde  $N_s$  es el número de vectores de soporte,  $\mathbf{S}_i$  son los vectores de soporte y  $\alpha_i$  son los multiplicadores óptimos de Lagrange. De este modo se puede evitar calcular  $\Phi(\mathbf{x})$  y usar el kernel  $K(\mathbf{S}_i, \mathbf{x}) = \Phi(\mathbf{S}_i) \cdot \Phi(\mathbf{x})$ . Por lo tanto, la ventaja de la función kernel es que la complejidad del problema de optimización sólo depende de la dimensionalidad del espacio de entrada y no del espacio de características. Así, es posible operar en un espacio de características teóricas de dimensión infinita.

En la presente Tesis se usó *kernel polinomial*.

SMO es un algoritmo simple que resuelve rápidamente el problema de QP en SVM sin usar una rutina iterativa (como se describió anteriormente), descomponiendo el problema de QP en subproblemas. A diferencia de los métodos anteriores de optimización de SVM, SMO elige resolver el problema de optimización más pequeño posible en cada paso. Así SMO escoge dos multiplicadores de Lagrange  $\alpha_i$  y  $\alpha_j$  con  $i \neq j$  para optimizarlos, encontrando los valores óptimos para estos dos parámetros dado que todos los otros están fijos y actualiza la SVM [118].

Debido a la restricción de igualdad lineal que implican los multiplicadores de Lagrange, el problema más pequeño posible involucra dos de tales multiplicadores. Entonces, para dos multiplicadores  $\alpha_1$  y  $\alpha_2$  la condición se reduce a  $0 \leq \alpha_1, \alpha_2 \leq C$  y  $\alpha_1 y_1 + \alpha_2 y_2 = k$ . Donde  $k$  es el negativo de la suma sobre el resto de términos en la restricción de igualdad, que se fija en cada iteración.

La SMO finalmente opera de la siguiente manera:

1. Encuentra un multiplicador de Lagrange  $\alpha_1$  que no respeta la condición de Karush-Kuhn-Tucker (KKT)
2. Elige un segundo multiplicador de Lagrange  $\alpha_2$  y optimiza el par  $(\alpha_1, \alpha_2)$
3. Repite el paso 1 y 2 hasta converger.

Como SMO evita cálculos iterativos y cada subproblema tiene una rápida solución, el problema de QP puede ser resuelto rápidamente. Por otro lado SMO no requiere almacenamiento de grandes matrices reduciendo las necesidades de almacenamiento y permitiendo de esta manera que grandes problemas de entrenamiento de SVM puedan ser solucionados con la capacidad de cómputo de cualquier computador portátil o estación de trabajo [118]. Estas características, sumado al hecho de poder generalizar el procedimiento al caso no lineal, lo hacen idóneo para la implementación de un sistema de BCI, por tal motivo se decidió incorporarlo como herramienta de clasificación en esta Tesis.

### 2.4.3. Clasificador Dagging

Dagging es un metaclasificador, éstos son una serie de métodos que permiten crear un conjunto de algoritmos de clasificación que trabajan unidos para realizar la clasificación final. Se trata, entonces, de combinar las predicciones del conjunto de algoritmos, normalmente por votación, para clasificar nuevos ejemplos. La precisión obtenida por esta combinación supera, generalmente, la precisión de cada componente individual del conjunto. A pesar que los métodos de metaclasificación se conocen hace bastante tiempo, no ha sido ampliamente usado en BCI [39, 56].

Los diversos métodos de metaclasificación permiten, por una parte, la generación de lo que se conoce como «ensamble», la unión de múltiples clasificadores débiles del mismo tipo, para la creación de uno general más preciso; o por otra parte, la combinación de diferentes métodos que dotan al metaclasificador de las características híbridas que todos los métodos simples utilizan. Gracias a su combinación son capaces de superar a un clasificador único, gracias a que se llega a un consenso para obtener el resultado final de la clasificación [108]. Es decir, toma la decisión final re-clasificando el resultado de cada clasificador de base o débil.

Cuando el muestreo se realiza en base a bootstrapping<sup>15</sup> el metaclasificador se llama Bagging, si el muestro se realiza a través de muestras disjuntas<sup>16</sup> el metaclasificador se llama Dagging [137].

Particularmente Dagging crea un número de carpetas estratificadas de muestras disjuntas (sin traslape), para generar múltiples conjuntos de datos de entrada. En cada uno de esos conjuntos, se ejecuta el mismo clasificador de base (como un SVM, Redes Neuronales, etc.) o función de regresión para obtener un modelo entrenado para cada uno de los conjuntos de entrenamiento. Las predicciones se hacen a través de voto mayoritario, ya que todos los

---

<sup>15</sup>En estadística, bootstrapping es cualquier prueba o métrica que se basa en muestreo aleatorio con reemplazo. Bootstrapping permite asignar medidas de precisión (definidas en términos de sesgo, varianza, intervalos de confianza, error de predicción o alguna otra medida de este tipo) a las estimaciones de la muestra.

Muestreo aleatorio con reemplazo: Corresponde a un modelo de selección de una muestra hecha al azar, donde cada muestra de tamaño  $n$  tiene la misma probabilidad de ser elegida. Si el muestreo se realiza con reemplazo, la muestra puede ser elegido en múltiples oportunidades.

El método consiste en: 1) crear un gran número de submuestras con reposición de los mismos datos, por ejemplo, 2000 muestras, y 2) calcular para cada muestra resultante el valor del estadístico en cuestión. Se obtiene así una aproximación a la distribución de muestreo del estadístico, a partir de la cual podemos construir un intervalo de confianza para dicho estadístico o realizar una prueba de significación. El procedimiento es sencillo y proporciona una estrategia bastante general para enfrentar diferentes tipos de situaciones, independientemente del tipo de datos y variables que son objeto de análisis.

<sup>16</sup>En matemáticas, dos conjuntos son disjuntos si no tienen ningún elemento en común. Equivalentemente, dos conjuntos son disjuntos si su intersección es vacía. En otras palabras, un elemento no puede estar en dos muestras.

clasificadores de base generan un voto de igual ponderación. Aquí la clase que recibe más votos es la elegida [137].

Al tratarse de un metaclasificador, hay que especificar un clasificador base, además de los parámetros que definen la selección de atributos (un método de búsqueda y un método de evaluación de subconjuntos de atributos). En la presente Tesis se trabajó con el metaclasificador Dagging y se utilizaron como clasificadores base (débiles) SVMs con kernel polinomial usando SMO (ver Capítulo 2.4.2), 10 carpetas y la predicción fue hecha vía promedio simple.

Se eligió el Clasificador Dagging debido a que puede incorporar como clasificadores de base a SVM con SMO, esto *a priori* contribuiría a mejorar la clasificación.

#### 2.4.4. Clasificador Logístico

El clasificador Logístico usa regresión logística multinomial para resolver problemas multiclase. La regresión logística multinomial se considera a menudo un análisis atractivo porque no asume normalidad ni linealidad. La regresión logística multinomial tiene supuestos, como la independencia entre las opciones de las variables dependientes. Este supuesto establece que la elección o pertenencia a una categoría no está relacionada con la elección o pertenencia a otra categoría (la variable dependiente). Es decir, es un modelo que se utiliza para predecir las probabilidades de los diferentes resultados posibles de una variable dependiente categóricamente distribuida, dado un conjunto de variables independientes.

En el clasificador Logístico las entradas son las características del conjunto de datos. Aquí se utiliza el mismo modelo que en regresión lineal, donde  $\mathbf{X}$  es el conjunto de entradas, el cual contiene todas las características (valores numéricos). Como en muchas otras técnicas de clasificación estadística, la idea es construir una función predictora lineal que permita generar un puntaje (o score) a partir de un conjunto de pesos que se combinan linealmente con las características de una observación dada usando un producto punto:

$$score(\mathbf{X}_i, k) = \beta_k \cdot \mathbf{X}_i \quad (2.56)$$

donde  $\mathbf{X}_i$  es el vector de características en la observación  $i$ ,  $\beta_k$  es el vector de peso o coeficiente de regresión<sup>17</sup> correspondiente a la salida (clase)  $k$  y  $score(\mathbf{X}_i, k)$  es el puntaje asociado con la asignación de la observación  $i$  a la clase  $k$ .

En particular, en el modelo logístico multinomial, el puntaje o *score* puede convertirse directamente en un valor de probabilidad, indicando de este modo la probabilidad de la observación  $i$  eligiendo la clase  $k$  dadas las características medidas de la observación. Por lo tanto la probabilidad para la clase  $k$  dado  $\mathbf{X}_i$

$$P_k(\mathbf{X}_i) = \frac{e^{(\mathbf{X}_i \beta_k)}}{1 + \sum_{k=1}^K e^{(\mathbf{X}_i \beta_k)}} \quad (2.57)$$

En la presente Tesis se decidió utilizar el Clasificador Logístico dado que se pretende clasificar 3 clases y responde a las características de no linealidad y de distribución no paramétrica propia de las señales de EEG.

---

<sup>17</sup>Para encontrar la matriz  $\beta$  se utiliza un Método cuasi-Newton. Los métodos cuasi-Newton son métodos utilizados para encontrar ceros o máximos y mínimos locales de funciones. El método completo de Newton requiere el Jacobiano para buscar ceros o el Hessiano para encontrar extremos.

## 2.4.5. Redes Neuronales

Las redes neuronales artificiales (RNA) son ampliamente utilizadas en el reconocimiento de patrones, porque son capaces de aprender de los datos de entrenamiento. Las RNA están compuesto por un gran número de elementos interconectados llamados neuronas. Cada neurona modela una neurona biológica y es capaz de realizar tareas computacionales simples.

Las RNA comprenden un conjunto de nodos y conexiones que se modifican durante el proceso de entrenamiento. Las RNA se alimentan de un conjunto de ejemplos de entrenamiento evaluando sistemáticamente la salida. Si la salida es incorrecta, los pesos internos son modificados por el algoritmo de entrenamiento para minimizar la diferencia entre la salida deseada y la real. Este entrenamiento continúa hasta que la red alcanza un estado estable, donde no se logra ninguna mejora significativa.

La red neural más utilizada en clasificación de tareas mentales a partir de registros de EEG es el *Perceptron Multicapa* (PM) [40, 81, 107]. En PM la red está organizada en tres capas, capa de entrada, capa oculta y capa de salida. La ventaja del PM es que su operación es rápida, es fácilmente implementado y requiere pequeños conjuntos de entrenamiento.

El PM se ha utilizado para clasificar dos o más clases [40, 81, 107]. Por otra parte, PM también se ha utilizado en el preprocesamiento de señales EEG antes de la extracción de características [37].

## 2.4.6. Clasificador de Bayes

La teoría de decisión de Bayes corresponde a un enfoque estadístico del problema de clasificación de patrones. Este enfoque se basa en determinar el costo asociado a las decisiones de clasificación usando probabilidad.

El método Bayesiano busca maximizar la probabilidad a posteriori de decidir una clase dada una observación. Este modelo probabilístico se puede escribir como

$$P(\omega_j | x) = \frac{p(x | \omega_j) P(\omega_j)}{p(x)} \quad (2.58)$$

donde  $\omega$  corresponde a la clase y  $x$  a la característica (feature) determinada, es decir: La probabilidad de corresponder a la clase  $\omega$  dado que tengo la característica  $x$ . Aquí

$$\omega \begin{cases} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_j \end{cases}$$

y

$$p(x) = \sum_{j=1}^C p(x | \omega_j) P(\omega_j) \quad (2.59)$$

Esto se puede describir informalmente de la siguiente manera [44]:

$$\text{Probabilidad a posteriori} = \frac{\text{verosimilitud x Probabilidad a Priori}}{\text{Evidencia}}$$

$P(\omega_j | x)$ : Probabilidad a posteriori  $\Rightarrow$  Probabilidad de que la clase sea  $\omega_j$  dado que se midió  $x$ .

$P(\omega_j)$ : Probabilidad a priori  $\Rightarrow$  Conocimiento previo del problema.

$p(x | \omega_j)$ : Verosimilitud  $\Rightarrow$  Probabilidad de la clase  $\omega_j$  respecto a  $x$ , cuanto mayor más probable que la verdadera clase sea  $\omega_j$ .

$p(x)$ : Evidencia  $\Rightarrow$  factor de escala para garantizar que las probabilidades a posteriori sumen 1.

Por otro lado la determinación de cuán costosa es cada acción es fundamental para convertir una probabilidad en una decisión, esto se define como *función de costo* [44].

Sea  $\{\omega_1, \dots, \omega_c\}$  el conjunto finito de estados o categorías y  $\{\alpha_1, \dots, \alpha_a\}$  el conjunto finito de  $a$  posibles acciones, la *función de pérdida*  $\lambda(\alpha_i | \omega_j)$  describe el costo de tomar la acción  $\alpha_i$  cuando la naturaleza del estado es  $\omega_j$ .

Sea  $\mathbf{x}$  un vector de características con  $d$  componentes y  $p(\mathbf{x} | \omega_j)$  sea la función de densidad de probabilidad condicional para  $\mathbf{x}$ .  $P(\omega_j)$  describe la probabilidad a priori que el estado sea  $\omega_j$ . Luego, la probabilidad a posteriori  $P(\omega_j | \mathbf{x})$  puede ser calculada desde  $p(\mathbf{x} | \omega_j)$  mediante

$$P(\omega_j | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_j) P(\omega_j)}{p(\mathbf{x})} \quad (2.60)$$

donde la evidencia ahora es

$$p(\mathbf{x}) = \sum_{j=1}^c p(\mathbf{x} | \omega_j) P(\omega_j) \quad (2.61)$$

Si se observa un vector  $\mathbf{x}$  particular y se toma la acción  $\alpha_i$ . Si el estado verdadero es  $\omega_j$ , por definición se incurre en la pérdida  $\lambda(\alpha_i | \omega_j)$ . Dado que  $P(\omega_j | \mathbf{x})$  es la probabilidad que el estado verdadero sea  $\omega_j$ , la pérdida esperada asociada a tomar la acción  $\alpha_i$  es

$$R(\alpha_i | \mathbf{x}) = \sum_{j=1}^c \lambda(\alpha_i | \omega_j) P(\omega_j | \mathbf{x}) \quad (2.62)$$

$R(\alpha_i | \mathbf{x})$  es denominado *Riesgo Condicional*.

Si se considera la ecuación 2.62 en el caso particular de la clasificación de dos clases, la acción  $\alpha_1$  corresponde a decidir que el verdadero estado es  $\omega_1$ , y la acción  $\alpha_2$  corresponde a decidir que el verdadero estado es  $\omega_2$ . Por simplicidad se tiene que  $\lambda_{ij} = \lambda(\alpha_i | \omega_j)$  es la pérdida incurrida al decidir  $\omega_i$  cuando el estado verdadero es  $\omega_j$ . De acuerdo a la ecuación 2.62 se tiene

$$R(\alpha_1 | \mathbf{x}) = \lambda_{11}P(\omega_1 | \mathbf{x}) + \lambda_{12}P(\omega_2 | \mathbf{x}) \quad (2.63)$$

$$R(\alpha_2 | \mathbf{x}) = \lambda_{21}P(\omega_1 | \mathbf{x}) + \lambda_{22}P(\omega_2 | \mathbf{x}) \quad (2.64)$$

La regla fundamental es decidir  $\omega_1$  si  $R(\alpha_1 | \mathbf{x}) < R(\alpha_2 | \mathbf{x})$ , en términos de probabilidad a posteriori, se decide  $\omega_1$  si

$$(\lambda_{21} - \lambda_{11})P(\omega_1 | \mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2 | \mathbf{x}) \quad (2.65)$$

empleando la fórmula de Bayes se pueden reemplazar las probabilidades a posteriori por las probabilidades a priori y la densidad de probabilidad condicional, por lo tanto se decide  $\omega_1$  si

$$(\lambda_{21} - \lambda_{11})p(\mathbf{x} | \omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22})p(\mathbf{x} | \omega_2)P(\omega_2) \quad (2.66)$$

de aquí, bajo el supuesto que  $\lambda_{21} > \lambda_{11}$  se decide  $\omega_1$  si

$$\frac{p(\mathbf{x} | \omega_1)}{p(\mathbf{x} | \omega_2)} > \frac{(\lambda_{12} - \lambda_{22}) P(\omega_2)}{(\lambda_{21} - \lambda_{11}) P(\omega_1)} \quad (2.67)$$

El clasificador de Bayes es útil para sistemas basados en series de tiempo y en términos generales no es muy popular en el uso de BCI, aun cuando sí se ha usado para clasificación de imaginería motora [84, 58, 59, 143]. En [58] se presenta un nuevo método llamado Redes Bayesianas Comunes (CBN) para discriminar multiclases de imaginería motora en señales de EEG, donde se evaluó la eficacia del método en la clasificación de cuatro clases (imaginería motora de mano izquierda, mano derecha, lengua y pie) en tres sujetos, encontrando una correcta clasificación en el 89% de los casos (incluso uno de los sujetos alcanzó un 98%). Para este fin los registros fueron obtenidos en un laboratorio con un equipo Neuroscan de 64 canales.

### 2.4.7. Clasificador Basado en Vecino más Cercano

El clasificador basado en el método de *vecino más cercano* corresponde a una forma práctica y de fácil aplicación para predecir la pertenencia a una clase o clasificar un nuevo dato. Cuando un nuevo dato se presenta al sistema de aprendizaje, éste se clasifica según el dato más cercano basado en observaciones conocidas.

La regla de clasificación más sencilla está determinada por la *distancia mínima*<sup>18</sup>. Brevemente, sea  $E$  un conjunto que posee una cierta métrica  $d : E \times E \rightarrow \mathbb{R}$ . Si se tienen  $m$  prototipos  $p_1, p_2, \dots, p_m \in E$ , tal que  $p_i \in c_i$  representa a la clase  $c_i$  ( $i = 1 \dots m$ ), es decir un prototipo por clase. Una clasificación por distancia mínima respecto a  $p_1, p_2, \dots, p_m$  puede definirse para cualquier prototipo  $x \in E$  como:

$$x \in c_i \Leftrightarrow d(x, p_i) \leq d(x, p_j); \quad i, j = 1 \dots m, \quad i \neq j \quad (2.68)$$

donde  $d$  es la métrica definida en  $E$ .

En este tipo de clasificador, la fase de aprendizaje consiste únicamente en la elección de un buen representante de cada clase en el conjunto de prototipos que pertenecen a la misma, donde la muestra  $x$  se clasifica en la clase cuyo representante se encuentra a menor distancia. En la figura 2.4 se puede ver que para un ejemplo de tres clases  $c_i, c_j$  y  $c_k$ , representadas por los prototipos  $p_i, p_j$  y  $p_k$  la muestra  $x$  se clasificará en la clase  $c_k$ .

<sup>18</sup>Las medida de distancia más usadas es la Distancia Euclidiana  $d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$

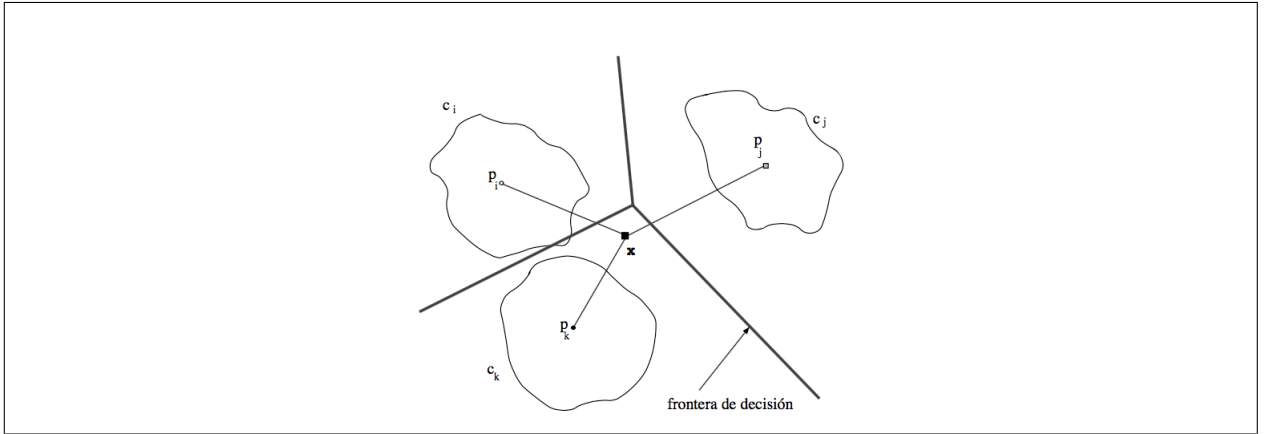


Figura 2.4: Clasificador por distancia mínima. La muestra se clasifica en la clase para la cual la distancia a su representante es mínima, en este caso se clasifica en la clase  $c_k$ .

Las fronteras de decisión en este tipo de espacios que separan las clases son los hiperplanos mediatrices de los segmentos formados por todos los pares de prototipos. Estas fronteras de decisión son un caso especial de clasificador lineal. El conjunto de regiones definidas por las fronteras de decisión asociadas a un clasificador de distancia mínima se denomina *diagrama de Voronoi*<sup>19</sup>. Todas las regiones de Voronoi de los puntos  $p_i, i = 1 \dots N$  forman una partición del espacio  $\mathbb{R}^k$  de la forma  $C = \{c_i, i = 1 \dots N\}$ .

Una generalización del clasificador por distancia mínima corresponde a la *regla del vecino más cercano*. Este caso permite que haya más de un prototipo por clase.

Aquí la distancia entre muestras y clases se define como

$$d_{vmc}(x, c_i) = \min_{\forall p \in c_i} d(x, p); \quad x \in E, \quad c_i \in C \quad (2.69)$$

y la muestra  $x$  se clasificará en la clase que tiene el prototipo más cercano (figura 2.5).

<sup>19</sup>El diagrama de Voronoi de un conjunto de puntos en el plano es la división de dicho plano en regiones, de tal forma que a cada punto le asigna una región del plano formada por los puntos que son más cercanos a él que a ninguno de los otros objetos. Dicho de otra manera, lo que hace dicho diagrama es dividir el plano en tantas regiones como puntos tengamos, de tal forma que a cada punto le asignemos la región formada por todo lo que está más cerca de él que de ningún otro.

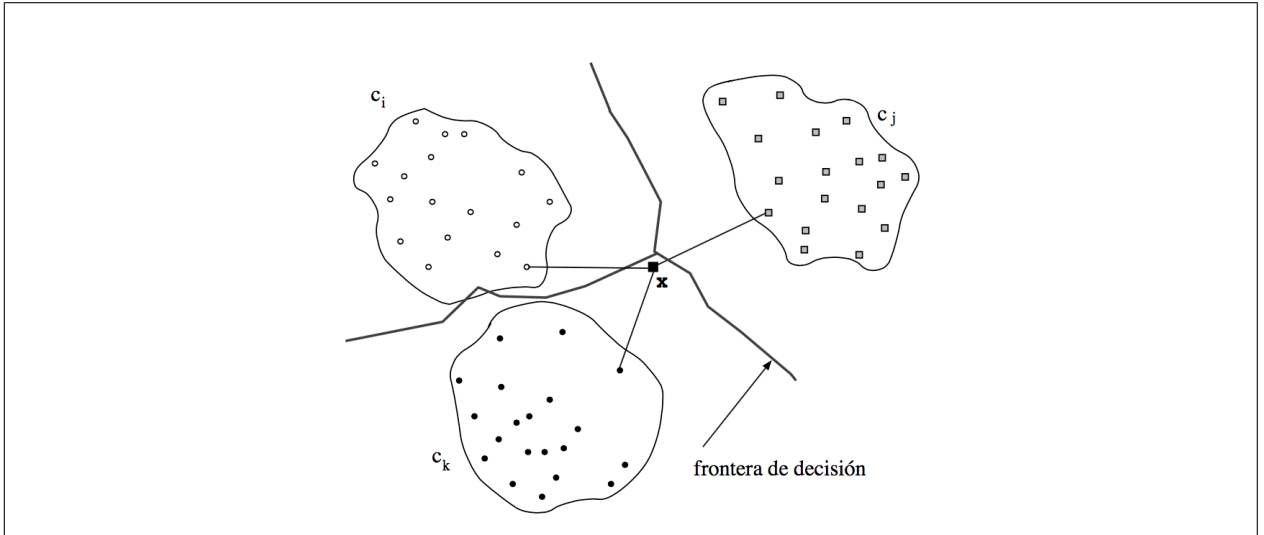


Figura 2.5: Clasificación por la regla del vecino más cercano. La muestra  $x$  se clasifica en la clase para la que la distancia a uno de sus puntos es mínima. En este ejemplo el prototipo más cercano pertenece a la clase  $c_k$ .

Una extensión de la regla del vecino más cercano es la regla de *k-vecino más cercano* o *K-nearest neighbour* (*k-NNC*). Esta regla clasifica  $x$  asignándole la etiqueta más frecuentemente representada entre las  $k$  muestras más cercanas. En otras palabras, se toma una decisión examinando las etiquetas de los  $k$  vecinos más próximos. Corresponde a un método no paramétrico y se basa en el principio de organización de clusters, donde diferentes clases deberían agruparse en un espacio de características determinado, aquí los vecinos cercanos pertenecerán a la misma clase. Este método toma  $k$  distancias entre las características de las muestras de prueba y las clases más cercanas con el fin de clasificar un vector de características de prueba. Las distancias son una medida de las similitudes entre las características del vector de prueba y las características de cada clase. La ventaja de tomar en cuenta  $k$  vecinos en la clasificación es que se reduce la probabilidad de error en la decisión. Algunas muestras de entrenamiento pueden verse afectadas por el ruido y los artefactos, los que pueden influir seriamente en los resultados de la clasificación. Dado que pueden existir varias clases  $k$  más cercanas, se requerirá un algoritmo de votación para decidir entre opciones competitivas. Considerando que las distribuciones de esos vecinos no son homogéneas, el método *k-NNC* tiene que asignar rangos diferentes a los vecinos más cercanos, de acuerdo a sus distancias respecto al ejemplo de prueba. Por lo tanto, *k-NNC* necesita definir una función de peso  $\omega$ , que varía con la distancia de tal manera que el valor de salida disminuye a medida que aumenta la distancia entre el vector de características de prueba y el vecino [107].

$$\omega^{(i)} = \begin{cases} \frac{d^{(k)} - d^{(i)}}{d^{(k)} - d^{(1)}} & \text{if } d^{(k)} \neq d^{(1)} \\ 1 & \text{if } d^{(k)} = d^{(1)} \end{cases} \quad (2.70)$$

donde  $d^{(i)}$  denota la distancia del  $i$ -ésimo vecino más cercano, esto implica que  $d^{(1)}$  corresponde al vecino más cercano y  $d^{(k)}$  al más lejano. La regla de decisión de *k-NNC* asigna los ejemplos desconocidos a las clases con la mayor suma de pesos entre sus  $k$  vecinos más cercanos.



A pesar de lo simple del método, éste tiene la desventaja de ser muy sensible al ruido y a características irrelevantes o redundantes [81].

## 2.5. Clasificación Usando Bases de Datos públicas

Una estrategia ampliamente usada para la evaluación de algoritmos de clasificación está dada por el uso de bases de datos públicas. Las bases de datos públicas son conjuntos de datos obtenidos a partir de registros de señales cerebrales en los laboratorios líderes en tecnología BCI, las cuales son puestas a disposición de los diferentes grupos de investigación a nivel mundial para la validación de sus propios métodos de procesamiento. A pesar que este método sólo puede ser usado de manera offline, y no aborda el problema del uso en tiempo real, ofrece la ventaja de poder comparar bajo similares condiciones diferentes métodos de procesamiento, selección de características y clasificación para su posterior uso en la interfaz propiamente tal.

En consecuencia, las bases de datos públicas cobran vital importancia al momento de evaluar el rendimiento de una metodología en particular, ya que a partir de los mismos datos es posible medir el efecto de uno u otro método.

Por el contrario, los métodos online proporcionan un medio de evaluación de sistemas BCI en un entorno real, sea controlado o no. Los datos se procesan de una manera causal y los algoritmos se prueban en un entorno en el que éstos pueden cambiar producto de diferentes condiciones que experimentan los usuarios, por ejemplo: cambios en la motivación, fatiga o respuesta a estímulos visuales o auditivos.

Las bases de datos más utilizadas en publicaciones científicas para uso en BCI son: BCI Competition<sup>20</sup> [89, 16, 106, 53, 55, 86, 33, 85, 122, 123, 100, 29, 130, 142, 17, 141, 154, 156, 94, 92, 119, 83, 3, 101, 113, 160] y Physionet<sup>21</sup> [127, 90, 68, 7, 66, 111, 4, 6, 1, 157, 145, 128, 91, 79, 5]<sup>22</sup>.

La base de datos BCI Competition está enmarcada en una iniciativa cuyo objetivo inicial fue que los investigadores de todo el mundo pudiesen ajustar sus métodos y obtener una salida de los algoritmos que permita una correcta clasificación. Está constituida por cuatro conjuntos (I, II, III y IV) con sus respectivos subconjuntos de datos. Originalmente todas las bases de datos se dividían en dos partes: una parte de los datos era etiquetada y constituía el conjunto de entrenamiento y otra parte de los datos no era etiquetada y constituía el conjunto de validación, con el cual los organizadores evaluaban el método propuesto por los concursantes. El uso de esta base de datos se ha extendido hasta nuestros días y ha sido ampliamente usada como herramienta de validación de los métodos de procesamiento propuestos por diferentes grupos de investigadores, utilizando para ello la amplia base de datos etiquetada (originalmente constituida como base de datos para entrenamiento). En [89, 16, 106, 53, 55, 86, 33, 85, 122, 123, 100, 29, 130, 142, 17, 141, 154, 156, 94, 92, 119, 83, 3, 101, 113, 160] se presentan estudios publicados desde 2010 a 2017 que usan la base de datos BCI Competition para la validación de diferentes métodos de procesamiento de señales de EEG durante tareas de imaginación motora.

Los tres concursos de la BCI Competition se organizaron en 2001, 2002 y 2004. El interés creciente en tales concursos se refleja en el número de presentaciones que subieron de 10 a 57 y

---

<sup>20</sup><http://www.bbci.de/competition/>

<sup>21</sup><https://physionet.org/physiobank/database/eegmmidb/>

<sup>22</sup>Sólo se consideraron las referencias publicadas a partir de 2010.

a 92 en los respectivos años del concurso. La competencia incluyó problemas de transferencia de sesión a sesión, no estacionariedad, pequeños conjuntos de entrenamiento, transferencia de sujeto a sujeto, datos de prueba continuos sin estructura de prueba, paradigmas asíncronos y estados inactivos [23].

Por otro lado la base de datos Physionet fue creada por los desarrolladores del sistema de instrumentación BCI2000 [125]. El sistema BCI2000 es un proyecto dirigido por el Programa de I+D de BCI en el Centro Wadsworth del Departamento de Salud del Estado de Nueva York en Albany, Nueva York, EEUU, con contribuciones sustanciales del Instituto de Psicología Médica y Comportamiento del Departamento de Neurobiología en la Universidad de Tübingen, Alemania. Además, muchos laboratorios de todo el mundo también han desempeñado un papel importante en el desarrollo del proyecto. Estos laboratorios incluyen el BrainLab en Georgia State University en Atlanta, Georgia, y Fondazione Santa Lucia en Roma, Italia. Los objetivos del proyecto BCI2000 son: 1) crear un sistema de software que facilite la implementación de una amplia gama de experimentos que dependen de la adquisición de datos, procesamiento en tiempo real y presentación de estímulos; 2) incorporar en este sistema el soporte para los métodos más comúnmente usados de BCI; y 3) difundir el sistema y la documentación asociada a otros laboratorios [54].

Al igual que lo ocurrido con la base de datos BCI Competition, la base de datos Physionet de imaginaria motora ha sido ampliamente usada hasta hoy con el fin de validar los diferentes métodos de procesamiento de señales de EEG durante tareas de imaginaria motora. En [127, 90, 68, 7, 66, 111, 4, 6, 1, 157, 145, 128, 91, 79, 5] se presentan estudios publicados desde 2010 a 2017 que usan esta base de datos como herramienta de validación metodológica.

Para efectos de validar la hipótesis planteada en esta Tesis se usará la base de datos pública Physionet de imaginaria motora.

# Capítulo 3

## Metodología Propuesta

Para evaluar la hipótesis de investigación se usó la base de datos pública Physionet. Esta base de datos fue creada por los desarrolladores de BCI2000 y ha sido ampliamente utilizada en diversas publicaciones [7, 111, 112, 127, 90, 68, 66, 4, 6, 1, 157, 145, 128, 91, 79, 5]<sup>1</sup>, constituyendo una base de datos vigente en la actualidad. La base de datos Physionet forma parte de PhysioBank<sup>2</sup>, un banco de datos de señales fisiológicas para uso en investigaciones biomédicas.

La base de datos Physionet consta de 1526 registros obtenidos de 109 voluntarios, donde cada voluntario cuenta con 14 registros de diferentes tareas, tanto motoras como imaginarias. Cada registro se hizo con 64 canales de EEG. Todos los registros tienen una frecuencia de muestreo de 160 Hz y una duración de uno a dos minutos.

Los registros de esta base de datos contienen las siguientes actividades: Actividad de EEG basal con ojos abiertos, actividad de EEG basal con ojos cerrados, actividad de EEG con tarea motora y de imaginaria de apertura y cierre del puño izquierdo o derecho, actividad de EEG con tarea motora y de imaginaria de apertura y cierre de ambos puños o pies (se considera cierre de pies a la flexión de los ortejos). Cada tarea mental se realizó después de la presentación de un estímulo en una pantalla.

Para la presente Tesis se consideraron tres tareas de imaginaria: Reposo, Imaginar abrir y cerrar el puño izquierdo e imaginar abrir y cerrar el puño derecho.

### 3.1. Metodología Experimental

Todos los sujetos que participaron de la base de datos realizaron tareas motoras y de imaginaria. En cada registro se usaron 64 canales de EEG utilizando el software BCI2000<sup>3</sup>. Cada sujeto realizó 14 ensayos experimentales: dos ensayos de un minuto de actividad basal de EEG (uno con los ojos abiertos y otro con los ojos cerrados) y tres ensayos de dos minutos de cada una de las siguientes cuatro tareas:

---

<sup>1</sup>Las referencias citadas corresponden a publicaciones desde 2013 a la fecha.

<sup>2</sup>Actualmente PhysioBank incluye bases de datos de señales cardiopulmonares, neurales y biomédicas de múltiples parámetros en sujetos sanos y pacientes con una variedad de condiciones con relevancia para la salud pública. PhysioBank ahora contiene más de 75 bases de datos que se pueden descargar libremente. <https://www.physionet.org/physiobank/>

<sup>3</sup>BCI2000 es un software de propósito general para la investigación en interfaz cerebro-computador (BCI). También puede utilizarse para la adquisición de datos, presentación de estímulos y aplicaciones de monitoreo cerebral. <http://www.bci2000.org>

1. Abrir y cerrar el puño (Tarea 1): El sujeto abre y cierra el puño izquierdo o derecho de acuerdo al lugar de aparición de un estímulo visual que se presenta en una pantalla (izquierda o derecha). La tarea se realiza hasta que el estímulo desaparece, luego el sujeto se relaja.
2. Imaginar abrir y cerrar el puño (Tarea 2): El sujeto imagina abrir y cerrar el puño izquierdo o derecho de acuerdo al lugar de aparición de un estímulo visual que se presenta en una pantalla (izquierda o derecha). La tarea se realiza hasta que el estímulo desaparece, luego el sujeto se relaja.
3. Abrir o cerrar ambos puños o ambos pies (Tarea 3): El sujeto abre y cierra ambos puños si aparece un estímulo visual en la parte superior de la pantalla, si el estímulo aparece en la parte inferior el sujeto abre y cierra ambos pies. La tarea se realiza hasta que el estímulo desaparece, luego el sujeto se relaja.
4. Imaginar abrir y cerrar ambos puños o ambos pies (Tarea 4): El sujeto imagina abrir y cerrar ambos puños si aparece un estímulo visual en la parte superior de la pantalla, si el estímulo aparece en la parte inferior el sujeto imagina abrir y cerrar ambos pies. La tarea se realiza hasta que el estímulo desaparece, luego el sujeto se relaja.

Por lo tanto, los 14 ensayos experimentales estuvieron constituidos por un ensayo de un minuto de actividad basal de EEG con los ojos abiertos, un ensayo de un minuto de actividad basal de EEG con los ojos cerrados, tres ensayos de dos minutos de la tarea abrir y cerrar el puño, tres ensayos de dos minutos de la tarea imaginar abrir y cerrar el puño, tres ensayos de dos minutos de la tarea abrir o cerrar ambos puños o ambos pies y tres ensayos de dos minutos de la tarea imaginar abrir y cerrar ambos puños o ambos pies. La secuencia de los ensayos se realizó de acuerdo a lo presentado en la figura 3.1.

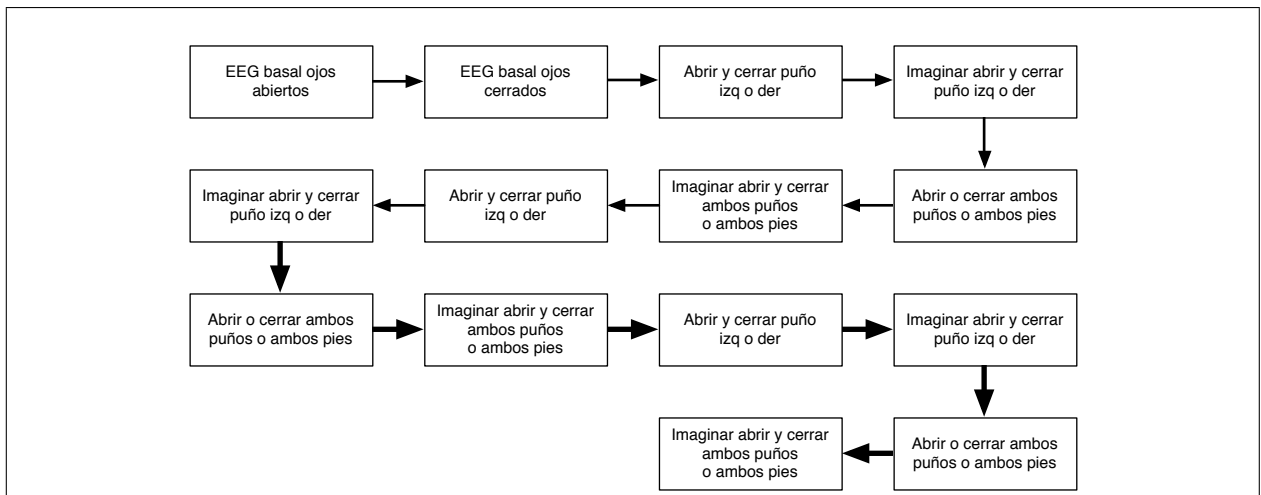


Figura 3.1: Secuencia de los 14 ensayos experimentales utilizados en la base de datos Physio-net.

Los datos de cada registro fueron almacenados en formato EDF<sup>4</sup>, conteniendo 64 señales de EEG (cada una de ellas muestreada a 160 Hz) además de un canal de anotación que

<sup>4</sup>El European Data Format (EDF) es un formato simple y flexible para el intercambio y almacenamiento de

permitió identificar cada tarea. Cada anotación incluye uno de tres códigos (T0, T1 o T2), donde

- T0 corresponde a reposo
- T1 corresponde al inicio del movimiento (real o imaginario) de
  - El puño izquierdo (en los ensayos 3, 4, 7, 8, 11 y 12)
  - Ambos puños (en los ensayos 5, 6, 9, 10, 13 y 14)
- T2 corresponde al inicio del movimiento (real o imaginario) de
  - El puño derecho (en los ensayos 3, 4, 7, 8, 11 y 12)
  - Ambos pies (en los ensayos 5, 6, 9, 10, 13 y 14)

Las clases de interés a clasificar fueron: reposo, imaginar abrir y cerrar el puño izquierdo e imaginar abrir y cerrar el puño derecho. Por lo tanto, se debió detectar las épocas de presentación en los ensayos 4, 8 y 12, los cuales correspondían a las clases de interés. Es decir, en los ensayos 4, 8 y 12 de cada uno de los 109 sujetos se detectaron todas las anotaciones T0, T1 y T2.

Las señales de la base de datos fueron exportadas a Matlab donde se realizó el procesamiento de ellas. Inicialmente las señales fueron procesadas usando un filtro pasa banda de [2,50] Hz del tipo Butterworth de quinto orden. Esto debido a que las bandas de frecuencia de mayor interés se encuentran entre las bandas Theta (4-8 Hz) y las bandas Gamma (36-44 Hz), particularmente las bandas  $\mu$  (8-12 Hz). Posteriormente las señales fueron centradas en cero y normalizadas para evitar efectos indeseados producto de la diferencia de magnitud de las señales. El primer segundo del registro fue descartado debido a que fue considerado como fase de preparación de la actividad de imaginación motora. Por lo tanto, se obtuvo 3 conjuntos de señales de EEG, donde cada conjunto contenía todas las épocas T0, T1 y T2.

## 3.2. Procesamiento de la Señal de EEG con Método Propuesto

Esta Tesis propone el uso de ICA y EMD en la etapa de procesamiento de la señal previo a la extracción de características. En la figura 3.2 se observa el diagrama de bloques del método propuesto.

---

señales biológicas de múltiples canales. EDF fue publicado en 1992 en *Electroencephalography and Clinical Neurophysiology*. Desde entonces, EDF se convirtió en el estándar de facto para los registros de EEG en equipos comerciales y proyectos de investigación multicéntricos.

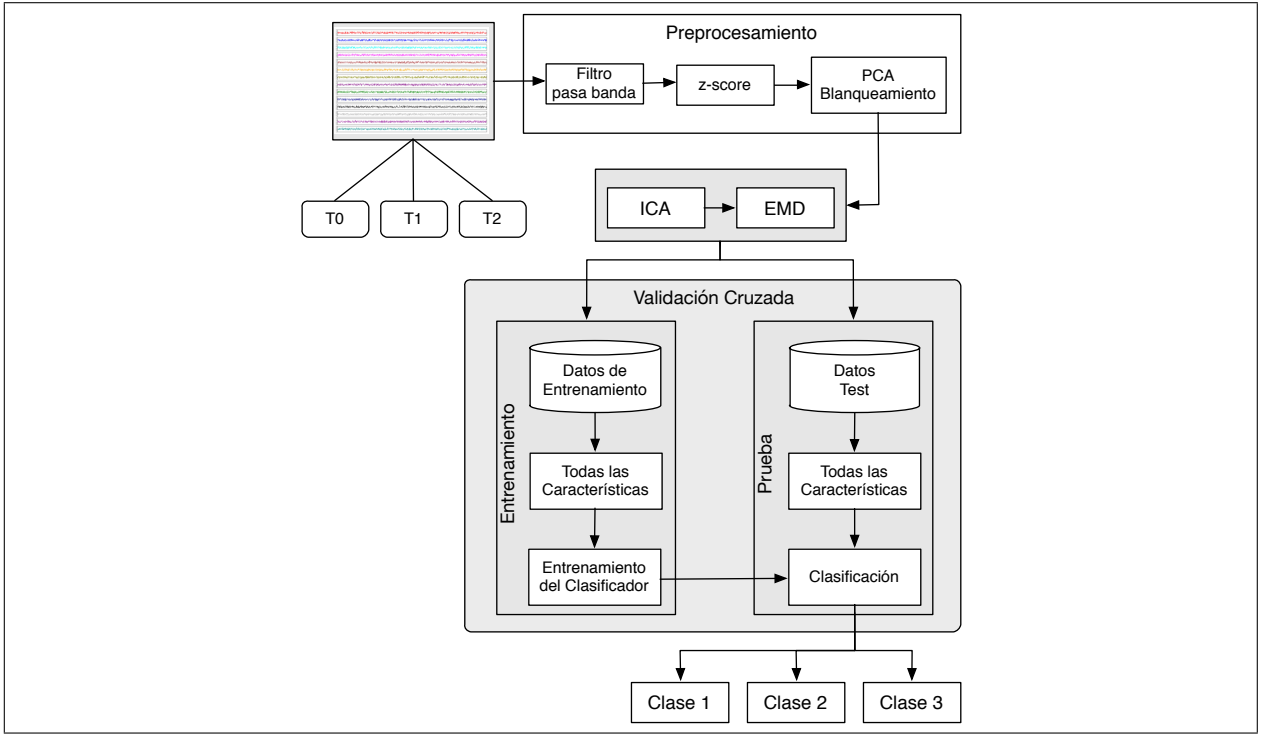


Figura 3.2: Diagrama de bloque del método empleado. La etapa de preprocesamiento incluyó z-score y blanqueamiento de los datos.

A continuación se describen las diferentes etapas de la metodología propuesta.

### 3.2.1. Preprocesamiento

Previo al procesamiento de las señales de EEG se realizó un preprocesamiento, éste incluyó el centrado en cero de la señal y la normalización de las amplitudes mediante el uso de z-score, el cual mide la distancia de cada punto a la media de los datos en términos de la desviación estándar. También se conoce como estandarización de los datos, aquí se convierten todas las amplitudes a una escala común con promedio cero y desviación estándar igual a uno<sup>5</sup>.

Posteriormente se utilizó PCA como estrategia de «blanqueamiento» de los datos antes de utilizar ICA. Aquí se transforma linealmente el vector de entrada  $\mathbf{x}$  (posterior a la aplicación de z-score) con el fin de obtener un nuevo vector  $\tilde{\mathbf{x}}$ , el cual se caracteriza porque sus componentes no están correlacionados y sus varianzas son igual a la unidad, en otras palabras la covarianza de la matriz  $\tilde{\mathbf{x}}$  es igual a la matriz identidad  $\Sigma \{ \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \} = \mathbf{I}$ .

Brevemente, si la matriz de covarianza de los datos de entrada  $\Sigma \{ \mathbf{x}\mathbf{x}^T \} = \mathbf{E}\mathbf{D}\mathbf{E}^T$ , donde  $\mathbf{E}$  es la matriz ortogonal de vectores propios de  $\Sigma \{ \mathbf{x}\mathbf{x}^T \}$  y  $\mathbf{D}$  es la matriz diagonal de sus valores propios. Luego, el blanqueamiento puede ser hecho mediante

$$x_{PCA} = \mathbf{E}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T\mathbf{x} \quad (3.1)$$

donde la matriz  $\mathbf{D}^{-\frac{1}{2}}$  es calculada de manera simple como  $\mathbf{D}^{-\frac{1}{2}} = \text{diag}(d_1^{-\frac{1}{2}}, \dots, d_n^{-\frac{1}{2}})$ , ahora se puede verificar  $\Sigma \{ \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \} = \mathbf{I}$ .

<sup>5</sup>z-score  $z = \frac{(x-\mu)}{\sigma}$ ; donde  $x$  es el dato,  $\mu$  es la media de los datos y  $\sigma$  la desviación estándar de los datos.

El uso de blanqueamiento fue fundamental previa aplicación de ICA (ver ecuación 3.2), pues en este caso se utilizó para asegurar la convergencia del algoritmo ICA, disminuir el tiempo de cómputo, pero no para reducir la dimensión de los datos, y para asegurar que no existiera correlación entre los datos de entrada. De esta manera, y de acuerdo a la ecuación 2.17 se tiene una mezcla lineal de  $n$  componentes independientes.

### 3.2.2. Procesamiento ICA y EMD

El objetivo del uso de ICA fue obtener las señales originales a partir de las mezclas derivadas de diversas fuentes independientes. Dado que ICA implica independencia estadística de las señales, en la presente Tesis se aplicó blanqueamiento de los datos de entrada (posterior a z-score), como se explicó en el Capítulo 3.2.1.

Para el método presentado en esta Tesis se tiene que, de acuerdo a la ecuación 3.1, el blanqueamiento transforma la matriz de entrada de la ecuación 2.18 en una nueva mezcla

$$x_{ICA} = \mathbf{E} \mathbf{D}^{-\frac{1}{2}} \mathbf{E}^T \tilde{\mathbf{A}} \mathbf{s} \quad (3.2)$$

La nueva matriz  $\tilde{\mathbf{A}}$  es ortogonal, lo que se puede comprobar a través de

$$\Sigma \{ \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T \} = \tilde{\mathbf{A}} \Sigma \{ \mathbf{s} \mathbf{s}^T \} \tilde{\mathbf{A}}^T = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T = \mathbf{I} \quad (3.3)$$

El blanqueamiento reduce el número de parámetros a ser estimados, en lugar de estimar  $n^2$  parámetros (elementos de la matriz original  $\mathbf{A}$ ), ahora sólo se calculan  $n(n-1)/2$ .

Por otra parte es necesario destacar que ICA no puede determinar las varianzas ni el orden de los componentes independientes, por tal motivo la eliminación de componentes independientes debe ser efectuado bajo algún método determinado y no por simple inspección visual [72, 114, 14]. Para evitar el sesgo propio que genera la dependencia de un evaluador experto para identificar qué componentes independientes deben ser eliminados, para la posterior reconstrucción de la señal sin la presencia de artefactos (u otra contaminación indeseada), se propuso un método de automatización en la selección de estos componentes independientes a eliminar.

Para identificar qué componentes independientes serían eliminados se dividió cada uno ellos en un número igual de segmentos. Se espera que la varianza de cada segmento en cada componente independiente sea la misma, excepto en los segmentos que contienen artefactos [22]. Se estableció como criterio de eliminación un umbral superior a la media de los datos más dos desviaciones estándar. Una vez identificados los componentes que contenían artefactos, éstos fueron eliminados. Por lo tanto, la señal de EEG se reconstruyó sin los componentes no deseados.

Posterior al uso de ICA se realizó descomposición de modo empírico en las señales de EEG reconstruidas sin los componentes independientes no deseados.

La descomposición de modo empírico (EMD) está basada en la Transformada de Hilbert-Huang (HHT) [63]. Este método descompone cualquier serie de tiempo en un set de oscilaciones espectralmente independientes.

EMD resulta útil para descomponer datos no estacionarios y no lineales [62] (ver capítulo 2.2), además ha mostrado ser útil en procesamiento de imaginaria motora [38]. Particularmente el algoritmo de EMD es un método para descomponer adaptativamente una señal en un

número finito y habitualmente pequeño de funciones, llamadas funciones de modo intrínseco (IMFs por su sigla en inglés). Esta función debe satisfacer dos condiciones:

- Condición 1: En toda la serie de tiempo el número de extremos y el número de cruces por cero deben ser iguales o diferir como máximo en 1. Notar que los extremos son mínimos locales o máximos locales<sup>6</sup>.
- Condición 2: En cualquier punto en la serie de tiempo, el valor medio de la envolvente definida por los máximos locales (envolvente superior) y la envolvente definida por los mínimos locales (envolvente inferior) es cero.

Por lo tanto, si la señal descompuesta cumple con estas condiciones, puede considerarse un IMF.

El método para determinar IMFs se describe a continuación (ver figura 3.3):

1. Sea  $x_{ICA}(t)$  la señal de entrada, luego de ser reconstruida sin los componentes independientes que fueron eliminados al aplicar ICA, y  $k$  el número de IMFs estimados

$$S(t) = x_{ICA}(t) \quad (3.4)$$

$$k = 0 \quad (3.5)$$

2. Se determina la envolvente superior  $ES(t)$  e inferior  $EI(t)$  de  $S(t)$  en base a los máximos locales y mínimos locales respectivamente.
3. El promedio de la envolvente  $m(t)$  es calculado como la media aritmética entre  $ES(t)$  y  $EI(t)$

$$m(t) = \frac{ES(t) + EI(t)}{2} \quad (3.6)$$

4. Un candidato  $h(t)$  a IMF es estimado como la diferencia entre  $S(t)$  y  $m(t)$ .
5. Si  $h(t)$  no satisface las condiciones para definir un IMF (condiciones 1 y 2) se reasigna como  $S(t)$  y se repiten los pasos 2 a 4. De otro modo, si  $h(t)$  es un IMF entonces se continua al paso 6.

---

<sup>6</sup>Si  $\mu \in \mathbb{R}$ , la función  $f: \mu \subset \mathbb{R}^n \rightarrow \mathbb{R}$  es diferenciable y  $x_0 \in \mu$  es un extremo local, entonces  $Df(x_0) = 0$ , esto es  $x_0$  es un punto crítico de  $f$ .

Si  $t$  alcanza su máximo local en  $x_0$ , entonces para cualquier  $h \in \mathbb{R}^n$  la función  $g(t) = f(x_0 + th)$  tiene un máximo local en  $t = 0$ . Así  $g'(0) = 0$ , pues como  $g(0)$  es máximo local,  $g(t) \leq g(0)$  para  $t > 0$  pequeño. Por lo tanto

$$g'(0) = \lim_{t \rightarrow t_0^+} \frac{g(t) - g(0)}{t} = 0$$

Análogamente para  $t < 0$  pequeño

$$g'(0) = \lim_{t \rightarrow t_0^-} \frac{g(t) - g(0)}{t} = 0$$

Ahora por regla de la cadena  $g'(0) = Df(x_0)h$ . Así  $Df(x_0)h = 0 \forall h$  de modo que  $Df(x_0) = 0$ . En resumen si  $x_0$  es un extremo local, entonces  $\frac{\partial f}{\partial x_i}(x_0)$ ,  $i = 1, \dots, n$ . En otras palabras  $\nabla f(x_0) = 0$ .



6. Si  $h(t)$  es un IMF, ahora  $c_k = h(t)$ , donde  $k$  es el  $k$  - *ésimo* componente.
7. El error medio cuadrático  $mse_{IMF}$  entre dos IMFs consecutivos  $c_{k-i}$  y  $c_k$  es calculado y este valor es comparado con la condición de término (usualmente un valor muy pequeño,  $10^{-5}$ ).

$$mse_{IMF} = \frac{1}{n} \sum_{i=1}^n (c_k - c_{k-i})^2 \quad (3.7)$$

8. Si la condición de término no se logra, el residuo parcial  $r_k$  es estimado como la diferencia entre un residuo parcial previo  $r_{k-1}$  y  $c_k$ , y su contenido es asignado a la variable  $S(t)$  y se repiten los pasos 2 a 4.

$$r_k = r_{k-1} - c_k \quad (3.8)$$

9. Si la condición de término se logra, entonces el algoritmo es finalizado y el residuo final  $r_{final}$  puede ser estimado como la diferencia entre  $S(t)$  y la suma de todos los IMFs.

Cuando el algoritmo se detiene, la señal de entrada  $x_{ICA}(t)$  puede ser representada como

$$x_{ICA}(t) = \sum_{k=1}^n c_k + r_{final} \quad (3.9)$$

donde  $n$  es el número de IMFs,  $c_k$  es el  $k$  - *ésimo* IMF y  $r_{final}$  es el residuo final.

En el presente estudio se consideró los 8 primeros IMFs [38, 12], producto de la aplicación de EMD, para la reconstrucción de la señal  $x_{EMD}(t)$  de acuerdo a la ecuación 3.10, ver figura 3.3.

$$x_{EMD}(t) = \sum_{k=1}^8 c_k \quad (3.10)$$

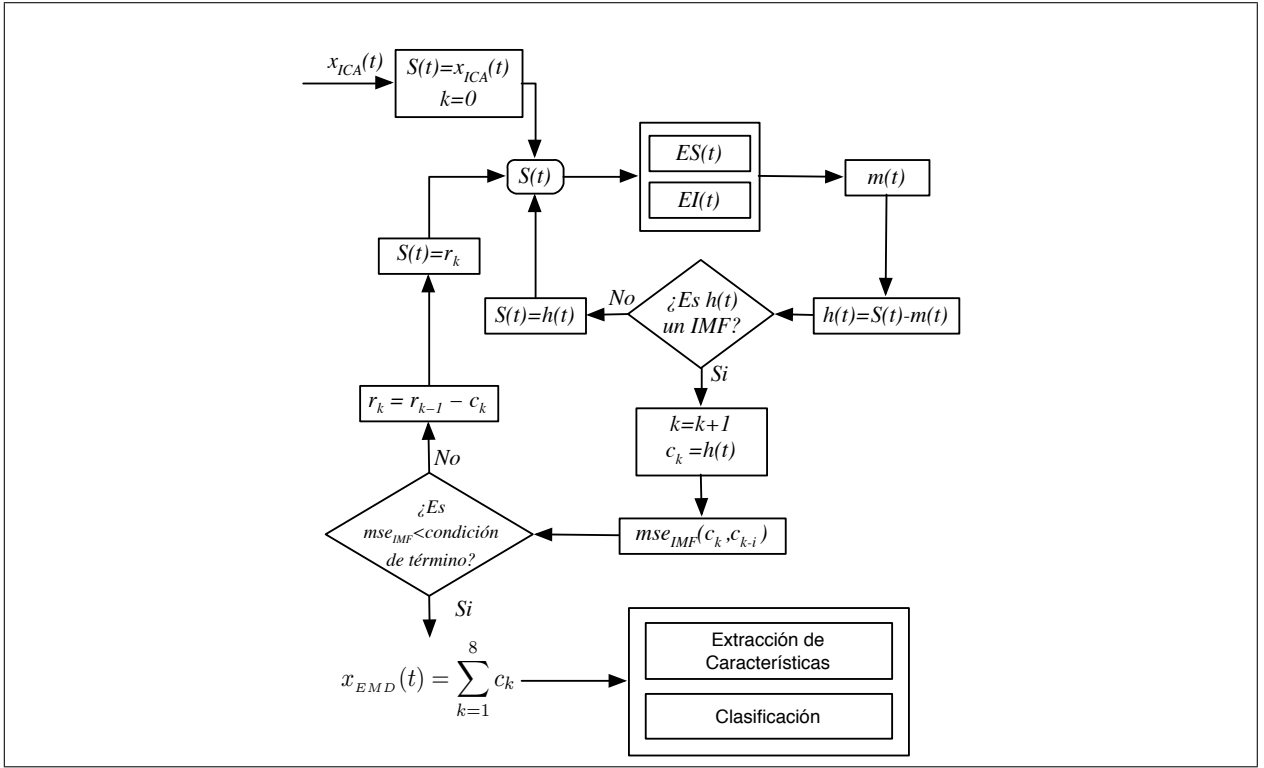


Figura 3.3: Método para determinar IMFs. Modificado de [12].

En resumen, se plantea el uso de z-score y blanqueamiento en la etapa de preprocesamiento, mientras que en la etapa de procesamiento se propone realizar ICA y EMD a las señales de EEG previo a la etapa de extracción de características y clasificación (ver figura 3.4). Destacando que la selección de los componentes independientes de ICA está dado por un criterio estadístico y no observador dependiente [22]; y la selección de los IMF en EMD está determinada por las funciones que contienen mayor información respecto a la señal de entrada [38, 12].

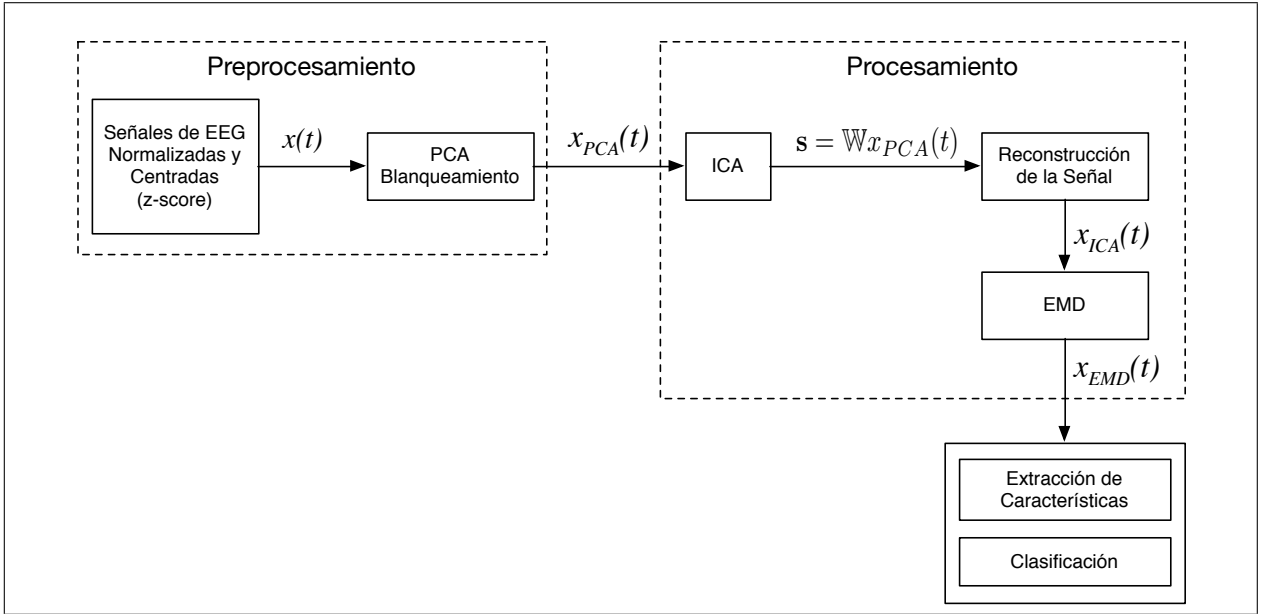


Figura 3.4: Diagrama de bloque que resume la metodología de preprocesamiento y procesamiento de la señal de EEG.

### 3.3. Extracción de Características

Para la extracción de características se usaron ventanas de 128 muestras en la señal  $x_{EMD}(t)$ . Esto debido a la necesidad de obtener un tiempo de respuesta que no genere un retardo significativo en caso de una implementación en tiempo real. Se decidió trabajar en el dominio del tiempo y de la frecuencia. Las características fueron establecidas basándose en estudios publicados desde 2010 a la fecha.

Se consideraron características estadísticas en dominio temporal debido a que entregan buena información acerca de la distribución de los datos que forman la señal de EEG [130]. Las características en el dominio del tiempo fueron definidas de acuerdo a los siguientes estudios [26, 32, 10, 78, 102, 15, 130, 87, 109, 80],

Además, se ha planteado que la clasificación de tareas mentales está determinada por la detección de patrones oscilantes dentro de las señales de EEG, por lo tanto se decidió usar también características espectrales. Las características en el dominio de la frecuencia fueron definidas de acuerdo a los siguientes estudios [26, 32, 102, 120, 15, 36, 34, 2, 35]. A continuación se describe cada característica usada en esta Tesis:

#### 1. Dominio del tiempo

- a) Kurtosis: La kurtosis es una medida de la dispersión que tiene una distribución respecto a su zona central.
- b) Media: Corresponde al promedio estándar de las observaciones.
- c) Skewness: Es es una medida de la asimetría de los datos alrededor de la media muestral.
- d) Varianza: Es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

- e) Rango: Es el intervalo entre el valor máximo y el valor mínimo. Permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango, más dispersos están los datos de un conjunto.

## 2. Dominio de la Frecuencia

- a) Frecuencia Dominante: Componente frecuencial de mayor amplitud en Transformada Discreta de Fourier
- b) Raíz Media Cuadrática de la Amplitud: La raíz media cuadrática o RMS (por su sigla en inglés) es una medida estadística de la magnitud de una cantidad variable, en este caso de la amplitud en el dominio de la frecuencia
- c) Media de la Amplitud en Dominio de la Frecuencia: Corresponde al promedio estándar de las observaciones de la amplitud en dominio de la frecuencia
- d) Amplitud Máxima en Dominio de la Frecuencia: Corresponde al valor máximo que presenta la amplitud en el dominio de la frecuencia
- e) Diferencia de Fase entre Canales Homolaterales: Corresponde a la diferencia expresada en radianes, en la posición de la frecuencia dominante, entre los canales homolaterales derecho e izquierdo

Como se mencionó anteriormente, para la construcción de cada vector de características se usó una ventana móvil de 128 muestras, donde cada vector de características fue de tamaño 133, obtenido a partir del producto del número de canales de registro por el número de características (salvo diferencia de fase), esto es:  $[(14 \text{ canales} \times 9 \text{ características}) + 7 \text{ diferencias de fase}]$ .

## 3.4. Clasificación

Dado que el objetivo de la Tesis es «*Optimizar la clasificación de tareas de imaginación motora mediante el uso automatizado de ICA y EMD en la fase de procesamiento de señales de EEG obtenidas de una base de datos pública*», no se realizó una selección de características, pues se pretende evaluar el efecto del uso de ICA y EMD en el procesamiento de la señal de EEG y no establecer qué características determinan una mejor clasificación<sup>7</sup>. Para la identificación de las tareas de imaginación motora se propone el uso de cuatro clasificadores estadísticos, Bayes [67], Logistic [82], Máquina de Soporte Vectorial usando Optimización Mínima Secuencial [118] y Dagging [137]. Cada uno de estos clasificadores fue evaluado mediante validación cruzada con 5 carpetas.

En cada uno de los 109 sujetos se evaluaron los ensayos 4, 8 y 12, obteniéndose parámetros estadísticos para cada uno de ellos. El promedio de los ensayos constituyó el resultado de la clasificación individual. Por lo tanto, se evaluó 3 veces a cada uno de los 109 sujetos para obtener, en primera instancia un valor representativo de cada sujeto y luego uno del total de la muestra.

Los parámetros estadísticos evaluados fueron:

1. Porcentaje correctamente clasificado

---

<sup>7</sup>En [138, 50, 16, 95, 10, 78, 32, 26, 102, 2, 120, 35, 15, 130, 36, 87, 34, 109, 159, 103, 5, 13, 4] se evalúan diferentes estrategias de extracción y selección de características.

2. Estadístico Kappa
3. Error medio absoluto
4. Error medio cuadrático
5. Tasa verdaderos positivos
6. Tasa falsos positivos
7. Sensibilidad
8. Especificidad
9. Valor predictivo positivo
10. Valor predictivo negativo

# Capítulo 4

## Resultados

A continuación se presentan los resultados de la clasificación de tres tareas de imaginaria motora en la base de datos Physionet. De un total de 109 sujetos, se analizaron 102, pues 7 archivos estaban incompletos. Como se mencionó en el capítulo 3, cada sujeto tuvo 14 ensayos, éstos estuvieron constituidos por un minuto de actividad basal de EEG con los ojos abiertos, un minuto de actividad basal de EEG con los ojos cerrados, tres ensayos de dos minutos de la tarea abrir y cerrar el puño, tres ensayos de dos minutos de la tarea imaginar abrir y cerrar el puño, tres ensayos de dos minutos de la tarea abrir o cerrar ambos puños o ambos pies y tres ensayos de dos minutos de la tarea imaginar abrir y cerrar ambos puños o ambos pies.

Dado que se decidió identificar las clases: reposo, imaginar abrir y cerrar el puño izquierdo e imaginar abrir y cerrar el puño derecho, se debió detectar las épocas de presentación de estas clases, las cuales fueron presentadas en los ensayos 4, 8 y 12. Es decir, en cada uno de estos ensayos se debió detectar reposo (identificado como T0), imaginar abrir y cerrar el puño izquierdo (identificado como T1) e imaginar abrir y cerrar el puño derecho (identificado como T2).

La metodología propuesta (capítulo 3) fue evaluada para los 102 sujetos en los ensayos 4, 8 y 12, ya que en éstos se realiza la presentación de las 3 clases. Luego se calculó el promedio de los tres ensayos para obtener una representación de la clasificación en cada sujeto, lo que arrojó 102 conjuntos de indicadores estadísticos (un conjunto por cada sujeto). De éstos se obtuvo el promedio de cada indicador como representación de toda la muestra.

Los resultados de los parámetros estadísticos obtenidos en la base de datos Physionet en la clasificación de tres clases usando los clasificadores Dagging, SMO, Bayes y Logistic, tanto con el método propuesto como sin el uso de éste, se presentan en la tabla 4.1.

Tabla 4.1: Comparación de los resultados de la clasificación para 3 tareas de imaginaria motora usando cuatro clasificadores con y sin el uso del método propuesto.

Parámetros Estadísticos en la Clasificación de Tres Tareas de Imaginería Motora

Indicadores	Sin ICA y EMD				Con ICA y EMD			
	Dagging	SMO	Bayes	Logistic	Dagging	SMO	Bayes	Logistic
Correctamente Clasificados [%]	87,5	93,6	54,0	96,8	93,9	97,7	58,8	98,8
Incorrectamente Clasificados [%]	12,5	6,4	45,9	3,2	6,1	2,3	41,2	1,2
Kappa	0,81	0,9	0,31	0,95	0,91	0,96	0,38	0,98
Error Medio Absoluto	0,26	0,24	0,31	0,03	0,24	0,23	0,28	0,01
Error Cuadrático Medio	0,32	0,3	0,54	0,13	0,3	0,28	0,51	0,08
Error Absoluto Relativo [%]	59,51	53,66	69,13	6,05	54,94	51,29	62,13	1,89
Error Cuadrático relativo [%]	68,26	63,47	113,34	27,29	62,69	59,82	107,69	15,92
Taza Verdaderos Positivos	0,87	0,94	0,54	0,97	0,94	0,98	0,59	0,99
Taza Falsos Positivos	0,06	0,03	0,23	0,02	0,03	0,01	0,21	0,01
F-Measure	0,87	0,94	0,52	0,96	0,93	0,98	0,57	0,99
Sensibilidad	0,87	0,94	0,54	0,97	0,94	0,98	0,59	0,99
Especificidad	0,94	0,97	0,77	0,98	0,97	0,99	0,79	0,99
Valor Predictivo Positivo	0,88	0,94	0,58	0,97	0,94	0,98	0,63	0,99
Valor Predictivo Negativo	0,94	0,97	0,79	0,98	0,97	0,99	0,81	0,99

Se realizó la prueba de Shapiro Wilk para evaluar el tipo de distribución de los resultados de los 102 sujetos estudiados, respecto al éxito de la clasificación. De acuerdo al test de Shapiro Wilk la hipótesis nula establece que la distribución de los datos es de tipo normal, por lo tanto para un nivel de significancia mayor de 0,05 se acepta esta hipótesis, de lo contrario se rechaza, estableciendo que la distribución de los datos no es normal. La tabla 4.2 muestra el resultado de la aplicación del test de Shapiro Wilk en el porcentaje de éxito de la clasificación.

Tabla 4.2: Evaluación del test de normalidad Shapiro Wilk para la distribución de los resultados de 102 sujetos evaluados con y sin el método propuesto.

Distribución de los Datos según Test de Shapiro Wilk

Shapiro Wilk	Sin ICA y EMD				Con ICA y EMD			
	Dagging	SMO	Bayes	Logistic	Dagging	SMO	Bayes	Logistic
P-value	0,1330	0,0017	0,9003	0,00003	0,1878	0,066	0,0096	0,0023
Distribución	Normal	No normal	Normal	No normal	Normal	Normal	No normal	No normal

Dado el tipo de distribución observada, se decidió evaluar si el método propuesto presenta diferencia estadísticamente significativa respecto al método tradicional en cuanto al porcentaje éxito en la clasificación. Para ello se aplicó la prueba de suma de rangos de Wilcoxon, demostrando que efectivamente existe diferencia estadísticamente significativa en la clasificación de tres tareas de imaginaria motora al utilizar el método propuesto. Los resultados se presentan en la tabla 4.3.

Tabla 4.3: Evaluación de significancia estadística en la clasificación de tres tareas de imaginación motora al utilizar procesamiento con ICA y EMD.

Resultado Test de Wilcoxon

Wilcoxon	Con ICA y EMD vs Sin ICA y EMD			
	Dagging	SMO	Bayes	Logistic
P-value	5,98e-16	9,10e-17	3,61e-4	6,48e-15
Significancia Estadística	Si	Si	Si	Si

En la figura 4.1 se observa la distribución de los valores de mediana, media, percentiles y valores extremos al comparar los resultados de la clasificación usando el método propuesto versus el método tradicional.

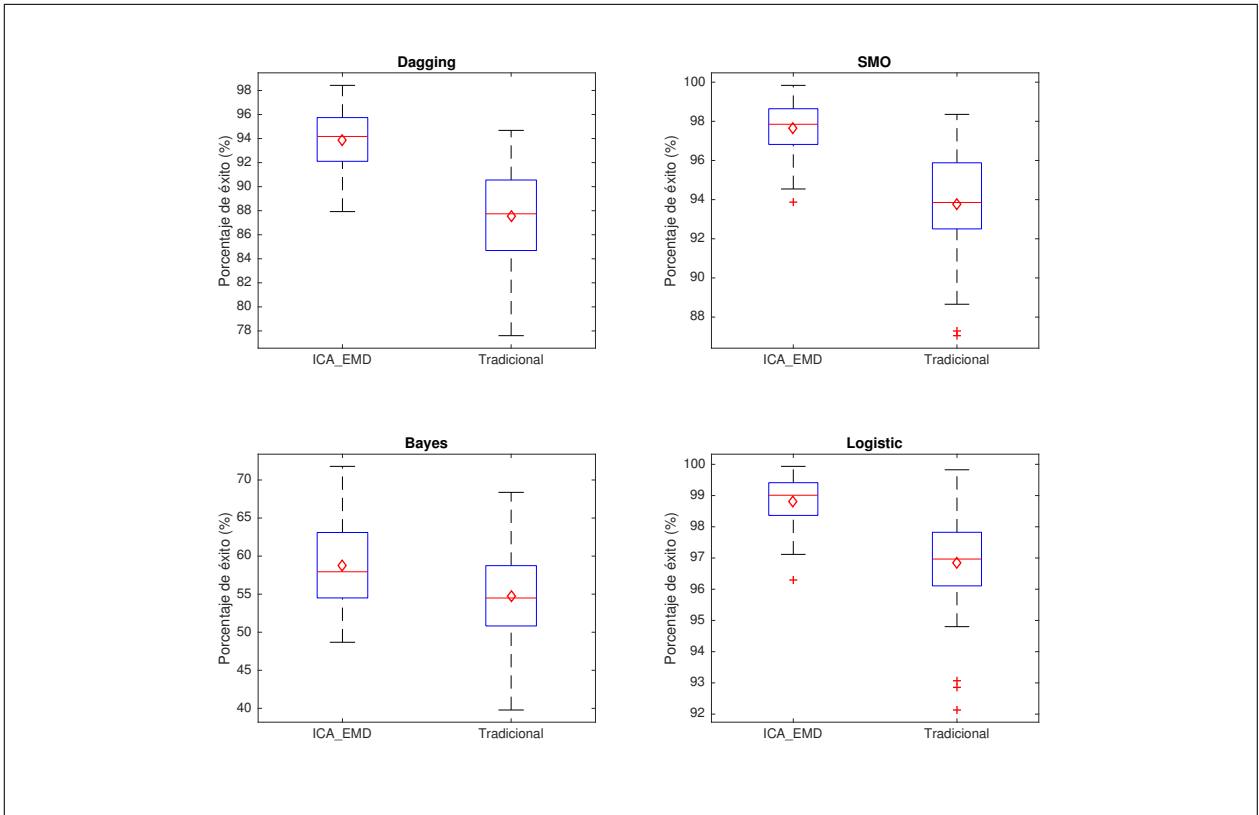


Figura 4.1: Comparación de gráfico de cajas respecto al resultado de la clasificación para los cuatro clasificadores usando el método propuesto versus el método tradicional. Cada caja muestra la distribución de los resultados entre el percentil 25 y el 75. La línea roja muestra la mediana y el rombo la media.



# Capítulo 5

## Conclusión y Discusión

Se propuso, implementó, evaluó y validó el uso automatizado de ICA y EMD en la optimización de la clasificación de tareas mentales de imaginación motora para señales de EEG en la base de datos pública *Physionet*, asumiendo como hipótesis que «El procesamiento automatizado de señales de EEG con ICA y EMD mejora el porcentaje de éxito en la clasificación de las tareas de imaginación motora». Para ello se usaron cuatro clasificadores: Clasificador de Bayes [67], Clasificador Logístico [82], Máquina de Soporte Vectorial usando Optimización Mínima Secuencial [118] y Clasificador Dagging [137].

Desde el punto de vista metodológico el objetivo de esta Tesis no fue establecer qué características determinan de mejor manera el patrón de una tarea de imaginación motora. Sin embargo, se considera relevante para futuros estudios, evaluar qué características definen mejor las tareas de imaginación motora y qué estrategias se deberían aplicar para la extracción de estas características. A continuación se presentan algunos estudios que abordaron esta problemática. Brevemente, en [65] se revisan diferentes técnicas de procesamiento de señales de EEG en BCI, incluyendo técnicas de extracción de características, aquí se establece que el uso de PCA o ICA puede contribuir a minimizar el número de características para maximizar el rendimiento en la clasificación. De manera similar, en [96] se plantea el uso de ICA como herramienta de preprocesamiento de la señal de EEG en BCI, aquí la extracción de características se realiza en el dominio del tiempo mediante ventaneos que detectan ERPs. En [88] se propone el uso de algoritmos basados en deep learning para la definición de características, particularmente usando Máquina de Boltzmann Restringida (RBM). En [20] se establece que los principales métodos de extracción de características se basan en actividad sensorio motora usando densidad espectral de potencia, representaciones tiempo-frecuencia, y modelos paramétricos. En [21] se plantea que las características usadas en sistemas de BCI que buscan controlar el movimiento de robots corresponden a características en el dominio del tiempo, como amplitudes de potenciales relacionados con evento, características en el dominio de la frecuencia, como densidad espectral de potencia, características en ambos dominios (tiempo-frecuencia) y parámetros autoregresivos. En [107] se presenta una revisión de diferentes estrategias de procesamiento de señales de EEG, destacando que los métodos de extracción de características más comúnmente usados son el ICA, el PCA y el CSP. En [149] se sostiene que las características más usadas en BCI son la amplitud en el tiempo, la densidad espectral de potencia, parámetros autoregresivos adaptativos y representaciones tiempo-frecuencia. Además señalan que los algoritmos más usados para la extracción de características espaciales incluyen principalmente PCA, ICA y CSP. En [43] se utilizó LDA y PCA para extraer características a partir de la base de datos pública BCI competition 2003

dataset Ia, obteniendo un porcentaje de 94,2% de éxito en la clasificación de dos tareas de imaginaria motora. En [59] se propone un método basado en redes Bayesianas para analizar tareas de imaginaria motora, extrayendo características en función de los bordes comunes en diferentes tareas (dada la diferencia existente en las diferentes tareas de imaginaria motora, los bordes comunes en estas posiciones podrían contener información de discriminación). El método fue evaluado en el conjunto de datos de BCI competition dataset IIIa y un conjunto de datos propio, obteniendo un porcentaje de éxito en la clasificación de 93% y 88% respectivamente.

Sin duda que la principal ventaja del uso automatizado de ICA y EMD durante el procesamiento de las señales de EEG es el mejor resultado en la clasificación. Sin embargo, considerando que una de las principales dificultades al usar ICA es definir qué componentes independientes deben ser eliminados para la posterior reconstrucción de la señal, el método propuesto ofrece una alternativa eficaz al automatizar este proceso. Una herramienta muy usada, pero que carece de evidencia científica, es la inspección visual de los componentes independientes. En este estudio no se consideró esta alternativa y se usó el criterio de eliminación basado en la diferencia de varianzas de segmentos generados en cada componente independiente, la varianza de cada segmento debería ser la misma excepto en los segmentos que contienen artefactos. Se estableció como criterio de eliminación un umbral superior a la media de los datos más dos desviaciones estándar. Esta técnica propuesta no está supeditada al sesgo del operador y por lo tanto no es operador dependiente.

En el caso de EMD se decidió trabajar con los primeros 8 IMFs. Sin embargo la determinación de los IMF más representativos es motivo de futuras investigaciones y está siendo investigado por el autor en la actualidad.

Finalmente, el uso de la base de datos pública Physionet para la validación de la metodología propuesta demostró que el uso de ICA y EMD es eficaz para mejorar la clasificación de tareas de imaginaria motora. Demostrando que existe diferencia estadísticamente significativa al comparar los resultados entre el método propuesto y el método tradicional, en los cuatro clasificadores usados.

Al comparar el resultado de los diferentes clasificadores, se observa que los mejores resultados se obtuvieron con Logistic, obteniendo un porcentaje de 98,8% de aciertos en la clasificación, seguido por SMO con 97,7%, Dagging con 93,9% y finalmente Bayes con 58,8%. Esto demuestra que los criterios de elección de los clasificadores fue correcto, sobre todo considerando las condiciones no lineales, no estacionarias y variante en el tiempo de las señales de EEG.

Si bien la diferencia entre Logistic y SMO es poca, en el caso de SMO el percentil 75 se encuentra en 98,64%, mientras que en Logistic se encuentra en 99,41%, para el caso del percentil 25 en SMO se encuentra en 96,82% y en Logistic en 98,36%. Esto implica que en el caso de Logistic la distribución entre el percentil 25 y 75 es sólo de 1,05% entorno a una mediana de 99,01% y una media de 98,82%, mientras que en el caso de SMO corresponde a 1,82% entorno a una mediana de 97,85% y una media de 97,65%. Cabe destacar que tanto en Logistic y SMO el percentil 25 obtenido con el método propuesto es superior al percentil 75 obtenido con el método tradicional.

A pesar que el clasificador Dagging obtuvo un rendimiento menor a Logistic y SMO, el efecto del método propuesto fue mayor, pues para este caso el percentil 25 fue de 92,11%, superior al percentil 75 obtenido con el método tradicional (90,55%). Además, para Dagging se obtuvo una mediana de 94,17% y una media de 93,89%, mientras que el caso del método tradicional la mediana fue de 87,74% y la media de 87,5%.

Por otro lado, queda en evidencia que el método de clasificación probabilística de Bayes no entregó buenos resultados, aun cuando éstos mejoraron con el método propuesto.

En la actualidad se ha dado especial énfasis a los métodos de procesamiento de la señal de EEG para una implementación exitosa en BCI, enfocándose en identificar las características que mejor definen las propiedades de la señal y qué clasificadores logran mejores resultados. Sin embargo, no se ha hecho énfasis en el procesamiento previo a la extracción de características y clasificación. Los resultados obtenidos en la presente Tesis muestran la relevancia del método propuesto como herramienta para optimizar el rendimiento de la clasificación. Sin embargo, también resulta fundamental la correcta selección de los clasificadores.

La tabla 5.1 presenta un resumen de los porcentajes de clasificación obtenidos en diferentes estudios. Para ello se consideraron aquellos que utilizaron bases de datos públicas<sup>1</sup>.

Tabla 5.1: Clasificación de tareas de imaginación motora usando bases de datos públicas. Revisión bibliográfica.

Clasificación de Tareas de Imaginería Motora Usando Base de Datos Públicas

Referencia	Base de Datos	Nº de clases	validación cruzada	% de éxito en clasificación
Alomari, M (2013) [7]	Physionet	2	No	88,9
Park, C (2013) [111]	Physionet	2	No	77,7
Sleight, J (2009) [131]	Physionet	2	Si	64
Park, C (2014) [112]	Physionet	2	Si	72,4
Alomari, M (2014) [5]	Physionet	2	No	89,1
Alomari, M (2014) [6]	Physionet	2	No	84,5
Loboda, A (2014) [91]	Physionet	1	No	82,5
Alomari, M (2014) [4]	Physionet	2	No	74,9
Li, Y (2016) [85]	Propia	3	Si	71,5
	BCI competition II-IV	2	No	85
	BCI competition II-Ib	2	No	60,56
Liu, Y (2016) [89]	BCI competition III-IVa;	2	Si	90
	Propia	2	Si	70
Baali, H (2015) [16]	BCI competition III-IIIa	4	Si	81,3
Qi, F (2015) [119]	BCI competition III-IVa	2	Si	82,15
	BCI competition III-IIIa	4	Si	93,3
	BCI competition IV-IIa	4	Si	82,61
Aghaei, A (2016) [3]	BCI Competition III-V	3	Si	65 %
	BCI competition IV-IIa	4	Si	47,98
Meng, J (2015) [101]	BCI Competition III-IVa	2	Si	90,7
Taran, S (2017) [134]	BCI Competition III-IVa	2	Si	97,5
Meng, J (2014) [100]	BCI competition IV-IIa	2	Si	76,7
Bashar, S (2016) [19]	BCI competition II-III	2	No	90,7
Dutta, S (2017) [46]	Colorado University	5	Si	83,33
<b>Método Propuesto</b>	<b>Physionet</b>	<b>3</b>	<b>Si</b>	<b>98,8</b>

<sup>1</sup>En la búsqueda se consideró sólo revistas (no conferencias) que expresaran claramente el porcentaje de éxito de la clasificación.

En [38] también se usó la base de datos Physionet para la clasificación de dos clases, sin embargo sólo se consignaron resultados de sensibilidad y especificidad, obteniendo 75,7% y 71,5% respectivamente.

Al comparar los estudios mencionados en la tabla 5.1 con los resultados obtenidos en esta Tesis se puede concluir que el método de procesamiento propuesto basado en el uso automatizado de ICA y EMD entrega mejores resultados que lo descrito en la literatura.

Finalmente, los buenos resultados de la metodología propuesta, comparada con los resultados obtenidos sin su uso, permiten aceptar la hipótesis planteada, es decir: «El procesamiento automatizado de señales de EEG con ICA y EMD mejora el porcentaje de éxito en la clasificación de las tareas de imagería motora».

Se espera que en futuras investigaciones se pueda implementar la metodología propuesta en un sistema de BCI que opere en tiempo real.

# Capítulo 6

## Contribución Científica y Trabajo Futuro

De acuerdo a [103], los enfoques existentes para remover artefactos en señales de EEG son eficientes bajo estrictas medidas de control en laboratorio, como salas aisladas, sistemas de adquisición de tipo clínico, rigurosos protocolos experimentales y precisos controles de estímulos, pero son impracticables en escenarios propios de la vida diaria. En este aspecto, la gran mayoría de estudios que usan EEG como método de registro para BCI han sido desarrollados en condiciones de laboratorio, incluso restringiendo el movimiento a los usuarios.

Por lo tanto se necesita investigar y desarrollar nuevas técnicas de procesamiento de las señales de EEG que permitan la aplicación de BCI en condiciones diferentes al laboratorio.

La presente Tesis presenta un nuevo modelo de procesamiento de la señal de EEG para su uso en BCI. A continuación se presentan las principales contribuciones de este estudio.

### 6.1. Optimización de la Clasificación de Tareas de Imaginería Motora

Sin duda que una de las principales contribuciones del método propuesto está dado por el mejor resultado obtenido al clasificar tres tareas de imaginería motora. Es decir, el uso secuenciado de ICA y EMD mejora el rendimiento de los clasificadores estudiados. Esto queda demostrado en la evaluación de la significancia estadística al comparar el uso del método propuesto con el método tradicional, aun cuando se usaron las mismas características y los mismos clasificadores. Además, el uso de una base de datos pública permitió validar los resultados.

Otro aspecto a considerar es el hecho de obtener un muy buen rendimiento en la clasificación de tres clases. En base a lo presentado en la presente Tesis, corresponde al mejor resultado obtenido en la literatura.

Cabe mencionar que los resultados obtenidos con el método tradicional, aun cuando estadísticamente son peores al compararlos con el método propuesto, sí presentan mejoría respecto a los resultados publicados en estudios previos, lo que demuestra que la automatización propuesta (capítulo 6.2) y el enfoque no lineal y no estacionario (capítulo 6.3) en el procesamiento de las señales permite optimizar la clasificación de las tareas de imaginería motora.

## 6.2. Automatización en la Eliminación de Componentes Independientes

Otra importante contribución está determinada por la automatización en el uso de ICA, eliminando el sesgo entregado por la dependencia del operador al definir qué componentes independientes deben ser eliminados para luego reconstruir la señal sin los elementos contaminantes. Como se mencionó en los capítulos 1.2 y 2.2.3, una de las principales dificultades en el uso de ICA es el reconocimiento de los componentes independientes que poseen contenido no relacionado con la señal que se desea procesar. Aquí cobra vital importancia la participación de un evaluador experto que, mediante la inspección visual, pueda determinar qué componentes independientes deben ser eliminados.

Evidentemente el primer sesgo en este método está determinados por la expertise del evaluador, haciéndolo evaluador dependiente. Luego, otro aspecto a considerar es que el uso de inspección visual impide la implementación de un sistema que opere en tiempo real. De esta manera, el método propuesto demostró, de acuerdo a los resultados obtenidos, que la automatización en el proceso de selección y eliminación de los componentes independientes no deseados mejora el rendimiento de los clasificadores, al identificar diferentes tareas de imaginación motora.

## 6.3. Enfoque No Lineal y No Estacionario en el Procesamiento de las Señales de EEG

Tradicionalmente, y por razones de conveniencia, las señales biológicas son consideradas estacionarias, lineales e invariantes en el tiempo. Sin embargo, éstas no cumplen esas condiciones.

Una importante contribución del presente estudio es el abordaje no lineal, variante en el tiempo y no estacionario de las señales de EEG, evidenciado al utilizar herramientas como EMD y los clasificadores propuestos.

El uso de EMD da cuenta del tratamiento de las señales de EEG considerando su naturaleza no lineal y no estacionaria. Se espera que en futuras investigaciones se pueda determinar el efecto aislado del uso de EMD durante el procesamiento de la señal de EEG en la clasificación de tareas de imaginación motora. Así también, se espera identificar un método apropiado para la selección de los IMFs relevantes previo a la etapa de extracción de características.

Otra evidencia relevante de lo anteriormente expuesto, está determinada por el excelente resultado obtenido por los clasificadores no lineales (Dagging, SMO y Logistic), mientras que el clasificador que obtuvo peor resultado fue el clasificador lineal (Bayes).

## 6.4. Clasificación de Tres Clases

El uso de un modelo que logra clasificar tres o más clases, permitirá un mejor uso en ambientes clínicos. Si consideramos la mayoría de los estudios publicados, en cuanto al número de tareas de imaginación motora clasificadas, se observa que éste corresponde a dos clases. Ésto implica que, de implementarse en un ambiente clínico, el usuario estará restringido a obtener respuestas de tipo binaria, por ejemplo: SI/NO. Sin embargo, el clasificar tres clases, con alto nivel de precisión, permitirá el uso de un menú contextual. Esto es, una respuesta puede tener

un determinado significado en un menú y la misma respuesta puede tener otro significado en un menú diferente. Para poder acceder al menú contextual, el usuario necesariamente debería poder responder más allá de un modelo binario.

## 6.5. Limitaciones y Trabajo Futuro

A pesar de los favorables resultados obtenidos, resulta evidente que una importante limitación del presente estudio es la falta de una validación en tiempo real del método propuesto.

Sin embargo, esta limitación abre una importante oportunidad para el desarrollo de una línea de investigación en sistemas de BCI que incluyan la metodología propuesta y que opere en tiempo real. Además abre las puertas para futuras investigaciones utilizando clasificadores no lineales.

Otro aspecto relevante de investigar es el tamaño del ventaneo para la determinación del vector de características, pues se busca evitar el retardo en la respuesta o actuación del sistema de BCI en una implementación en tiempo real.

Además se presenta la oportunidad de desarrollar nuevas líneas de investigación en sistemas de BCI que puedan funcionar en condiciones ambientales diferentes a las observadas en Laboratorios. Abordando variables que la base de datos pública utilizada no consignó, tales como: parámetros de concentración, condiciones del entorno, efecto de la experiencia o tiempo de entrenamiento en el éxito de la clasificación. Otro aspecto relevante y necesario de investigar es la estrategia de registro, esto es: ¿cuál es la mejor estrategia para entrenar las tareas de imaginación motora en un usuario de BCI?

Sin duda que este primer proyecto de investigación en BCI con la metodología propuesta no resuelve todas las problemáticas antes señaladas. Sin embargo, es el primero que aborda el uso secuenciado de ICA y EMD de manera automatizada.

El desarrollo de este estudio dio origen a nuevas interrogantes, y con ello a diversas líneas de investigación que se derivan de la metodología propuesta.

En resumen, el presente estudio permite el inicio de las siguientes líneas de investigación en BCI:

1. Desarrollo de BCI en tiempo real, tomando en cuenta las siguientes variables:
  - a) Tipo de Ventaneo y tamaño de ventana
  - b) Extracción de características
  - c) Selección de IMFs
  - d) Condiciones del entorno
2. Optimización de BCI mediante uso de clasificadores no lineales
3. Efecto de la experiencia y tiempo de entrenamiento en la clasificación de tareas mentales
4. Efecto de la concentración en la clasificación de tareas mentales.
5. Estrategia de registro de las tareas de imaginación motora
  - a) Número de canales
  - b) Localización de electrodos

- c) Tiempo de registro
- d) Estrategia de entrenamiento

## 6.6. Trabajo Relacionado-Estudio de Casos

La misma metodología propuesta en esta Tesis fue implementada en un estudio de casos, aquí las señales se obtuvieron con el sistema de registro de bajo costo Emotiv Epoc. Ésto con el fin de proyectar futuras aplicaciones de BCI en situaciones de la vida diaria y no restringidas a ser usadas sólo en un laboratorio. Además se privilegia un período breve de registro, simulando las condiciones en las que podría utilizarse un dispositivo de este tipo.

En [9] se evaluó la clasificación de dos tareas mentales en tres sujetos, concluyendo que el uso de Emotiv permitió obtener resultados comparables a dispositivos avanzados de EEG como BrainAmp®<sup>1</sup>, además Emotiv proporcionó mayor comodidad, ya que utiliza un canal de comunicación inalámbrico evitando la restricción propia del uso de cables. Por otro lado en [97] se establece que el desempeño de Emotiv Epoc es comparable al de dispositivos profesionales, particularmente Biosemi Active 2®, cuando se usa el mismo número de sensores y su posición. Sin embargo, en [155] se revisaron los avances en tecnologías de BCI en la industria comercial, concluyendo que Emotiv es útil para controlar bajos niveles de información (tal como movimiento hacia arriba, abajo, etc), pero no es suficientemente preciso ni confiable para controlar movimientos finos como los que pueda hacer un robot.

Lo antes expuesto permite considerar la posibilidad de desarrollar una interfaz cerebro computador utilizando un sistema de adquisición de bajo costo, lo que sumado a un adecuado método de procesamiento logre optimizar la clasificación de tareas de imaginería motora obtenidas en condiciones diferentes al laboratorio y con periodos breves de registro.

Brevemente, la adquisición de los datos se realizó en el Laboratorio de Visión Computacional del Departamento de Ingeniería Eléctrica de la Universidad de Chile en condiciones de funcionamiento normal, lo que implica que todos los integrantes del Laboratorio (quienes no estaban participando del estudio) realizaron sus actividades de manera regular sin tomar precauciones respecto a guardar silencio o restricciones en el tránsito al interior del recinto para facilitar la concentración de los participantes del estudio. Tampoco se tomó precaución en cuanto a factores contaminantes electromagnéticos de la señal, esto es: se asumió la contaminación de origen muscular, la asociada al movimiento de los ojos y pestañeo, así como la contaminación propia del cross-talk y volumen de conducción. Tampoco se restringió el movimiento de los voluntarios que participaron del estudio (figura 6.1).

---

<sup>1</sup>[http://www.brainproducts.com/products\\_by\\_apps.php?aid=5](http://www.brainproducts.com/products_by_apps.php?aid=5)





Figura 6.1: Sistema de registro en condiciones de funcionamiento habitual del Laboratorio de Visión Computacional de la Universidad de Chile. El usuario está expuesto a múltiples estímulos externos diferentes a las tareas de imaginación motora.

Para la adquisición de las señales de electroencefalografía se usó el dispositivo Emotiv EPOC de 14 canales y dos referencias. Esta configuración corresponde al sistema de registro 10-20. La frecuencia de muestreo fue de 128Hz. Es un dispositivo inalámbrico y en la actualidad es considerado uno de los mejores sistemas de registro de EEG de bajo costo en términos de aplicaciones [9]. La distribución de los electrodos, de acuerdo al sistema internacional 10/20 es: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 más los electrodos de referencia P3 (CMS) y P4 (DRL).

se utilizó la interfaz gráfica de Emotiv (Epoc Control Panel) para el desarrollo de las tareas de imaginación motora. Esta interfaz gráfica muestra al usuario un cubo flotando en un espacio visto en perspectiva como se observa en la figura 6.2. Además permite al usuario mover el cubo en diferentes direcciones.

Para el presente estudio se consideraron las siguientes clases de imaginación motora:

1. Imaginar movimiento de mano izquierda, realimentada positivamente mediante el movimiento del cubo a izquierda
2. Imaginar movimiento de mano derecha, realimentada positivamente mediante el movimiento del cubo a derecha
3. Imagina el reposo de ambas extremidades superiores, retroalimentado con la posición fija del cubo en el centro de la interfaz.

La figura 6.2 muestra el panel que se utilizó en el paradigma experimental.

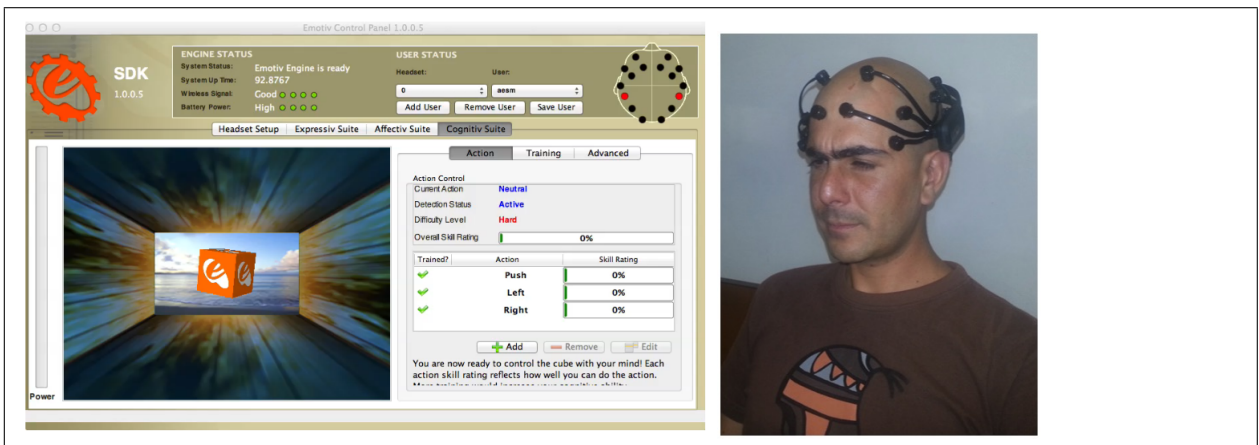


Figura 6.2: Izquierda: Panel de control utilizado en el paradigma experimental; Derecha: Sujeto utilizando el sistema de registro Emotiv Epoc.

Un total de 8 sujetos sanos participaron del estudio de casos, cada participante fue instruido para realizar tres tareas de imaginación motora, todas realimentadas a través del movimiento del cubo en la interfaz gráfica. Como se observa en la figura 6.3 cada registro incluyó 5 minutos de práctica o entrenamiento del sistema, 1 minuto de descanso, 10 segundos de registro de tarea 1 (reposo: equivalente a mantener el cubo inmóvil), 1 minuto de descanso, 10 segundos de registro de tarea 2 (imaginar movimiento de mano derecha: equivalente a mover el cubo a la derecha), 1 minuto de descanso, 10 segundos de registro de tarea 3 (imaginar movimiento de mano izquierda: equivalente a mover el cubo a la izquierda). Es importante destacar que durante la fase de entrenamiento se configura el panel gráfico para que el cubo se mueva a la derecha cuando el participante imagine mover su mano derecha, se mueva a la izquierda cuando el participante imagine mover su mano izquierda y permanezca inmóvil cuando no imagina movimiento de sus extremidades.

Cada sujeto realizó esta estrategia dos veces con un periodo de 10 minutos entre registros. Es importante destacar que no se retiró el dispositivo entre registros.

Los tiempos definidos en la metodología de registro se establecieron considerando que en condiciones reales se requieren periodos breves de entrenamiento y registro, evitando las limitaciones propias de los registros complejos que requieren largos periodos de entrenamiento para su aplicación.

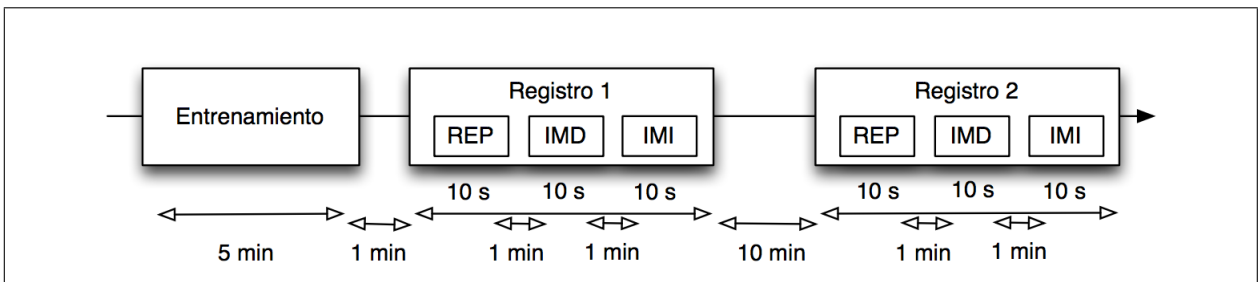


Figura 6.3: Metodología de Registro

Para el desarrollo de este estudio de casos, el Departamento de Ingeniería Eléctrica no se solicitó aprobación por comité de Bioética.

Las señales fueron adquiridas en formato bdf usando el software de adquisición de Emotiv, posteriormente las señales fueron exportadas a Matlab donde se realizó el mismo preprocesamiento y procesamiento descritos en la presente Tesis. Cabe destacar que el primer registro se usó como entrenamiento y el segundo fue usado para evaluar.

Los resultados preliminares muestran diferencias entre el método propuesto y el método tradicional (tabla 6.1).

Tabla 6.1: Comparación de resultados de clasificación para 3 tareas de imaginería motora usando cuatro clasificadores con y sin el uso del método propuesto en 8 sujetos voluntarios (resultado preliminar).

Parámetros estadísticos en Estudio de Casos para Clasificación de Tres Tareas de Imaginería Motora

Indicadores	Sin ICA y EMD				Con ICA y EMD			
	Dagging	SMO	Bayes	Logistic	Dagging	SMO	Bayes	Logistic
Correctamente Clasificados [%]	48,2	43,57	38,28	37,47	53,84	51,18	57,10	48,6
Incorrectamente Clasificados [%]	51,80	56,43	61,72	62,53	46,16	48,82	42,9	51,94
Kappa	0,22	0,15	0,07	0,06	0,31	0,27	0,36	0,22
Error Medio Absoluto	0,40	0,41	0,41	0,42	0,38	0,37	0,29	0,35
Error Cuadrático Medio	0,49	0,51	0,64	0,64	0,47	0,47	0,53	0,58
Error Absoluto Relativo [%]	90,20	92,21	92,61	93,87	84,92	83,35	64,34	77,91
Error Cuadrático relativo [%]	102,59	108,12	135,81	136,42	99,23	98,94	111,57	122,96
Taza Verdaderos Positivos	0,48	0,44	0,38	0,37	0,514	0,51	0,57	0,48
Taza Falsos Positivos	0,26	0,28	0,31	0,31	0,23	0,24	0,21	0,26
F-Measure	0,38	0,33	0,29	0,26	0,46	0,43	0,46	0,42
Sensibilidad	0,48	0,44	0,38	0,38	0,54	0,51	0,57	0,48
Especificidad	0,74	0,72	0,69	0,69	0,77	0,76	0,79	0,74
Valor Predictivo Positivo	0,6	0,55	0,42	0,5	0,51	0,51	0,70	0,54
Valor Predictivo Negativo	0,78	0,76	0,7	0,71	0,82	0,81	0,82	0,78

Aun cuando los resultados son preliminares, y que el tamaño de la muestra es pequeño, se aprecia un mejor rendimiento de la clasificación en las señales procesadas con ICA y EMD (figura 6.4).

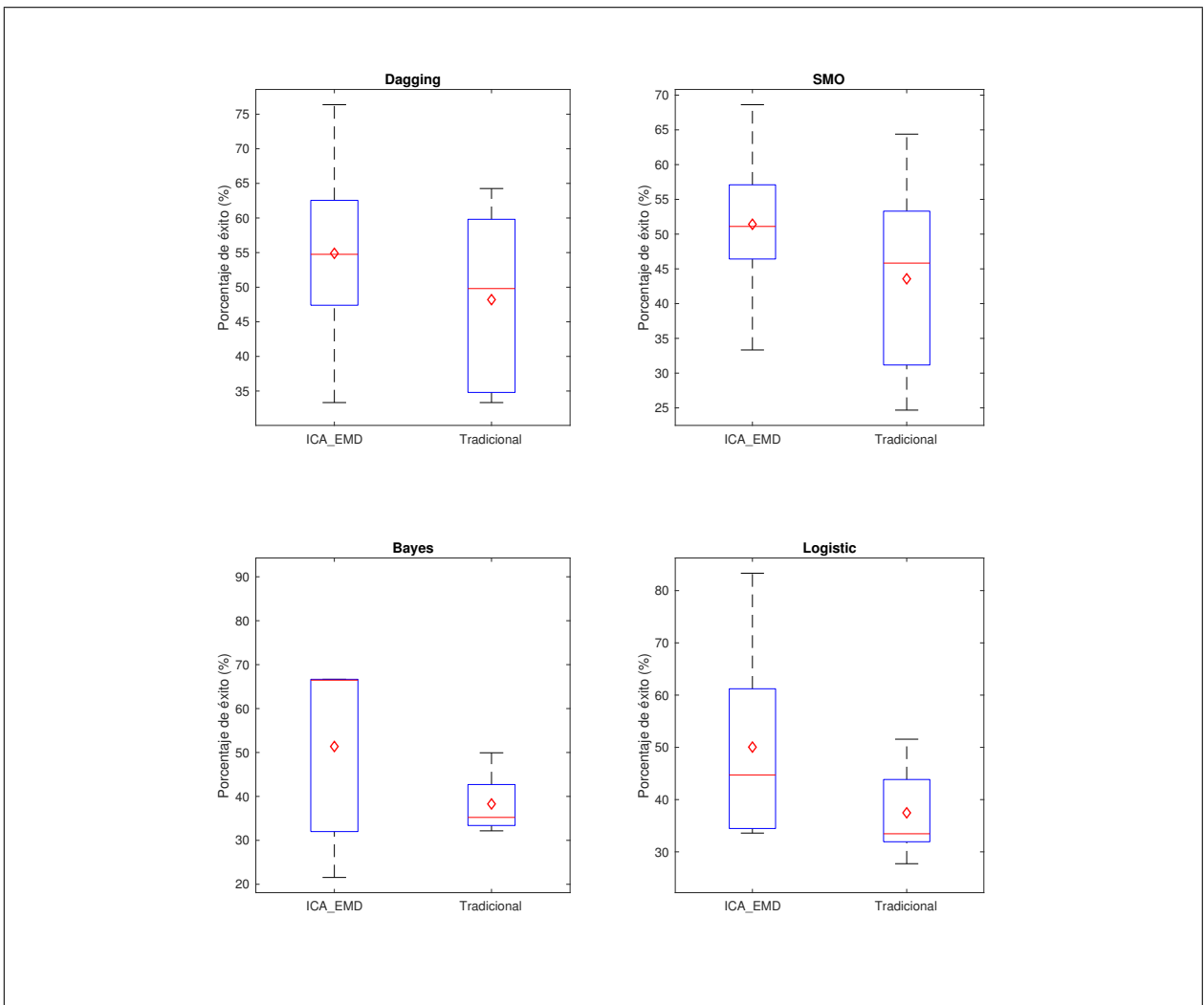


Figura 6.4: Resultados preliminares de un estudio de casos aplicando procesamiento con ICA y EMD para la clasificación de tareas de imaginación motora.

Sin duda, los resultados preliminares de este reporte de casos no permiten realizar un análisis estadístico riguroso. Sin embargo, presentan la misma tendencia de lo observado en la base de datos pública, en cuanto al mejor rendimiento de los clasificadores cuando se utiliza el método propuesto.

Como se mencionó en el capítulo 6, estas observaciones dan origen a nuevas interrogantes y generan oportunidades para el desarrollo de nuevas líneas de investigación.

# Bibliografía

- [1] Khald AI Aboalayon, Helen T Ocbagabir, and Miad Faezipour. Efficient sleep stage classification based on eeg signals. In *Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island*, pages 1–6. IEEE, 2014.
- [2] Faris A Abuhashish, Mohd S Sunar, Hoshang Kolivand, Farhan Mohamed, and Dzul-kiffi B Mohamad. Feature extracted classifiers based on eeg signals: A survey. *Life Science Journal*, 11(4), 2014.
- [3] Amirhossein S Aghaei, Mohammad Shahin Mahanta, and Konstantinos N Plataniotis. Separable common spatio-spectral patterns for motor imagery bci systems. *IEEE Transactions on Biomedical Engineering*, 63(1):15–29, 2016.
- [4] Mohammad H Alomari, Ayman Abubaker, Aiman Turani, Ali M Baniyounes, and Adnan Manasreh. Eeg mouse: A machine learning-based brain computer interface. *International Journal of Advanced Research in Computer Science and Applications*, 5(4):1–6, 2014.
- [5] Mohammad H Alomari, Emad A Awada, Aya Samaha, and Khaled Alkamha. Wavelet-based feature extraction for the analysis of eeg signals associated with imagined fists and feet movements. *Computer and Information Science*, 7(2):17, 2014.
- [6] Mohammad H Alomari, Ali M Baniyounes, and Emad A Awada. Eeg-based classification of imagined fists movements using machine learning and wavelet transform analysis. 2014.
- [7] Mohammad H Alomari, Aya Samaha, and Khaled AlKamha. Automated classification of l/r hand movement eeg signals using advanced feature extraction and machine learning. *International Journal of Advanced Computer Science and Applications*, 4(6), 2013.
- [8] Damian A Alvarez and Eduardo Giraldo. Ica aplicado a la extracción de características en imágenes. *Scientia Et Technica*, 3(40), 2009.
- [9] Hamzah S AlZu’bi, Nayel S Al-Zubi, and Waleed Al-Nuaimy. Toward inexpensive and practical brain computer interface. In *Developments in E-systems Engineering (DeSE), 2011*, pages 98–101. IEEE, 2011.
- [10] Zahra Amini, Vahid Abootalebi, and Mohammad T Sadeghi. Comparison of performance of different feature extraction methods in detection of p300. *Biocybernetics and Biomedical Engineering*, 33(1):3–20, 2013.

- [11] KL Anderson and M Ding. Attentional modulation of the somatosensory mu rhythm. *Neuroscience*, 180:165–180, 2011.
- [12] Adriano O Andrade, Slawomir Nasuto, Peter Kyberd, Catherine M Sweeney-Reed, and FR Van Kanijn. Emg signal filtering based on empirical mode decomposition. *Biomedical Signal Processing and Control*, 1(1):44–55, 2006.
- [13] ABM Aowlad, Md Wasiur Rahman, and Manjurul Ahsan Riheen. Left and right hand movements eeg signals classification using wavelet transform and probabilistic neural network. *International Journal of Electrical and Computer Engineering*, 5(1):92, 2015.
- [14] Fiorenzo Artoni, Danilo Menicucci, Arnaud Delorme, Scott Makeig, and Silvestro Micera. Relica: a method for estimating the reliability of independent components. *NeuroImage*, 103:391–400, 2014.
- [15] John Atkinson and Daniel Campos. Improving bci-based emotion recognition by combining eeg feature selection and kernel classifiers. *Expert Systems with Applications*, 47:35–41, 2016.
- [16] Hamza Baali, Aida Khorshidtalab, Mostefa Mesbah, and Momoh JE Salami. A transform-based feature extraction approach for motor imagery tasks classification. *IEEE journal of translational engineering in health and medicine*, 3:1–8, 2015.
- [17] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012.
- [18] Fanny Barlaam, Médéric Descoins, Olivier Bertrand, Thierry Hasbroucq, Franck Vidal, Christine Assaiante, and Christina Schmitz. Time–frequency and erp analyses of eeg to characterize anticipatory postural adjustments in a bimanual load-lifting task. *Frontiers in human neuroscience*, 5, 2011.
- [19] Syed Khairul Bashar and Mohammed Imamul Hassan Bhuiyan. Classification of motor imagery movements using multivariate empirical mode decomposition and short time fourier transform based hybrid method. *Engineering Science and Technology, an International Journal*, 19(3):1457–1464, 2016.
- [20] Ali Bashashati, Mehrdad Fatourehchi, Rabab K Ward, and Gary E Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R32, 2007.
- [21] Luzheng Bi, Xin-An Fan, and Yili Liu. Eeg-based brain-controlled mobile robots: a survey. *IEEE transactions on human-machine systems*, 43(2):161–176, 2013.
- [22] Dorothy Vera Margaret Bishop and Mervyn James Hardiman. Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophysiology*, 47(4):697–705, 2010.
- [23] Benjamin Blankertz, K-R Muller, Dean J Krusienski, Gerwin Schalk, Jonathan R Wolpaw, Alois Schlogl, Gert Pfurtscheller, Jd R Millan, M Schroder, and Niels Birbaumer. The bci competition iii: Validating alternative approaches to actual bci problems. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2):153–159, 2006.

- [24] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and K-R Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56, 2008.
- [25] Vladimir Bostanov. Bci competition 2003-data sets ib and iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Transactions on Biomedical engineering*, 51(6):1057–1061, 2004.
- [26] Soni Brijesh, Mishra Deepak, Richaria Gauri, and Beg Samiulla. Brain-computer-interface: A survey on feature extraction techniques for brain signal. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(11):18905–18908, 2016.
- [27] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.
- [28] John T Cacioppo, Louis G Tassinary, and Gary Berntson. *Handbook of psychophysiology*. Cambridge University Press, 2007.
- [29] Jose-Manuel Cano-Izquierdo, Julio Ibarrola, and Miguel Almonacid. Improving motor imagery classification with a new bci design using neuro-fuzzy s-dfasart. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(1):2–7, 2012.
- [30] Jose M Carmena, Mikhail A Lebedev, Roy E Crist, Joseph E O’Doherty, David M Santucci, Dragan F Dimitrov, Parag G Patil, Craig S Henriquez, and Miguel AL Nicolelis. Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS biol*, 1(2):e42, 2003.
- [31] P Carrión, J Ródenas, and J Rieta. Procesado de señales biomédicas. *Colección ciencia y técnica*, 53, 2007.
- [32] Sarah N Carvalho, Thiago BS Costa, Luisa FS Uribe, Diogo C Soriano, Glauco FG Yared, Luis C Coradine, and Romis Attux. Comparative analysis of strategies for feature extraction and classification in ssvep bcis. *Biomedical Signal Processing and Control*, 21:34–42, 2015.
- [33] Hubert Cecotti and Axel Graser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):433–445, 2011.
- [34] Yongwook Chae, Jaeseung Jeong, and Sungho Jo. Noninvasive brain-computer interface-based control of humanoid navigation. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 685–691. IEEE, 2011.
- [35] Min Hye Chang and Kwang Suk Park. Frequency recognition methods for dual-frequency ssvep based brain-computer interface. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 2220–2223. IEEE, 2013.
- [36] Damien Coyle, T Martin McGinnity, and Girijesh Prasad. Improving the separability of multiple eeg features for a bci by neural-time-series-prediction-preprocessing. *Biomedical Signal Processing and Control*, 5(3):196–204, 2010.

- [37] Damien Coyle, Thomas M McGinnity, and Girijesh Prasad. A multi-class brain-computer interface with sofnn-based prediction preprocessing. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 3696–3703. IEEE, 2008.
- [38] Simon RH Davies and Christopher J James. Using empirical mode decomposition with spatio-temporal dynamics to classify single-trial motor imagery in bci. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4631–4634. IEEE, 2014.
- [39] V Baby Deepa. Dagging meta classifiers with support vector machine. In *International Journal of Engineering Research and Technology*, volume 2. ESRSA Publications, 2013.
- [40] J del R Millan, Josep Mouriño, Marco Franzé, Febo Cincotti, Markus Varsta, Jukka Heikkonen, and Fabio Babiloni. A local neural classifier for the recognition of eeg patterns associated to mental tasks. *IEEE transactions on neural networks*, 13(3):678–686, 2002.
- [41] Arnaud Delorme, Terrence Sejnowski, and Scott Makeig. Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4):1443–1449, 2007.
- [42] Bruce H Dobkin. Brain–computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation. *The Journal of physiology*, 579(3):637–642, 2007.
- [43] Lijuan Duan, Menghu Bao, Jun Miao, Yanhui Xu, and Juncheng Chen. Classification based on multilayer extreme learning machine for motor imagery task from eeg signals. *Procedia Computer Science*, 88:176–184, 2016.
- [44] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.
- [45] Connie C Duncan, Robert J Barry, John F Connolly, Catherine Fischer, Patricia T Michie, Risto Näätänen, John Polich, Ivar Reinvang, and Cyma Van Petten. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, p300, and n400. *Clinical Neurophysiology*, 120(11):1883–1908, 2009.
- [46] Suman Dutta, Mandeep Singh, and Amod Kumar. Classification of non-motor cognitive task in eeg based brain-computer interface using phase space features in multivariate empirical mode decomposition domain. *Biomedical Signal Processing and Control*, 39:378–389, 2018.
- [47] Dario Farina, Omar Feix Do Nascimento, Marie-Françoise Lucas, and Christian Doncarli. Optimization of wavelets for classification of movement-related cortical potentials generated by variation of force-related parameters. *Journal of neuroscience methods*, 162(1):357–363, 2007.



- [48] M Fernandes, NS Dias, J Serrado Nunes, M El Tahchi, S Lanceros-Mendez, JH Correia, and PM Mendes. Wearable brain cap with contactless electroencephalogram measurement for brain-computer interface applications. In *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pages 387–390. IEEE, 2009.
- [49] Gerhard M Friehs, Vasilios A Zerris, Catherine L Ojakangas, Mathew R Fellows, and John P Donoghue. Brain-machine and brain-computer interfaces. *Stroke*, 35(11 suppl 1):2702–2705, 2004.
- [50] Suraj R Gaikwad and Shruti S Kshirsagar. A review: Analysis of eeg signal based brain-computer interface. *International Journal of Science and Research*, 3(6):2447–2451, 2014.
- [51] Pedro J García-Laencina, Germán Rodríguez-Bermudez, and Joaquín Roca-Dorda. Exploring dimensionality reduction of eeg features in motor imagery task classification. *Expert Systems with Applications*, 41(11):5285–5295, 2014.
- [52] Pramod Gaur, Ram Bilas Pachori, Hui Wang, and Girijesh Prasad. A multi-class eeg-based bci classification using multivariate empirical mode decomposition based filtering and riemannian geometry. *Expert Systems with Applications*, 2017.
- [53] Francesco Gianfelici and Dario Farina. An effective classification framework for brain-computer interfacing based on a combinatoric setting. *IEEE Transactions on Signal Processing*, 60(3):1446–1459, 2012.
- [54] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [55] Zhenghui Gu, Zhuliang Yu, Zhifang Shen, and Yuanqing Li. An online semi-supervised brain-computer interface. *IEEE Transactions on Biomedical Engineering*, 60(9):2614–2623, 2013.
- [56] Paul S Hammon and Virginia R de Sa. Preprocessing and meta-classification for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54(3):518–525, 2007.
- [57] Sepideh Hatamikia and Ali Motie Nasrabadi. Subject transfer bci based on composite local temporal correlation common spatial pattern. *Computers in biology and medicine*, 64:1–11, 2015.
- [58] Lianghua He, Die Hu, Meng Wan, Ying Wen, Karen M von Deneen, and MengChu Zhou. Common bayesian network for classification of eeg-based multiclass motor imagery bci. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(6):843–854, 2016.
- [59] Lianghua He, Bin Liu, Die Hu, Ying Wen, Meng Wan, and Jun Long. Motor imagery eeg signals analysis based on bayesian network with gaussian distribution. *Neurocomputing*, 188:217–224, 2016.

- [60] Thilo Hinterberger, Andrea Kübler, Jochen Kaiser, Nicola Neumann, and Niels Birbaumer. A brain–computer interface (bci) for the locked-in: comparison of different eeg classifications for the thought translation device. *Clinical Neurophysiology*, 114(3):416–425, 2003.
- [61] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. An efficient p300-based brain-computer interface for disabled subjects. *Journal of Neuroscience methods*, 167(1):115–125, 2008.
- [62] Tong Hong, Xu Zhang, Hongjun Ma, Yan Chen, and Xiang Chen. Fatiguing effects on the multi-scale entropy of surface electromyography in children with cerebral palsy. *Entropy*, 18(5):177, 2016.
- [63] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society, 1998.
- [64] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000.
- [65] Mohd Zaizu Ilyas, Puteh Saad, and Muhammad Imran Ahmad. A survey of analysis and classification of eeg signals for brain-computer interfaces. In *Biomedical Engineering (ICoBE), 2015 2nd International Conference on*, pages 1–6. IEEE, 2015.
- [66] PN Jadhav, D Shanamugan, A Chourasia, AR Ghole, AA Acharyya, and G Naik. Automated detection and correction of eye blink and muscular artefacts in eeg signal for analysis of autism spectrum disorder. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 1881–1884. IEEE, 2014.
- [67] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo, 1995. Morgan Kaufmann.
- [68] M Kalaivani, V Kalaivani, and V Anusuya Devi. Analysis of eeg signal for the detection of brain abnormalities. at *International Journal of Computer Applications® year*, 2014.
- [69] J Kalcher and G Pfurtscheller. Discrimination between phase-locked and non-phase-locked event-related eeg activity. *Electroencephalography and clinical neurophysiology*, 94(5):381–384, 1995.
- [70] Steven M Kay and Stanley L Marple. Spectrum analysis a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, 1981.
- [71] Andreas Keil, Stefan Debener, Gabriele Gratton, Markus Junghöfer, Emily S Kappenman, Steven J Luck, Phan Luu, Gregory A Miller, and Cindy M Yee. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, 51(1):1–21, 2014.

- [72] Robert E Kelly, George S Alexopoulos, Zhishun Wang, Faith M Gunning, Christopher F Murphy, Sarah Shizuko Morimoto, Dora Kanellopoulos, Zhiru Jia, Kelvin O Lim, and Matthew J Hoptman. Visual inspection of independent components: defining a procedure for artifact removal from fmri data. *Journal of neuroscience methods*, 189(2):233–245, 2010.
- [73] P R Kennedy and R AE Bakay. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport*, 9(8):1707–1711, 1998.
- [74] Philip R Kennedy. The cone electrode: a long-term electrode that records from neurites grown onto its recording surface. *Journal of neuroscience methods*, 29(3):181–193, 1989.
- [75] Philip R Kennedy and Roy AE Bakay. Activity of single action potentials in monkey motor cortex during long-term task learning. *Brain research*, 760(1):251–254, 1997.
- [76] Philip R Kennedy, Roy AE Bakay, Melody M Moore, Kim Adams, and John Goldwaithe. Direct control of a computer from the human central nervous system. *IEEE Transactions on rehabilitation engineering*, 8(2):198–202, 2000.
- [77] Philip R Kennedy, Suzanne S Mirra, and Roy AE Bakay. The cone electrode: ultrastructural studies following long-term recording in rat and monkey cortex. *Neuroscience letters*, 142(1):89–94, 1992.
- [78] Jasmin Kevric and Abdulhamit Subasi. Comparison of signal decomposition methods in classification of eeg signals for motor-imagery bci system. *Biomedical Signal Processing and Control*, 31:398–406, 2017.
- [79] Wael H Khalifa, Mohamed I Roushdy, and Abdel-Badeeh M Salem. User identification system based on eeg signals. In *proceeding of: The Sixth International Conference on Intelligent Computing and Information Systems, Cairo, Egypt*, pages 262–267, 2013.
- [80] D Hari Krishna, IA Pasha, and T Satya Savithri. Classification of eeg motor imagery multi class signals based on cross correlation. *Procedia Computer Science*, 85:490–495, 2016.
- [81] M Rajya Lakshmi, Dr TV Prasad, and Dr V Chandra Prakash. Survey on eeg signal processing methods. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(1):84–91, 2014.
- [82] S. le Cessie and J.C. van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201, 1992.
- [83] Hyekyoung Lee, Jiho Yoo, and Seungjin Choi. Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters*, 17(1):4–7, 2010.
- [84] Steven Lemm, Christin Schafer, and Gabriel Curio. Bci competition 2003-data set iii: probabilistic modeling of sensorimotor  $\mu$  rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering*, 51(6):1077–1080, 2004.
- [85] Junhua Li, Yijun Wang, Liqing Zhang, Andrzej Cichocki, and Tzyy-Ping Jung. Decoding eeg in cognitive tasks with time-frequency and connectivity masks. *IEEE Transactions on Cognitive and Developmental Systems*, 8(4):298–308, 2016.

- [86] Yan Li, Hiroyuki Kambara, Yasuharu Koike, and Masashi Sugiyama. Application of covariate shift adaptation techniques in brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 57(6):1318–1324, 2010.
- [87] Yan Li, Peng Paul Wen, et al. Modified cc-lr algorithm with three diverse feature sets for motor imagery tasks classification in eeg based brain–computer interface. *Computer Methods and programs in Biomedicine*, 113(3):767–780, 2014.
- [88] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.
- [89] Ye Liu, Hao Zhang, Min Chen, and Liqing Zhang. A boosting-based spatial-spectral model for stroke patients'eeg analysis in rehabilitation training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(1):169–179, 2016.
- [90] A Llera, Vicenç Gómez, and Hilbert J Kappen. Adaptive multiclass classification for brain computer interfaces. *Neural computation*, 26(6):1108–1127, 2014.
- [91] Ana Loboda, Alexandra Margineanu, Gabriela Rotariu, and Anca Mihaela Lazar. Discrimination of eeg-based motor imagery tasks by means of a simple phase information method. *International Journal of Advanced Research in Artificial Intelligence*, 3(10).
- [92] Fabien Lotte and Cuntai Guan. Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *IEEE Transactions on biomedical Engineering*, 58(2):355–362, 2011.
- [93] Martin Lotze and Ulrike Halsband. Motor imagery. *Journal of Physiology-Paris*, 99(4):386–395, 2006.
- [94] Haiping Lu, How-Lung Eng, Cuntai Guan, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. Regularized common spatial pattern with aggregation for eeg classification in small-sample setting. *IEEE Transactions on Biomedical Engineering*, 57(12):2936–2946, 2010.
- [95] Jun Lu, Kan Xie, and Dennis J McFarland. Adaptive spatio-temporal filtering for movement related potentials in eeg-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4):847–857, 2014.
- [96] Tyler C Major and James M Conrad. A survey of brain computer interfaces and their applications. In *SOUTHEASTCON 2014, IEEE*, pages 1–8. IEEE, 2014.
- [97] Juan-Antonio Martinez-Leon, Jose-Manuel Cano-Izquierdo, and Julio Ibarrola. Are low cost brain computer interface headsets ready for motor imagery applications? *Expert Systems with Applications*, 49:136–144, 2016.
- [98] Steven G Mason and Gary E Birch. A brain-controlled switch for asynchronous control applications. *IEEE Transactions on Biomedical Engineering*, 47(10):1297–1307, 2000.
- [99] Dennis J McFarland, Laurie A Miner, Theresa M Vaughan, and Jonathan R Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain topography*, 12(3):177–186, 2000.

- [100] Jianjun Meng, Xinjun Sheng, Dingguo Zhang, and Xiangyang Zhu. Improved semisupervised adaptation for a small training dataset in the brain-computer interface. *IEEE journal of biomedical and health informatics*, 18(4):1461–1472, 2014.
- [101] Jianjun Meng, Lin Yao, Xinjun Sheng, Dingguo Zhang, and Xiangyang Zhu. Simultaneously optimizing spatial spectral features based on mutual information for eeg classification. *IEEE Transactions on Biomedical Engineering*, 62(1):227–240, 2015.
- [102] Minmin Miao, Hong Zeng, Aimin Wang, Changsen Zhao, and Feixiang Liu. Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery eeg: An sparse regression and weighted naïve bayesian classifier-based approach. *Journal of neuroscience methods*, 278:13–24, 2017.
- [103] Jesus Minguillon, M Angel Lopez-Gordo, and Francisco Pelayo. Trends in eeg-bci for daily-life: Requirements for artifact removal. *Biomedical Signal Processing and Control*, 31:407–418, 2017.
- [104] Ganesh R Naik, Dinesh Kant Kumar, Vijay Pal Singh, and Marimuthu Palaniswami. Hand gestures for hci using ica of emg. In *Proceedings of the HCSNet workshop on Use of vision in human-computer interaction-Volume 56*, pages 67–72. Australian Computer Society, Inc., 2006.
- [105] Ganesh R Naik, Suvishamuthu Easter Selvan, Massimiliano Gobbo, Amit Acharyya, and Hung T Nguyen. Principal component analysis applied to surface electromyography: a comprehensive review. *IEEE Access*, 4:4025–4037, 2016.
- [106] Luis F Nicolas-Alonso, Rebeca Corralejo, Javier Gomez-Pilar, Daniel Álvarez, and Roberto Hornero. Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(4):702–712, 2015.
- [107] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279, 2012.
- [108] Javier Nieves, Igor Santos, and Pablo G Bringas. Enhancing the prediction stage of a model predictive control systems through meta-classifiers. *Dyna*, 88(3):290–298, 2013.
- [109] Farzan Majeed Noori, Noman Naseer, Nauman Khalid Qureshi, Hammad Nazeer, and Rayyan Azam Khan. Optimal feature selection from fnirs signals using genetic algorithms for bci. *Neuroscience Letters*, 647:61–66, 2017.
- [110] Jorge Baztarrica Ochoa. Eeg signal classification for brain computer interface applications. *Ecole Polytechnique Federale de Lausanne*, 2002.
- [111] Cheolsoo Park, David Looney, Naveed ur Rehman, Alireza Ahrabian, and Danilo P Mandic. Classification of motor imagery bci using multivariate empirical mode decomposition. *IEEE Transactions on neural systems and rehabilitation engineering*, 21(1):10–22, 2013.

- [112] Cheolsoo Park, Clive Cheong Took, and Danilo P Mandic. Augmented complex common spatial patterns for classification of noncircular eeg from motor imagery tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(1):1–10, 2014.
- [113] Sang-Hoon Park and Sang-Goog Lee. Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern. *IEEE Sensors Journal*, 17(10):2977–2983, 2017.
- [114] Vincent Perlbarg, Pierre Bellec, Jean-Luc Anton, Mélanie Péligrini-Issac, Julien Doyon, and Habib Benali. Corsica: correction of structured noise in fmri by automatic identification of ica components. *Magnetic resonance imaging*, 25(1):35–46, 2007.
- [115] Gert Pfurtscheller, Christa Neuper, Clemens Brunner, and F Lopes da Silva. Beta rebound after different types of motor imagery in man. *Neuroscience letters*, 378(3):156–159, 2005.
- [116] Gert Pfurtscheller, Christa Neuper, Alois Schlogl, and Klaus Lugger. Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE transactions on Rehabilitation Engineering*, 6(3):316–325, 1998.
- [117] Robert T Pivik, Roger J Broughton, Richard Coppola, Richard J Davidson, Nathan Fox, and Margreth R Nuwer. Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, 30(6):547–558, 1993.
- [118] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.
- [119] Feifei Qi, Yuanqing Li, and Wei Wu. Rstfc: A novel algorithm for spatio-temporal filtering and classification of single-trial eeg. *IEEE transactions on neural networks and learning systems*, 26(12):3070–3082, 2015.
- [120] Alimed Celecia Ramos, René González Hernández, and Marley Vellasco. Feature selection methods applied to motor imagery task classification. In *Computational Intelligence (LA-CCI), 2016 IEEE Latin American Conference on*, pages 1–6. IEEE, 2016.
- [121] Xiaomei Ren, Zhiguo Yan, Zhizhong Wang, and Xiao Hu. Noise reduction based on ica decomposition and wavelet transform for the extraction of motor unit action potentials. *Journal of neuroscience methods*, 158(2):313–322, 2006.
- [122] Jaime F Delgado Saa and Mujdat Cetin. Discriminative methods for classification of asynchronous imaginary motor tasks from eeg data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(5):716–724, 2013.
- [123] Simanto Saha, Khawza I Ahmed, Raqibul Mostafa, Ahsan H Khandoker, and Leontios Hadjileontiadis. Enhanced inter-subject brain computer interface with associative sensorimotor oscillations. *Healthcare Technology Letters*, 4(1):39–43, 2017.

- [124] Thomas Schack, Kai Essig, Cornelia Frank, and Dirk Koester. Mental representation and motor imagery training. *Frontiers in human neuroscience*, 8, 2014.
- [125] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.
- [126] Daniel Senkowski and Christoph S Herrmann. Effects of task difficulty on evoked gamma activity and erps in a visual discrimination task. *Clinical Neurophysiology*, 113(11):1742–1753, 2002.
- [127] Rinkal G Shah. A review on body movement classification using motor imagery eeg. *spectrum*, 5(2), 2016.
- [128] Rinkal G Shah and Rutu Nayak. Hand movement classification using motor imagery eeg. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(6), 2016.
- [129] Hesham Sheikh, Dennis J McFarland, William A Sarnacki, and Jonathan R Wolpaw. Electroencephalographic (eeg)-based communication: Eeg control versus system performance in humans. *Neuroscience letters*, 345(2):89–92, 2003.
- [130] Siuly Siuly and Yan Li. Improving the separability of motor imagery eeg signals using a cross correlation-based least square support vector machine for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(4):526–538, 2012.
- [131] Jason Sleight, Preeti Pillai, and Shiwali Mohan. Classification of executed and imagined motor movement eeg signals. *Ann Arbor: University of Michigan*, pages 1–10, 2009.
- [132] Xiaomu Song and Suk-Chung Yoon. Improving brain-computer interface classification using adaptive common spatial patterns. *Computers in biology and medicine*, 61:150–160, 2015.
- [133] Andrej Stancák Jr and Gert Pfurtscheller. Effects of handedness on movement-related changes of central beta rhythms. *Journal of clinical Neurophysiology*, 14(5):419–428, 1997.
- [134] Sachin Taran, Varun Bajaj, Dheeraj Sharma, Siuly Siuly, and A Sengur. Features based on analytic imf for classifying motor imagery eeg signals in bci applications. *Measurement*, 2017.
- [135] Dawn M Taylor, Stephen I Helms Tillery, and Andrew B Schwartz. Direct cortical control of 3d neuroprosthetic devices. *Science*, 296(5574):1829–1832, 2002.
- [136] Paul Thagard. *Mind: Introduction to cognitive science*, volume 4. MIT press Cambridge, MA, 1996.
- [137] K. M. Ting and I. H. Witten. Stacking bagged and dagged models. In D. H. Fisher, editor, *Fourteenth international Conference on Machine Learning*, pages 367–375, San Francisco, CA, 1997. Morgan Kaufmann Publishers.

- [138] Dalila Trad, Tarik Al-Ani, and Mohamed Jemni. A feature extraction technique of eeg based on emd-bp for motor imagery classification in bci. In *Information & Communication Technology and Accessibility (ICTA), 2015 5th International Conference on*, pages 1–6. IEEE, 2015.
- [139] Jan BF Van Erp, Fabien Lotte, and Michael Tangermann. Brain-computer interfaces: beyond medical applications. *Computer-IEEE Computer Society-*, 45(4):26–34, 2012.
- [140] Garrick L Wallstrom, Robert E Kass, Anita Miller, Jeffrey F Cohn, and Nathan A Fox. Automatic correction of ocular artifacts in the eeg: a comparison of regression-based and component-based methods. *International journal of psychophysiology*, 53(2):105–119, 2004.
- [141] Haixian Wang. Multiclass filters by a weighted pairwise criterion for eeg single-trial classification. *IEEE Transactions on Biomedical Engineering*, 58(5):1412–1420, 2011.
- [142] Haixian Wang, Qin Tang, and Wenming Zheng. L1-norm-based common spatial patterns. *IEEE Transactions on Biomedical Engineering*, 59(3):653–662, 2012.
- [143] Hua Wang, Yanchun Zhang, et al. Detection of motor imagery eeg signals employing naïve bayes based learning process. *Measurement*, 86:148–158, 2016.
- [144] Yijun Wang, Shangkai Gao, and Xiaornog Gao. Common spatial pattern method for channel selection in motor imagery based brain-computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5392–5395. IEEE, 2006.
- [145] Christian Ward and Iyad Obeid. Feasibility of identity vectors for use as subject verification and cohort retrieval of electroencephalograms. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2016 IEEE*, pages 1–5. IEEE, 2016.
- [146] Nienke W Willigenburg, Andreas Daffertshofer, Idsart Kingma, and Jaap H van Dieën. Removing ecg contamination from emg recordings: A comparison of ica-based and other filtering procedures. *Journal of electromyography and kinesiology*, 22(3):485–493, 2012.
- [147] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain-computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.
- [148] Jonathan R Wolpaw and Dennis J McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–17854, 2004.
- [149] WANG Xing-Yu, Jin Jing, Yu ZHANG, and WANG Bei. Brain control: Human-computer integration control based on brain-computer interface approach. *Acta Automatica Sinica*, 39(3):208–221, 2013.
- [150] Xuxian Yin, Baolei Xu, Changhao Jiang, Yunfa Fu, Zhidong Wang, Hongyi Li, and Gang Shi. Nirs-based classification of clench force and speed motor imagery with the use of empirical mode decomposition for bci. *Medical engineering & physics*, 37(3):280–286, 2015.



- [151] Xinyang Yu, Pharino Chum, and Kwee-Bo 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*, 125(3):1498–1502, 2014.
- [152] Jingwei Yue, Jun Jiang, Zongtan Zhou, and Dewen Hu. Smr-speller: A novel brain-computer interface spell paradigm. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on*, volume 3, pages 187–190. IEEE, 2011.
- [153] Asmat Zahra, Nadia Kanwal, Shoaib Ehsan, Klaus D McDonald-Maier, et al. Seizure detection from eeg signals using multivariate empirical mode decomposition. *Computers in Biology and Medicine*, 88(C):132–141, 2017.
- [154] Hong Zeng and Aiguo Song. Optimizing single-trial eeg classification by stationary matrix logistic regression in brain-computer interface. *IEEE transactions on neural networks and learning systems*, 27(11):2301–2313, 2016.
- [155] Biao Zhang, Jianjun Wang, and Thomas Fuhlbrigg. A review of the commercial brain-computer interface technology from perspective of industrial robotics. In *2010 IEEE International Conference on Automation and Logistics*, pages 379–384. IEEE, 2010.
- [156] Haihong Zhang, Zheng Yang Chin, Kai Keng Ang, Cuntai Guan, and Chuanchu Wang. Optimum spatio-spectral filtering network for brain-computer interface. *IEEE Transactions on Neural Networks*, 22(1):52–63, 2011.
- [157] Xiang Zhang, Lina Yao, Chaoran Huang, Quan Z Sheng, and Xianzhi Wang. Enhancing mind controlled smart living through recurrent neural networks. *arXiv preprint arXiv:1702.06830*, 2017.
- [158] Yu Zhang, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. *Journal of neuroscience methods*, 255:85–91, 2015.
- [159] Yu Zhang, Guoxu Zhou, Jing Jin, Qibin Zhao, Xingyu Wang, and Andrzej Cichocki. Sparse bayesian classification of eeg for brain-computer interface. *IEEE transactions on neural networks and learning systems*, 27(11):2256–2267, 2016.
- [160] Yu Zhang, Guoxu Zhou, Qibin Zhao, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Spatial-temporal discriminant analysis for erp-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2):233–243, 2013.