



UNIVERSIDAD NACIONAL DE COLOMBIA

Extracción y Selección de Características Discriminantes para la Detección de TDAH en Registros de Potenciales Evocados Cognitivos

Paola Alexandra Castro Cabrera

Universidad Nacional de Colombia - Sede Manizales
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales, Colombia
2011

Extracción y Selección de Características Discriminantes para la Detección de TDAH en Registros de Potenciales Evocados Cognitivos

Paola Alexandra Castro Cabrera

Tesis presentada como requisito parcial para optar al título de:
Magister en Ingeniería - Automatización Industrial

Director:
Ph.D. César Germán Castellanos Domínguez

Línea de Investigación:
Procesamiento Digital de Señales
Grupo de Investigación:
Grupo de Procesamiento y Reconocimiento de Señales

Universidad Nacional de Colombia - Sede Manizales
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales, Colombia
2011

Extraction and Selection of Discriminative Features for ADHD Detection on Cognitive Evoked Potentials Recordings

Paola Alexandra Castro Cabrera

Thesis submitted as partial requirement to apply for degree of:
Master of Engineering - Industrial Automation

Supervisor:
Ph.D. César Germán Castellanos Domínguez

Research Area:
Digital Signal Processing
Research Group
Processing and Recognition Signal Group

Universidad Nacional de Colombia - Manizales Branch
Faculty of Engineering and Architecture
Department of Electrical, Electronic and Computer Engineering
Manizales, Colombia
2011

Este trabajo se realiza en el marco de los proyecto *“Identificación Automática del Trastorno por Déficit de Atención y/o Hiperactividad sobre Registros de Potenciales Evocados Cognitivos”*, financiado por COLCIENCIAS - convocatoria 452 de 2008 - y *“Sistema de Diagnóstico Asistido para la Identificación de TDAH sobre Registros de Potenciales Evocados Cognitivos”*, avalado por la Dirección de Investigaciones DIMA de la Universidad Nacional de Colombia sede Manizales - convocatoria nacional de investigación 2008.

Quisiera dedicar este trabajo a:

Al Señor, por iluminar mi camino hacia el entendimiento y fortalecer mi voluntad.

A mis padres, por ser la voz de aliento que siempre me reconforta.

A mis hermanas, por estimularme diariamente a ser mejor persona.

A mi novio, por convertir mi vida en una hermosa realidad.

...Gracias a ustedes por ser incondicionales.

Agradecimientos

En este arduo trayecto de la academia que a la sombra del Señor he recorrido, han intervenido un sinnúmero de personas que de una u otra manera aportaron valiosas enseñanzas para mi formación personal y profesional. Ahora que estoy próxima a ver cumplido un nuevo objetivo académico, quisiera aprovechar la oportunidad para ofrecer un sencillo reconocimiento y expresar mi sincero agradecimiento a ellas, quienes con palabras de aliento, explicaciones y aportes en la materia, contribuyeron en la edificación de mi ser en el conocimiento, la integridad y la fe.

En primera instancia, debo agradecer al principal gestor de este logro, al dueño y dador de la vida y todo existente. Gracias a ti Padre Santo porque sin tu presencia en las sendas de mi vida y tu generoso amor, no hubiese sido posible dar por concluida esta laboriosa tarea. Gracias Señor por mantener tu mano extendida para levantarme una y otra vez e impulsarme a no abandonar mis sueños. Y gracias por rodearme de personas extraordinarias que están siempre prestas a colaborar y compartir sus conocimientos y demás virtudes.

En el ámbito académico, quisiera dar las gracias a mi director de tesis, el profesor Germán Castellanos Domínguez, por sus importantes aportes y continuo acompañamiento en el desarrollo de este trabajo investigativo. Los debates generados en torno a los inconvenientes encontrados a lo largo de las diferentes etapas de esta tesis, fueron muy pertinentes y beneficiosos; además, fomentaron en mí un carácter más inquieto que me estimularon a seguir indagando y aprendiendo más.

Quisiera también agradecer a cada uno de los miembros del Grupo de Control y Procesamiento Digital de Señales, que con sus conocimientos y orientación hicieron una gran contribución a la realización de esta tesis. De manera especial, quiero agradecer a algunos compañeros que hicieron de mi estancia en el grupo una experiencia enriquecedora y provechosa, y contribuyeron desinteresadamente en el desarrollo de algoritmos y la comprensión de temas que no eran de mi dominio: gracias a Diego Peluffo, Santiago Murillo, a Leonardo Duque, a Lina Sepúlveda, a Juan David Martínez, a Jorge Iván Padilla, a Luis David Avendaño y a Jorge Andrés Gómez. También quisiera agradecer a otros compañeros que, aunque ya no están en el grupo, fueron en su momento un gran apoyo académico y emocional para mí: gracias a Johanna Paola Carvajal, a Carolina Ospina, a Genaro Daza, a Andrés Felipe Quiceno y a Milton Orlando Sarria. A ellos muchas gracias por el tiempo dedicado a resolver mis dudas y por la continua motivación a no desistir de la búsqueda mis objetivos. Además, quisiera dar las gracias a otros profesores miembros del grupo, cuyas orientaciones en determinadas áreas fueron indispensables para comprender y dar solución a problemas planteados en este

trabajo; en particular, gracias a los profesores Julio Fernando Suárez y Mauricio Orozco.

Por otra parte, quisiera extender mis agradecimientos a las instituciones y grupos de investigación que intervinieron en el desarrollo de este trabajo. Gracias al grupo de Neuroaprendizaje de la Universidad Autónoma de Manizales, que en cabeza de los doctores Francia Restrepo y Oscar Moscoso contribuyó a la construcción de la base datos utilizada en esta tesis. De igual manera, agradezco al equipo de trabajo de la Universidad de Manizales, que realizó una dispendiosa labor resumida en el diseño y aplicación de un riguroso protocolo clínico aplicado a los niños que participaron en este estudio.

De igual forma, debo dar las gracias a las instituciones que con su apoyo financiero facilitaron la realización de este trabajo. Gracias al Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología “Francisco José de Caldas” (COLCIENCIAS) y a la Dirección de Investigaciones de Manizales (DIMA), por el aval ofrecido mediante el programa Jóvenes Investigadores e Innovadores 2008 y la financiación de proyectos relacionados a esta tesis.

Por último, quisiera agradecer a los verdaderos artífices de este trabajo: mi familia, que son la verdadera razón de los esfuerzos y sacrificios realizados por alcanzar este meta. A mis padres, Álvaro Iván y Flor Aydee, muchas gracias por el inmesurable amor que me brindan; sus palabras motivates, consejos y enseñanzas son el norte que orientan cada paso que doy. Ruego al Señor me dé larga vida para poder retribuirles todo lo bueno que de su parte he recibido. A mi hermana mayor, Carolina, gracias por ser mi amiga fiel y confidente, por tener siempre palabras positivas para alentarme a continuar mis proyectos, y por ser un ejemplo de perseverancia, empeño y tesón. A Sandrita, mi hermanita menor, gracias por ser la motivación de los actos que me mueven a buscar la excelencia, a superar mis miedos y dificultades para llegar a ser un modelo a seguir. A mi novio Mauricio, gracias por ser el pilar que soporta mi lucha diaria, la inspiración de mis mejores pensamientos y acciones, y sobretodo, muchas gracias por nunca perder la confianza en mí, y siempre recordarme que podría lograrlo. A todos ellos, mil gracias por acompañarme incondicionalmente en este andar. Y ojalá el Señor me permita compartir muchos nuevos triunfos con las personas que más amo.

Paola Alexandra Castro C.

Resumen

La adecuada representación de eventos fisiológicos subyacentes, es una condición fundamental en la detección automática de patologías que ha motivado la búsqueda de una representación efectiva de datos. Asimismo, hallar patrones descriptivos convenientes de señales como los potenciales relacionados a eventos (ERPs) puede marcar la diferencia entre la correcta o errada discriminación de sujetos sanos y pacientes que presentan trastorno por déficit de atención y/o hiperactividad (TDAH). En este trabajo se presenta un estudio comparativo de diferentes técnicas de extracción y selección de características que permite evaluar la calidad de los atributos representativos de las señales ERP, con el fin de definir un conjunto adecuado de patrones que conlleve a la identificación de sujetos sanos y enfermos. El aporte principal de este estudio es la identificación de las condiciones bajo las cuales un conjunto conformado por características morfológicas y espectrales, puede alcanzar una alta capacidad discriminante, que conduce a una mayor separabilidad de las dos clases evaluadas.

Las pruebas efectuadas en este estudio se llevaron a cabo sobre una única base de datos conformada por niños entre 4 y 15 años pertenecientes a instituciones educativas del municipio de Manizales, Colombia, evaluados en las modalidades auditiva y visual, bajo el paradigma *oddball*. Estas pruebas comprenden la aplicación de métodos supervisados y no supervisados que incluyen técnicas como *clustering*, que permite evaluar la calidad de agrupamientos formados sobre el espacio de características previamente definido; además, se emplea el error de clasificación para cuantificar la capacidad discriminante del conjunto inicial de características. Como aporte adicional significativo de esta investigación se incluye, en primera instancia, el desarrollo de una metodología de reducción de dimensión, que se basa en una búsqueda heurística sobre una nueva representación de los datos originales para determinar las variables relevantes. Asimismo, en esta tesis se propone la implementación de un algoritmo para la localización temporal del componente P300 basado en criterios médicos reales, y la definición de una medida de desempeño no supervisada que estima la distancia relativa que separa los agrupamientos generados por un algoritmo de *clustering*.

Palabras claves: TDAH, señales ERPs, características wavelet - morfológicas - espectrales, agrupamiento, reducción de dimensión, selección/extracción de características, función objetivo.

Abstract

Adequate representation of underlying physiological events is a fundamental condition for the automatic detection of pathologies, which have motivated the search for effective representation of data. Likewise, find convenient descriptive patterns of signals as event related potentials (ERPs) can make the difference between right or wrong discrimination of healthy subjects and patients with attention deficit hyperactivity disorder (ADHD). This work presents a comparative study of different extraction and selection techniques of features to evaluate the quality of representative attributes of the ERP signals, in order to define an appropriate set of patterns that lead to the identification of healthy and ill subjects. The main contribution of this study is the identification of conditions under which a set consisting of morphological and spectral characteristics, can achieve high discriminatory power, leading to greater separability of two evaluated classes.

Tests done in this work were carried out on an unique database composed of children between 4 and 15 years belonging to educational institutions in the city of Manizales, Colombia. These tests consider the implementation of supervised and unsupervised methods that include techniques such as clustering, which evaluates the quality of clusters formed on the previously defined feature space. Also, classification error was used to quantify the ability to discriminate the initial set of features. As a significant additional contribution of this research is included the development of a methodology that performs a heuristic search in a new representation of the original data to determine the relevant variables. Additionally, this master's thesis presents the implementation of an algorithm for on time location of P300 component based on real medical criteria, and the definition of an unsupervised performance measure that estimates the relative distance that separates the clusters generated by an clustering algorithm.

Keywords: ADHD, ERP signals, wavelet - morphological - spectral features, clustering, dimension reduction, feature selection/extraction, objective function)

Índice General

Agradecimientos	XI
Resumen	XIII
Abstract	XIV
Símbolos y Notaciones	XXII
Abreviaturas	XXIII
I. Preliminares	1
1. Introducción	2
Planteamiento del problema	3
Justificación	4
Pertinencia de la tesis	4
Antecedentes	5
Organización de documento	6
Objetivos	8
II. Contextualización	9
2. Preliminares Fisiológicos y Clínicos	10
2.1. Clínica de los potenciales evocados	10
2.1.1. Definición	10
2.1.2. Nomenclatura	11
2.1.3. Clasificación de los potenciales evocados	11
2.1.4. Generación de los EPs	13
2.1.5. Características generales de los potenciales evocados	13
2.1.6. Componentes de los potenciales evocados cognitivos	15
2.1.7. Uso clínico de los potenciales evocados	19

2.2.	Trastorno por déficit de atención y/o hiperactividad	23
2.2.1.	Definición	23
2.2.2.	Estadísticas	24
2.2.3.	Diagnóstico del TDAH	25
2.2.4.	Prevalencia	26
2.3.	Potenciales evocados y cognición	27
2.3.1.	Alteraciones de los ERPs a causa del TDAH	27
3.	Estado del Arte en la Detección Automática del TDAH	29
3.1.	Diagnóstico asistido de patologías	29
3.1.1.	Estado del arte	29
3.1.2.	Requerimientos del diagnóstico asistido	31
3.1.3.	Diseño de un sistema automático de diagnóstico	32
3.1.4.	Detección automática de TDAH	33
3.2.	Técnicas aplicadas al análisis de los potenciales evocados cognitivos	35
3.2.1.	Análisis mediante distribuciones tiempo - frecuencia	35
III.	Marco Teórico	38
4.	Caracterización y Preprocesamiento de Datos	39
4.1.	Caracterización de la señal ERP	39
4.1.1.	Características morfológicas	40
4.1.2.	Características frecuenciales de la señal	44
4.1.3.	Características basadas en transformada wavelet	45
4.2.	Preprocesamiento de datos	49
4.2.1.	Identificación de datos faltantes	50
4.2.2.	Detección de datos atípicos	50
4.2.3.	Verificación de gaussividad	51
4.2.4.	Depuración de base de datos	51
5.	Reducción de Dimensionalidad y Selección de Características	53
5.1.	Reducción de dimensión	53
5.1.1.	Conceptos generales	54
5.1.2.	Relevancia	54
5.2.	Selección de características	55
5.2.1.	Estrategias de generación	56
5.2.2.	Funciones de evaluación	57

6. Transformación de Variables y Clasificación	61
6.1. Proyección lineal	61
6.1.1. Análisis de componentes principales	62
6.2. Clasificación no supervisada	65
6.2.1. Agrupamiento basado en centroides	65
6.2.2. Medidas de desempeño del agrupamiento	67
 IV. Marco Experimental	 71
7. Esquema de Trabajo	72
7.1. Descripción de la base de datos	72
7.1.1. Sujetos	72
7.1.2. Pruebas diagnósticas neuropsicopedagógicas	73
7.1.3. Adquisición de los registros	74
7.2. Conjunto de características	75
7.3. Sintonización de parámetros	76
7.3.1. Caracterización	76
7.4. Diseño de experimentos	77
7.4.1. Experimento 1	77
7.4.2. Experimento 2	79
7.4.3. Experimento 3	82
7.4.4. Experimento 4	84
7.4.5. Experimento 5	87
 8. Resultados y Discusión	 89
8.1. Experimento 1	
Resultados y discusión	89
8.2. Experimento 2	
Resultados y discusión	93
8.3. Experimento 3	
Resultados y discusión	96
8.4. Experimento 4	
Resultados y discusión	99
8.4.1. Sintonización de la función wavelet	100
8.4.2. Optimización del espacio tiempo-frecuencia	101
8.4.3. Clasificación	102
8.5. Experimento 5	
Resultados y discusión	102

V. Comentarios finales	105
9. Conclusiones y Trabajo Futuro	106
9.1. Conclusiones	106
9.2. Trabajo futuro	107
10. Discusión académica	109
A. Anexo: Criterios Diagnósticos del TDAH, según DSM-IV	110
B. Anexo: Algoritmo K-medias	112
C. Anexo: Métodos de Selección de Características	114
C.1. Algoritmos de búsqueda para la generación de subconjuntos	114
C.1.1. Búsqueda exhaustiva	114
C.1.2. Búsqueda heurística	115
C.2. Funciones de evaluación	116
C.2.1. Análisis multivariado de varianza	116
Bibliografía	118

Índice de Figuras

2-1. Forma de onda del potencial evocado cognitivo.	16
2-2. Ubicación de los electrodos según sistema internacional 10-20	21
2-3. Extracción de EPs por promediación de señales EEG.	23
2-4. Presentación del componente P300 dado un estímulo infrecuente	24
4-1. Descripción gráfica del algoritmo de localización del P300.	41
4-2. Resolución de la transformada wavelet.	47
4-3. Análisis multiresolución de una señal	49
4-4. Descomposición wavelet generalizada	49
7-1. Wavelet madre spline biortogonal, empleada en el análisis multi-resolución .	77
7-2. Descomposición wavelet de una señal ERP	78
7-3. Señales patrón de las clases control y casos	83
7-4. Alineamiento de la señal ERP de acuerdo a la señal patrón	84
7-5. Esquema de ajuste para construcción de la WT	85
7-6. Reconstrucción por niveles del análisis multiresolución	88
8-1. Comparaciones de matrices de covarianza por clases	91
8-2. Sintonización de clasificador k -nn usado sobre base de datos depurada, después de aplicar SFFS con clasificador bayesiano como función de evaluación y sin PCA	93
8-3. Esquema de la metodología aplicada en experimento 2	94
8-4. Estabilidad de la metodología 2	94
8-5. Rendimiento de clasificación por subconjuntos de características	95
8-6. Agrupamientos formados en el espacio de características original: latencia vs otras características morfológicas	98
8-7. Agrupamientos formados en el espacio de características que excluye la latencia	99
8-8. Agrupamientos formados en espacio de características generado por las señales alineadas	100
8-9. Rendimiento de clasificación por subconjuntos de características	103

Índice de Tablas

5-1. Comparación de diferentes conceptos de relevancia	60
6-1. Medidas de desempeño empleadas	70
7-1. Conjunto de características empleadas en el análisis de señales ERPs	75
8-1. Remoción de variables en el preproceso	90
8-2. Comparación del preprocesamiento y depuración de la base de datos (BD) en los resultados de clasificación	91
8-3. Comparación de la extracción de características de la base de datos depurada, seguida de SFFS.	92
8-4. Comparación del preprocesamiento y depuración, seguida de la extracción de características de la base de datos en los resultados de clasificación	92
8-5. Rendimiento de clasificación por grupos de características	95
8-6. Medidas de rendimiento de clasificación para base de datos sin depurar	96
8-7. Medidas de rendimiento de clasificación para base de datos depurada	96
8-8. Vectores de valores relativos.	97
8-9. Coeficientes de los operadores predictivos	100
8-10. Medidas de rendimiento de clasificación usando <i>customized wavelet transform</i>	102
8-11. Medidas de rendimiento de clasificación usando características informativas	103

Índice de Algoritmos

1.	Algoritmo de localización del componente P300	42
2.	Algoritmo de detección multivariada de datos atípicos	51
3.	Algoritmo de MANOVA progresivo	79
4.	Algoritmo de re-caracterización de señales ERPs mediante medidas de disimilitud	81
5.	Algoritmo de K-medias	113

Símbolos y Notaciones

Variables

Notación	Significado
t, f	Tiempo, frecuencia
$x(t), s(t)$	Señal en tiempo continuo
\mathbf{X}	Matriz de datos
\mathbf{x}_i	i -ésima fila de la matriz \mathbf{X} (observación)
$\boldsymbol{\xi}_j$	j -ésima columna de la matriz \mathbf{X} (variable o característica)
\mathbb{R}	Dominio de los reales
$S(f)$	Densidad espectral de potencia de $s(t)$
$\dot{s}(t)$	Derivada de la señal
$\boldsymbol{\Sigma}_{\mathbf{X}}$	Matriz de covarianza
\mathbf{Q}	Matriz de centroides
\mathbf{q}_j	j - ésimo centroide

Operadores matemáticos

Notación	Significado
$d(\cdot, \cdot)$	Operador de distancia
$ \cdot $	Valor absoluto
$\mu(\cdot)$	Operador de media
$\sigma(\cdot)$	Operador de desviación estándar

Abreviaturas

Notación	Significado
EP	Potencial evocado
ERP	Potencial relacionado a eventos
EEG	Señal electroencefalográfica
EMG	Electromiografía
PCA	Análisis de componentes principales
WT	Transformada wavelet
FT	Transformada de Fourier
STFT	Transformada de Fourier de tiempo corto
CWT	Transformada wavelet continua
DWT	Transformada wavelet discreta
TDAH	Trastorno por déficit de atención y/o hiperactividad
LDC	Clasificador discriminante lineal
SVM	Máquinas de soporte vectorial
N100, N200, P300	Componentes de los ERPs
SNR	Relación señal-ruido
DSM-IV	Manual diagnóstico y estadístico de los trastornos mentales
SVM	Máquinas de soporte vectorial
k -NN	k -vecinos más cercanos
LDC	Clasificador lineal discriminante

Parte I.

Preliminares

1. Introducción

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) se define como un desorden conductual que se caracteriza por un patrón persistente de comportamientos problemáticos que reflejan desatención y desinhibición conductual (impulsividad e hiperactividad) que no se explica por cualquier otro trastorno del desarrollo, del pensamiento o afectivo. La sintomatología interfiere con el funcionamiento del paciente en por lo menos dos ambientes distintos (casa, escuela, trabajo, etc.) [1]. De acuerdo a la American Academy of Pediatrics (AAP), el TDAH es el desorden neuro-comportamental más común en la niñez, con un estimado del 3 % al 5 % de prevalencia en edad pediátrica a nivel mundial [2].

Entre las diferentes técnicas empleadas para el estudio del TDAH se encuentran la neuroimagen (tomografía, resonancia magnética), estudios neuroquímicos, neurofisiológicos y tests neuropsicológicos. Una de las técnicas de investigación neurofisiológica aplicada actualmente al estudio del TDAH es la que registra la actividad eléctrica cerebral a través de los potenciales cerebrales asociados a estímulos específicos, denominados potenciales relacionados a eventos (ERPs), también llamados potenciales evocados cognitivos. Estos potenciales son registros de la actividad eléctrica del cerebro, que se generan en respuesta a un estímulo específico que puede ser auditivo, somatosensorial o visual. Se presentan como oscilaciones enmascaradas por la señal electroencefalográfica (EEG), y se describen generalmente en términos de sus magnitudes máximas y mínimas (amplitudes pico) y de su duración relativa respecto al estímulo (latencia).

El empleo de técnicas de procesamiento de señales en el análisis de la actividad cerebral mediante potenciales evocados cognitivos, podría generar una herramienta de apoyo al diagnóstico médico que pueda ser empleada en la identificación de la patología y el mejoramiento de los tratamientos que deben formularse a pacientes que presentan TDAH. Aunque la señal presenta una dinámica de cambio temporal, la cual refleja de manera directa el proceso cerebral subyacente, no se han generado metodologías de caracterización de dicha dinámica que puedan ser tenidas en cuenta por parte de los especialistas médicos en el diagnóstico y tratamiento del TDAH. El presente estudio es una primera aproximación a los requerimientos anteriormente planteados.

Planteamiento del problema

En el ámbito social, la falta de diagnóstico del TDAH representa un serio problema que puede generar graves implicaciones tanto en la infancia como en la edad adulta. Se estima que más del 80 % de los niños que lo presentan continuarán expresándolo en la adolescencia, etapa en la cual quienes lo padecen están mucho más propensos a retirarse de la escuela (32-40 %), a comprometerse en actividades antisociales (40-50 %), a hacer uso de drogas ilícitas, y a presentar embarazos a una edad más temprana (40 %). Además, entre el 30 % y el 65 % de quienes fueron diagnosticados con esta enfermedad en la infancia lo presentarán también en la edad adulta, etapa en la cual se caracteriza por dificultades en las relaciones interpersonales, laborales, sociales y por su alta co-morbilidad con trastorno de personalidad disocial, farmacodependencia y alcoholismo [3].

La valoración médica del paciente con TDAH mediante potenciales evocados se realiza a través de la medición de marcadores neurofisiológicos de la señal, específicamente, se estima su amplitud y latencia [4][5]. La literatura refiere un importante número de investigaciones que relacionan directamente la valoración de estos marcadores biológicos al diagnóstico de TDAH, sin embargo, muchas de ellas muestran resultados contradictorios que impiden consolidar un estándar de valores de normalidad para dichos parámetros de los ERPs. Así, por ejemplo, en relación a la latencia del subcomponente P300 de la señal, Robaey et al., en [6], con una tarea visual obtiene una latencia más corta en niños con TDAH en comparación con niños control; por su parte, Satterfield et al. y Klorman et al., en [7] y [8], usando un protocolo con tareas auditivas y visuales, sugieren que no existen diferencias en la latencia de los niños con TDAH y los niños control, mientras que Strandburg et al. y Taylor et al., en [9] y [10], respectivamente, encuentran que los niños con TDAH presentan una latencia más larga que los niños control.

La falta de concordancia en los estudios realizados a grupos de casos y controles, repercute directamente en la falta de elaboración de conceptos médicos de alta confiabilidad, lo cual genera cierto grado de incertidumbre en la correcta etiquetación de los registros. Por esta razón, en este estudio se ha considerado la depuración de la base de datos mediante un *pre-clustering*; además se ha contemplado la aplicación de técnicas no supervisadas en las que se obvia la etiqueta de cada registro, y se forman grupos asociados a las clases normal y patológica para estudiar la calidad de los patrones por los cuales se representan las señales ERP.

Por otra parte, dentro de las diferentes etapas de un sistema de reconocimiento de patrones, y en el caso específico del trabajo con bioseñales, una de las principales restricciones que se debe superar inicialmente está relacionada con el tipo de parámetros que deben medirse, acorde a los procesos físicos que dichas señales están reflejando, con el fin de obtener características que proporcionen niveles de discriminación aceptables en etapas posteriores del proceso. Además, la gran variabilidad intra y entre clase de la morfología de los registros plantea otra restricción en la búsqueda de patrones que proporcionen capacidad de generalización

al sistema de clasificación y se pueda garantizar una optima separabilidad entre las clases. Gran parte de los esfuerzos realizados en esta tesis están orientados a dar solución a las limitaciones previamente mencionadas, en especial, se enfoca en la definición de un grupo de parámetros que, tras el análisis de su capacidad discriminante mediante técnicas de extracción y selección de características, permita identificar correctamente si un sujeto es sano o padece del trastorno.

Justificación

El Trastorno por Déficit de Atención y/o Hiperactividad (TDAH) es una de las psicopatologías más frecuentes en la infancia y su prevalencia actual, a nivel mundial, se estima en un 5.29 % . Sin embargo, estudios realizados en Colombia basados en evaluación psiquiátrica estructurada, pruebas clínicas y test neuropsicológicos, estiman que alrededor del 17 % de la población, entre los 4 y los 17 años de edad, presenta la psicopatología [11]. En el ámbito regional, estudios previos han mostrado, que se presenta antes de los 7 años de edad, con una incidencia cercana al 10 % en la población infantil preescolar y escolar de Manizales [12]. De hecho, el TDAH es el problema de conducta más frecuente en la población escolar y está relacionado con problemas académicos, sociales, familiares y vocacionales, tanto en los períodos de infancia y adolescencia, como en la adultez. La ausencia del respectivo diagnóstico correcto y precoz se asocia en la niñez con tasas más altas de deserción y fracaso escolar; mientras, en la edad adulta, con trastornos de conducta, alcoholismo y fármaco dependencia.

Con respecto al estudio de los ERPs para la identificación del TDAH, se ha podido mostrar que la naturaleza no estacionaria de los registros no se refleja suficientemente en valores puntuales, por lo que se hace necesario la generación de características en forma de arreglos (contornos, espectrogramas, matrices, etc.), que muestren la dinámica de cambio de las mismas sobre un intervalo del dominio de análisis, por ejemplo, el tiempo, la frecuencia, etc. Un estudio de los diferentes eventos presentes en la señal ERP que contemple el análisis de la dinámica completa del proceso, aplicando técnicas de reconocimiento de patrones y de procesamiento de señales, podría generar mayor información discriminante en el diagnóstico de TDAH.

Pertinencia de la tesis

A pesar de la importancia de este trastorno y del creciente número de estudios al respecto, hasta la última década no se han definido criterios de diagnóstico unívocos; no existen marcadores biológicos ni pruebas concluyentes para su diagnóstico, dado que éste sólo se basa en los criterios clínicos DSM-IV (Manual diagnóstico y estadístico de los trastornos mentales) o CIE-10 (Clasificación internacional de enfermedades), soportados por las conductas reseñadas en cuestionarios por parte de padres y profesores. Además, la evaluación visual de

los ERPs por parte de los especialistas como herramienta de apoyo diagnóstico, está sujeta a una alta subjetividad y variabilidad del concepto médico. De ahí, surge la necesidad de crear un sistema de detección automática del TDAH basado en el análisis de los ERPs, y este trabajo es una primera aproximación al cumplimiento de este fin.

Inicialmente, una de las principales restricciones está relacionada con la estimación y selección adecuada de los parámetros de medida, que reflejen con mayor precisión los respectivos procesos físicos, y que proporcionen niveles de discriminación aceptables en etapas posteriores del entrenamiento automático de reconocimiento. Lo anterior demuestra la necesidad de hallar patrones descriptivos de las señales de potenciales relacionados a eventos, que sirvan como primera etapa en la creación de un sistema automático de identificación del TDAH.

La evaluación rigurosa de la conducta de un paciente en el diagnóstico de TDAH requiere la participación de un conjunto amplio de profesionales en el área de la salud, incluyendo pediatras, psicólogos, trabajadores sociales, neuropsicólogos, entre otros; lo cual demanda altos costos para el paciente. Por esta razón, gran parte de la población colombiana de estratos 1,2 y 3 no puede acceder a estos servicios. La pertinencia de esta tesis, radica principalmente en el impacto real del servicio médico que se puede producir al crear un sistema de apoyo al diagnóstico del TDAH, debido a la falta de consenso médico, la variabilidad morfológica de la señal, entre otras razones que hacen complejo su diagnóstico. Este trabajo de investigación es una primera aproximación a la implementación de este sistema de apoyo al diagnóstico médico.

Antecedentes

El Grupo de Control y Procesamiento Digital de Señales (clasificación A en Colciencias) adscrito al Departamento de Ingeniería Eléctrica, Electrónica y Computación de la Universidad Nacional de Colombia sede Manizales ha enfocado su trabajo al análisis automatizado de la actividad biológica con una serie de trabajos y proyectos (financiados por la Universidad Nacional, Colciencias, entidades nacionales e internacionales, etc.), entre otras, en las áreas de desarrollo de sistemas de medición (ECG de 12 canales), detección de patologías de emisión vocal, caracterización y reconocimiento automático de señales de ECG (electrocardiografía) y FCG (fonocardiografía), procesamiento de imágenes médicas, diseño de sistemas automático de identificación de patologías.

El grupo de investigación incluye estudiantes de maestría y doctorado que trabajan en nuevos enfoques sobre los métodos de análisis multivariado, variables dinámicas de tipo cuantitativo, técnicas de análisis con variables cualitativas que permitan ampliar la clase de tareas de entrenamiento automático a aplicaciones en procesos de bioseñales. La producción académica del grupo en los últimos dos años incluye la presentación de trabajos y ponencias en el área de bioingeniería en más de 30 eventos internacionales, la aceptación de más de 10 tra-

bajos publicados en revistas indexadas internacionales, la publicación de tres libros y varios capítulos de libros internacionales.

Recientemente, nuestro grupo ha enfocado su interés al estudio de señales EEG (electroencefalografía) mediante la aplicación de técnicas tradicionales y de aquellas que son de dominio para los estudiantes de grupo. En particular, el análisis de los registros EEG se ha orientado a la detección de patologías como la epilepsia, el parkinson, el estrés y el TDAH. En este sentido, algunos de los proyectos que se han desarrollado en conjuntos con otras instituciones y universidades son:

- *“Sistema de identificación de fuentes localizadas epiletogénicas empleando modelos espaciotemporales de representación inversa.”*
- *“Identificación automática del trastorno por déficit de atención y /o hiperactividad sobre registros de potenciales evocados Cognitivos.”*
- *“Sistema automatizado de clasificación de eventos fisiológicos a partir de patrones bioeléctricos como soporte en el tratamiento de la enfermedad de Parkinson y otros desórdenes neurológicos.”*
- *“Implementación y efectividad de un sistema basado en inteligencia artificial como herramienta para el tratamiento psicológico de personas con trastorno de estrés post-traumático.”*

Organización del documento

Este documento está compuesto por 9 capítulos distribuidos en cuatro partes: preliminares, contextualización, marco teórico y marco experimental.

La primera parte está constituida por el Capítulo 1, el cual contiene la introducción del presente trabajo, incluyendo el planteamiento del problema, su justificación, la pertinencia de la tesis, y además se definen los objetivos general y específicos a cumplir.

La parte de contextualización está conformada por los Capítulos 2 y 3. En el Capítulo 2 se presentan los preliminares fisiológicos y clínicos de las señales de potenciales relacionados a eventos; además, se hace la definición de la patología a identificar (TDAH) y su relación con este tipo de potenciales. En el Capítulo 3 se expone el estado del arte de las técnicas aplicadas al análisis de los potenciales evocados cognitivos, como también se definen los principales requerimientos que incurren en los sistemas automáticos de identificación de patologías.

Los Capítulos 4, 5 y 6 hacen parte del denominado marco teórico. En el Capítulo 4 se define el espacio inicial de características utilizado como patrones descriptivos de los potenciales evocados cognitivos, haciendo una explicación teórica de los conceptos involucrados. Por su

parte, en el Capítulo 5 se consideran algunos conceptos empleados en la reducción de dimensión, análisis de relevancia y selección de características para comprender la importancia de la representación de datos multidimensionales en un espacio reducido de variables. En el Capítulo 6 se describen los procesos de transformación de variables y medidas de desempeño usadas en la cuantificación de las tareas de clasificación.

El marco experimental está constituido por los Capítulos 7 y 8. En el Capítulo 7 se describe la base de datos empleada en este estudio, además de los procedimientos implementados para el desarrollo de los experimentos y las estrategias de validación utilizadas con el fin de comparar el desempeño de las diferentes técnicas de extracción y selección de características analizadas. El Capítulo 8 contiene los resultados obtenidos en los diferentes experimentos descritos en el capítulo anterior, en los cuales puede apreciarse la capacidad discriminante de las características propuestas.

Finalmente, en el Capítulo 9 se presentan las conclusiones y los aportes generados en esta investigación; además, se plantea el posible trabajo futuro relacionado al estudio de los potenciales evocados cognitivos mediante técnicas de procesamiento digital de señales.

Objetivos

Objetivo general

Evaluar técnicas de extracción y selección de características sobre un conjunto de patrones descriptivos de la señal de potenciales evocados cognitivos (ERPs), con el fin de hallar la representación adecuada que conduzca a la clasificación de pacientes diagnosticados con TDAH y sanos.

Objetivos específicos

- Determinar un espacio inicial de características que contenga información temporal, espectral y morfológica de la señal, a partir del cual se pueda hacer una representación de los procesos cognitivos subyacentes presentes en las señales de los potenciales evocados cognitivos.
- Implementar y comparar algoritmos representativos de extracción y selección de características, para determinar la técnica que genere la representación apropiada de los datos en términos de la tarea de clasificación.
- Cuantificar la calidad de la representación de los datos a través del rendimiento de clasificadores supervisados y no supervisados en la discriminación entre pacientes con TDAH y sujetos sanos.

Parte II.

Contextualización

2. Preliminares Fisiológicos y Clínicos

A fin de contextualizar al lector, en el presente capítulo se hace una revisión minuciosa de la fisiología de las señales de los potenciales evocados, que incluye la descripción de sus componentes, nomenclatura usada en la literatura, las características morfológicas usualmente empleadas en neurofisiología para el análisis de los registros, entre otros. Además, se expone de manera resumida algunas generalidades del trastorno por déficit de atención y/o hiperactividad (TDAH) y la relación de los potenciales evocados con los procesos cognitivos y la detección de este desorden comportamental.

2.1. Clínica de los potenciales evocados

2.1.1. Definición

Los potenciales evocados (*evoked potentials* - EPs) se definen como respuestas eléctricas del sistema nervioso a la estimulación sensorial, es decir, una manifestación de la recepción del cerebro de una respuesta a un estímulo externo. Estas formas de onda consisten en una secuencia de picos y deflexiones caracterizadas por parámetros medidos sobre la señal como latencia, amplitud y otras características descritas en detalle en la sección 2.1.5. En pruebas clínicas, los EPs son generados por estimulación visual o auditiva, o por estimulación eléctrica de los nervios sensoriales. Estos potenciales son registrados con electrodos desde el cuero cabelludo o la superficie de la piel sobre la médula espinal o los nervios periféricos. Para propósitos investigativos, los EPs pueden ser provocados por estimulación eléctrica de puntos en el sistema nervioso central con electrodos insertados quirúrgicamente [4].

En este documento, el término *potencial evocado* es usado arbitrariamente para indicar sólo el promedio de las respuestas individuales; el término *respuesta* es usado para cualquier registro seguido de un estímulo simple. Los términos *pico* u *onda*, se usan para denotar las deflexiones que van hacia arriba o hacia abajo y que forman un EP. El nombre *potencial relacionado a un evento* (en inglés *event-related potential* - ERP) es comúnmente usado para denotar los EPs y otros tipos de potenciales que son el resultado de procesos cognitivos seguidos de un estímulo.

2.1.2. Nomenclatura

Aunque existe cierta diversidad en la denominación de los componentes electrofisiológicos de los EPs, en general, se observan ciertos principios básicos relacionados fundamentalmente con sus polaridades y latencias promedios. La polaridad de un componente está determinada por la dirección del trazado (registro) sobre una línea de base determinada (muchas veces sólo virtual). Si el trazado se representa “hacia arriba”, el componente se registra como una negatividad (polaridad N), y si se representa “hacia abajo”, el componente se registra como una positividad (polaridad P). Es necesario recordar que en electrofisiología clásica se reconoce como “negativo” lo que se muestra graficado hacia arriba y viceversa.

La latencia de un componente electrofisiológico expresa su relación temporal con el estímulo que lo genera y suele medirse en milisegundos. De modo que cuando se denomina un componente como P100, quiere decir que se trata de un componente de polaridad “positiva”, que aparece aproximadamente 100 ms después de la aplicación del estímulo. Otra manera de identificar los componentes de los EPs es numerando las polaridades en secuencia. Los picos positivos y negativos se etiquetan con las letras N y P respectivamente (de acuerdo a los principios de electrofisiología clásica arriba mencionados), seguidos por un número que indica el orden en que se van presentando entre los picos de la misma polaridad. Así, el primer pico positivo se etiquetará como N1, el segundo como N2, etc. [13].

2.1.3. Clasificación de los potenciales evocados

Existen numerosas clasificaciones de los EPs. Algunas de las más utilizadas por su aporte informativo son [14][15]:

1. Según su naturaleza, los EPs pueden subdividirse en:
 - *EPs exógenos*. Son aquellas respuestas evocadas cuyos componentes electrofisiológicos se modifican fundamentalmente dependiendo de las propiedades físicas del estímulo. Suelen ser respuestas evocadas que ocurren de manera temprana, esto es, durante los primeros 200 milisegundos después de la aplicación del estímulo, y están relacionadas principalmente con el procesamiento de estímulos sensoriales (EPs sensoriales).
 - *EPs endógenos*. A diferencia de la anterior, son respuestas evocadas cuyos componentes electrofisiológicos son independientes de las propiedades físicas del estímulo, y por el contrario se modifican fundamentalmente de acuerdo con variables psicológicas como el estado de conciencia, atención, motivación, significación, etc, las cuales están relacionadas con el estado del sujeto explorado y de la tarea a realizar. Suelen ser más tardías y ocurren 200-250 milisegundos posterior a la aplicación del estímulo, y están vinculados principalmente con el procesamiento de eventos neurocognitivos (EPs cognitivos).

2. Según la modalidad sensorial, los EPs pueden subdividirse en:

- *EPs auditivos*. Son las respuestas evocadas que se obtienen mediante la aplicación de un estímulo acústico o sonoro.
- *EPs visuales*. Son las respuestas evocadas que se obtienen mediante la aplicación de un estímulo visual, fótico o luminoso.
- *EPs somatosensoriales*. Son las respuestas evocadas que se obtienen mediante la aplicación de estimulación eléctrica (puede incluirse el láser) a nervios periféricos o áreas dermatómicas.
- *EPs motores*. Son aquellas respuestas evocadas que se obtienen mediante la aplicación de estimulación eléctrica o magnética sobre vías motoras e incluso áreas de la corteza motora.

3. Dependiendo de su relación temporal con el estímulo (latencia) los EPs pueden subdividirse en:

- *EPs de latencia corta*. Estas respuestas evocadas aparecen en los primeros 10 milisegundos posteriores a la aplicación del estímulo.
- *EPs de latencia media*. Se registran entre 10 y 50 milisegundos posteriores al estímulo.
- *EPs de latencia larga*. Se registran entre 50 y 200-250 milisegundos posteriores al estímulo.
- *EPs de latencia muy larga o tardíos*. En este caso las respuestas evocadas se registran más allá de 250 milisegundos posteriores a la aplicación del estímulo, y están casi siempre relacionadas con EPs cognitivos.

4. Según su relación con la frecuencia de estimulación (frecuencia de repetición del estímulo), los EPs se subdividen en:

- *EPs transientes*. Son aquellas respuestas evocadas que se obtienen cuando ocurre un cambio o modificación en el estímulo.
- *EPs sostenidos*. Son aquellas respuestas evocadas que se obtienen sólo cuando se mantiene la continuidad del estímulo. Ambos tipos de EPs se obtienen con frecuencias de estimulación suficientemente lentas como para permitir el registro de una respuesta evocada completa antes de que se presente el siguiente estímulo.
- *EPs de estado estable*. Son aquellas respuestas evocadas que se obtienen sólo con frecuencias de estimulación tan rápidas que la respuesta transiente a un estímulo se superpone con la respuesta transiente que genera el siguiente estímulo. Esto origina un tipo de respuesta evocada periódica (casi sinusoidal) cuyos nuevos componentes electrofisiológicos (amplitud y fase) están relacionados armónicamente con la frecuencia de repetición del estímulo.

2.1.4. Generación de los EPs

El fenómeno eléctrico que se lleva a cabo a nivel neuronal para la generación de un potencial consiste básicamente en el cambio de polaridad de la membrana de una neurona. El impulso nervioso, también llamado potencial de acción, es una onda de naturaleza eléctrica que se autopropaga a lo largo de la superficie de la membrana de la neurona y de sus prolongaciones (axones). La neurona en reposo, cuando no se encuentra transmitiendo mensajes, se encuentra polarizada, es decir, la parte externa de su membrana tiene una carga eléctrica distinta de la interna y ello se debe a que los iones de sodio y potasio (los iones son partículas con carga eléctrica) tienen distinta concentración en el interior de la membrana que afuera de ella. Al llegar un estímulo a la célula nerviosa o a sus prolongaciones se altera el orden de las moléculas que hay a uno y otro lado de la membrana, ya que se hace más permeable, penetrando al interior los iones de sodio y saliendo al exterior iones de potasio. Esto hace que la superficie interna de la membrana celular se vuelva positiva en relación con la superficie externa. Este cambio electrónico induce a su vez a la alteración de la estructura molecular de la siguiente sección de la membrana, que de esta forma también se despolariza [16][17].

Este tipo de eventos eléctricos pueden registrarse mediante métodos no invasivos, esto es, ubicando electrodos sobre el cuero cabelludo. Los cambios en la actividad eléctrica de una región particular pueden utilizarse para determinar si una estructura interviene en las diferentes funciones cognitivas superiores, especialmente, la atención. Los registros se llevan a cabo en situaciones controladas y se realizan durante la presentación de estímulos, la toma de decisiones o las actividades motoras [18].

2.1.5. Características generales de los potenciales evocados

Los potenciales evocados consisten en una serie de picos u ondas, que se caracterizan por determinados parámetros medidos a partir de la forma de onda; estos son: la polaridad eléctrica, el número de ondas en secuencia, latencia desde la aparición del estímulo o desde un pico precedente, la amplitud con respecto a la línea base o a picos de polaridad opuesta precedentes anteriores o subsecuentes, y la forma de onda.

Latencia

La latencia es un concepto temporal, e implica el intervalo de tiempo que transcurre desde el momento de la aplicación del estímulo hasta un punto específico en la morfología de la respuesta evocada. En estudios de potenciales evocados, los picos en la forma de onda son usados como puntos de medición. El tiempo de separación entre dos picos se denomina *latencia inter-onda* o *latencia inter-pico*.

Existen diversas formas de medir la latencia; la más utilizada en los registros de EPs es considerarla hasta el punto de máxima amplitud del componente evaluado, esto es lo que

se conoce por el término de latencia absoluta. Además de la anterior, se conoce otra forma particular de evaluar este parámetro: la latencia interpicos o intercomponentes, que resulta de la medición de la diferencia temporal entre dos componentes relacionados. Las mediciones de latencias se realizan a través de los cursores (marca movibles identificables sobre el modo gráfico) y su unidad de medida son los milisegundos [4].

La latencia suele representar la velocidad de conducción de las fibras nerviosas que participan en la conformación de la respuesta evocada, así como también nos da información indirecta sobre la posible eficiencia de los diferentes relevos sinápticos que tienen lugar en la transmisión nerviosa.

Amplitud

Este parámetro cuantitativo considera el tamaño relativo de un determinado componente. Existen varias formas de medir la amplitud, pero la más utilizada en los registros EPs es aquella que se considera como amplitud absoluta, y consiste en medir del punto de máxima amplitud de un componente o pico de una polaridad al punto siguiente (o precedente) de máxima amplitud pero de polaridad opuesta. Esto es lo que se conoce como amplitud pico-pico. En algunos equipos comerciales puede evaluarse la amplitud de un componente, con relación a una línea de base, que mide la distancia vertical que explica la diferencia de voltaje entre un pico y un nivel de referencia que representa una amplitud cero. Su unidad de medida son los microvoltios (μV). Los valores de amplitud son mucho más variables que los de latencia, pues no suelen presentar una distribución estadística normal y se requieren entonces de alternativas estadísticas más complejas para definir su normalidad.

La amplitud refleja la cantidad de elementos neurales que participan en la transmisión nerviosa, y por tanto, también tiene que ver con la sincronización con que esta se produce. Es de importancia enfatizar que en numerosas ocasiones, fundamentalmente dependiendo del momento y del tipo de daño neurofuncional que caracterice al proceso patológico, estos elementos se interrelacionan y no se observan entonces alteraciones electrofisiológicas puras.

Polaridad

La polaridad se refiere a la positividad y negatividad entre dos electrodos conectados a la entrada del sistema de adquisición. La relación entre los cambios en los potenciales eléctricos en los electrodos y las deflexiones ascendentes y descendentes del trazado de un EP depende de cuáles electrodos estén conectados en las dos entradas y de qué convención de polaridad es usado por el sistema de adquisición. Por lo tanto, las deflexiones ascendentes y descendentes en un potencial evocado deben ser explicadas en ambos términos.

Distribución topográfica

La apariencia de los potenciales evocados en diferentes regiones de la cabeza no sólo depende de la modalidad sensorial, sino también del estímulo usado. En general, las respuestas evocadas suelen ser muy simétricas. Evaluar este parámetro, por tanto, consiste en considerar el grado de similitud o relación interhemisférica de las respuestas evocadas. Pueden establecerse relaciones de simetría para cada uno de los marcadores electrofisiológicos ya referidos. De esta manera, podemos evaluar simetrías interhemisféricas de morfología, latencias y amplitudes.

Existen estudios normativos para casi todos los tipos de EPs, los cuales caracterizan con criterios estadísticos, las variaciones normales de casi todos los marcadores electrofisiológicos. El experto en su estrategia para la interpretación correcta del EP, utiliza la comparación de los valores del paciente, para cada uno de los diferentes marcadores, con sus valores normativos ya conocidos. Los estudios normativos se realizan sobre una muestra de sujetos sanos o "normales", utilizando siempre las mismas condiciones de registro y obtención del EP, con que se examinarán a los pacientes.

Morfología

La morfología de una respuesta evocada la constituye su forma de onda (patrón) característica. Esta forma de onda resulta de la disposición gráfica típica que toman sus diferentes componentes (estructura multicomponente). Por supuesto, que cualquier patrón morfológico incluye las interconexiones de polaridad, latencia, amplitud, distribución topográfica, generadores y función de cada uno de sus componentes. Con alguna frecuencia, estos patrones típicos pueden presentarse con algunas modificaciones que no afectan su esencia general, esto es lo que se conoce como variaciones morfológicas normales. El reconocimiento del patrón morfológico del EP suele ser la primera estrategia visual del experto para su correcta interpretación. Las características del patrón morfológico del EP nos informan acerca del grado de sincronía o sincronización, así como de la orientación de los dipolos generadores de los elementos neurales que participan en la respuesta evocada.

La figura 2-1 muestra el registro de la forma de onda de un potencial evocado auditivo normal. La gráfica señala los componentes que conforman la señal y los parámetros de morfología más significativos en los estudios neurofisiológicos de los EPs (latencia y amplitud), medidos con respecto a la presentación del estímulo y el pico P300.

2.1.6. Componentes de los potenciales evocados cognitivos

A través de paradigmas experimentales, de mapeos topográficos de las señales promediadas y de análisis de las fuentes eléctricas, ha sido posible recopilar una considerable cantidad de información sobre los potenciales evocados que ha mostrado la existencia de varios compo-

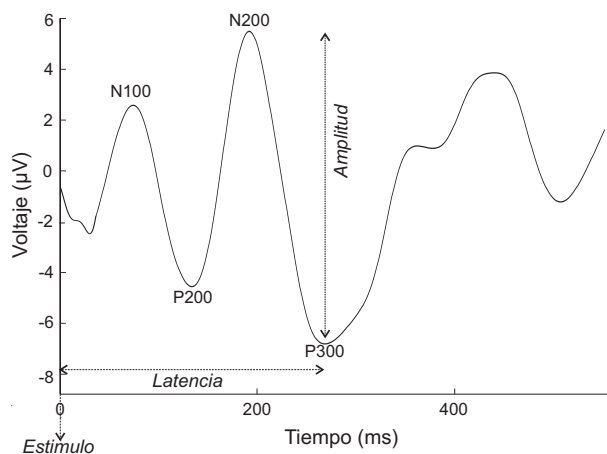


Figura 2-1.: Forma de onda del potencial evocado cognitivo.

nentes de los ERPs que están asociados con diferentes fases y operaciones de la cognición humana. A continuación se hace una breve descripción de los principales componentes de los ERPs y de mayor interés en el área de la neurofisiología relacionado al análisis de múltiples enfermedades neurocognitivas y cerebrales.

P100

Alrededor de los 100 ms después de la presentación del estímulo se presenta una deflexión denominada P1 o P100. Este pico puede verse mejor definido en los electrodos del área occipital, la cual está muy próxima al área visual primaria. Por esta razón, el componente P100 representa una respuesta sensorial que se evoca por un estímulo visual sin importar la tarea que esté desarrollando el sujeto (las variaciones en las tareas pueden influir en la amplitud de la P1, pero no necesariamente una tarea en particular es la que provoca dicha onda). Este componente puede también ser observado en electrodos centrales y frontales pero no muy bien definidos, y en ubicaciones posteriores aparece de forma más tardía [19].

Aunque el componente P1 es discernible en la mayoría de los casos y domina en el hemisferio cerebral ipsilateral al músculo que se contrae, no se puede observar con movimientos bilaterales simultáneos, lo cual sugiere que P1 está relacionado con la inhibición al movimiento de imitación de la mano opuesta (movimiento espejo) [20].

N100

El componente N100 o N1 generalmente representa la extracción inicial de información a partir del análisis sensorial del estímulo o también puede ser interpretada como la excitación asociada a la asignación de un canal de procesamiento de la información de la corteza primaria [5], así el componente N100 muestra una amplitud mayor ante estímulos que el sujeto atiende, ya sea de forma pasiva o activa. Anatómicamente, la N100 es generada por múltiples

áreas del plano temporal superior, con una mayor contribución de las áreas cerebrales denominadas Planum Temporale y la circunvolución de Heschl. Fisiológicamente, la aparición de patrones sencillos o dobles en la N100 parece estar determinada por interacciones no lineales entre las respuestas dadas a dos eventos acústicos sucesivos [21].

Otro componente dentro del intervalo de los 100 – 200 ms es la denominada *negatividad del procesamiento*, que puede solaparse con la N100 y que aparece particularmente en tareas de atención selectiva. En promedio, la amplitud de estos componentes varía entre 2 y 4 μV , pero ésta se hace mayor cuando más complicado resulta discriminar entre dos estímulos que se presentan simultáneamente. Un tercer componente encontrado en este intervalo de tiempo se conoce como *la negatividad de desemparejamiento* (*mismatch negativity* en inglés), y se produce cuando en la secuencia de estímulos que el sujeto atiende aparece alguno que difiere del resto (sobre todo físicamente) [22].

P200

Otro componente positivo que sigue a la onda N100 es el denominado pico P200 o P2, que puede distinguirse en las ubicaciones central y anterior del cuero cabelludo. Este componente muestra mayor amplitud ante estímulos que contiene características denominadas diana (*target* en inglés), y su efecto aumenta cuando los *target* son relativamente infrecuentes. En este sentido, la P200 de la zona anterior es similar a la onda P300. Sin embargo, la P2 anterior sólo ocurre cuando el *target* está definido por estímulos de características muy simples, mientras que los efectos de la P300 pueden ocurrir por categorías de *targets* complejos que aparecen arbitrariamente. En sitios posteriores, la onda P200 frecuentemente es difícil de distinguir debido al solapamiento que puede ocurrir con las ondas N1, N2 y P3; en consecuencia, la bibliografía no reporta mucha información sobre la generación de este componente en tal ubicación [23]. Sin embargo, en [18] se muestra que en general la P200 parece reflejar la relevancia intrínseca que el estímulo posee, y hace referencia a la importancia del estímulo independientemente de la tarea encomendada.

N200

La N200 (o N2) es una negatividad resultante de una desviación en forma de un estímulo predominante, típicamente evocado entre los 180 y 325ms después de la presentación de un estímulo auditivo o visual. La generación de esta respuesta puede ser alcanzada mediante el paradigma experimental *oddball*, en el cual los sujetos son expuestos a una sucesión continua de dos tipos de estímulos, uno que se presenta regularmente y otro que se manifiesta esporádicamente. En este experimento, la N200 es típicamente evocada antes de la respuesta motora, lo que sugiere su relación con el proceso cognitivo de distinción e identificación de estímulos [24].

Distintos tipos de potenciales N200 han sido caracterizados [25]: un conjunto que refleja el

procesamiento involuntario, y otro, evocado mediante el procesamiento activo. En la presentación repetitiva del estímulo, la *N2a* es una distribución cortical anterior evocada por la atención consciente o inconsciente de un estímulo desviado; la *N2b* es una negatividad de la distribución cortical central que sólo es observada durante un estímulo de la atención consciente; la *N2c* se presenta frontal y centralmente durante tareas de clasificación. Además, los estímulos presentados en tareas de búsqueda visual pueden evocar una deflexión denominada *N2pc*, como índice de un cambio atencional, en la región occipital-temporal de la corteza contralateral [24].

P300

La onda P300 ha sido objeto de numerosos estudios relativos a los procesos cognitivos, procesamiento de la información y la atención selectiva [5], debido en parte a su amplitud relativamente grande y fácil obtención en contextos experimentales. Este componente, de polaridad “positiva” y latencia promedio de 300 ms, es evocado principalmente en tareas discriminativas con estimulación auditiva, visual o somatosensorial (paradigma *oddball*), mediante la aparición inesperada de un estímulo infrecuente (también llamados estímulos *diana* o *raros*), que tiene una probabilidad significativamente inferior a la de los estímulos frecuentes (estímulos *no diana* o *comunes*) [15].

La onda P300 depende enteramente de las tareas ejecutadas por el sujeto y no está directamente influenciada por propiedades físicas de la generación del estímulo. La P3 es por tanto, un componente tipo endógeno debido a su dependencia a factores internos y no externos [23]. En varias patologías relacionadas a la cognición, la onda P300 muestra alguna alteración que refleja anormalidad de la respuesta, tal como sucede en la depresión, la esquizofrenia, la demencia, entre otras [26] [27].

Se desconocen las estructuras anatómicas responsables de la distribución topográfica de la P300. El hipocampo y áreas adyacentes del sistema límbico se activan al mismo tiempo que se registra la P300; sin embargo, una lobectomía temporal unilateral no afecta la distribución topográfica de la P300. Es posible entonces que la P300 represente actividad generada tanto en sistema límbico como en áreas asociativas parietales. Aunque los orígenes neurales de la P300 continúan en debate, los registros con electrodos profundos en humanos indican que al menos alguna porción de la P300 se genera muy probablemente en el lóbulo temporal medio.

N400

Este componente ha despertado un especial interés en diferentes áreas de investigación debido a los signos de procesamiento del lenguaje que muestra. En [28] [29], Kutas y Hillyard describen una deflexión negativa que se presenta entre los 300 y los 500 ms después de la estimulación (haciendo referencia a la N400), la cual está correlacionada con la aparición de palabras semánticamente anómalas en oraciones con sentido contrario. Lo anterior refleja

la *memoria semántica*, es decir, la previsibilidad de una palabra basada en el contenido semántico de la oración anterior [19].

2.1.7. Uso clínico de los potenciales evocados

Los EPs son usados principalmente en tests de conducción de los sistemas somatosensorial, auditivo y visual, especialmente en la parte central de estos sistemas. Los potenciales evocados son tan sensibles que pueden llegar a detectar lesiones que no fueron descubiertas clínicamente o por otras técnicas de laboratorio.

Determinación de límites de normalidad

Dado que la interpretación clínica de los EPs tiene que ser confiable y consistente, debe haber un método definido para determinar si los resultados de las pruebas de los pacientes están o no dentro del rango de normalidad. En la mayoría de casos, esto requiere que cada laboratorio ejecute en al menos 35 sujetos normales la misma prueba que será usada en pacientes. Se debe enfatizar que exactamente los mismos parámetros se deben medir tanto en los sujetos normales como en los pacientes a quienes se aplicará la prueba, debido a que muchos de estos factores tienen efectos significativos en las latencias y amplitudes de los EPs. Por ejemplo, en los potenciales visuales, pequeños cambios en los patrones de iluminación podrían repercutir en considerables alteraciones en la latencia de la P100, los cuales podrían sugerir un estado de normalidad en una prueba realizada bajo un nivel de iluminación, o de anormalidad bajo otro nivel. Por esta razón, no tiene sentido poner a prueba sujetos normales en un cuarto con una estimulación, y poner a prueba pacientes en otro cuarto con una estimulación diferente.

La mayoría de parámetros tienen una distribución gaussiana o normal. En este caso, las desviaciones estándar pueden resultar eficientes para describir los límites esperados de normalidad. La desviación estándar usada como límite de normalidad debe incluir al menos el 98 % de la población que ha sido estudiada; la mayoría de laboratorios usa 2.5 (98.8 %) o 3 (99.7 %) desviaciones estándar. Es incorrecto usar 2.0 o menos desviaciones estándar para definir el límite superior de normalidad para propósitos de interpretación clínica. Otras medidas estadísticas usadas son las distribuciones de frecuencia acumulada y los límites de confianza, los cuales ilustran de manera más comprensible la relación entre los resultados de una prueba de un paciente dado y la población normal. Los métodos no paramétricos se usan cuando los valores normales tienen una distribución no gaussiana. Por ejemplo, muchas medidas de amplitud muestran curvas más pronunciadas en el lado de las bajas amplitudes que en el lado de las altas amplitudes; en estos casos, hay cierta validez en tomar como límites de normalidad, los valores extremos registrados para sujetos normales.

Registro de los potenciales evocados

Los potenciales evocados son señales eléctricas que se encuentran embebidas en el electroencefalograma (EEG), el cual corresponden a una respuesta espontánea y altamente no estacionaria, a diferencia de los potenciales. El EEG mide la actividad eléctrica promedio del cerebro en diferentes puntos de la cabeza.

La adquisición de los potenciales evocados requiere de cinco elementos básicos para llevar a cabo su registro:

El paciente Debido a que los electrodos no solo registran los EPs y EEG generados a nivel cerebral, sino también el EMG (electromiografía) de la musculatura subyacente a la posición del electrodo, es indispensable que el paciente se encuentre en reposo, relajado y lo más confortable posible dentro de las limitaciones del examen; de lo contrario se aumentará considerablemente la actividad eléctrica de fondo que deberá ser reducida por los procesos de filtrado y promediación de la señal [30].

Electrodos La manera más común de obtener el registro del EEG (y por ende de los EPs) es mediante métodos no invasivos, esto es, ubicando electrodos sobre la superficie del cuero cabelludo o de la piel; los registros intracraneales son rara vez utilizados en humanos, en ocasiones se usan en pacientes con epilepsia para la localización de focos epilépticos [19].

Los electrodos utilizados para el registro de EEG son copas metálicas con un orificio central, elaboradas en diferentes materiales conductores como oro, plata o estaño, y con un diámetro de 4 – 10 mm. La ubicación de los electrodos se hace de acuerdo al sistema internacional 10 – 20 (Figura 2-2), y depende del tipo de potencial evocado que se desee registrar; la clara definición de la forma de onda y la considerable amplitud de los picos obedece directamente a la correcta localización de los electrodos. La actividad de cada electrodo está referenciada por uno o un par de electrodos en común, ubicados generalmente en los lóbulos de las orejas [19]. Puesto que el electrodo es la interfaz entre la piel y el sistema de registro, se requiere que éste presente una impedancia eléctrica mínima (en lo posible de 5 k Ω o menor), para lo cual es indispensable que la superficie de la piel donde se ubicarán los electrodos esté libre de suciedad, grasa y células muertas; para tal fin se emplean diferentes métodos que incluyen técnicas de limpieza con productos químicos y pastas conductoras [4].

Estimulación En cada una de las modalidades sensoriales evaluadas, los potenciales evocados varían dependiendo no sólo del tipo de estímulo aplicado sino también de los métodos específicos de generación de estímulos. La intensidad del estímulo aplicado es determinante en la forma de onda del potencial evocado: si se aplica un estímulo muy débil, entonces se requiere hacer una promediación de una gran cantidad de respuestas para aumentar su amplitud; por el contrario, un aumento desmedido del estímulo

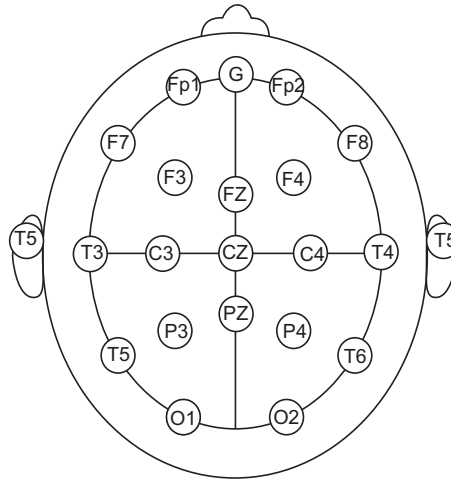


Figura 2-2.: Ubicación de los electrodos según sistema internacional 10-20

provocaría un incremento en la amplitud y la latencia puede decrecer, cambiando así la forma de la señal y aumentando más picos al EP. Por otra parte, debe elegirse una frecuencia de estimulación óptima para cada modalidad sensorial y de acuerdo al tipo de respuesta que se desea registrar, ya que para todos los tipos de estímulos la respuesta del sistema nervioso varía de acuerdo a la frecuencia de presentación del estímulo; así, la amplitud del EP puede disminuir con el aumento de la frecuencia de estimulación, y la sensibilidad de los componentes de la respuesta a los cambios de la frecuencia dependerá de la modalidad de estimulación y del nivel del sistema nervioso en que se origina [31].

Amplificación y filtrado En los sistemas modernos de adquisición de EPs, los amplificadores, unidades de calibración, filtros y demás dispositivos de procesamiento de la señal, conforman un sólo módulo. La amplificación es un procedimiento necesario durante el registro de los potenciales evocados con el fin de aumentar la amplitud de la señal eléctrica adquirida, la cual está en el orden de los microvoltios. Los amplificadores usados en los sistemas de adquisición son de tipo diferencial, el cual amplifica la diferencia de voltaje entre dos de las tres entradas y cancela los componentes en común adquiridos por los tres electrodos.

Una segunda etapa del acondicionamiento de la señal (que debe hacerse previamente a la promediación) es el filtrado análogo de las respuestas, cuyo objetivo es la reducción, y en lo posible, eliminación del ruido de fondo de la señal adquirida, sin que la amplitud y morfología de la onda sufran modificaciones significativas, mejorando así una baja relación señal-ruido (SNR). Para tal propósito, se dispone de filtros pasabanda en diferentes configuraciones, dadas de acuerdo al tipo de EPs que se desee registrar (y a las bandas de frecuencia que se pretenden restringir). La selección de filtros ade-

cuados es igualmente determinante en la morfología de la onda, dado que la incorrecta determinación de las frecuencias de corte, pueden generar efectos significativos sobre las latencias de los picos [30].

Promediación La promediación es el método comúnmente utilizado para la obtención de los potenciales evocados. Consiste en hacer la promediación de las respuesta individuales obtenidas en la repetición de un mismo experimento; de esta manera, el ruido esporádico conformado por el EEG, el EMG, y la interferencia eléctrica, es reducido substancialmente y los EPs se manifiestan de forma más clara. Esta técnica se realiza bajo la presunción que el EP es invariante en el tiempo, ya que el EEG tiene por lo menos una distribución estadística simétrica. Además, para que esta técnica resulte confiable, debe asumirse que cada vez que un estímulo dado se presente, provocará esencialmente el mismo potencial evocado; de esta manera, la amplitud y polaridad que exhiben los EPs serán razonablemente estables (en tiempos fijos) en relación con el estímulo y no se cancelarán al promediar el número de repeticiones, lo cual si ocurriría con el ruido de fondo al presentar amplitudes y polaridades eventuales en cualquier instante que sigue al estímulo.

La Figura 2-3 muestra cómo el procedimiento de promediación de registros EEG, permite ver con claridad los potenciales evocados en la respuesta promediada. En esta gráfica, se muestran 16 registros individuales del electrodo occipital izquierdo de un sujeto normal y su correspondiente señal promedio. El triángulo ubicado en la parte superior marca el tiempo de estimulación. Nótese como las respuestas evocadas se ven claramente después de promediar, pero difícilmente pueden ser identificadas en los registros individuales

Paradigmas de evocación de potenciales

El método más común para la provocación de potenciales relacionados a eventos es el paradigma *oddball* (*tareas de ejecución continua*). Éste consiste en la presentación aleatoria de una secuencia de dos tipos de estímulos; el primero de ellos aparece de forma frecuente, y se denomina estímulo estándar o frecuente. El otro tipo de estímulo (en inglés, denominado *target* o *oddball*), por ser raro o infrecuente (diana) aparece con menor frecuencia y de manera inesperada. Los estímulos raros y comunes pueden ser tonos a diferentes frecuencias, o figuras de diferentes colores, formas, etc. La tarea del sujeto evaluado consiste en clasificar el estímulo en una de las dos categorías, y se pide entonces que cuente el número de veces que percibe un estímulo *target* en una sesión o que presione un botón cuando esto ocurra. Si el estímulo presentado y captado por el sujeto pertenece a la categoría de infrecuentes, entonces se habrá evocado la onda P300 y la deflexión positiva se manifestará en el registro del ERP [19][32].

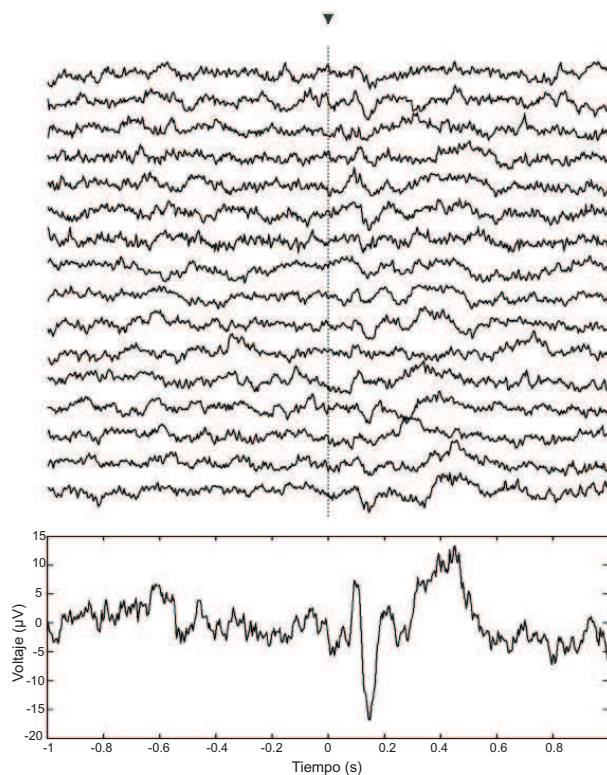


Figura 2-3.: Extracción de EPs por promediación de señales EEG.

La figura 2.4(b) muestra un potencial evocado visual promediado mediante el paradigma oddball. En esta gráfica se observa la respuesta promedio a un estímulo frecuente, mientras que en la gráfica 2.4(a) se muestra la respuesta promedio a un estímulo raro o infrecuente. Nótese que la depresión acentuada que se da alrededor de los 300 ms después de la estimulación, se presenta solamente por un estímulo infrecuente.

2.2. Trastorno por déficit de atención y/o hiperactividad

2.2.1. Definición

El TDAH es un cuadro sintomático muy heterogéneo desde el punto de vista clínico, el cual se caracteriza por un patrón persistente de desatención, impulsividad e inquietud motriz exagerada e inapropiada para la edad de desarrollo. Aunque tradicionalmente se ha relacionado con una situación clínica típica de la infancia y la adolescencia, en la actualidad el TDAH se define como un trastorno crónico sintomáticamente evolutivo, que debe manifestarse por un lapso de tiempo no inferior a 6 meses, y presentarse por lo menos en dos ámbitos distintos. Los pacientes afectados se muestran inquietos y dispersos en los primeros años de edad, e incluso, en los primeros meses de vida. El cuadro se hace especialmente notable a partir de los tres primeros años, mostrando una diversidad clínica e intensa a partir de los seis años

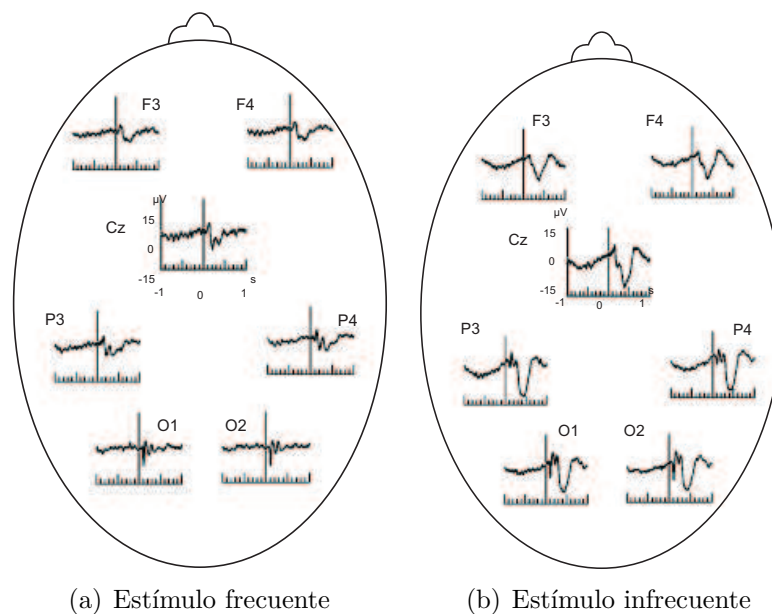


Figura 2-4.: Presentación del componente P300 dado un estímulo infrecuente

de edad, durante la etapa escolar [33][12].

2.2.2. Estadísticas

El trastorno por déficit de la atención con hiperactividad (TDAH) se considera uno de los desórdenes del comportamiento más comunes de la infancia; según la Asociación Psiquiátrica Americana, afecta entre 3 y 5 % de los niños en edad escolar. Se calcula que más de 80 % de los niños que lo presentan en la infancia, lo manifestarán también en la adolescencia. Entre el 32 y 40 % de quienes lo padecen en la adolescencia, están más propensos a retirarse de sus estudios; y sólo entre el 5 y 10 % de los que continúan completan el colegio, pero el 50 – 70 % de ellos no tiene buenas relaciones interpersonales, tienen pocos amigos o no los tienen. Entre el 40 y 50 %, se involucran en actividades antisociales y hacen uso de drogas ilícitas; el 40 % de las adolescentes son más propensas a quedar en embarazo a una edad temprana, y el 16 % presentan mayor riesgo de contraer enfermedades de transmisión sexual. Quienes conducen vehículos tienden a exceder los límites de velocidad y a tener múltiples accidentes de tránsito [34]. Por último, alrededor del 50 % de los sujetos que son diagnosticados en la infancia, presentan manifestaciones en la edad adulta [35], etapa que se caracteriza por dificultades en las relaciones interpersonales, laborales, sociales, entre otras. Además, este trastorno tiene una alta comorbilidad con entidades psiquiátricas como el trastorno de conducta disocial y oposicional desafiante (entre el 40 y 50 %), problemas de ansiedad (25 %), depresión (30 – 33 %), síndrome de Gilles de la Tourette (2 %), fallas en el aprendizaje (20 – 25 %), farmacodependencia, alcoholismo, entre otros trastornos de la conducta [36].

2.2.3. Diagnóstico del TDAH

El diagnóstico del trastorno es clínico y se basa en 18 síntomas definidos en los criterios del DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, fourth edition) (ver Anexo A) y en el CIE-10 (Clasificación Internacional de Enfermedades, Décima revisión). En la actualidad, a partir de los criterios del DSM-IV, se aceptan tres formas principales: de predominio en inatención, de predominio hiperactivo y la forma combinada. La primera se suele asociar con trastornos del aprendizaje, del lenguaje y/o de la coordinación motora, mientras que las dos últimas se relacionan directamente con los trastornos de la conducta [37]. En la población colombiana la forma más frecuente es el tipo combinado con 9,2 %, le sigue el predominio en desatención con 6,2 % y el hiperactivo con 1 % [38].

Aportes de Pruebas neurofisiológicas en el diagnóstico y seguimiento

Electroencefalograma. Entre la anomalías comúnmente encontradas en niños con afecciones de TDAH se encuentra la disminución de la velocidad difusa o intermitente de la actividad eléctrica cerebral fuera de los límites de la normalidad para la edad cronológica. En la actualidad no se ha logrado determinar un patrón electroencefalográfico específico para el TDAH, debido a la posible inmadurez eléctrica por la presencia de una mayor densidad de ondas lentas o a que el trazado puede ser normal aún cuando padece el trastorno.

El EEG no debe usarse de manera indiscriminada, ya que puede condicionar ciertas actitudes terapéuticas debido a una inadecuada interpretación de los fenómenos bioeléctricos; este debe ser usado en ciertos diagnósticos diferenciales y en determinados casos [39][40].

El EEG cuantificado puede llegar a ser aún de más utilidad, ya que da el valor preciso de la frecuencia dominante y una medida de similitud entre dos señales que se han registrado desde derivaciones simétricas al mismo tiempo o en tiempos distintos. El principal propósito del EEG computarizado es apoyar las pruebas electroencefalográficas con datos objetivos en forma numérica o gráfica. Se han realizado estudios de análisis espectral mediante el EEG computarizado para hacer el seguimiento del tratamiento con Metilfenidato; en tal caso se han demostrado mejoras en la respuesta eléctrica cerebral, sustituyendo las ondas lentas por ondas rápidas o beta [41].

Registros poligráficos de sueño. Los estudios relacionados al sueño en niños con TDAH aún no han generado datos determinantes que permitan definir el sueño como un marcador biológico para el proceso de maduración del sistema nervioso central. Estos estudios han aportado hallazgos relacionados a la estructura del sueño en estos niños, su relación con los movimientos periódicos durante el sueño, el menor umbral auditivo para el despertar, etc.

En este sentido, en [42][43] Busby señala que la estructura de sueño de los pacientes hipercinéticos no muestra mayor alteración; sin embargo, hallazgos estadísticos significativos refieren alargamiento de la latencia durante la primera fase de sueño y la mayor cantidad absoluta y

relativa de tiempo en movimiento. Además se ha demostrado que durante la fase 2 de sueño, los niños con TDAH no medicados presentan un umbral auditivo de alertamiento más bajo.

Por su parte, Picchietti [44] señala una mayor prevalencia de movimientos periódicos de las extremidades durante el sueño en niños con TDAH con relación a niños sanos. La mayoría de los movimientos periódicos fueron asociados a alertamientos, dados durante un tiempo total de sueño menor que los niños control, con un alargamiento de la latencia de sueño y problemas en el mantenimiento del sueño.

Potenciales evocados de corta latencia. En niños afectos de TDAH se han encontrado alteraciones en los potenciales evocados auditivos de tronco que consisten en incrementos significativos en las latencias de las ondas III y V (generadas en el complejo olivar superior y en el colículo inferior, respectivamente); asimismo los tiempos de transmisión de tallo mostraron incrementos de las ondas I (correspondiente al potencial del nervio auditivo) a la III en niñas y de la I a la V en niños, en comparación con niños normales. Estos hallazgos sugieren la existencia de una conducción asimétrica de los estímulos en el tallo que puede repercutir en el curso del trastorno [33].

Potenciales cognitivos de larga latencia. En niños con TDAH se han descrito anomalías en la amplitud, alargamiento de latencia de la P300 o ambas, probablemente por un retraso de la maduración en estos niños de los procesos de expectación, atención selectiva y elaboración sensorial de los estímulos auditivos que mejorarían con la medicación. Así los potenciales evocados cognitivos pueden tener una gran utilidad clínica en la predicción de la respuesta a psicoestimulantes, especialmente al Metilfenidato, pudiendo utilizarse como test predictivo previo a la introducción a largo plazo de medicación [33].

2.2.4. Prevalencia

Debido a las diferencias entre los criterios diagnósticos empleados, los métodos y las fuentes de información, establecer con exactitud la prevalencia del TDAH puede resultar complejo. De esta forma, la prevalencia de este trastorno se sitúa entre el 1 % y el 24 %, dependiendo de las series y los métodos diagnósticos empleados. Las frecuencias más altas pueden ser observadas en estudios generalmente observacionales basados en entrevistas estructuradas y pobremente confirmadas. Se debe anotar que entre el 40 y 50 % de los varones, son etiquetados por profesores y padres como inquietos o “hiperactivos”.

Finalmente, es importante señalar que la utilización de criterios diagnósticos diferentes al DSM, como por ejemplo el CIE-10, pueden contribuir a esta variabilidad en la prevalencia de esta patología. Incluso, dependiendo de la versión del DSM empleada, se obtienen prevalencias diferentes en una misma población; diferentes estudios parecen reflejar que el DSM-IV-TR es más permisivo que las versiones anteriores, recogiendo mayor prevalencia en

sus estudios en una misma muestra [33].

2.3. Potenciales evocados y cognición

Dado que los procesos cerebrales pueden estar relacionados a movimientos voluntarios y a procesos psicológicos independientes de estímulos, el término *potenciales evocados* no es lo suficientemente general para ser aplicado a todos los fenómenos EEG relacionados a procesos motores-sensoriales. Por otra parte, los eventos psicológicos suficientemente prominentes o distintivos pueden servir como referencias de tiempo para hacer un promedio, además de estímulos y respuestas motoras. El término *potenciales relacionados a eventos* (*event-related potentials* - *ERPs* en inglés) se propuso para designar la clase general de potenciales que muestran relaciones de tiempo estables con un evento de referencia definido. La mayoría de investigaciones en neurociencia cognitiva se refiere a ellos como ERPs, aunque pueden también encontrarse en la literatura referidos como “respuestas evocadas” o “potenciales de respuesta evocada” [23].

Dado que el interés de este trabajo está centrado en la detección del TDAH a partir de registros de potenciales evocados, este estudio apunta al análisis de los potenciales evocados de larga latencia (que en adelante serán llamados ERPs por las razones arriba expuestas), ya que éstos son provocados por la elaboración sensorial del individuo frente al estímulo, cuyas respuestas se relacionan con funciones cognitivas superiores, especialmente la atención.

2.3.1. Alteraciones de los ERPs a causa del TDAH

Desde la década de los 70's los estudios de los ERPs han revelado valiosa información sobre el procesamiento de la información en el TDAH [45][46]. La mayoría de estos estudios se han enfocado en pruebas de funcionamiento de los sistemas de atención visual y auditivo, junto a otros exámenes de procesos preliminares, y en los llamados “procesos ejecutivos” bajo el control del lóbulo frontal, tales como la inhibición.

Mientras muchos estudios comportamentales sugieren que una capacidad reducida [47] o disfuncional [48] de los sistemas de atención no es la mayor causa de los síntomas y/o comportamientos en el TDAH, los estudios con ERPs señalan que varias etapas del procesamiento cognitivo y sensorial son atípicos. Estas aparentes discrepancias pueden ser consecuencia de las diferencias básicas en la metodología y el enfoque, puesto que los estudios comportamentales analizan el resultado de procesos cognitivos internos, mientras que los estudios con ERPs examinan las correlaciones eléctricas de tales procesos cerebrales [5].

Numerosos estudios han puesto de manifiesto la existencia de alteraciones en los ERPs en niños con TDAH, como una disminución de la amplitud y un incremento de la latencia del componente P300 [49][7]. Sin embargo, no todos los niños con TDAH muestran una homogeneidad sintomática. Podemos encontrar niños con una afectación específica de la atención

y con escasa afectación de los otros dos síntomas claves del TDAH, la hiperactividad y la impulsividad [50][51]. La mayor parte de estos estudios se realizaron con grupos heterogéneos de niños con TDAH, sin diferenciar los diferentes subtipos del trastorno según el DSMIV [52].

3. Estado del Arte en la Detección Automática del TDAH

En este capítulo se expone el estado del arte de las técnicas aplicadas al análisis de los potenciales evocados cognitivos, los avances realizados en materia investigativa en el procesamiento digital de estas señales y la adquisición de la información relevante y subyacente de estos registros. Por otra parte, se define brevemente la importancia y los requerimientos básicos para el diagnóstico asistido y los sistemas de identificación automática orientados a la detección de patologías.

3.1. Diagnóstico asistido de patologías

El diagnóstico de enfermedades por parte de los médicos se hace mediante la detección de síntomas u otras señales orgánicas que determinan el estado de funcionalidad del organismo. Para tal fin, los médicos se valen de la percepción sensorial, ayudados con instrumentos de medición que facilitan la emisión de un juicio médico sobre el grado de “normalidad” o “anormalidad” del sujeto estudiado. Aún cuando este estándar de dictamen médico es universalmente aceptado, posee ciertas limitaciones que han llevado a la búsqueda de otras alternativas diagnósticas: la primera de estas restricciones y la de mayor cuidado está relacionada a la alta subjetividad de un diagnóstico al depender de las destrezas y conocimientos del especialista. Asimismo, la alta variabilidad del concepto médico y la imposibilidad de realizar una réplica del protocolo aplicado para un futuro análisis, disminuyen la confiabilidad del diagnóstico. Por otra parte, desde el punto de vista técnico, el hecho de no poder almacenar automáticamente la información que se genera durante un examen médico, representa otro impedimento para realizar un estudio clínico fidedigno de un paciente. De ahí surge la necesidad de utilizar instrumentación avanzada, equipos y otras ayudas tecnológicas que faciliten la exploración y reconocimiento de indicios que sugieran enfermedad y se mejore de manera objetiva la calidad de decisiones de tipo clínico.

3.1.1. Estado del arte

Los esfuerzos realizados para poder aplicar el aprendizaje automático en la resolución de problemas complejos bajo condiciones de incertidumbre y con información faltante no son

recientes. En la década de 1950, un artículo pionero que apareció en la revista Science [53] describió cómo las computadoras podrían ayudar en el proceso diagnóstico. Un software diseñado por Warner et al. en 1961 [54] para el diagnóstico de cardiopatías congénitas alcanzó con éxito un nivel equiparable al de cardiólogos expertos. Sin embargo, el progreso en el desarrollo de sistemas de expertos clínicamente útiles que soporten el diagnóstico médico en general estuvo estancado hasta la década de 1990, debido a expectativas poco realistas de la comunidad médica y las limitaciones de los primeros algoritmos determinísticos de toma de decisiones [55].

Actualmente, los sistemas de apoyo en decisiones están proliferando en muchas áreas de la actividad humana, favoreciendo así la resolución de problemas complejos tales como el diagnóstico médico. Los algoritmos de sistemas expertos comprenden una combinación de sistemas basados en reglas, técnicas de aprendizaje de máquina (*machine learning*) como las redes neuronales artificiales, la teoría de decisión bayesiana, árboles de decisión, máquinas de soporte vectorial y lógica difusa. Algunos de estos sistemas intentan imitar el proceso de decisión humano, subrayando la presunción errónea que el rendimiento del hombre es indefectiblemente óptimo. Otros sistemas proporcionan predicciones basados en el análisis y la minería de datos para ayudar a los médicos en la elaboración de un diagnóstico o a decidir sobre el curso de un tratamiento o terapia [56].

Algunos autores que promueven el intento por emular los procesos de decisión humano se refieren a este como *soft computing*. Afirman que la ingeniería del conocimiento usa técnicas simples de inteligencia artificial para simular las acciones de los expertos. Este planteamiento resulta desconcertante para los investigadores que estudian los procesos de decisión humana, ya que han descubierto muchas deficiencias robustas en el razonamiento humano. Partir de la comprensión de este hecho puede ayudar a identificar estrategias para mejorar el rendimiento humano [57].

A pesar de la proliferación de los sistemas de decisión en el diagnóstico médico, investigaciones han demostrado que entre 98,000 y 115,000 de los pacientes hospitalizados mueren cada año en Estados Unidos debido a algún tipo de error médico. Estos incluyen una amplia gama de eventos adversos de los cuales sólo algunos pueden ser atribuidos a un error humano. Existe cierta controversia en las cifras, así como en la definición y el cálculo del “error evitable”. En una revisión del 2005 sobre 14 estudios de errores médicos publicados entre 1991 y 2004, se encontró que los errores relacionados con el diagnóstico representaron el 30,10 % de todos los errores registrados; otros investigadores han estimado que estos errores representan hasta el 76 % de los errores médicos, y que el *gold standard* de un mal diagnóstico obtenido de autopsias, han producido una tasa de errores en el diagnóstico del 40 % en los últimos 65 años [56]. Las anteriores cifras muestran la inminente necesidad de generar nuevas técnicas en el diseño de sistemas de apoyo para el diagnóstico médico o de refinar las ya utilizadas, a fin de aumentar el grado de confiabilidad de estos sistemas y asegurar el diagnóstico emitido por un médico.

Desde hace algunas décadas se tiene la concepción que los sistemas de ayuda para el diagnóstico médico pueden aumentar considerablemente la precisión del dictamen médico. Por su parte, muchos investigadores de varias disciplinas han apoyado la convicción de que los procedimientos matemáticos podrían superar en confiabilidad las decisiones tomadas por los médicos. En este sentido, Koziellecki [58] afirma que “si el médico está dispuesto a definir el problema clínico en términos de acciones, resultados, probabilidades y utilidades, entonces él podría delegar a un software la tarea de interpretar las consecuencias de sus juicios”. Por esta razón, algunas personas se atreven a afirmar que esta es la era post-doctor o del diagnóstico sin médicos. Pensar en obviar el dictamen médico podría ser contraproducente, pues difícilmente una máquina podrá disponer de todas las habilidades manuales y mentales de un médico, pero con certeza se puede afirmar que sistemas MDSS (*Diagnostic Decision Support System*) pueden proporcionar al médico una valiosa herramienta en la toma de decisiones diagnósticas.

Un sistema de ayuda en el diagnóstico médico está creado bajo las siguientes presunciones:

1. El conocimiento médico es un dominio bien organizado y estructurado de la ciencia y puede ser manipulado de tal manera que podría combinarse sin problema alguno con un programa computacional. Una vez la información necesaria es recogida, el programa producirá las respuestas precisas.
2. El médico mantiene una necesidad de información que un software especializado, matemáticamente apoyado, está en capacidad de proveer.
3. Cuando un médico reconoce los grandes alcances de estas aplicaciones informáticas, donde la calidad de la atención médica ofrecida está en juicio, entonces justificará y favorecerá su uso en la práctica clínica diaria en el menor tiempo posible.

Los anteriores supuestos están sujetos a revisiones y cambios, dado que el conocimiento médico universalmente aceptado es una controversia debido a las considerables diferencias de conceptos de enfermedad entre los distintos países, culturas y regiones. [59]

3.1.2. Requerimientos del diagnóstico asistido

El diagnóstico asistido presenta las siguientes características, que deben ser tenidas en cuenta al momento de diseñar un sistema de diagnóstico asistido [60]:

- *Naturaleza subjetiva.* Dado que es el médico quien juzga la información recogida en un examen o auscultación, la calidad del diagnóstico puede cambiar de especialista a especialista. Un estudio de concordancia entre los dictámenes de varios especialistas se considera aceptable alrededor del 60 – 70 %, mostrando así la disparidad entre los conceptos médicos emitidos por los especialistas en consenso. Lo anterior aporta un

alto grado de incertidumbre que afecta el entrenamiento de los sistemas automáticos de diagnóstico e impiden alcanzar una óptima efectividad.

- *Naturaleza cualitativa.* Gran parte de la información y evidencia recogida por el médico debe ser parametrizada, y a partir de los valores obtenidos se definen unos rangos cualitativos de juzgamiento; sin embargo, existe un amplio margen subjetivo en la escala de calificación final en el dictamen clínico. La generación de características debe acompañarse de una etapa donde se definan valores representativos para cada estado funcional, que cumplan con las condiciones de consistencia y efectividad impuestas sobre los procedimientos de estimación, a efectos de que los resultados puedan ser reproducibles y transportables.
- *Naturaleza Local.* El juicio de normalidad o anormalidad de un estado funcional del cuerpo humano no puede ser generalizado o universal, ya que está sujeto a la locación donde fue realizada la recogida de información. El resultado de un examen clínico puede variar de acuerdo a factores externos como la naturaleza étnica de la población, la actividad humana, el origen social, entre otros, por lo cual se entiende que los rangos de normalidad de un examen no sean los mismo entre países de diferentes continentes. Por otra parte, dado que el organismo es un *sistema complejo* y posee mecanismos dinámicos de compensación, la medición de las variables fisiológicas cambian con el tiempo, debido a la naturaleza no estacionaria del funcionamiento del cuerpo humano.
- *Trascendencia heurística.* La valoración de muchas de las variables diagnósticas usadas aún en la actualidad han sido desarrolladas de forma empírica, basadas en la percepción sensorial del médico, y por lo tanto no son fáciles de modelar ni parametrizar. Este aspecto dificulta la generación de modelos automáticos de diagnóstico, ya que siempre es preferible que la característica de análisis tenga un sentido físico y sea de fácil comprensión al personal médico.

3.1.3. Diseño de un sistema automático de diagnóstico

Desde el punto de vista técnico también se deben tener en cuenta los siguientes aspectos para el diseño de un sistema automático de diagnóstico [60]:

- *Modelos irregulares de representación.* En la práctica médica es frecuente que a varios estados de normalidad o anormalidad del cuerpo humano, le correspondan modelos idénticos de representación; es decir, el paciente se diagnostica por un único estado de funcionalidad, pero en realidad presenta varias irregularidades, o viceversa, el diagnóstico entregado presenta varias disfuncionalidades cuando realmente sólo presenta una de ellas.
- *Asimetría en la distribución poblacional de las clases.* Una dificultad común que se presenta al entrenar sistemas automáticos de detección de patologías es que la población

en general no presenta igual número de individuos de las clases evaluadas. A lo anterior, se suma el inconveniente de la consecución del número suficiente de personas por clase para el aprendizaje de las máquinas. Como resultado surge la necesidad de aprovechar al máximo la información brindada por cada observación clínica, o de generar estrategias de registro que impliquen la consecución y extracción de mayor información sobre cada caso.

- *Orientabilidad del diagnóstico.* De acuerdo a la categorización por niveles del servicio de salud en nuestro país, los centros de atención primaria, por su ubicación en zonas remotas y de difícil acceso a tecnologías avanzadas, están orientados a la atención de requerimientos básicos de salubridad; por tal motivo, para estos centros es preferible el desarrollo de sistemas de diagnóstico automáticos que apunten al tamizaje y a la prevención de enfermedades, con énfasis en la detección con alta precisión de los estados de funcionamiento anormal. Por su parte, los centros de tercer y cuarto nivel ubicados en las grandes ciudades, que presentan menores restricciones en cuanto a la dotación de equipos, exigen de los sistemas automáticos un mayor énfasis en la diferenciación con alta precisión entre varios estados de anormalidad de naturaleza muy similar.
- *Complejidad y alto costo.* La implementación de un sistema automático de diagnóstico supone el desarrollo y aplicación de un riguroso protocolo de registro y procesamiento de datos que generan altos costos. En primera instancia, la adquisición de las señales implica gastos relacionados a la ubicación, transporte y preparación de los pacientes a quienes se les tomarán las muestras (bioseñales), además de los gastos que se puedan generar en la consecución de los equipos médicos necesarios para tal fin. Asimismo, el etiquetamiento de los registros y el acompañamiento de los especialistas durante el desarrollo del aplicativo, también adicionan una carga considerable al presupuesto de una investigación. Por otra parte, la elaboración y sintonización de los algoritmos de proceso requiere de personal calificado y máquinas de alto rendimiento que igualmente debe ser tenido en cuenta en el importe del proyecto investigativo.
- *Normatividad en el área de la salud.* Es importante tener en cuenta las restricciones y limitaciones de la normatividad colombiana en el sector de la salud al momento de plantear los alcances del desarrollo de un sistema de detección automática. A saber, la resolución número 1995 de 1999 establece las normas para el manejo de la historia clínica, incluyendo su forma de presentación electrónica. La transmisión y almacenamiento de bioseñales e imágenes médicas tiene como estándar de facto la norma DICOM, y así sucesivamente.

3.1.4. Detección automática de TDAH

Si bien, el número de estudios relacionados a los potenciales evocados cognitivos se ha incrementado substancialmente en las últimas décadas, permitiendo avanzar en el conocimiento de

su complejidad y amplia aplicabilidad, también es cierto que existen ciertas áreas específicas de estudio que apenas si han sido exploradas, como lo es el caso de la detección automática de patologías mediante el análisis de ERPs. Son pocos los documentos de carácter científico que al respecto han sido publicados, lo que supone una gran área aún por estudiar por parte del procesamiento digital de señales y el análisis automático. Algunos de los pocos documentos reportados en el estado del arte son los siguientes:

En [61] se propone un sistema de apoyo en decisión (*Decision Support System* - DSS) para la clasificación de pacientes a partir de sus potenciales relacionados a eventos. Este sistema consta de dos etapas: en la primera, se hace la extracción de características mediante la implementación de un modelo de auto-regresión en conjunto con un método de optimización global, para la selección de las características óptimas de los ERPs. La tarea de clasificación es llevada a cabo con una red neuronal de tres capas simples, entrenada con un algoritmo de propagación hacia atrás. El DSS ha sido probado en datos de pacientes con desórdenes obsesivo-compulsivo, depresivos y drogadictos, alcanzando porcentajes de clasificación superiores al 95 %.

Los siguientes artículos presentan un enfoque para el diagnóstico automático del TDAH y otros desórdenes neurológicos, pero no basado en las señales ERPs sino a partir de su sintomatología. Por una parte, Delavarian et al. propone en [62] un enfoque que hace la clasificación automática de diferentes desórdenes comportamentales que presentan sintomatología semejante, que sirve como una útil herramienta para los psiquiatras al momento de realizar el diagnóstico de un paciente. El estudio consiste en diferenciar y clasificar 306 niños con síntomas muy similares pero con diferentes desórdenes de comportamiento diagnosticados como TDAH, depresión, ansiedad y comorbilidades de las dos anteriores. La clasificación se basó en los síntomas (y su severidad) más significativos de cada patología, y formando con ellos el vector de características, se probó con 16 diferentes clasificadores, alcanzado con uno de ellos una precisión de clasificación de un 96,92 %.

De igual manera, en un estudio similar, Delavarian et al. [63] aplica el mismo enfoque trabajado en [62] y realiza la clasificación de niños con diferentes desórdenes de comportamiento basándose en su sintomatología. En esta ocasión, utiliza redes neuronales de función de base radial y multicapa, alcanzado rendimientos de clasificación del 96, 62 % y 95,5 %, respectivamente.

Como puede observarse, los resultados obtenidos en algunos de los trabajos mencionados son superiores a los alcanzados en esta investigación (ver Capítulo 8); esto se debe principalmente a las diferencias en los protocolos y experimentos llevados a cabo para la adquisición de los registros de potenciales evocados cognitivos, dado que tales discrepancias generan marcados contrastes en las formas de onda de los ERPs [64]. Por otra parte, en los artículos previamente mencionados, fueron utilizadas las redes neuronales como clasificadores, lo que explica el alto porcentaje de acierto alcanzado, dada la gran recursividad y aplicabilidad de este prototipo de sistemas. Sin embargo, vale la pena resaltar que el objetivo principal de esta tesis está

enfocado a la exploración de la calidad de las características y no se concentra enfáticamente en la etapa de clasificación, prueba de ello es el uso de clasificadores básicos como se verá en el capítulo que expone los resultados obtenidos.

3.2. Técnicas aplicadas al análisis de los potenciales evocados cognitivos

Los potenciales evocados son definidos como cambios de voltaje en el curso del EEG debidos a un estímulo externo. El hecho de ser respuestas embebidas en el electroencefalograma y de presentar una pequeña amplitud presume la necesidad de aplicar técnicas de preprocesamiento para la extracción de los ERPs y el uso de otras herramientas matemáticas para el análisis de los mismos. A continuación se presenta un breve resumen del estado del arte de las principales técnicas empleadas en el análisis de los potenciales evocados cognitivos.

3.2.1. Análisis mediante distribuciones tiempo - frecuencia

La transformada wavelet (*wavelet transform* - *WT*) es una útil herramienta de uso común en el análisis de potenciales relacionados a eventos. Mediante la representación tiempo-frecuencia, los finos detalles de la estructura de los ERPs pueden ser examinados a una mayor escala; en particular, permite una resolución en tiempo teóricamente ilimitada para la detección de picos de corta duración, y admite una flexible selección de funciones bases wavelet para el análisis de diferentes tipos de ERPs. Entre las aplicaciones más comunes del análisis wavelet en los ERPs se encuentran:

- la ubicación precisa en el tiempo de la ocurrencia de solapamiento entre picos en respuestas evocadas auditivas de tallo cerebral;
- la extracción de registros ERPs individuales (sin promediar) a partir del ruido de fondo del EEG;
- la descomposición del ERP promediado en funciones de detalle ortogonal que desliga el comportamiento experimental de la forma de onda en bandas de frecuencia ortogonales;
- el uso de los coeficientes wavelet extrae de forma concisa información importante de los ERPs que permiten predecir el rendimiento de la detección de las señales humanas.

Eliminación de ruido EEG

Aunque el tradicional método de promediación de registros individuales para la extracción de los ERPs da estructura a la forma de onda, este procedimiento presume la pérdida de información que podría representar componentes adicionales en el registro de la señal. Por

tal motivo, en la última década muchos investigadores se han enfocado en la búsqueda de otras alternativas que permitan la eliminación de ruido EEG con la menor pérdida de información posible. Partiendo desde este punto de vista, en [65] la definición convencional de eliminación de ruido implica un criterio de umbralización en el dominio wavelet. La señal se recupera de los datos con ruido poniendo en cero los coeficientes wavelet que están por debajo de cierto umbral o con el uso de una transformación más suave. Sin embargo, este procedimiento no es óptimo para extraer los potenciales evocados dado que éstos son de un orden mucho menor que el EEG en curso. Para contrarrestar esta limitación, en trabajos posteriores se ha planteado la eliminación de ruido basado en la localización específica en tiempo y en frecuencia de las respuestas evocadas.

En este orden de ideas, en [66] y [67] se propone una metodología basada en un esquema de descomposición wavelet, a fin de obtener las señales de EPs a partir de la señal EEG en curso. El método permite la visualización de un único registro de potenciales evocados, haciendo énfasis en la variabilidad de su amplitud y latencia entre una prueba y otra. Básicamente, el procedimiento propuesto consiste, en primera instancia, en descomponer el EP promediado (en la literatura conocido como *grand average*) usando una descomposición wavelet multiresolución, y aquellos coeficientes no correlacionados con tal registro promedio son puestos en cero. Posteriormente se aplica la transformada wavelet inversa para obtener la señal sin ruido. Dado que el método mostrado es no-paramétrico y no debe ser adaptado a características particulares de cada registro, puede ser de gran utilidad en la práctica clínica.

Diferenciación de respuesta a estímulos *target* y *no-target*

Los potenciales relacionados a eventos son especialmente utilizados en *brain-computer interfaces (BCI)* (interface cerebro - computador), el cual consiste en un sistema de comunicación que traduce la actividad cerebral en comandos para una computadora u otros dispositivos. En busca de este objetivo, la WT ha sido aplicada en la distinción de respuesta a estímulos *target* y a estímulos frecuentes. En [68] se usa como wavelet madre la función B-spline para descomponer los ERPs en cuatro octavas. El análisis realizado indica que varios componentes de las frecuencias delta, theta y alpha en el rango de latencia del pico P300 difiere en latencia y amplitud entre las respuestas *target* y las *no-target*. Este hallazgo demuestra que la onda P300 está compuesta de múltiples componentes funcionales que predominan ante un estímulo infrecuente. Estos resultados se confirman también en [69], donde se usa un filtrado digital adaptativo en respuestas P300 (obtenidas mediante el paradigma *oddball*) para generar las respectivas bandas de frecuencia: la respuesta delta domina en los ERPs, mientras la respuesta theta se muestra prolongada en una segunda ventana tardía.

Por su parte, en [70] se presenta la aplicación de una medida de entropía definida a partir de la transformada wavelet, denominada entropía wavelet, para el estudio de los ERPs. La WT es calculada sobre los registros ERPs de sujetos sanos con tres diferentes tipos de estímulos. Un decrecimiento significativo de la entropía fue correlacionado con las respuestas a estímu-

los *target* (infrecuentes o raros), mostrando así que tales respuestas corresponden a estados más “ordenados” que el registro espontáneo de EEG. Además se propone la entropía wavelet como una medida cuantitativa para tal transición entre el EEG (“estado desordenado”) y el ERP (“estado ordenado”).

En [71] Aviyente et al. presentan un nuevo enfoque para el análisis de ERPs basado en distribuciones tiempo-frecuencia y medidas de información. La base de datos de ERPs adquirida pertenece a dos grupos de sujetos: aracnofóbicos (fobia a las arañas) y sin fobias diagnosticadas. El método propuesto radica en la cuantificación (sobre representaciones tiempo-frecuencia) de las diferencias de la actividad cerebral antes y después de la percepción de un estímulo. Para tal fin, se introduce una nueva medida de distancia con base en la teoría de la información, que depende de la entropía de Rényi. Los resultados ilustran la eficiencia de usar medidas de distancia combinadas con distribuciones tiempo-frecuencia en la diferenciación de dos clases de sujetos y las diferentes regiones cerebrales.

Detección de componentes en los ERPs

En [72] se presenta una aplicación del filtrado wavelet para el análisis del componente P300 en registros ERPs individuales dada una tarea cognitiva (en este caso en particular, el tiempo de reacción entre dos elecciones). Para tal objetivo se emplea la transformada wavelet discreta con la función Daubechies como wavelet madre. El filtrado wavelet se usa para la remoción de ruido y de componentes frecuenciales no deseados; mientras que los coeficientes wavelet del *grand average* de todos los sujetos participantes se utilizan para computar la máscara de filtrado. Los hallazgos muestran las bondades de tal máscara wavelet aplicada en épocas de registros individuales de ERPs, la cual proporciona una buena localización de los componentes de los potenciales relacionados a eventos tanto en tiempo como en escala. Dadas las limitaciones producto de la naturaleza no dinámica del filtro wavelet diseñado, se sugiere la aplicación de filtros adaptativos.

Parte III.

Marco Teórico

4. Caracterización y Preprocesamiento de Datos

En el presente capítulo se define el espacio inicial de características utilizado como patrones descriptivos de los potenciales evocados cognitivos. Inicialmente, se hace un breve estado del arte del conjunto de parámetros que fue particularmente empleado en este trabajo. Posteriormente, se hace la descripción del algoritmo de localización del componente P300, necesario para definir el grupo de características morfológicas. Además, se definen las características espectrales (de frecuencia y coeficientes wavelet) enmarcadas dentro de la teoría de la transformada wavelet. Por último, se explican los procedimientos llevados a cabo durante el preprocesamiento de las señales y los datos.

4.1. Caracterización de la señal ERP

Aunque el estado del arte no define un conjunto de características típicas a medir sobre la forma de onda de la señal de ERP (como sí sucede en otras bioseñales como ECG - electrocardiograma - y voz), para este trabajo se ha convenido la utilización de características que han mostrado un óptimo desempeño en estudios previos relacionados al análisis de los potenciales evocados cognitivos.

Por un lado, en [73], estas características fueron utilizadas para la discriminación de pacientes depresivos y sujetos sanos, usando el componente P600 de la señal de ERP. Estos parámetros fueron también empleados en el diseño de un clasificador SVM (máquina de soporte vectorial) óptimo para cada carga (electrodo), y por voto mayoritario se asignaba a cada sujeto la clase de pertenencia. El máximo desempeño de clasificación alcanzado fue de 94 % usando todas las cargas, y del 92 % y 84 % usando las cargas del lado izquierdo y derecho de la cabeza, respectivamente.

Por otra parte, en Abootalebi et al [74]. utiliza estas mismas características en la identificación de sujetos culpables e inocentes, basados en la detección de la onda P300 bajo el paradigma del test de conocimiento de culpabilidad (*Guilty Knowledge Test - GKT*). Además de las características morfológicas, en este artículo se adicionan coeficientes wavelet y características frecuenciales de señal al conjunto de parámetros empleados en [73]. De esta manera, se obtuvo una tasa de acierto en la detección de culpables e inocentes de un 86 %.

Es importante mencionar que varios de estos parámetros fueron aplicados según indican los artículos citados, aunque algunos otros han sido redefinidos a partir de un criterio médico (de los especialistas consultados) de evaluación visual de los ERPs.

4.1.1. Características morfológicas

El primer grupo de características consta de 16 parámetros medidos sobre un segmento de la señal, los cuales están relacionados a la distribución en el tiempo de la forma de onda (morfología de la bioseñal). La mayoría de ellos corresponden a áreas bajo el segmento de señal, distancias entre determinados puntos, medición de las pendientes formadas por los componentes de los ERPs, entre otros.

Antes de entrar a definir los parámetros empleados en la caracterización de los ERPs, es necesario hacer claridad en otro aspecto del cual dependen tales características: la identificación y localización del componente P300.

Localización del componente P300

Dado que varias de las características utilizadas en este trabajo de investigación dependen directamente de la localización de la onda P300, la correcta identificación y ubicación de todos los componentes de la señal (P100, N200, P300, etc) es una tarea determinante a la hora de caracterizar los registros de ERPs. Teóricamente, la latencia se define como el tiempo que tarda en aparecer el mayor valor de la señal después del estímulo en una ventana de tiempo de interés. Dado que tal ventana se centra en los 300 ms, en muchos casos ese valor coincide con el componente P300 [4]; sin embargo, varios autores refieren que esta onda no necesariamente se presenta en el mismo instante de tiempo, dado que la latencia puede presentar alteraciones debidas a algún tipo de desorden neurológico [32].

Por otra parte, los especialistas consultados para etiquetar los registros, señalan que no en todos los casos el P300 corresponde al pico más cercano a los 300 ms, y que en ocasiones su ubicación obedece a otros criterios, como por ejemplo el número de componentes que la precede, amplitud de tales ondas, profundidad del posible componente P300, entre otros. La aplicación de estos criterios en el análisis visual de los ERPs puede variar de un especialista a otro y alterar el juzgamiento médico del paciente, lo cual amplía el margen de subjetividad y variabilidad de diagnóstico.

En respuesta a esta restricción, en este trabajo de investigación se propone un algoritmo para la localización del pico P300 basado en la derivada de la señal, aplicada en una ventana de tiempo, a fin de automatizar la identificación del punto P300 de acuerdo a criterios médicos reales.

En general, el algoritmo propuesto consta de las siguiente etapas. Primero, se establece una ventana de análisis $T = (t_1, t_2)$ que contenga el componente ERP de interés, definiendo

así un segmento de señal denotado como $\tilde{s}(t)$. Posteriormente, se localizan los mínimos en $\tilde{s}(t)$ representados por $\mathbf{p} = [p_1, \dots, p_M]$, y se calcula la señal derivada $\dot{s}(t)$. Después, se ubican todos los máximos y mínimos de $\dot{s}(t)$; tales puntos serán representados por $\tilde{\mathbf{p}} = [\tilde{p}_1, \dots, \tilde{p}_N]$. Los puntos \mathbf{p} son localizados sobre $\dot{s}(t)$ con el fin de determinar la pendiente asociada. Finalmente, para cada p_i se calcula la distancia entre los puntos de $\tilde{\mathbf{p}}$ de $\dot{s}(t)$ que inmediatamente lo anteceden y preceden. Así, el componente P300 se determina como el mínimo de $s(t)$ asociado a la mayor distancia obtenida en $\dot{s}(t)$.

El anterior procedimiento se describe gráficamente en la Figura 4-1 y de manera matemática en el Algoritmo 4.1.1.

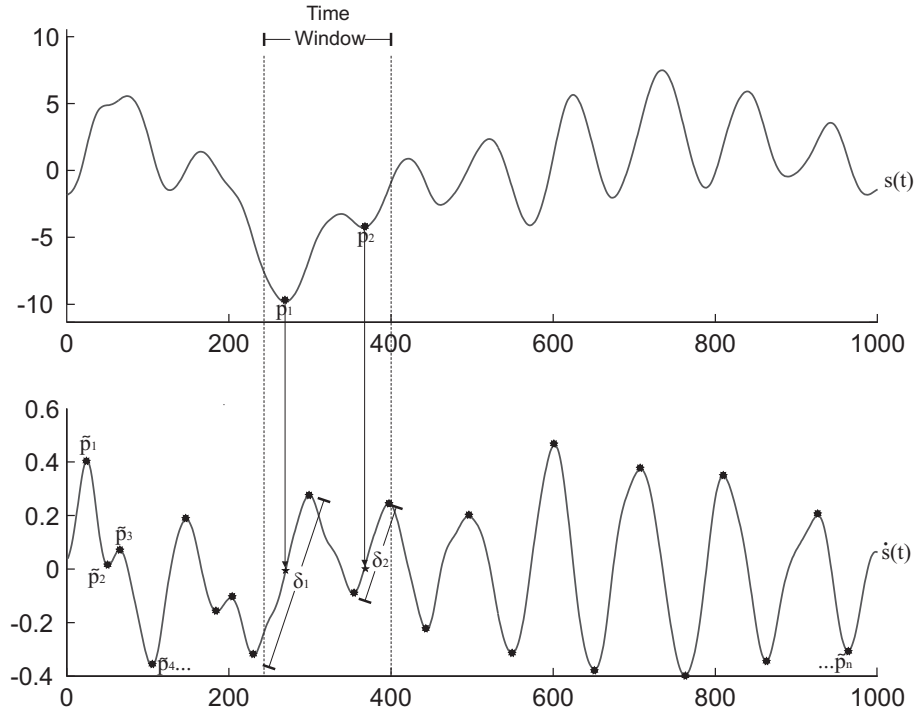


Figura 4-1.: Descripción gráfica del algoritmo de localización del P300.

Definición de las características morfológicas

A continuación se hace la definición matemática de las características morfológicas empleadas en la parametrización de las señales ERPs, que fueron calculadas a partir de la identificación del punto P300. Es importante mencionar que todos los parámetros relacionados a la morfología de la señal fueron estimados dentro de una ventana de t_a, t_b

1. *Latencia* (LAT, t_{sP300}). Se define como el tiempo en el que aparece la onda P300 previamente identificada:

$$t_{sP300} = \{t | s(t) = s_{P300}\} \quad (4-1)$$

Algoritmo 1 Algoritmo de localización del componente P300

Dada una señal $s(t)$.

1. Seleccionar una ventana de análisis $T = (t_1, t_2)$.
2. Localizar los mínimos locales $\mathbf{p} = [p_1, \dots, p_M]$ de $s(t)$ dentro del intervalo (t_1, t_2) , donde M es el número de picos detectados en tal intervalo.
3. Calcular $\dot{s}(t)$ como la derivada de la señal $s(t)$.
4. Localizar en $\dot{s}(t)$ todos los picos (máximos y mínimos) $\tilde{\mathbf{p}} = [\mathbf{p}_1, \dots, \mathbf{p}_N]$, donde N es el total del número de picos de $\dot{s}(t)$.
5. Determinar la localización temporal de todos los puntos de \mathbf{p} en $\dot{s}(t)$.
6. Para cada localización en el tiempo ejecutada en el paso 5, calcular la distancia euclidiana entre los picos inmediatamente anterior y posterior, como sigue:

$$\delta = \sqrt{(\tilde{s}(\tilde{p}_j) - \tilde{s}(\tilde{p}_{j-1}))^2 + (\tilde{p}_j - \tilde{p}_{j-1})^2}$$

7. Conformar el vector de distancias $\Delta = \{\delta_i, i = 1, \dots, M\}$.
 8. Escoger el componente P300 como el $\max_i\{\Delta\}$.
-

donde $s(t)$ es la señal ERP en la ventana de tiempo definida durante la localización del P300, y s_{P300} es el valor de la señal en el punto P300.

2. *Amplitud* (AMP, s_{P300}). Se refiere al valor de la señal en el punto identificado como P300.

$$AMP = s(P300) \tag{4-2}$$

3. *Razón latencia/amplitud* (LAR).

$$LAR = \frac{t_{s_{P300}}}{s_{N200-P300}} \tag{4-3}$$

4. *Amplitud absoluta* (AAMP).

$$AAMP = |s_{N200-P300}| \tag{4-4}$$

5. *Área positiva* (PAR, A_p). Se define como la suma de los valores positivos de la señal:

$$A_p = \frac{1}{2} \sum_{t=t_a}^{t_b} (s(t) + |s(t)|) \tag{4-5}$$

donde (t_a, t_b) es la ventana de análisis.

6. *Área negativa* (NAR, A_n). Valores negativos de la señal:

$$A_n = \frac{1}{2} \sum_{t=t_a}^{t_b} (s(t) - |s(t)|) \quad (4-6)$$

7. *Área total* (TAR, A_{pn}).

$$A_{pn} = A_p + A_n \quad (4-7)$$

8. *Área total absoluta* (ATAR).

$$ATAR = |A_{pn}| \quad (4-8)$$

9. *Área absoluta total* (TAAR, $A_{p|n|}$).

$$A_{p|n|} = A_p + |A_n| \quad (4-9)$$

10. *Pendiente de la señal absoluta promedio* (AASS, $|\bar{s}|$).

$$|\bar{s}| = \frac{1}{n} \sum_{t=t_a}^{t_b-\tau} \frac{1}{\tau} |s(t+\tau) - s(t)| \quad (4-10)$$

donde τ es el intervalo de muestreo de la señal y n es el número de muestras de la señal digital.

11. *Valor pico a pico* (PP, $s_{N200-P300}$). Definido como la diferencia de voltaje (valor pico a pico) entre el pico P300 y el componente N200. La onda N200 se define como la positividad con amplitud significativa inmediatamente anterior al componente P300.

$$s_{N200-P300} = s_{N200} - s_{P300} \quad (4-11)$$

12. *Ventana de tiempo pico-pico* (PPT, t_{pp}). Es el tiempo transcurrido entre la aparición de la onda N200 y la P300.

$$t_{pp} = t_{N200} - t_{P300} \quad (4-12)$$

13. *Pendiente pico-pico* (PPS, \dot{s}_{pp}).

$$\dot{s}_{pp} = \frac{s_{N200-P300}}{t_{pp}} \quad (4-13)$$

14. *Cruces por cero* (ZC, n_{ZC}). Es el número de veces t que $s(t)=0$ en la ventana de tiempo previamente definida.

$$n_{ZC} = \sum_{t=t_{sN200}}^{t_{sP300}} \delta_s \quad (4-14)$$

donde t_{sN200} y t_{sP300} son los tiempos en los que aparece la N200 y la P300, respectivamente; y $\delta_s = \delta(s(t))$, donde $\delta(\cdot)$ es un delta de Dirac, es decir, $\delta_s = 1$ cuando $s(t) = 0$ y $\delta_s = 0$ en otro caso.

15. *Densidad de cruces por cero* (ZCD, d_{ZC}). Es el número de cruces por cero por unidad de tiempo en la ventana de análisis.

$$d_{ZC} = \frac{n_{ZC}}{t_{pp}} \quad (4-15)$$

16. *Alteraciones de signo de las pendientes* (SSA, n_{sa}). Es el número de alteraciones de signo de las pendientes de dos puntos adyacentes de la señal ERP.

$$n_{sa} = \frac{1}{2} \sum_{t=t_a+\tau}^{t_b-\tau} \left| \frac{s(t-\tau) - s(t)}{|s(t-\tau) - s(t)|} + \frac{s(t+\tau) - s(t)}{|s(t+\tau) - s(t)|} \right| \quad (4-16)$$

4.1.2. Características frecuenciales de la señal

En este apartado se definen las características frecuenciales de los ERPs. Los siguientes son parámetros medidos sobre el espectro de potencia de la señal:

1. *Moda de la frecuencia* (f_{mode}). Es la frecuencia con la mayor concentración de energía en el espectro de la señal; así, la máxima amplitud de la densidad del espectro de potencia de la señal está en tal frecuencia:

$$S(f_{mode}) = \text{Max}_f \{S(f)\} \quad (4-17)$$

donde S es la densidad espectral de potencia de la señal y f es la frecuencia en Hz.

2. *Mediana de la frecuencia* (f_{median}). La mediana de la frecuencia separa el espectro de la potencia en dos áreas bajo la curva iguales. Entonces, el valor de f_{median} satisface:

$$\int_0^{f_{median}} S(f) df = \int_{f_{median}}^{\infty} S(f) df \quad (4-18)$$

3. *Media de la frecuencia* (f_{mean}). Representa el centroide del espectro y se calcula a partir del promedio ponderado de las frecuencias en la densidad espectral de potencia

de la señal. El valor de f_{mean} puede calcularse como sigue:

$$f_{mean} = \frac{\int_0^{\infty} f \cdot S(f) df}{\int_0^{\infty} S(f) df} \quad (4-19)$$

4.1.3. Características basadas en transformada wavelet

Un tercer grupo de características está conformado por los coeficientes wavelet de la señal. Esta forma de parametrización está basada en la transformada wavelet discreta y el análisis multi-resolución de la forma de onda. A continuación, se hace una breve documentación de esta útil herramienta y sus ventajas sobre otro tipo de transformaciones.

Transformada wavelet

Una señal puede ser considerada como estacionaria si sus características estadísticas no cambian con el tiempo. Las señales estacionarias pueden ser analizadas usando métodos clásicos como la transformada de Fourier (FT), en los cuales la señal se puede expandir en las funciones de base ortogonal (ondas seno y coseno). Esto permite una mejor visualización de las periodicidades de la señal, especialmente cuando varias frecuencias están superpuestas; sin embargo, la FT no da información sobre la localización temporal de tales periodicidades, lo cual supone la asunción de estacionariedad de la señal. La mayoría de señales biomédicas son no estacionarias y tienen características tiempo-frecuencia altamente complejas.

En la práctica, la condición de estacionariedad para las señales no estacionarias puede ser alcanzada dividiendo la señal en bloques de segmentos cortos en los que el segmento de la señal puede ser asumido como estacionario. Este método, llamado transformada de Fourier de tiempo corto (STFT), fue propuesto por Gabor en 1946. La STFT da una óptima representación tiempo-frecuencia, pero presenta una limitación debida al principio de incertidumbre: la longitud del segmento deseado. Elegir una ventana de análisis pequeña puede causar una pobre resolución en la frecuencia; por otra parte, una ventana de análisis grande puede mejorar la resolución en frecuencia pero compromete la suposición de estacionariedad dentro de la ventana. Por tal motivo, la STFT no es aconsejable para el análisis de señales que involucren diferentes rangos de frecuencias, como es el caso de los ERPs [67].

En 1984, Grossmann y Morlet [75] introducen la transformada wavelet para superar esta restricción. La principal ventaja de las wavelet es su variabilidad en el tamaño de las ventanas deslizantes, siendo anchas para las frecuencias bajas y angostas para las altas; de esta manera se lleva a una óptima resolución tiempo-frecuencia en todos los rangos de frecuencia. Dado que las ventanas se adaptan a los transitorios de cada escala, las wavelet no requieren la condición de estacionariedad.

Transformada wavelet continua La transformada wavelet continua (*Continuous Wavelet Transform - CWT*) de la señal $x(t) \in L^2(\mathbb{R})$ está definida como el producto interno entre la señal y la función wavelet $\psi_{a,b}(t)$ [76] [77]:

$$CWT(a, b) \equiv C_{a,b} = \langle x(t), \psi_{a,b}(t) \rangle = |a|^{-1/2} \int_{-\infty}^{\infty} x(t) \psi * \left(\frac{t-b}{a} \right) dt \quad (4-20)$$

donde $C_{a,b}$ son los coeficientes wavelet y $\psi_{a,b}(t)$ son las versiones trasladadas y dilatadas (o contraídas) de una única función wavelet $\psi_{a,b}(t)$ denominada *wavelet madre* (el resto de funciones generadas se denominan simplemente *wavelets*), que se define como:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi * \left(\frac{t-b}{a} \right) \quad (4-21)$$

donde a, b son los parámetros de escala y traslación en el tiempo, respectivamente. El factor $|a|^{-1/2}$ se introduce con el fin de asegurar que todas las funciones escaladas $|a|^{-1/2} \psi\left(\frac{t-b}{a}\right)$, con $a \in \mathbb{R}$, tengan la misma energía.

Puesto que la función de análisis $\psi(t)$ se escala y no se modula como la función de ventaneo de la STFT, el análisis wavelet se llama tiempo-escala y no tiempo-frecuencia. Una variación en el retraso de tiempo b y/o en el parámetro de escala a no tiene efecto en la forma del *kernel* de transformación de la transformada wavelet. Sin embargo, la resolución en tiempo y frecuencia de la transformada wavelet depende de a . Así, para frecuencias altas (a pequeño) se tiene una buena localización en el tiempo, pero una pobre resolución en frecuencia. Por otro lado, en frecuencias bajas, se tiene buena resolución en frecuencia, pero baja resolución en el tiempo.

La localización en frecuencia es logarítmica, es decir, proporcional al nivel de frecuencia. Como consecuencia, la localización en el tiempo es más detallada para las frecuencias más altas tal como se puede observar en la Figura 4-2. La gráfica muestra como la resolución de la CWT en el tiempo es buena para frecuencias altas, pero como el ancho de banda de la wavelet madre se hace grande para estas frecuencias, la resolución en frecuencia no es buena. Y en las frecuencias bajas, la wavelet madre es mucho más ancha en el tiempo y tiene una concentración en frecuencias altas, lo cual permite identificar correctamente componentes de baja frecuencia con el costo de una resolución menor en el tiempo.

Transformada wavelet discreta (DWT) La CWT mapea una señal de una variable independiente t a una función de dos variables independientes (a, b) . Este procedimiento es redundante y poco eficiente al momento de implementar algoritmos. En consecuencia, es

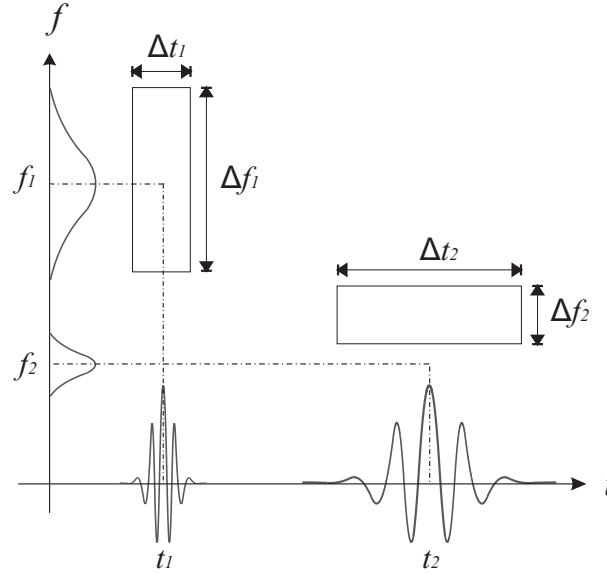


Figura 4-2.: Resolución de la transformada wavelet.

más práctico definir la transformada wavelet sólo en escalas discretas a y tiempos discretos b , empleado así una rejilla diádica donde a toma el valor de 2^j . En este caso la transformada viene dada por [78]:

$$CWT(2^i, b) = |2^i|^{-1/2} \int_{-\infty}^{\infty} x(t) \psi \left(\frac{t-b}{2^i} \right) dt \quad (4-22)$$

Cada i se denomina octava o escala, y consiste en cada uno de los niveles en los que se descompone la señal. Las escalas bajas tienen en cuenta las frecuencias bajas, y las frecuencias mayores son tenidas en cuenta por las escalas altas.

Cuando se usan wavelets ortonormales (desde el punto de vista de las funciones de cuadrado integrable L^2), lo habitual es usar un procedimiento denominado *decimation* (diezmar). Consiste en descomponer la señal en un número de coeficientes proporcional a la escala analizada. Esto hace que la señal tenga distinto número de coeficientes en cada escala. Físicamente esto refleja el hecho de que las frecuencias menores de una señal necesitan menos coeficientes para ser representadas. Una Transformada de Wavelet diezmada es:

$$CWT(2^i, 2^i \cdot n) = |2^i|^{-1/2} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t}{2^i} - n \right) dt \quad (4-23)$$

A partir de la anterior ecuación, es posible definir la versión discretizada de la transformada

wavelet, como sigue:

$$DWT(2^i, 2^i \cdot n) = |2^i|^{-1/2} \sum_k x(k) \psi^* \left(\frac{k}{2^i} - n \right) \quad (4-24)$$

Análisis multi-resolución El análisis multi-resolución (MRA) es una herramienta que hace uso de la DWT para representar una señal de tiempo variante en términos de sus componentes de frecuencia. Esencialmente, este método consiste en mapear una señal unidimensional del tiempo a una señal bidimensional de tiempo y frecuencia.

La idea de MRA es aproximar una señal $s(n)$ como un límite de aproximaciones sucesivas; las diferencias de las aproximaciones sucesivas entre la resolución 2^{j-1} y 2^j dan los detalles de la señal en la resolución 2^j . Lo anterior se refiere a que, después de elegir una resolución inicial J cualquier señal $s(n) \in L^2(\mathbb{R})$ puede ser expresada como [79]:

$$s(n) = \sum_{k \in \mathbb{Z}} c_{J,k} \varphi_{J,k}(n) + \sum_{j=J}^{\infty} \sum_{k \in \mathbb{Z}} d_{j,k} w_{j,k}(n) \quad (4-25)$$

donde los detalles o coeficientes de wavelets $\{d_{j,k}\}$ están definidos por:

$$d_{j,k} = 2^{-j/2} \int_{-\infty}^{\infty} s(n) w_{j,k}(2^{-j}n - k) dn \quad (4-26)$$

las aproximaciones ó coeficientes de escala $\{c_{j,k}\}$ se expresan de la siguiente manera:

$$c_{j,k} = 2^{-j/2} \int_{-\infty}^{\infty} s(n) \varphi_{j,k}(2^{-j}n - k) dn \quad (4-27)$$

De esta manera es como se lleva a cabo el MRA de la señal $s(n)$, expresada en términos de los coeficientes de detalle y de los coeficientes de aproximación.

El algoritmo empleado en este trabajo para hacer MRA consta de las siguientes etapas [80]: Dada una señal s de longitud n , la DWT consta máximo de $\log_2 n$. La primera etapa produce, empezando a partir de s , dos conjuntos de coeficientes: los coeficientes de aproximación CA1, y coeficientes de detalle CD1. Estos vectores son obtenidos por la convolución de s con un filtro pasa-bajo LoD para obtener los coeficiente de aproximación, y con un filtro pasa-alto HiD para los coeficiente de detalle, seguido de una decimación diádica (*downsampling*). Este primer paso se muestra de forma gráfica en la figura 4-3.

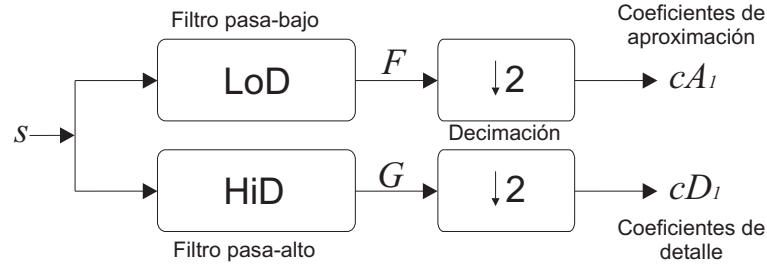


Figura 4-3.: Análisis multiresolución de una señal

La longitud de cada filtro es igual a $2n$. Ahora, si $N = \ell(s)$ (dado que s cambiará de longitud después de cada decimación), entonces las señales F y G son de tamaño $N + 2n - 1$ y los coeficientes cA_l y cD_l son iguales a $\left(\frac{N-1}{2}\right) + n$, aproximando $\left(\frac{N-1}{2}\right)$ al entero más cercano.

En la siguiente etapa se dividen los coeficientes de aproximación cA_l en dos partes usando el mismo esquema previamente explicado, pero reemplazando a s por cA_l , generando así un nuevo nivel de descomposición con coeficientes de aproximación cA_2 y coeficientes de detalle cD_2 , como se muestra en la figura 4-4 donde se generaliza la descomposición a un nivel j .

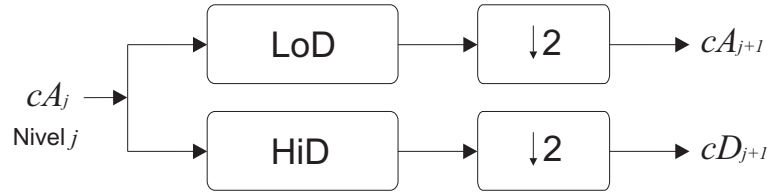


Figura 4-4.: Descomposición wavelet generalizada

4.2. Preprocesamiento de datos

El procedimiento previo a la caracterización de las señales ERPs consiste en la normalización y centralización de las mismas, que respectivamente se refiere a la estandarización de valores de la señal (reducir heterogeneidad) y la eliminación de componentes AC, a fin de garantizar coherencia en las fluctuaciones de las señales de la base de datos. Para tal efecto, se aplica la siguiente fórmula:

$$\mathbf{x}_i \leftarrow \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} \quad (4-28)$$

Después de obtener patrones representativos de las señales mediante la caracterización, se procede a realizar pruebas de preprocesamiento de datos con el fin de robustecer posteriores procedimientos sobre éstos. Este proceso básicamente consiste en la búsqueda de datos fal-

tantes o no convergentes, la identificación de datos atípicos y la verificación de gaussividad de cada variable.

4.2.1. Identificación de datos faltantes

En primera instancia, se debe verificar en cada una de las variables la presencia de datos faltantes (que puede deberse a mediciones no captadas o mal registradas en algunos pacientes) y de datos no convergentes (generadas como resultados de operaciones indeterminadas en la estimación de parámetros). Cuando esto ocurra, se debe estudiar cada caso en particular, pues si la cantidad de muestras es relativamente pequeña, no es recomendable simplemente eliminar la observación, debido a que esto puede conllevar a estimaciones estadísticas con menor nivel de significancia. Para esta base de datos en particular, se optó por eliminar la variable cuando el número de datos faltantes o no convergentes superaba el 10 % del total de observaciones.

4.2.2. Detección de datos atípicos

Para la aplicación de diferentes técnicas de análisis multivariado es importante asegurar la homogeneidad de la muestra mediante la eliminación de observaciones que parecen tener un comportamiento diferente a las demás de una misma clase o haber sido generados de forma diferente al resto de los datos. Para cumplir este propósito, se debe hacer un estudio univariado en cada clase para identificar los valores anómalos y contemplar la posibilidad de eliminarlos. Sin embargo, al igual que para el caso anterior debe tenerse cuidado con retirar observaciones de la base de datos, porque es posible que la muestra resultante no posea suficientes observaciones para trabajar con estimadores estadísticos. En este caso en particular, se decidió descartar aquellas variables que tienen más de un 10 % de valores atípicos mediante el análisis del coeficiente de kurtosis, que se define como sigue:

Sea \mathbf{X} la matriz de datos de dimensión $n \times p$, cuyas filas corresponden a las observaciones y las columnas a las variables; x_{ij} es un elemento de esta matriz, donde $i = 1, \dots, n$ representa la observación y $j = 1, \dots, p$ representa la variable.

Así el coeficiente de kurtosis es,

$$K_j = \frac{1}{n} \frac{\sum (x_{ij} - \bar{x}_j)^4}{s_j^4} \quad (4-29)$$

donde \bar{x}_j es la media de la variable x_j y s_j es su respectiva desviación estándar.

El coeficiente de kurtosis $K \geq 1$ mide la relación entre la variabilidad de las desviaciones y la desviación media, y en distribuciones gaussianas $K_j \rightarrow 3$. Así, aquellos coeficientes alejados de $K_j = 3$ pueden ser descartados por falta de homogeneidad, dado que un coeficiente de

kurtosis menor que 2 (que tienda a uno) puede significar que la variable refleja la mezcla de dos o más poblaciones, y un $K_j > 7$ indica que la variabilidad será grande y los datos tenderán a alejarse del resto [81].

Para la detección de datos atípicos de forma multivariada, en este trabajo se ha aplicado el Algoritmo 2 propuesto en [82]:

Algoritmo 2 Algoritmo de detección multivariada de datos atípicos

- 1: Buscar p direcciones ortogonales de máxima kurtosis.
- 2: Buscar p direcciones ortogonales de mínima kurtosis.
- 3: Eliminar provisionalmente los datos extremos en estas direcciones.
- 4: Calcular la media y la matriz de covarianza con los datos no sospechosos.
- 5: Finalmente identificar los datos atípicos como aquellos que son extremos con la distancia de Mahalanobis (ec. (4-30)) calculada con las estimaciones no contaminadas, sobre toda la muestra.

$$d_i = \left[(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right]^{1/2} \quad (4-30)$$

4.2.3. Verificación de gaussividad

Debido a las multiples restricciones que se pueden presentar en el análisis multivariado de una muestra en cuanto al tipo de distribución a la que deben aproximarse las variables, la verificación de gaussividad de dichas variables puede resultar muy significativa al momento de definir las técnicas para el estudio y tratamiento de las variables. Este procedimiento puede ser llevado a cabo mediante la prueba de Kolmogorov-Smirnov o por medio del análisis de Kurtosis explicado anteriormente.

4.2.4. Depuración de base de datos

En la construcción de bases de datos de señales biomédicas, orientadas a la creación de sistemas automáticos de detección de patologías, ocurren errores que impiden un buen análisis estadístico de los datos y limitan el rendimiento en la identificación de dichas patologías. Algunos de estos errores tienen origen en la adquisición de las señales debido a razones como: la incorrecta manipulación de los instrumentos, el mal etiquetado de las señales a la hora de almacenar los registros (al no tener un ambiente controlado en la adquisición de la señal), o por el dictamen de un diagnóstico médico equivocado. Adicional a lo anterior, en muchas ocasiones el dictamen del profesional no coincide con la realidad o inclusive los criterios médicos de quien evalúa son diferentes al de otro profesional. Este tipo de inconsistencias impiden conformar un conjunto de datos lo suficientemente robusto para asegurar que los sistemas automáticos de detección de patologías funcionen correctamente.

Resulta de gran importancia para el reconocimiento de patrones y los sistemas de diagnóstico automático de patologías partir de una base de datos altamente consistente. La gran dificultad para la depuración de este tipo de bases de datos ocurre en razón a los múltiples factores que pueden generar problemas en su conformación, por tanto es necesario encontrar una herramienta que al tiempo identifique valores atípicos, registros mal etiquetados, señales con poca calidad en su adquisición, registros con datos faltantes, entre otros.

Teniendo en cuenta el planteamiento anterior, en todos los experimentos desarrollados en este trabajo, se ha aplicado una metodología de depuración de bases de datos, basada en las técnicas de clustering e identificación de registros por error de clasificación, que permite al mismo tiempo eliminar todas las inconsistencias mencionadas. El planteamiento detallado de la metodología de pre-clustering aplicada en este trabajo puede verse en [83]. El sustento teórico de esta metodología puede encontrarse en [84].

5. Reducción de Dimensionalidad y Selección de Características

En el presente capítulo se hace una definición general sobre los conceptos de reducción de dimensión, análisis de relevancia y selección de características, los cuales ayudan a comprender la importancia de la representación de datos multidimensionales en un espacio reducido de variables como una etapa previa a la clasificación en el reconocimiento de patrones.

5.1. Reducción de dimensión

Las técnicas de aprendizaje de máquina han sido empleadas de manera satisfactoria en un amplio rango de aplicaciones como el control de procesos, análisis de datos, toma de decisiones, entre otras. El reconocimiento de patrones es una de las tareas principales consideradas dentro del aprendizaje de máquina y está directamente relacionado con asignar etiquetas de pertenencia a determinadas categorías (conocidas como *clases*) de algún fenómeno, basándose en el análisis de un conjunto de datos previamente dado [85]. El conjunto de variables empleadas para medir las observaciones se conoce como *espacio de características*, y la cantidad de variables es la *dimensión* de este espacio.

Cuando la dimensión del espacio de características es alta, se debe procurar remover la redundancia de datos o información irrelevante para reducir la dimensión; algunas de las principales razones para hacerlo son: facilitar el análisis de los datos, obtener mayor precisión en las tareas de clasificación, disminución del costo computacional, hacer clara cualquier estructura oculta (si existiera), obtener una representación gráfica que facilite la interpretación de datos, entre otras.

El objetivo de la reducción de dimensión consiste en encontrar la representación sobre alguna variedad en un sistema coordenado, que permita obtener una representación alterna y compacta con la menor dimensión posible. En general, la reducción de dimensiones se entiende como la representación de patrones de alta dimensión en un subespacio de menor dimensión basado en alguna transformación que optimiza una *función criterio o de evaluación* específica en el subespacio y está regida por un concepto encargado de dirigir el contexto de representación denominado *relevancia* [86].

5.1.1. Conceptos generales

Función de evaluación

Dado el conjunto inicial de características $\mathbf{X} = [\boldsymbol{\xi}_1 | \cdots | \boldsymbol{\xi}_p]$, $\boldsymbol{\xi}_j$ representa la j -ésima variable o el vector asociado a la característica j . De otro modo, los datos pueden representarse $\mathbf{X} = [\mathbf{x}_1^\top | \cdots | \mathbf{x}_n^\top]^\top$, contenido en el $\text{span}\{\mathbf{X}\}$ donde \mathbf{x}_i es la i -ésima observación. A partir de \mathbf{X} , aplicando transformaciones lineales, se obtienen espacios de representación $\mathbf{Z}_{\hat{h}} = \mathcal{G}_{\hat{h}}\{\mathbf{X}\} = \{\mathbf{z}_i \in \mathbb{R}^m : i = 1, \dots, n\}$, compuestos por los conjuntos de características o variables de representación $\hat{\boldsymbol{\xi}}_{\hat{h}} = \{\zeta_j : j = 1, \dots, m\}$, de menor dimensión, $m \leq p$, donde $\hat{h} \in \mathbb{N}$. Sea $\mathbf{k} = \{k_r : r \in \mathbb{N}\}$ el conjunto de etiquetas de clase, de forma que a cada observación \mathbf{z}_i le corresponda una etiqueta de clase.

Una función de evaluación es aquella que, teniendo en cuenta un criterio o una métrica, encuentra un valor real que represente a los datos evaluados. Esta función puede ser definida como:

$$\begin{aligned} f_{\hat{\boldsymbol{\xi}}_{\hat{h}}} : \mathbb{N} \times \mathcal{H}_{\mathbf{z}} &\rightarrow \mathbb{R} \\ (\mathbf{k}, \mathbf{Z}_{\hat{h}}) &\mapsto f_{\hat{\boldsymbol{\xi}}_{\hat{h}}}(\mathbf{k}, \mathbf{Z}_{\hat{h}}) \end{aligned}$$

donde $\mathcal{H}_{\mathbf{z}} \subseteq \mathbb{R}^m$. Por facilidad de notación se hará: $f_{\hat{\boldsymbol{\xi}}_{\hat{h}}}(\mathbf{k}, \mathbf{Z}_{\hat{h}}) \equiv f_{\hat{\boldsymbol{\xi}}_{\hat{h}}}(\mathbf{k}, \hat{\boldsymbol{\xi}}_{\hat{h}})$.

5.1.2. Relevancia

Sea el conjunto de observaciones $\mathbf{Z} = \{\mathbf{z}_i \in \mathbb{R}^m : i = 1, \dots, n\}$ de m variables, $\hat{\boldsymbol{\xi}}$. Sea $\mathbf{k} = \{k_r : r \in \mathbb{N}\}$ el conjunto de etiquetas de clase. Sea δ un valor umbral de significación de acuerdo a alguna métrica (geométrica, estadística, etc).

El conjunto \mathbf{Z} es relevante, de acuerdo a la métrica de significación, si existe una función de evaluación, $f_{\hat{\boldsymbol{\xi}}}$ de forma que el valor de $f_{\hat{\boldsymbol{\xi}}}(\mathbf{k}, \hat{\boldsymbol{\xi}})$ sea menor que el umbral, es decir, $f_{\hat{\boldsymbol{\xi}}}(\mathbf{k}, \hat{\boldsymbol{\xi}}) > \delta$.

Entre los tipos de métricas que existen para la reducción de dimensión, se encuentran [87]:

- *Métricas geométricas*: Sirven para determinar los subespacios geométricos de menor dimensión en donde las variables se descartan si no brindan capacidad de discriminación entre clases dentro del espacio geométrico de representación.
- *Métricas estadísticas*: Estas métricas generan modelos de representación basados en estructuras estadísticas para describir el comportamiento aleatorio con un número reducido de variables.

- *Métricas informativas*: Estas métricas emplean modelos para capturar la informatividad de las variables, con el fin de rechazar las variables que incrementen la incertidumbre, complejidad y costo computacional.

5.2. Selección de características

El proceso de selección de características puede entenderse como una forma particular de reducción de dimensión. Consiste en que la medida obtenida con la función de evaluación $f_{\hat{\xi}}$, correspondiente a alguna métrica, se incluye en el criterio de relevancia de la función de mapeo \mathfrak{F} . Así se obtiene, de acuerdo a la métrica, un subconjunto efectivo de variables ζ correspondientes al espacio inicial de variables o características ξ que mejora la capacidad de representación y reduce el costo computacional.

El subconjunto son las variables del espacio inicial de características que tienen mayor capacidad de representación. A este conjunto de variables iniciales se le conoce con el nombre de características relevantes.

Definición 5.2.1 (Característica relevante). Sea $\mathbf{k} = \{k_r : r \in \mathbb{N}\}$ el conjunto de clases, donde cada clase está compuesta por un determinado número de observaciones. Sea $\xi = \{\xi_j : j = 1, \dots, p\}$ el conjunto de características del cual se obtiene el subconjunto $\hat{\xi}_j = \xi \cap \overline{\xi_j}$ cuando se extrae una característica, siendo $\overline{\xi_j}$ el complemento de ξ_j en ξ y $\hat{\xi}_j \subset \xi$.

A continuación se presentan algunas definiciones relacionadas con relevancia:

1. *Fuertemente relevante*. Una variable de representación ξ_i se dice fuertemente relevante con respecto a la función de evaluación $f_{\hat{\xi}}$, si y sólo si,

$$f_{\xi}(\mathbf{k}, \xi) \neq f_{\hat{\xi}_j}(\mathbf{k}, \hat{\xi}_j) \quad (5-1)$$

2. *Débilmente relevante*. Una variable de representación ξ_j es débilmente relevante con respecto a la función de evaluación $f_{\hat{\xi}}$, si y sólo si,

$$f_{\xi}(\mathbf{k}, \xi) = f_{\hat{\xi}_j}(\mathbf{k}, \hat{\xi}_j) \wedge \exists \hat{\xi}'_j \subset \hat{\xi}_j \text{ tal que, } f_{\{\xi_j, \hat{\xi}'_j\}}(\mathbf{k}, \{\xi_j, \hat{\xi}'_j\}) \neq f_{\hat{\xi}'_j}(\mathbf{k}, \hat{\xi}'_j) \quad (5-2)$$

3. *Característica irrelevante*. Una variable de representación ξ_j es irrelevante, si y sólo si

$$\forall \hat{\xi}'_j \subseteq \hat{\xi}_j, \quad f_{\{\xi_j, \hat{\xi}'_j\}}(\mathbf{k}, \{\xi_j, \hat{\xi}'_j\}) = f_{\hat{\xi}'_j}(\mathbf{k}, \hat{\xi}'_j) \quad (5-3)$$

El problema de selección de características consiste en encontrar un subconjunto $\hat{\xi}_j \subseteq \xi$, tal que, si el cardinal de $\hat{\xi}_j$ es m , y en ξ están todos los subconjuntos de cardinal m , se procura

hallar el subconjunto $\hat{\xi}_j$, para el cual se optimice la función criterio $f_{\hat{\xi}_j}$ [88][89],

$$f_{\hat{\xi}_j}(\hat{\xi}_j) = \max_{\xi \subseteq \xi} f_{\xi}(\xi) \quad (5-4)$$

La selección de características contempla las siguientes dos etapas básicas de proceso [90]:

1. *Estrategias de generación:* se origina cada nuevo subgrupo \hat{h} de variables que van a ser analizadas, $\hat{\xi}_{\hat{h}} = \{\xi_j : j = 1, \dots, m\}$, $m \leq p$, $\hat{\xi}_{\hat{h}} \subset \xi$, tomadas directamente del espacio inicial de características, o bien, se generan subespacios conseguidos a través de algún principio de transformación del espacio inicial de entrenamiento, $\hat{\xi}_{\hat{h}} = \mathcal{G}\{\xi\}$.
2. *Función de evaluación:* mide la efectividad respecto a alguna métrica asociada a un criterio de relevancia del subconjunto \hat{h} que se está analizando, $f_{\hat{\xi}_{\hat{h}}}(\mathbf{k}, \hat{\xi}_{\hat{h}})$, la cual es comparada con la mejor obtenida para los subespacios anteriormente evaluados. La función de evaluación incluye la *condición de parada*, que corresponde a la restricción impuesta sobre los valores umbrales de efectividad, cuya aparición implica la detención en la búsqueda de un siguiente posible grupo sub-óptimo de características.

5.2.1. Estrategias de generación

Para la generación de subgrupos de características se encuentran los siguientes procedimientos [90][91]:

1. **Exhaustiva.** El generador de subconjuntos realiza la búsqueda sobre todas las posibles combinaciones que se pueden formar con el conjunto completo de características para encontrar el subconjunto óptimo de acuerdo a la función de evaluación dada. La búsqueda es *exhaustiva*, debido a que para un conjunto de características con dimensión p , el número total de subconjuntos candidatos, inducirían una complejidad equivalente a $\mathcal{O}(2^p)$, lo cual puede ser muy costoso computacionalmente.
2. **Heurística.** Del conjunto inicial de características (p variables) se selecciona alguno de los posibles grupos sub-óptimos con dimensión m , tal que $m < p$, y supere un umbral de aceptación para una función de evaluación a priori dada [91]. Estos métodos están basados en técnicas de búsqueda que operan en concordancia con reglas empíricas y están destinados a reducir la complejidad computacional, sin disminuir el rendimiento del sistema. Además, requieren de una condición de parada para prevenir que la búsqueda de subconjuntos se vuelva exhaustiva. La generación de subgrupos mediante criterios heurísticos puede tener variaciones en cuanto al espacio de búsqueda, por cuanto ésta puede ser realizada directamente del conjunto inicial de características o desde el nuevo espacio obtenido a través de algún principio de transformación $\hat{\xi}_{\hat{h}} = \mathcal{G}\{\xi\}$. Inicialmente, se determina el punto (o puntos) de partida en el espacio inicial de características, los

cuales darán la dirección de la búsqueda y se usan procedimientos basados en técnicas estadísticas o evolutivas para generar los estados siguientes.

3. **Aleatoria.** Este procedimiento halla aleatoriamente el espacio inicial de búsqueda y luego usando algoritmos basados en elecciones probabilísticas, o medidas de consistencia de los datos, es guiado a una solución óptima respecto a una función de evaluación dada [92]. Si bien el espacio de búsqueda es $\mathcal{O}(2^p)$, estos métodos típicamente buscan en un número más reducido de conjuntos que 2^p ; para ello establecen un número máximo de iteraciones posible. En esta categoría el hallazgo del subconjunto óptimo de características depende de los recursos disponibles, pues aunque la generación de los subconjuntos en un principio se hace de manera aleatoria, después se basa en procedimientos matemáticos para la continua generación de los subconjuntos siguientes, es por esto que este método requiere el valor de algunos parámetros o de una función objetivo para la continua construcción de subconjuntos.

Los tres procedimientos expuestos para la generación de subconjuntos se diferencian básicamente en la conformación del subconjunto inicial al momento de empezar la búsqueda. Es decir, el procedimiento denominado *exhaustivo* hace un barrido por todas las posibles maneras de conformación de subconjuntos usando algún orden pre-establecido comenzando desde el primero hasta el último; la estrategia denominada *heurística* usa métodos como pueden ser los estadísticos o evolutivos para inicializar la búsqueda con un subconjunto sub-óptimo y luego continuar la búsqueda mediante reglas heurísticas; finalmente la estrategia *aleatoria* elige el primer subconjunto de búsqueda aleatoriamente, es por esto que, como puede encontrarse el subconjunto óptimo rápidamente, puede llegar a ser exhaustivo, o no efectivo, en casos cuando el número de características es muy grande [90].

5.2.2. Funciones de evaluación

Estas funciones evalúan los subconjuntos de características usando alguna métrica que relaciona la capacidad que tienen para discriminar entre las clases existentes de los datos de entrenamiento.

Relativo a la dependencia que existe entre la función de evaluación y la etapa de clasificación que se aplica finalmente al subconjunto de características seleccionado, los métodos de selección de características se agrupan en dos tipos:

- *Métodos tipo filtro*, cuando son independientes de la etapa de clasificación, y se evalúan parámetros o funciones para lo cual se asume que influyen indirectamente en el rendimiento total del clasificador.
- *Métodos de envoltante (wrapper)*, cuando se evalúa directamente el rendimiento total de clasificación [93]

La selección de características orientada a la clasificación se puede llevar a cabo por alguno(s) de los siguientes criterios de efectividad incluidos en la función de evaluación:

1. *Similitud en la estructura de aleatoriedad de los conjuntos analizados*, que puede tener en cuenta, la independencia estadística de las características o la carga informativa de las mismas. Por ejemplo, la selección entrega un subconjunto con dimensión reducida, tal que su clase de distribución estadística sea lo más cercana posible a la clase de distribución original del espacio inicial de características [94].
2. *Error de reconstrucción de las observaciones*. Dado el subespacio con menor dimensión de características obtenido a través de algún principio de transformación $\mathcal{G}\{\cdot\}$, se analiza la capacidad de reconstrucción del espacio inicial de características ξ , mediante un criterio de error de reconstrucción ϵ establecido previamente.
3. *Rendimiento de proceso*. Cuando el subconjunto de características mejora la tasa de rendimiento del clasificador o disminuye el tamaño de la estructura sin reducir significativamente la precisión del clasificador construido, usando solamente las características seleccionadas [94].
4. *Coste computacional*. Cuando se obtiene un subconjunto sub-óptimo en un determinado número de iteraciones, o con un número de características previamente fijado.

A continuación, de acuerdo a los diferentes criterios de efectividad anteriormente expuestos, se describen algunos conceptos de *relevancia* asociada a una métrica, por medio de los cuales se obtiene la implementación de procedimientos para la reducción de dimensiones de un sistema de entrenamiento:

- *Relevancia geométrica*. Se asocia a una métrica geométrica relacionada con medidas de distancia en espacios de representación. Se prefiere la característica ξ_1 con respecto a ξ_2 , en un problema de clasificación binario, si ξ_1 induce a una mayor separación geométrica en el plano de representación para las dos clases. Cuando la diferencia es cero, entonces ξ_1 y ξ_2 se toman como indistinguibles.
- *Relevancia estadística*. Se asocia a una métrica que cuantifica atributos estadísticos como pueden ser, entre otros, la dispersión de los datos, la oblicuidad, el apuntamiento y, en general, medidas de dependencia estadística. Por ejemplo, el coeficiente de correlación valora la capacidad de predecir el valor de una variable a partir de otra(s) y es una medida convencional de dependencia estadística existente entre características y clases. Si la correlación entre la característica ξ_1 y la clase k es mayor que la correlación entre la característica ξ_2 y la clase k , entonces se prefiere la característica ξ_1 a la característica ξ_2 . Una variación en la aplicación de esta métrica corresponde a la determinación del valor de dependencia de una característica con respecto a las otras; este valor indica el grado de redundancia de las características [95].

- *Relevancia informativa.* Una característica ξ_j puede considerarse relevante a nivel informativo si existe una métrica con la cual se pueda obtener una medida relacionada con la cantidad de información que ofrece, y ésta posee alguna significación o importancia en el comportamiento, desempeño y capacidad de agrupamiento de un fenómeno aleatorio en particular. El eje central del modelo de información es el concepto de *información relevante*, también denominado *concepto de pertinencia*, el cual se obtiene a través de procedimientos estadísticos basados en *probabilidades*.

En general, a menor conocimiento implícito de acontecer un suceso (probabilidad de ocurrencia) o de conocer los datos de una variable correspondiente a un proceso para el cual se quiere obtener un modelo, existe una mayor relevancia informativa en el caso de acontecer dicho suceso o de revelarse los datos de dicha variable, ya que el sistema de entrenamiento valora como especialmente informativos los datos que desconoce previamente y la relación existente entre el estado de conocimiento a priori y a posteriori del sistema de entrenamiento determinará la relevancia informativa que poseen los datos de un suceso o variable.

Así, la cantidad de información respecto a la ocurrencia de un evento, es inversamente proporcional a su probabilidad. Una probabilidad de ocurrencia grande, ofrece una cantidad de información menor, y por ello implica menor relevancia informativa [96]. Por ejemplo, una característica ξ_j le ofrece mayor carga informativa a una característica ξ_i que a una característica ξ_k , si $I_G(\xi_i|\xi_j) > I_G(\xi_k|\xi_j)$, donde I_G es la *ganancia de información* que se puede expresar como [97]:

$$I_G(\xi_i|\xi_j) = H(\xi_i) - H(\xi_i|\xi_j), \quad i \neq j; i, j = 1, \dots, p \quad (5-5)$$

El valor $H(\xi_j)$ se denomina *entropía* de la variable aleatoria ξ_{ji} para $i = 1, \dots, n$ observaciones y corresponde a la medida cuantitativa de su incertidumbre. La entropía se puede considerar como la principal característica de una fuente de información, entre más alto sea el valor de la entropía mayor es la información contenida en ξ_i y se expresa como:

$$H(\xi_j) = - \sum_{i=1}^n p(\xi_{ji}) \ln(p(\xi_{ji})) \quad (5-6)$$

- *Relevancia por clasificación.* Se asocia a una métrica de error o precisión de clasificación. Los procedimientos que usan en su función de evaluación este tipo de relevancia corresponden a los métodos de envoltente (es decir, el clasificador es la función de evaluación), por lo tanto, los subconjuntos de características son evaluados mediante una medida de error o precisión de clasificación. El uso de este tipo de métrica es suficientemente preciso debido a que implica el análisis del propio clasificador para el cual se diseña la selección de características, pero la exigencia computacional comparada

con otras medidas puede ser bastante costosa [98].

Una descripción cualitativa del desempeño de diferentes conceptos de relevancia para tareas concernientes a la clasificación, se muestra en la Tabla 5-1, donde se tienen en cuenta los siguientes parámetros:

1. *Capacidad de generalización*. Indica el nivel de aplicabilidad a variados conjuntos de características (sin importar su origen) y las facilidades que ofrece como criterio de separación para diferentes clasificadores, esto es, que no se restrinja el buen rendimiento a un sólo tipo de clasificadores.
2. *Coste computacional*. El tiempo empleado en el procedimiento que ejecuta la selección del subconjunto óptimo o sub-óptimo de características.
3. *Precisión*. Indica el mayor rendimiento posible del clasificador usando el subconjunto seleccionado.

Tabla 5-1.: Comparación de diferentes conceptos de relevancia

<i>Función de evaluación</i>	<i>Capacidad de generalización</i>	<i>Costo computacional</i>	<i>Precisión</i>
Relevancia geométrica	Sí	Bajo	—
Relevancia estadística	Sí	Bajo	—
Medida informativa	Sí	Bajo	—
Relevancia por clasificación	No	Alto	Muy alto

La notación (—) significa que no se puede concluir nada acerca de la precisión del clasificador con respecto a la respectiva función de evaluación.

En cuanto a la *condición de parada*, existen diferentes formas para establecerla, por ejemplo, se puede dejar de adicionar o remover características cuando ninguno de los subconjuntos mejore la precisión de clasificación; aunque se puede continuar evaluando los subconjuntos de características hasta que la precisión no se degrade; o se puede seguir generando subconjuntos mientras se alcanza el final del espacio de búsqueda y entonces se selecciona el mejor. Una condición simple de parada es detenerse cuando cada combinación de valores para las características seleccionadas converjan a valores simples de clase, pero esto asume datos de entrenamiento libres de ruido. Una alternativa más robusta ordena las características de acuerdo con algún puntaje de relevancia, luego se fija un umbral para determinar el punto de parada final.

6. Transformación de Variables y Clasificación

En este capítulo se describen los procesos de transformación de variables y clasificación empleados. En la Sección 6.1 se presenta la proyección lineal aplicada que se basa en análisis de componentes principales. En la Sección 6.2 se describe el método de clasificación, en este caso, de tipo no supervisado. El método de agrupamiento no supervisado se describe en 6.2.1. Además se presenta un método de inicialización para métodos particionales de agrupamiento (Sección 6.2.1) y las medidas de desempeño que serán aplicadas (6.2.2).

6.1. Proyección lineal

El análisis de relevancia se ha dado un lugar importante en el área de reconocimiento de patrones, debido a la necesidad de encontrar las variables o características relevantes tratando de evitar la redundancia de la información y mejorar el desempeño de los clasificadores. El proceso de análisis de relevancia implica selección y/o extracción de características [99]. La selección consiste en determinar las características más relevantes en términos de algún criterio de representación o clasificación, mientras que la extracción consiste en obtener nuevas características a partir de los datos originales. La selección y extracción están estrechamente relacionados, los métodos de selección pueden obtenerse a partir de un problema de optimización orientado a la extracción [99] y viceversa [100]. El proceso de extracción también puede ser interpretado como la obtención de una nueva representación de los datos de acuerdo a un criterio de representación. Usualmente, dicha representación se hace en un espacio de menor dimensión, entonces un proceso de reducción de dimensión está implícito en la extracción.

Existen diversas alternativas para solucionar esta tarea y la escogencia de determinado método depende de las condiciones del problema y de la naturaleza de los datos [101]. Debido a su naturaleza no paramétrica, facilidad de implementación y versatilidad, el análisis de componentes principales (PCA) representa una buena alternativa, entre otras razones. Diversos estudios han comprobado su aplicabilidad como técnica de mapeo, extracción de características y reducción de dimensionalidad en diferentes contextos como procesamiento de señales biomédicas [102], [103], segmentación de imágenes [104], entre otros.

6.1.1. Análisis de componentes principales

Este método consiste en determinar el número de elementos representativos dentro de un conjunto de datos que contengan información de la variabilidad de dichos datos [101]. Por tal razón, en PCA se busca una proyección en donde los datos se vean mejor representados en términos de mínimos cuadrados (SM), dicha proyección corresponde a la varianza acumulada de cada observación. En otras palabras, PCA reduce la dimensionalidad de un conjunto de datos, reteniendo aquellos atributos o características del conjunto de datos que contribuyen más a su varianza y por tanto las características escogidas son las que presentan mayor separabilidad de acuerdo a un criterio de SM [105].

Aplicar PCA consiste en realizar una proyección lineal de los datos originales de manera que se genere un nuevo sistema de coordenadas en donde la mayor varianza del conjunto de datos es capturada en el primer eje, la segunda varianza más grande en el segundo eje, y así sucesivamente; donde la medida de varianza la define una matriz de covarianza de los datos [101]. Entonces, el objetivo es determinar los elementos que minimizan el error cuadrático medio de la proyección de los datos sobre los vectores propios de la matriz de covarianza, sujeto a una condición de ortonormalidad.

Proyección ortonormal

Sea \mathbf{x}_i la i -ésima observación de p atributos o características y $\mathbf{X} \in \mathbb{R}^{n \times p} = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)^T$ la matriz de datos.

Por facilidad de notación, el vector \mathbf{x} de dimensión p denotará cualquier observación. Dicho vector puede escribirse como una combinación lineal de los elementos de una base ortonormal, así:

$$\mathbf{x} = \sum_{i=1}^p c_i \mathbf{u}_i \quad (6-1)$$

donde $\mathbf{U} = (\mathbf{u}_1^T, \dots, \mathbf{u}_p^T)$ representa la base ortonormal y $\mathbf{c} = (c_1, \dots, c_p)$ son los pesos de la combinación lineal.

Pensando en la reducción de dimensión, la proyección ortonormal $\tilde{\mathbf{x}}$ se realiza en un espacio q -dimensional ($q < p$), que mejor represente a \mathbf{x} :

$$\tilde{\mathbf{x}} = \sum_{i=1}^q c_i \mathbf{u}_i \quad (6-2)$$

El error cuadrático medio de la proyección ortonormal con respecto del vector original, se

puede escribir como:

$$\overline{e^2} = \mathbf{E} \left\{ (\mathbf{x} - \tilde{\mathbf{x}})^\top (\mathbf{x} - \tilde{\mathbf{x}}) \right\} \quad (6-3)$$

Reemplazando las expresiones (6-1) y (6-2) en (6-3), el error se puede re-escribir como:

$$\overline{e^2} = \mathbf{E} \left\{ \left(\sum_{i=1}^p c_i \mathbf{u}_i - \sum_{i=1}^q c_i \mathbf{u}_i \right)^\top \left(\sum_{i=1}^p c_i \mathbf{u}_i - \sum_{i=1}^q c_i \mathbf{u}_i \right) \right\} = \mathbf{E} \left\{ \left(\sum_{i=q+1}^p c_i \mathbf{u}_i \right)^\top \left(\sum_{i=q+1}^p c_i \mathbf{u}_i \right) \right\} \quad (6-4)$$

Transformación PCA

Minimizar el error de la ecuación (6-4), es equivalente a maximizar el complemento del mismo, es decir:

$$\mathbf{E} \left\{ \left(\sum_{i=1}^q c_i \mathbf{u}_i \right)^\top \left(\sum_{i=1}^q c_i \mathbf{u}_i \right) \right\} = \mathbf{E} \left\{ \sum_{i=1}^q c_i^2 \right\} = \text{tr}(\mathbf{c}_q^\top \mathbf{c}_q) = \text{tr}(\mathbf{\Sigma}_q) \quad (6-5)$$

donde \mathbf{c}_q es un vector compuesto por los q primeros elementos de \mathbf{c} , $\mathbf{\Sigma}_q = \mathbf{c}_q^\top \mathbf{c}_q$ y $\text{tr}(\cdot)$ representa la traza de su matriz argumento.

Se puede apreciar que $\mathbf{\Sigma}_q$ es una matriz de $q \times q$ simétrica y semi-positiva definida. En el caso de PCA, para realizar la proyección de todos los vectores \mathbf{x}_i , la matriz $\mathbf{\Sigma}$ es de $p \times p$ y corresponde a la matriz de covarianza, que puede ser estimada como:

$$\mathbf{\Sigma} = \frac{1}{n} \mathbf{X}^\top \mathbf{X} \quad (6-6)$$

La ecuación previa se aplica una vez los datos se hayan centrado en la media de cada observación, es decir:

$$\mathbf{x}_i \leftarrow \mathbf{x}_i - \mu(\mathbf{x}_i), \quad i = 1, \dots, n \quad (6-7)$$

donde $\mu(\cdot)$ representa la media.

Aplicando el criterio de invariancia ortonormal [106], puede plantearse el siguiente problema

de optimización:

$$\max \frac{1}{n} \text{tr}(\mathbf{U}_q^T \mathbf{X}^T \mathbf{X} \mathbf{U}_q) = \sum_{j=1}^q \lambda_j \quad (6-8)$$

$$\text{s. a. } \mathbf{U}_q^T \mathbf{U}_q = \mathbf{I}_q \quad (6-9)$$

donde \mathbf{I}_d representa una matriz identidad de dimensión d , \mathbf{U}_q representa los q primeros elementos de \mathbf{U} y λ_j los valores propios de $\mathbf{\Sigma}$.

La matriz $\mathbf{\Sigma}$ es simétrica, por tanto existe una base completa de vectores propios de la misma y la transformación lineal que proyecta los datos sobre esta base corresponde a la representación de los datos que se utiliza para la reducción de la dimensionalidad [105].

Los elementos de la base ortogonal se denominan componentes principales y la proyección lineal de los datos se escribe:

$$\mathbf{Y} = \mathbf{X}\mathbf{U} \quad (6-10)$$

En el procedimiento estándar de PCA, se considera \mathbf{U} como la matriz de vectores propios de $\mathbf{\Sigma}$ ordenados de forma descendente, es decir:

$$\begin{aligned} [\mathbf{U}, \mathbf{\Lambda}] &= \text{eig}(\mathbf{\Sigma}), \quad \mathbf{\Lambda} = \text{diag}(\lambda), \\ \lambda &= (\lambda_1, \dots, \lambda_p) \text{ y } \mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_p) \text{ con } \lambda_1 > \dots > \lambda_n \end{aligned} \quad (6-11)$$

donde $\text{eig}(\cdot)$ representa la descomposición en valores y vectores propios, y $\text{diag}(\cdot)$ denota una matriz diagonal formada por el vector de su argumento.

Por último, se escogen los q primeros componentes principales como los elementos relevantes, es decir, los que mejor representan a \mathbf{X} . El valor de q , puede definirse a través de algún criterio de varianza acumulada o evaluando iterativamente los resultados de un clasificador.

Considerando que el valor de la función objetivo de PCA se asocia directamente a la suma de los valores propios de la matriz de covarianza, como se ve en la ecuación (6-8), se puede definir un criterio de selección del valor de q como sigue.

El criterio de varianza acumulada se aplica sobre el valor de la función objetivo normalizado, así:

$$\nu^* = \text{diag}(\mathbf{U}^T \mathbf{X}^T \mathbf{X} \mathbf{U}) / \text{tr}(\mathbf{U}^T \mathbf{X}^T \mathbf{X} \mathbf{U}) \quad (6-12)$$

El vector $\boldsymbol{\nu}^*$ es un indicador de la variabilidad de cada componente, entonces, para un criterio del $N\%$ se consideran los q elementos de \mathbf{Y} que correspondan a un valor de varianza acumulada del $N\%$, es decir,

$$\sum_{i=1}^q \nu_i^* \approx N/100.$$

6.2. Clasificación no supervisada

En general, la clasificación no supervisada se refiere a los métodos denominados discriminativos, en los que no se requiere de un conocimiento a priori de las clases para la clasificación. Entonces la tarea de clasificación se lleva a cabo a través de la información contenida en los datos únicamente, comúnmente empleando búsquedas heurísticas [107]. Estos métodos sólo requieren de algún parámetro de inicialización como la cantidad de grupos resultantes para empezar el proceso de agrupamiento. Entonces, la clasificación no supervisada consiste en agrupar patrones homogéneos sin ninguna información acerca de la naturaleza de las clases presentes en el conjunto de datos. Los subconjuntos de datos homogéneos se generan a partir de algún criterio basado en distancias, disimilitudes o medidas estadísticas.

Existen diversas razones por las que los métodos no supervisados son de interés, entre ellas, porque convergen rápidamente y con buen desempeño en caso de que las características cambien poco en el tiempo, permiten categorizar elementos, son útiles cuando el etiquetado de un conjunto de datos no es factible o no es confiable, entre otras. Sin embargo, la solución generada por un sistema de análisis no supervisado puede verse afectada por factores, como parámetros iniciales no adecuados, que pueden generar una mala convergencia.

En la literatura se denominan algoritmos particionales de agrupamiento a aquéllos en los que se agrupan patrones homogéneos a partir de una división inicial de los datos que se refina, comúnmente, de forma iterativa. Por ejemplo, el agrupamiento basado en centroides es de tipo particional.

6.2.1. Agrupamiento basado en centroides

El agrupamiento basado en centroides (CBC) es la técnica clásica del agrupamiento no supervisado [108]. La idea general del agrupamiento basado en centroides (CBC) es minimizar una función de costo, la cual define la calidad de la solución del agrupamiento. Dicha solución se obtiene iterativamente a través de la actualización de centroides, y la partición resultante por cada iteración corresponde a la asignación de los elementos al subconjunto cuyo centroide sea el más cercano [107].

Agrupamiento basado en la mínima suma de cuadrados

Este método es uno de los ejemplos básicos de CBC. En el agrupamiento basado en la mínima suma de cuadrados (MSSC), explicado ampliamente en [107], el problema de optimización es:

$$\min_{\rho_k \in \mathbf{P}_k} \sum_{j=1}^k \sum_{\mathbf{x}_l \in \mathbf{C}_l} \|\mathbf{x}_l - \mathbf{q}_j\|^2 \quad (6-13)$$

donde $\|\cdot\|$ denota la norma euclidiana, \mathbf{P}_k denota el conjunto de todas las particiones del conjunto de datos $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ en el espacio euclídeo \mathbb{R}^d , $\mathbf{x}_i = (\mathbf{x}_{1i}, \dots, \mathbf{x}_{di})^\top$ es el vector de atributos de la i -ésima observación, $\mathbf{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_k\}$ representa el conjunto de grupos, $\mathbf{Q} = (\mathbf{q}_1, \dots, \mathbf{q}_k)^\top$ es el conjunto de centroides y \mathbf{q}_j es el centroide correspondiente al grupo j , dado por:

$$\mathbf{q}_j = \frac{1}{n_e(\mathbf{C}_j)} \sum_{l: \mathbf{x}_l \in \mathbf{C}_j} \mathbf{x}_l, \quad j = 1, \dots, k \quad (6-14)$$

donde $n_e(\cdot)$ denota la cantidad de elementos de su grupo argumento y k es el número de grupos.

El objetivo de este método de agrupamiento es encontrar la partición de los datos que minimice la distancia entre los elementos contenidos en cada grupo y su respectivo centroide, es decir, minimizar la varianza intra-clase. Este mismo principio lo aplica el algoritmo básico de *K-medias*, que se muestra en el Anexo B.

Criterios de inicialización

Una de las desventajas de los algoritmos de agrupamiento es que son sensibles a la inicialización, por tanto, si la partición inicial no es adecuada, pueden tener mala convergencia [109]. Una forma de solucionar este problema es garantizar una buena inicialización de los centroides; para esta tarea existen herramientas eficaces como el algoritmo *máx-mín* y *J-medias*. En este trabajo se aplica únicamente el algoritmo *máx-mín*, por lo cual sólo nos centraremos en la explicación de éste.

Criterio máx-mín El método máx-mín, descrito en [109], busca en el conjunto \mathbf{X} , los k elementos más alejados con un criterio de distancia preestablecido, mejorando la cantidad de grupos necesarios para separar las clases y el valor de convergencia. Este algoritmo comienza con un punto arbitrario de \mathbf{X} como el primer centroide \mathbf{q}_1 y el resto se escogen siguiendo una estrategia en la que el elemento elegido en la i -ésima iteración es aquél cuyo elemento más cercano entre los $i - 1$ ya elegidos, se encuentra más alejado [109]. El segundo centroide \mathbf{q}_2 se

obtiene como el elemento que presenta la máxima distancia entre \mathbf{q}_1 y los puntos restantes $\{\mathbf{X} - \mathbf{q}_1\}$.

Con estos dos centroides se pueden determinar los demás centroides aplicando el criterio máx -mín, así:

$$f(\mathbf{x}_i) = \max_{\mathbf{x}_i \in \{\mathbf{X} - \mathbf{Q}\}} \left\{ \min_{\mathbf{q}_j \in \mathbf{Q}} \|\mathbf{x}_i - \mathbf{q}_j\|^2 \right\}, \quad j = 1, \dots, k \quad (6-15)$$

donde $\|\cdot\|$ representa la norma euclidiana.

6.2.2. Medidas de desempeño del agrupamiento

Medidas supervisadas

Para evaluar los resultados del agrupamiento se usaron medidas de sensibilidad (Se), especificidad (Sp) y porcentaje de clasificación (CP), las cuales se definen como sigue, de acuerdo a los siguientes parámetros:

- *Verdaderos positivos-TP*. Número de observaciones de la clase de interés (CI) que son clasificados correctamente.
- *Verdaderos negativos-TN*. Número de observaciones diferentes de la clase de interés (NCI) que son clasificados correctamente.
- *Falsos positivos-FP*. Número de observaciones *NCI* clasificadas como *CI*.
- *Falsos negativos-FN*. Número de observaciones *CI* clasificadas como *NCI*.

Sensibilidad y especificidad

$$Se = \frac{TN}{TN + FP} \times 100 \quad (6-16)$$

$$Sp = \frac{TP}{TP + FN} \times 100 \quad (6-17)$$

Porcentaje de clasificación En función de los parámetros descritos previamente, el desempeño de clasificación puede ser estimado con la siguiente ecuación:

$$CP = \frac{TN + TP}{TN + TP + FN + FP} \quad (6-18)$$

La sensibilidad y especificidad miden, respectivamente, la proporción de observaciones NCI y la proporción de observaciones CI , clasificados correctamente. Estas medidas se aplican sobre todas las clases presentes en el registro y se usan para medir el desempeño del sistema, pero no tienen implicación en la sintonización de los parámetros del proceso de agrupamiento (todos los criterios de sintonización son estrictamente no supervisados debido a las razones discutidas anteriormente).

Dado que no es de utilidad un número de grupos relativamente grande, se aplica un factor de penalización sobre las anteriores medidas, con el fin de que un agrupamiento correcto realizado con un número de grupos considerablemente grande no registre un desempeño del 100 %. El factor de penalización aplicado en este trabajo, descrito en [100], es de la forma

$$e^{-\eta k_r / k_a} \quad (6-19)$$

donde k_r es la cantidad de grupos resultantes del agrupamiento, k_a es el valor admisible de la cantidad de grupos y η es un factor de ajuste. Con esto, la medida \mathbf{m} (que en este caso, puede ser Se , Sp o CP), se pondera de modo que

$$\mathbf{m} \leftarrow \begin{cases} \mathbf{m} e^{-\eta k_r / k_a} & k_r > k_a \\ \mathbf{m} & k_r \leq k_a \end{cases} \quad (6-20)$$

El término η se fija de acuerdo al rigor de penalización que se determine, teniendo en cuenta que debe ser estrictamente mayor que 0, además, es aconsejable que sea menor que 1 para que la penalización no sea tan rigurosa: $0 < \eta \leq 1$.

Medidas no supervisadas

Además de describir las medidas de desempeño supervisadas ya empleadas en otros estudio de *clustering*, en este trabajo se introduce una nueva medida denominada *vector de valores relativos*, mediante la cual se estima la distancia relativa que separa a un agrupamiento de otro. A continuación se describen estas medidas de rendimiento.

Desviación de la función objetivo. En [110] se introduce un índice de desempeño que corresponde a la medida complementaria al radio de dispersión o desviación del valor de la función objetivo, que se denota como f_1/f_2 y corresponde a la relación entre el valor óptimo de la función objetivo f_1 y el valor calculado usando la partición resultante f_2 . La función objetivo evaluada para la estimación de este índice debe ser coherente con el método de agrupamiento empleado. Dado que el mínimo valor de f_2 es f_1 , esto es, $f_2 \geq f_1$, esta medida indica un buen agrupamiento cuando su valor se aproxima a 1.

Coherencia de los grupos. A partir de lo discutido en [111], se puede formular una medida

no supervisada del desempeño del agrupamiento, que cuantifica la coherencia de los grupos empleando un criterio de particionado, así:

$$\varepsilon_M = \frac{1}{k} \sum_{l=1}^k \frac{\mathbf{M}_l^T \mathbf{W} \mathbf{M}_l}{\mathbf{M}_l^T \mathbf{D} \mathbf{M}_l} \quad (6-21)$$

donde \mathbf{M} es una matriz binaria que indica la pertenencia de los elementos a los grupos (1 si pertenece, 0 no pertenece), \mathbf{M}_l denota la l -ésima columna de la matriz \mathbf{M} (asociada al grupo l), \mathbf{W} es la matriz de afinidad y \mathbf{D} es una matriz diagonal que representa el grado de la matriz \mathbf{W} , definida como $\mathbf{D} = \text{Diag}(\mathbf{W}\mathbf{1}_n)$, donde $\mathbf{1}_n$ es vector de unos de dimensión n . La matriz de afinidad es simétrica y definida positiva de dimensión $n \times n$ y se escoge de acuerdo al problema de clasificación [112].

Debido a la normalización con respecto del grado de la matriz de afinidad, el máximo valor de ε_M es 1, por tanto indica un buen agrupamiento si su valor es próximo a 1 [110].

Vector de valores relativos. El vector de valores relativos \mathbf{v} es un indicador de la calidad del agrupamiento, dado que tiene en cuenta las diferencias entre las distancias intra y entre clases. Este vector está asociado a una matriz de distancias $\widetilde{\mathbf{D}}$ que se define como sigue:

A partir de los centroides obtenidos mediante un algoritmo de agrupamiento, se define una matriz de distancias $\mathbf{D} = [d_{ij}] \in \mathbb{R}^{n \times k}$, formada por los elementos:

$$d_{ij} = d(\mathbf{x}_i, \mathbf{q}_j) \quad (6-22)$$

donde \mathbf{q}_j denota el j -ésimo centroide, con $i = 1, \dots, n$ y $j = 1, \dots, k$. En este caso $k = 2$.

Una matriz de distancias $\widetilde{\mathbf{D}} \in \mathbb{R}^{k \times k}$ es obtenida a partir de la matriz \mathbf{D} , cuya matriz diagonal está constituida por la suma de las distancias entre los centroides \mathbf{C}_k y los datos de su respectivo agrupamiento k , y los elementos por fuera de la diagonal son la suma de las distancias entre los centroides del grupo k y los datos pertenecientes a los grupos restantes. De esta manera, matricialmente $\widetilde{\mathbf{D}}$ se define como:

$$\widetilde{\mathbf{D}} = \begin{bmatrix} \sum_{i \in \mathbf{C}_1} d(\mathbf{x}_i, \mathbf{q}_1) & \sum_{i \in \mathbf{C}_1} d(\mathbf{x}_i, \mathbf{q}_2) \\ \sum_{i \in \mathbf{C}_2} d(\mathbf{x}_i, \mathbf{q}_1) & \sum_{i \in \mathbf{C}_2} d(\mathbf{x}_i, \mathbf{q}_2) \end{bmatrix} = \begin{bmatrix} \tilde{d}_{11} & \tilde{d}_{12} \\ \tilde{d}_{21} & \tilde{d}_{22} \end{bmatrix} \quad (6-23)$$

donde $\mathbf{C}_k \in \mathbb{R}^{n_k \times k}$ representa el k -ésimo agrupamiento y n_k es su correspondiente número de datos.

De esta manera, el vector de valores relativos asociados a la matriz $\tilde{\mathbf{D}}$ es calculado con la siguiente ecuación:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} |\tilde{d}_{11} - \tilde{d}_{12}|/\tilde{d}_{11} \\ |\tilde{d}_{21} - \tilde{d}_{22}|/\tilde{d}_{22} \end{bmatrix} \quad (6-24)$$

Esta medida de desempeño presenta una ventaja adicional: como se observa en la ecuación (6-24), el cociente entre la diferencia de elementos de la matriz \mathbf{D} evita la sensibilidad a la magnitud de los valores, ya que la diferencia entre \tilde{d}_{k1} y \tilde{d}_{k2} en normalizada con respecto a \tilde{d}_{kk} .

En la Tabla 6-1 se relacionan las medidas de desempeño consideradas en este estudio con su respectiva notación y expresión matemática.

Denominación	Notación	Expresión
Sensibilidad	Se	$\frac{TN}{TN + FP} \times 100$
Especificidad	Sp	$\frac{TP}{TP + FN} \times 100$
Porcentaje de clasificación	CP	$\frac{TN + TP}{TN + TP + FN + FP} \times 100$
Desviación de la función objetivo	f_1/f_2	$\frac{f_1}{f_2}$
Coherencia de los grupos	ε_M	$\frac{1}{k} \sum_{l=1}^k \frac{\mathbf{M}_l^T \mathbf{W} \mathbf{M}_l}{\mathbf{M}_l^T \mathbf{D} \mathbf{M}_l}$
Vector de valores relativos	\mathbf{v}	$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \tilde{d}_{11} - \tilde{d}_{12} /\tilde{d}_{11} \\ \tilde{d}_{21} - \tilde{d}_{22} /\tilde{d}_{22} \end{bmatrix}$

Tabla 6-1.: Medidas de desempeño empleadas

Parte IV.

Marco Experimental

7. Esquema de Trabajo

En este capítulo se describe la base de datos utilizada en esta investigación, las configuraciones de los experimentos implementados y las estrategias de validación empleadas con el fin de comparar el desempeño de las diferentes técnicas de extracción y selección de características analizadas. Además, bajo los criterios de precisión del sistema en la etapa de clasificación y la reducción del espacio inicial de características, se evalúan la capacidad discriminante del conjunto de características estáticas descrito en la Sección 4.1, y la capacidad de reducción del método de selección de características aplicado, respectivamente.

Cabe anotar que los algoritmos implementados para llevar a cabo los experimentos que se describen a continuación, fueron elaborados en la herramienta matemática y de programación Matlab®.

7.1. Descripción de la base de datos

La base de datos empleada en esta tesis ¹ obedece al desarrollo de un riguroso protocolo de selección de la muestra y de adquisición de registros, el cual fue llevado a cabo en conjunto con los grupos de *Neuroaprendizaje* de la Universidad Autónoma de Manizales y de *Desarrollo Infantil* de la Universidad de Manizales. La estimación inicial del número de muestras requeridas para el desarrollo del estudio fue de 400 sujetos, sin embargo, a pesar que a la fecha se han intervenido más de 500 niños, únicamente 180 de ellos han cumplido con todos los criterios de inclusión y condiciones de pareamiento. Los datos y registros almacenados en la base de datos fueron recopilados durante 2009 y 2011.

7.1.1. Sujetos

Los individuos participantes en el estudio fueron niños entre 4 y 15 años pertenecientes a instituciones educativas del área urbana del Municipio de Manizales. La definición de los grupos de casos (niños con TDAH) y controles (niños sanos) se hizo, respectivamente, a partir del cumplimiento o rechazo de criterios clínicos del desorden, y la ponderación de pruebas ceñidas a un protocolo de evaluación neuropsicopedagógicas, como se detalla en la sección 7.1.2. Las pruebas diagnósticas y demás registros realizados en ambos grupos, fueron

¹La construcción de esta base de datos se llevó a cabo exclusivamente para el desarrollo de los proyectos de investigación en los cuales está enmarcado este trabajo

llevadas a cabo con previo consentimiento informado de los padres o acudientes.

En relación al diagnóstico clínico, se registró la historia clínica de cada sujeto, haciendo énfasis en los síntomas y signos presentados, antecedentes de riesgos prenatales (infecciones, contacto con tóxicos, antecedentes de ingesta de medicación, sangrados vaginales, preeclampsia, prematuridad) y post natales (hipoxia neonatal, infecciones) y de otras patologías. Además, se tiene en cuenta la historia académica del paciente, el momento de aparición del trastorno y los contextos de presentación, y antecedentes familiares cercanos con TDAH. El examen clínico se ha enfatizado en la búsqueda de signos neurológicos blandos y el estado de la visión y la audición.

7.1.2. Pruebas diagnósticas neuropsicopedagógicas

El protocolo diseñado para la búsqueda, evaluación y clasificación de los sujetos sanos y con TDAH está compuesto por métodos empleados en el campo de la psiquiatría y la psicología, además de cuestionarios comportamentales diseñados para la evaluación de la sintomatología asociada a categorías o tipologías diagnósticas. El algoritmo de investigación aplicado consiste inicialmente en el rastreo de casos con sospecha de TDAH y de controles, a partir de criterios generales de búsqueda; esta labor se lleva a cabo por parte de profesores y padres de familia en las instituciones educativas intervenidas. Después de identificar los grupos de posibles casos y controles, se procede a la etapa de tamizaje, que consiste en la aplicación de cuestionarios de rastreo a padres y maestros y de un test breve de capacidad intelectual a los niños. Del paso anterior, se define el grupo de sujetos que continúan y aquellos que por tener CI (capacidad intelectual) menor a 85 se excluyen del estudio.

Posteriormente, se procede a realizar la entrevista psiquiátrica semiestructurada MINIKID (*Mini International Neuropsychiatric Interview for Children and Adolescents*) que se considera el *Gold Standard* en el diagnóstico de TDAH: así, quienes obtuvieron MINIKID positiva conforman el grupo de casos y aquellos con MINIKID negativa forman el grupo de controles. A partir de ese momento se inicia la etapa de evaluación médica de los sujetos, en la cual se determina la ausencia de trastornos de tipo sistemático, neurológico o sensorial; de esta manera se filtra nuevamente la muestra de estudio, dado que los sujetos que presenten alguno de los trastornos mencionados también deben excluirse de la investigación.

Finalmente, se lleva a cabo el protocolo neuropsicopedagógico que consiste en la valoración neuropsicológica mediante pruebas de memoria verbal y visual, atención visual y auditiva, flexibilidad cognoscitiva, lenguaje, entre otras, y la valoración de habilidades académicas en lectura y escritura. Concluidos los anteriores procedimientos, se debe realizar la evaluación neurofisiológica, en la cual se realiza la adquisición de potenciales evocados cognitivos, como se explica en la sección [7.1.3](#).

7.1.3. Adquisición de los registros

El registro de las pruebas neurofisiológicas se realizó mediante un equipo de adquisición de potenciales evocados de cuatro canales de *Cadwell Laboratories* y el software *Sierra Wave*™ versión 6.0.33. El paradigma aplicado en la adquisición de registros fue el *oddball* en modalidad auditiva y visual; en ambos casos, el experimento consta de 200 estímulos, de los cuales el 80 % corresponden a estímulos comunes y el 20 % restante, a los estímulos infrecuentes.

El registro de la actividad bioeléctrica cerebral se ha realizado siguiendo los parámetros definidos por el Sistema Internacional 10 – 20 [113] mediante electrodos de superficie convencionales con una impedancia de 5Ω , ubicados en los puntos Fz, Cz, y Pz, con referencia a los procesos mastoides. La adquisición de los registros se hizo con los sujetos sentados y ubicados en una habitación aislada de ruidos externos y en penumbra.

Después de adquirido, el registro es pasado por un preamplificador y filtros pasabajas con frecuencias entre 0,3 y 100 Hz. La tasa de muestreo de la señal digital es de 640 Hz, con una grabación de 2 s por señal (1 s - 640 puntos - pre y post estímulo).

Los criterios de exclusión aplicados en el examen clínico fueron: existencia de déficit visual o auditivo, evidencia de retardo mental, coexistencia con trastornos psiquiátricos, y antecedentes de enfermedad neurológica, tales como síndrome convulsivo, tumores, enfermedades degenerativas, infecciones del sistema nervioso central, entre otras. Adicionalmente, se solicitó la suspensión de la administración de fármacos desde el día anterior a los niños con TDAH que estaban medicados.

Potenciales evocados visuales

Para el caso de la modalidad visual, se solicita al sujeto que observe un monitor ubicado en frente y separado a un metro de distancia. Durante el desarrollo de la prueba, en el monitor se muestra una imagen constituida por un damero monocromático en patrón reversible constituido por 16 cuadros como estímulo frecuente; el sujeto debe estar atento a la aparición del estímulo infrecuente que consiste en un *target* en el centro de la pantalla manteniendo como fondo el damero. El sujeto debe oprimir un botón de un contador de estímulos cada vez que aparezca el estímulo infrecuente.

Potenciales evocados auditivos

En la modalidad auditiva, los sujetos se vendan los ojos para minimizar el parpadeo y los movimientos oculares, y los estímulos se presentan binauralmente a través de auriculares. Estos consisten en tonos de 80 dB y de 50 ms de duración. Los estímulos frecuentes se presentan a una frecuencia de 1,000 Hz, y los infrecuentes a 3,000 Hz. El orden de aparición de los estímulos es aleatorio y se presenta uno cada 1,500 ms.

7.2. Conjunto de características

El conjunto de características \mathbf{X} utilizado en este trabajo corresponde a parámetros empleados en estudios previos, como se ha descrito en la Sección 4.1. Después de realizar los respectivos cálculos, la matriz de datos ha quedado conformada de la siguiente manera: las primeras 16 corresponden a características morfológicas, las siguientes tres, son características de tipo espectral, y las características restantes corresponden a coeficientes wavelet. La Tabla 7-1 resume el conjuntos de características mencionadas, relacionando su ubicación dentro de la matriz \mathbf{X} y la forma de cálculo implementada en esta tesis.

N°	Tipo	Característica	Descripción
1	Morfológicas	Latencia	$t_{sP300} = \{t s(t) = s_{P300}\}$
2		Amplitud	$s_{N200-P300} = s_{N200} - s_{P300}$
3		Razón latencia/amplitud	$LAR = \frac{t_{sP300}}{s_{N200-P300}}$
4		Amplitud absoluta	$AAMP = s_{N200-P300} $
5		Área positiva	$A_p = \frac{1}{2} \sum_{t=t_a}^{t_b} (s(t) + s(t))$
6		Área negativa	$A_n = \frac{1}{2} \sum_{t=t_a}^{t_b} (s(t) - s(t))$
7		Área total	$A_{pn} = A_p + A_n$
8		Área total absoluta	$ATAR = A_{pn} $
9		Área absoluta total	$A_{p n } = A_p + A_n $
10		Pendiente de señal absoluta promedio	$ \bar{s} = \frac{1}{n} \sum_{t=t_a}^{t_b-\tau} \frac{1}{\tau} s(t+\tau) - s(t) $
11		Valor pico a pico	$s_{N200-P300} = s_{N200} - s_{P300}$
12		Ventana de tiempo pico-pico	$t_{pp} = t_{N200} - t_{P300}$
13		Pendiente pico-pico	$\dot{s}_{pp} = \frac{s_{N200-P300}}{t_{pp}}$
14		Cruces por cero	$n_{ZC} = \sum_{t=t_{sN200}}^{t_{sP300}} \delta_s$
15		Densidad de cruces por cero	$d_{ZC} = \frac{n_{ZC}}{t_{pp}}$
16		Alteraciones de signo de las pendientes	$n_{sa} = \frac{1}{2} \sum_{t=t_a+\tau}^{t_b-\tau} \left \frac{s(t-\tau)-s(t)}{ s(t-\tau)-s(t) } + \frac{s(t+\tau)-s(t)}{ s(t+\tau)-s(t) } \right $
17	Espectrales	Moda de la frecuencia	$S(f_{mode}) = \text{Max}_f \{S(f)\}$
18		Mediana de la frecuencia	$\int_0^{f_{median}} S(f) df = \int_{f_{median}}^{\infty} S(f) df$
19		Media de la frecuencia	$f_{mean} = \frac{\int_0^{\infty} f \cdot S(f) df}{\int_0^{\infty} S(f) df}$
20,...,83	Wavelet	Coeficientes wavelet	$CWT(2^i, b) = 2^i ^{-1/2} \int_{-\infty}^{\infty} x(t) \psi * \left(\frac{t-b}{2^i}\right)$

Tabla 7-1.: Conjunto de características empleadas en el análisis de señales ERPs

7.3. Sintonización de parámetros

7.3.1. Caracterización

Localización temporal de la onda P300

Como se indica en la Sección 4.1.1, para llevar a cabo la localización del componente P300 es necesario definir una ventana de tiempo móvil $T = (t_1, t_2)$, en la cual podría encontrarse la onda P300. Inicialmente, esta ventana se fija en $t_1 = 260$ ms y $t_2 = 400$ ms, que corresponde al intervalo de tiempo en el que usualmente se presenta el pico P300. Debido a las bien conocidas alteraciones en la latencia y la amplitud que sufren las señales ERPs a causa del TDAH, es necesario considerar un intervalo más amplio en el que puede ocurrir la presentación del pico P300. Considerando esta restricción, si el algoritmo de localización no halla una deflexión significativa en el intervalo de tiempo inicial, la ventana se desplazará a la derecha y luego a la izquierda hasta encontrar un punto de inflexión significativamente profundo. Así, t_1, t_2 podría tomar los siguientes valores:

1. $t_1 = 400$ ms, $t_2 = 500$ ms, si no hay deflexión, entonces
2. $t_1 = 500$ ms, $t_2 = 600$ ms, si no hay deflexión, entonces
3. $t_1 = 230$ ms, $t_2 = 260$ ms

Estos valores fueron definidos heurísticamente de acuerdo al estudio visual que previamente se hizo a la base de datos por parte de los especialistas médicos.

Características morfológicas y espectrales

- Para realizar el cálculo de las características morfológicas, fue necesario la selección de una ventana de análisis (t_a, t_b) debido a los retardos típicos de los componentes ERPs (especialmente el P300). Los valores de $t_a = 250$ ms y $t_b = 450$ ms fueron definidos de acuerdo al rango de normalidad hallado según se recomienda en la Sección 2.1.7, aplicando $\mu + 2,5\sigma$, donde μ es la media de la latencia obtenida para las clases normales y σ es la desviación estándar. En este orden de ideas, $t_a \approx \mu$ y $t_b \approx \mu + 2,5\sigma$.
- Para estimar las características frecuenciales, la densidad espectral de potencia $S(f)$ se calculó a través de transformada de Fourier discreta sobre la señal discretizada, de forma que: $S(f) = 2F(f)F^*(f)$, donde $F(f)$ es la transformada de Fourier de $s(t)$ y $F^*(f)$ es la función conjugada de $F(f)$.
- Las características basadas en transformada wavelet se estimaron con una función madre de tipo *spline biortogonal*, con 3 momentos de desvanecimiento, y que utiliza un filtro pasa-bajo en descomposición de orden 8, un filtro pasa-alto en descomposición

de orden 4, un filtro pasa-bajo en reconstrucción de orden 4 y un filtro pasa-alto en reconstrucción de orden 8. En la figura 7-1 se muestra la wavelet madre empleada, que fue aplicada como sugiere el estado del arte dada su similitud con los potenciales evocados cognitivos.

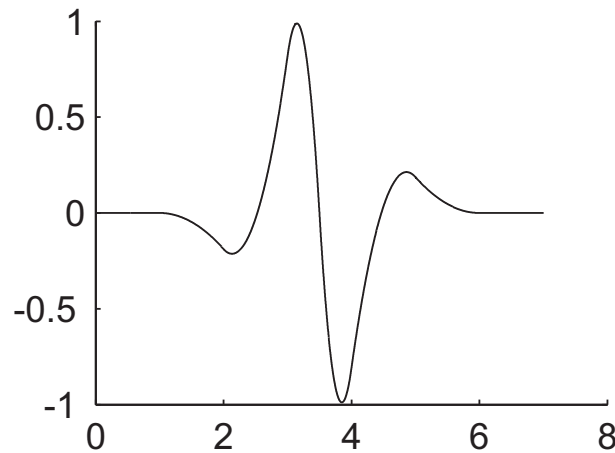


Figura 7-1.: Wavelet madre spline biortogonal, empleada en el análisis multi-resolución

- Para el caso de la transformada discreta se re-muestreó la señal a una frecuencia de 1024 Hz y se emplearon 7 niveles de descomposición con el fin de ajustar de forma aproximada las bandas de frecuencia de los niveles a los ritmos cerebrales delta (0,2 a 3,5 Hz), theta (3,5 a 7,5 Hz), alfa (7,5 a 13 Hz) y beta (13 a 28 Hz). En la Figura 7-2 se observan los niveles de descomposición generados mediante transformada wavelet discreta y análisis multi-resolución (ver Sección 4.1.3), junto a su correspondiente banda de frecuencia y número de coeficientes obtenidos.
- De los 7 niveles de descomposición obtenidos, fueron seleccionadas como características wavelet los coeficientes A7, D7, D6 y D5 (según Figura 7-2). Para justificar la selección de estos coeficientes se utilizó un criterio de informatividad basado en la entropía de *Shanon* [114] acumulada mayor a 60 %.

7.4. Diseño de experimentos

7.4.1. Experimento 1

En el primer experimento se presenta una comparación de técnicas de extracción y selección de características en términos del desempeño de clasificación de un sistema de identificación automático de TDAH. Las técnicas de selección evaluadas emplean funciones de costo tipo filtro y wrapper (análisis multivariado de varianza y clasificador bayesiano sobre distribuciones gaussianas, respectivamente), junto con algoritmos de búsqueda heurística. Además,

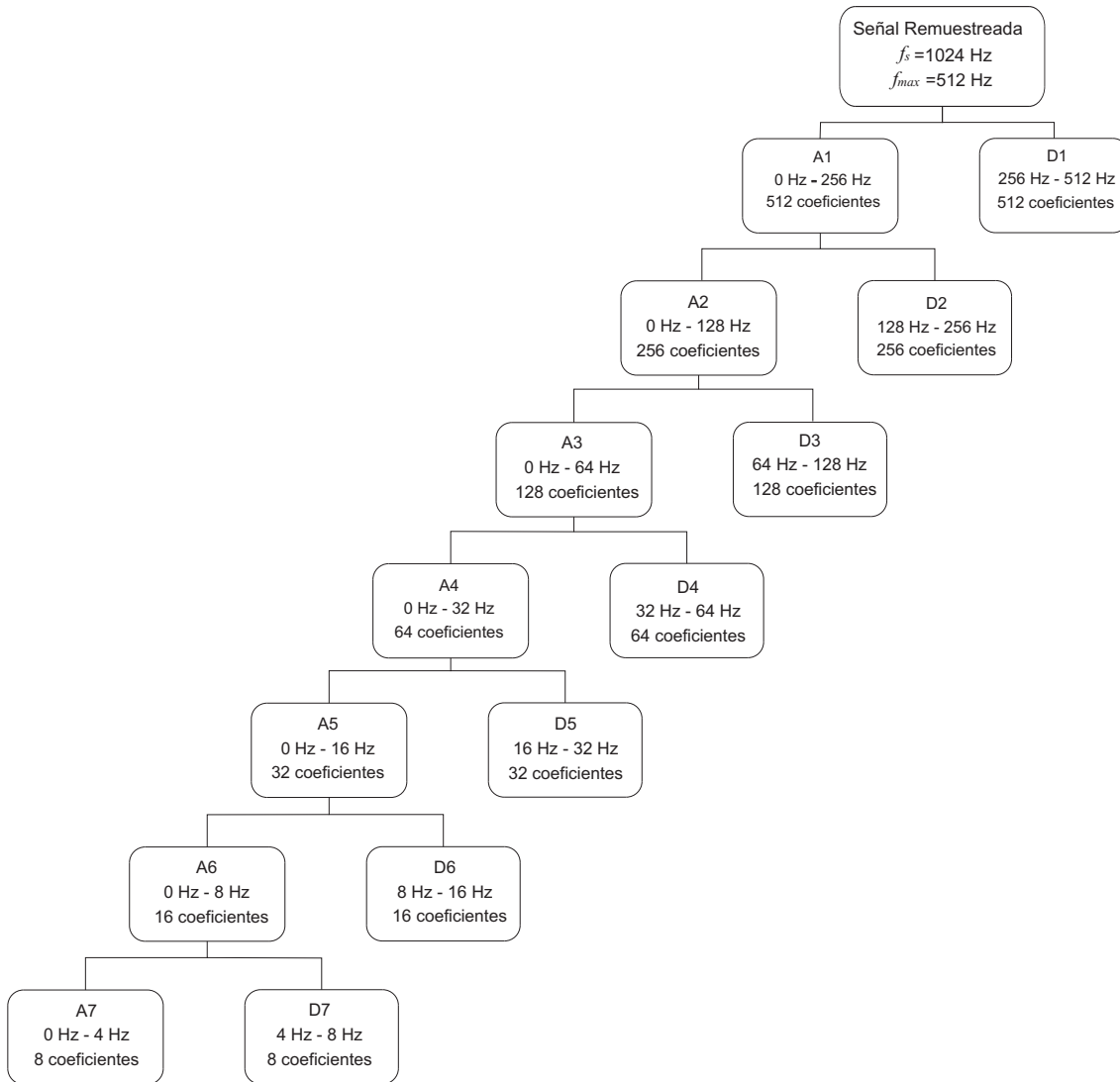


Figura 7-2.: Descomposición wavelet de una señal ERP

se aplica el análisis de componentes principales como técnica de extracción de características. Estas técnicas son comparadas con base en el error de validación de un clasificador de k -NN y uno discriminante lineal (LDC).

A continuación, se presenta una breve explicación de las etapas ejecutadas durante el desarrollo de este experimento.

Algoritmo filtro con base en análisis multivariado de varianza

La evaluación sucesiva de subconjuntos de variables sobre un espacio inicial de características p se lleva a cabo empleando análisis multivariado de varianza (MANOVA – Multivariate Analysis of Variance) como función de costo. Los subconjuntos de características se construyen empleando la técnica denominada selección secuencial hacia adelante (*Sequential*

Forward Selection-SFS) (ver Sección C.1.2).

El siguiente algoritmo, implementado de forma heurística, realiza una búsqueda secuencial empleando MANOVA como función de costo. En general, consiste en comparar los conjuntos de variables de diferente dimensión mediante la evaluación de la probabilidad acumulada sobre la distribución F .

Algoritmo 3 Algoritmo de MANOVA progresivo

- 1: Se calcula el estadístico F sobre los subconjuntos con una sola característica, y se selecciona aquella cuyo estadístico F fue mayor (denotado por ξ^1). Luego, se estima su correspondiente valor de probabilidad acumulada (\hat{P}_1) sobre la distribución F .
 - 2: Se construyen subconjuntos de dos características, combinando la característica seleccionada en el paso anterior con cada una de las restantes (al aplicar SFS). Se actualiza el valor de estadístico F , evaluando cada subconjunto formado mediante la prueba de Wilks.
 - 3: De los subconjuntos formados en el *paso 2*, se escoge aquel con mayor valor en el estadístico F (representado por ξ^2) y se calcula su probabilidad acumulada (\hat{P}_2). Si $\hat{P}_2 > \hat{P}_1$, entonces el ξ^2 pasará a ser analizado, de lo contrario, la búsqueda concluye.
 - 4: Se repite el paso 2, pero ahora se construyen subconjuntos de tres características, partiendo de ξ^2 .
 - 5: Se actualiza el nuevo subconjunto usando el mismo criterio expuesto en el paso 3. Se repite el paso 4, actualizado una y otra vez el número de características del subconjunto del que se parte.
 - 6: El algoritmo se detiene cuando las características agregadas no incrementan el valor de probabilidad acumulada.
 - 7: El subconjunto final tiene un tamaño p' , donde $p' \leq p$.
-

Algoritmo wrapper con base en clasificador bayesiano

La selección de características que emplea una función de costo tipo *wrapper* utiliza el porcentaje de acierto de validación de un clasificador bayesiano, y al igual que en la prueba anteriormente descrita, usa como procedimiento iterativo de búsqueda heurística una selección secuencial hacia adelante. En ambos casos, tanto en el algoritmo tipo filtro como el tipo *wrapper*, el objetivo es maximizar el criterio de evaluación mediante un subconjunto de características seleccionado por una estrategia de búsqueda.

7.4.2. Experimento 2

Este experimento corresponde a la evaluación de una nueva metodología propuesta en este trabajo, que consiste en una variación del método de reducción de dimensiones y visualización presentado en [115]. En general, el método propuesto consiste en realizar una búsqueda secuencial flotante hacia adelante en una nueva representación de los datos originales para

determinar las variables relevantes. La nueva representación de los datos se realiza en dos etapas: en primera instancia se aplica análisis de componentes principales (PCA) con el fin obtener una proyección de los datos en un nuevo espacio donde ocurra la mayor variabilidad de los mismos. Posteriormente, se aplica un algoritmo de agrupamiento no supervisado basado en centroides para obtener una nueva representación de las observaciones a través de las distancias de los datos y los centroides. Finalmente, esta nueva definición de patrones descriptivos de las señales ERP, son sometidos a una fase de clasificación supervisada a fin de determinar la capacidad discriminante de esta re-caracterización en la identificación del TDAH.

A continuación se describen en detalle cada una de las etapas.

Proyección basada en PCA

Con el fin de obtener una representación más compacta de las clases en términos de mínimos cuadrados, se aplica una proyección de los datos basada en PCA (ver Sección 6.1.1) de forma que se obtiene una nueva representación de la forma $\mathbf{Y} = \mathbf{XV}$. Esta proyección se realiza en un espacio de menor dimensión, es decir, usando $\widehat{\mathbf{V}} \in \mathbb{R}^{p \times q}$ donde $q < p$. Por tanto, la proyección truncada es

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\mathbf{V}} = (\hat{\mathbf{y}}_1^T, \dots, \hat{\mathbf{y}}_n^T)^T \quad (7-1)$$

El parámetro q es escogido con un criterio de varianza acumulada mayor a 90 %.

Clustering

Los datos proyectados son agrupados empleando un método de agrupamiento basado en centroides, k-medias (ver Anexo B), con el fin de obtener una nueva representación de los datos $\mathbf{B} = \{b_{ij}\}$, en donde cada observación es representada a través de su distancia con los centroides de sus respectivos grupos y con los restantes, es decir,

$$b_{ij} = d(\hat{\mathbf{y}}_i, \mathbf{q}_j), \quad i = 1, \dots, n; \quad j = 1, \dots, k \quad (7-2)$$

donde k es el número de grupos, \mathbf{q}_j denota el j -ésimo centroide y $d(\cdot, \cdot)$ es un operador de distancia.

De esta manera, se ha hecho una nueva caracterización de los datos, tomando las características morfológicas y espectrales como espacio inicial de representación.

Búsqueda heurística

Después de la re-caracterización de los registros obtenida mediante el cálculo de las distancias entre los datos y los centroides, se aplica un algoritmo de búsqueda heurística; en este caso en particular, se emplea una selección secuencial flotante hacia adelante (*sequential forward floating selection* - SFFS). Esto se lleva a cabo con el fin de realizar una reducción supervisada que conlleve a encontrar el menor número de características que permita de manera suficiente la clasificación de patrones.

El algoritmo SFFS implementado utiliza como función de evaluación la correcta clasificación alcanzada con un clasificador bayesiano, dado que cada función de densidad de probabilidad está modelada de forma gaussiana. Además, el método empleado fue mejorado mediante una prueba de hipótesis *t-test* y una evaluación de pérdida de información [116].

El siguiente es el algoritmo que resume las etapas previamente explicadas de la metodología aplicada en este experimento.

Algoritmo 4 Algoritmo de re-caracterización de señales ERPs mediante medidas de disimilitud

Entrada: $\mathbf{X}_{n \times p}$.

1. Se aplica un preclustering sobre la matriz de datos: $\hat{\mathbf{X}} = \text{preclustering}\{\mathbf{X}\}$, donde la matriz $\hat{\mathbf{X}}$ es de dimensiones $h \times p$ donde $h < n$.
2. Estimar la matriz de covarianza $\Sigma_{\mathbf{X}}$.
3. Hallar los autovalores $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ y autovectores $\mathbf{V} = [\mathbf{v}_1 | \dots | \mathbf{v}_p]$ de $\Sigma_{\mathbf{X}}$ ordenados de forma descendente, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.
4. Se determina q ($q < p$) mediante un criterio de varianza acumulada mayor del 90 %.
5. Se obtiene la proyección truncada de los datos $\hat{\mathbf{Y}} \in \mathbb{R}^{h \times q} = \hat{\mathbf{X}}\hat{\mathbf{V}}$.
6. Se aplica agrupamiento de los datos y se obtiene la matriz de centroides $\mathbf{Q} = [\mathbf{q}_1^T | \dots | \mathbf{q}_p^T] = \text{kmedias}(\hat{\mathbf{Y}})$.
7. Se re-caracterizan los datos: $\mathbf{B} \in \mathbb{R}^{h \times k} = \{b_{ij}\} = \{d(\hat{\mathbf{y}}_i, \mathbf{q}_j)\}, i = 1, \dots, n; j = 1, \dots, k$, siendo $d(\cdot, \cdot)$ la distancia euclidiana.
8. $\hat{\mathbf{B}} \in \mathbb{R}^{h \times m} = \text{SFFS}\{\mathbf{B}\}$, m es el número de variables relevantes, $m < k$.
9. Validación: clasificadores k -nn, LDC y SVM (70 % para entrenamiento y 30 % para validación).

Salida: $\hat{\mathbf{B}} = \{ \text{conjunto efectivo de características} \}$

7.4.3. Experimento 3

En el experimento N° 3, se desea evaluar los efectos de la latencia en la formación de agrupaciones bien definidas y compactas a través de técnicas de agrupamiento no supervisado. Para este fin, se tiene en cuenta únicamente el grupo de características morfológicas que fue definido en la Sección 4.1.1. En el desarrollo de las pruebas fueron utilizadas, por una parte, las señales alineadas de acuerdo a la ubicación del componente P300, y por otra, las señales ERP originales únicamente con su respectivo preprocesamiento aplicado. Para lograr el alineamiento de las señales, se introduce un nuevo enfoque para la ubicación temporal del punto P300 (que se describe en detalle en la Sección 4.1.1) mediante el estudio de la morfología de la señal derivada. Las pruebas realizadas utilizan diferentes configuraciones de la matriz de datos calculadas a partir de las señales alineadas y originales. Para la tarea de agrupamiento, se emplea una técnica tradicional de agrupamiento mediante la aplicación del algoritmo de k-medias. Para evaluar la calidad de los grupos formados y la relevancia de las características basadas en la latencia, se ha definido una medida de separabilidad y compacidad mediante valores relativos de distancia entre los centroides y los datos de cada agrupación.

En los siguientes apartados se describen con más detalle los procedimientos llevados a cabo en este experimento.

Alineamiento de las Señales

Para analizar la separabilidad y compacidad de los grupos formados a partir de las características morfológicas, todos los registros de la base de datos fueron alineados mediante el remuestreo por partes de la señal para localizar el componente P300 en el mismo punto de acuerdo a una señal patrón previamente seleccionada.

Selección de la señal patrón La señal patrón se selecciona entre todos los registros que componen la base de datos, escogiendo una por cada clase. Esta señal se define como aquella con el más alto coeficiente de correlación calculado entre los registros de cada clase, el cual es almacenado en un vector de correlación promediado, y se estima de la siguiente manera:

Sea \mathbf{R} una matriz de correlación triangular superior, definida como:

$$\mathbf{R}_c = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & 0 \\ \vdots & \vdots & \dots & 0 \\ r_{n1} & 0 & 0 & 0 \end{bmatrix} = [\mathbf{r}_1 | \dots | \mathbf{r}_n] \quad (7-3)$$

donde $r_{ij} = \text{corr}(\mathbf{s}_i, \mathbf{s}_j)$, \mathbf{s}_i representa la señal asociada al i -ésimo sujeto de la clase $c = 1, 2$ (correspondientes a las clases normal y patológica), y $\text{corr}(\cdot, \cdot)$ es un operador

estándar de correlación.

Así, el vector de correlación promediado $\rho \in \mathbb{R}^n$ puede ser definido como:

$$\rho = [\mu(\mathbf{r}_1), \dots, \mu(\mathbf{r}_n)], \quad (7-4)$$

donde \mathbf{r}_k es el k -ésimo vector columna.

Por tanto, se escoge como \mathbf{s}_p a la señal que corresponde al $\max\{\rho\}$.

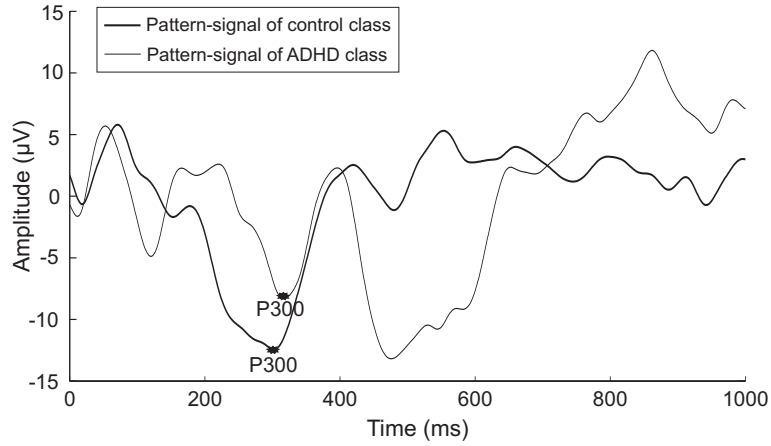


Figura 7-3.: Señales patrón de las clases control y casos

En la Figura 7-3 se muestran las señales patrón seleccionadas para cada una de las dos clases evaluadas. De acuerdo al procedimiento explicado arriba, estos son los registros que más se asemejan al resto de señales de las respectivas clases.

Después de definir dicha señal patrón, su componente P300 (en adelante llamado *patrón-P300*) es ubicado sobre la forma de onda mediante el algoritmo de localización descrito en la Sección 4.1.1. Posteriormente, los registros restantes son alineados haciendo coincidir sus ondas P300 (identificadas también mediante el procedimiento descrito en la Sección 4.1.1) con el patrón-P300. Para alcanzar tal fin, todas las señales son divididas en dos segmentos: segmentos \mathbf{a}_s y \mathbf{b}_s , como se muestra en la figura 7-4. Estos dos segmentos son remuestreados a igual longitud que sus correspondientes segmentos (\mathbf{a}_p y \mathbf{b}_p) a una frecuencia de muestreo f_a y f_b , respectivamente, las cuales se definen como:

$$f_a = \ell(\mathbf{a}_p)/\ell(\mathbf{a}_s); \quad f_b = \ell(\mathbf{b}_p)/\ell(\mathbf{b}_s), \quad (7-5)$$

donde $\ell(\cdot)$ denota el número de muestras de su señal argumento.

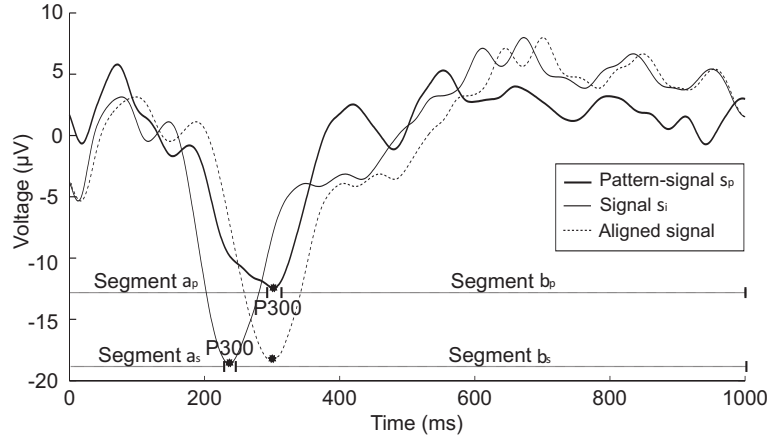


Figura 7-4.: Alineamiento de la señal ERP de acuerdo a la señal patrón

Agrupamiento no supervisado

Para evaluar la compacidad y separabilidad del conjunto de datos \mathbf{X} , dividido en dos grupos homogéneos (cada uno de ellos asociado a una clase) es necesario el uso de una técnica no supervisada de agrupamiento. Dado que el interés principal de este experimento es mostrar la capacidad discriminante del conjunto de características evaluados en términos de separabilidad y compacidad entre los grupos formados, se ha empleado una técnica de agrupamiento básica; en este caso en particular se ha aplicado el algoritmo de *k-medias*, como se explica en el Anexo B e implementado como muestra el Algoritmo 5.

Cuatro diferentes configuraciones de la matriz de datos fueron desarrolladas con el fin de probar la influencia de la latencia en la formación de grupos compactos y diferenciables. La primera matriz de datos \mathbf{X}_1 está formada usando únicamente la latencia como vector de características; la segunda configuración \mathbf{X}_2 forma una matriz de datos constituida por las 16 características morfológicas descritas en la Sección 4.1.1, incluyendo la latencia; la tercera configuración de \mathbf{X} , \mathbf{X}_3 , excluye la latencia del conjunto de características usado en \mathbf{X}_2 . Finalmente \mathbf{X}_4 utiliza la misma configuración que \mathbf{X}_2 pero se usa una matriz de características obtenida a partir de las señales ERP alineadas mediante la técnica descrita en la Sección 7.4.3.

Se aplica entonces el algoritmo de *k-medias* a las diferentes matrices de datos para generar las matrices \mathbf{D} y $\tilde{\mathbf{D}}$ como se indica en la Sección 6.2.2 y así poder calcular los valores relativos de distancia; de esta manera, se cuantifica la separación relativa entre los agrupamientos formados y entre los centroides y los datos de su correspondiente grupo.

7.4.4. Experimento 4

En este experimento se ha aplicado una metodología para la detección del TDAH basada en *customized wavelet*, que se ha llevado a cabo según se indica en [117] y [118]. Para obtener

nuevas características discriminantes a partir de los potenciales relacionados a eventos, se ha optimizado el espacio tiempo-frecuencia de la transformada wavelet packet (WPT), mediante la *customización* de la función wavelet para esta aplicación en particular. A continuación se detallan las etapas desarrolladas en esta metodología:

Customización de la función wavelet

Para customizar la función wavelet, se aplica el esquema de ajuste (*lifting scheme* - LS) mostrado en la figura 7-5, con el cual se construye la transformada wavelet (el LS es una representación polifásica de un clásico banco de filtros). En esta figura se observa sólo un paso de descomposición de la señal de entrada $x[n]$. En tal esquema, $x[n]$ es dividida en muestras pares $x_e[n]$ e impares $x_o[n]$; entonces, los coeficientes de detalle son obtenidos como el error de predicción de $x_o[n]$ a partir de $x_e[n]$, usando el operador p . Este operador es introducido para obtener una estimación de las muestras impares convolucionando $x_e[n]$ con los coeficientes del predictor. Asimismo, los coeficientes de aproximación son obtenidos actualizando los coeficientes de detalle con otro operador u y retomando las muestras pares. Del mismo modo, este operador estima los componentes de baja frecuencia de la señal de entrada, convolucionando los coeficientes de detalle con los valores del operador.

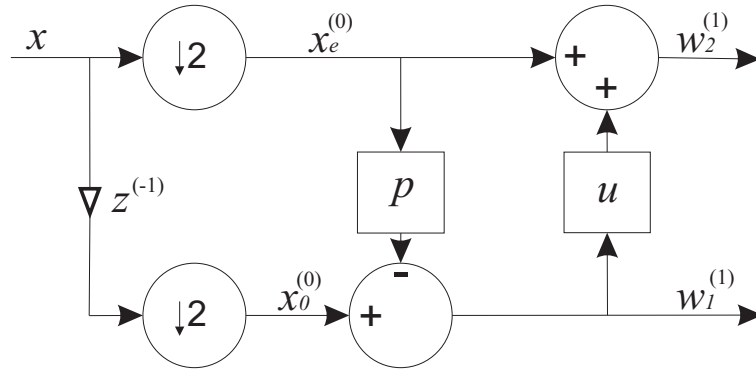


Figura 7-5.: Esquema de ajuste para construcción de la WT

Los operadores u y p , están asociados con funciones wavelet con características únicas. En esta metodología, los algoritmos genéticos (GE) son empleados para customizar tales operadores y hallar los coeficientes que satisfagan la función de ajuste. Adicionalmente, el número de parámetros que los GA pueden optimizar, es reducido usando la normalización y restricciones lineales de los operadores del LS. Estas restricciones garantizan el carácter biortogonal de la transformada wavelet resultante. En el reconocimiento de patrones biológicos es deseable emplear una función de ajuste que tenga en cuenta la forma y las distancias de las clases en el espacio de características, además de contar con una herramienta que maximice esa relación. En este caso, la función de ajuste seleccionada es la distancia euclídea, como se propone en [117].

Optimización del árbol de descomposición

La WPT tiene una única característica relacionada con la resolución tiempo-frecuencia del espacio wavelet, donde la señal de entrada es proyectada. En esta transformada, los coeficientes de detalle y de aproximación se descomponen en un conjunto redundante de subespacios donde las señales pueden ser representadas con múltiples representaciones tiempo-frecuencia. En general, con j niveles de descomposición hay $2^{(j+1)-1}$ combinaciones de subespacios con diferentes configuraciones tiempo-frecuencia. Para seleccionar la configuración más adecuada para esta aplicación, se computa el valor de energía normalizado en cada subespacio para el conjunto de patrones de las ERPs (conjunto de entrenamiento). Ahora, la distancia euclídea se calcula para medir la distancia entre los centroides de las clases, respecto a la dispersión entre clases de cada subespacio. Aquellos subespacios que obtienen las mayores distancias son seleccionados como óptimos.

Selección de características

Usando los coeficientes de los subespacios óptimos evaluados en la optimización del espacio tiempo-frecuencia, el espacio tiempo-frecuencia wavelet debe ser reconstruido. Sin embargo, debido a la acción de la división en muestras pares e impares del esquema de ajuste LS y a la naturaleza discreta de la transformada, algunas consideraciones deben ser tenidas en cuenta para tal construcción. En primera instancia, los subespacios deben ser remuestreados de acuerdo al nivel de descomposición dado por la relación 2^j . Por ejemplo, dada una señal de 256 puntos, descompuesta en $j = 4$ niveles, entonces uno de los subespacios en el último nivel contendría solo 16 coeficientes que cubren todo el dominio del tiempo. Así, para compensar las 240 muestras restantes, tales coeficientes deben ser remuestreados por 2^4 . En segunda instancia, para hacer más suave el espacio tiempo-frecuencia, los coeficientes remuestreados deben ser interpolados.

Una vez el espacio tiempo-frecuencia es construido, se extraen diferentes métricas a partir del mismo. En particular, en este trabajo se muestra que el contorno de frecuencia de este espacio contiene valiosa información [119]. Este contorno de frecuencia se calcula proyectando el espacio sobre el eje de la frecuencia. Usualmente, la función de proyección es una medida estadística tal como el valor medio o la varianza. En esta tesis se emplea el valor medio para extraer el contorno mencionado.

Entonces, se prueba un conjunto de métricas estadísticas que se han sugerido para el análisis de otras bioseñales. Estas métricas se dividen en dos grupos: uno, extraído del contorno, y otro extraído a partir de la forma de onda de la señal ERP. El primer grupo está compuesto por cuatro momentos alrededor de la media (kurtosis) y la desviación estándar. El segundo grupo está conformado por el número de picos de la serie temporal.

Sintonización de la función wavelet

Para customizar la función wavelet y adaptarla a la aplicación actual, el procedimiento de optimización basada en GA debe contar con un conjunto de entrenamiento que contenga registros de ambas clases. En este caso, se empleó el 50 % de la base de datos para este fin. Los parámetros para la sintonización de los GA y para la optimización del espacio tiempo-frecuencia se definen con detalle en la Sección 8.4.

7.4.5. Experimento 5

El quinto experimento consiste en la caracterización y clasificación de los ritmos cerebrales, que son señales de la actividad eléctrica del cerebro que corresponden a bandas de frecuencia asociadas a algunas función cognitiva. Dichas señales son experimentalmente obtenidas como resultado de la reconstrucción de las señales asociadas a cada uno de los niveles de la descomposición wavelet, como se describe en las Secciones 4.1.3 y 7.3.1. A partir de los coeficientes wavelet estimados para la definición del tercer grupo de características (ver Sección 4.1.3), se hace la reconstrucción de los niveles en los cuales inicialmente se hizo la descomposición wavelet de la señal $s(t)$. En la figura 7-6 se observan las ondas generadas por el proceso de reconstrucción de los coeficientes de detalle de los 7 niveles y los de aproximación del último nivel.

Para la caracterización de las ondas obtenidas, se han empleado sólo algunas de las características definidas en la Sección 4.1, dado que muchas de éstas dependen de la localización temporal del componente P300, condición que no se puede cumplir en las ondas de baja frecuencia (como las generadas por los coeficientes cA7 y cD7) ni en las de muy alta frecuencia (como las estimadas por los coeficientes cD1, cD2 y cD3). Las características utilizadas en este experimento son de tipo morfológico (estimadas considerando toda la señal y no sólo un segmento de la misma como se hizo en la caracterización de los experimentos 1, 2 y 3), tales como el área total de la señal y el área positiva de la señal. Además, se han empleado otras características de tipo informativo como lo son la entropía de Shannon, la entropía basada en norma euclídea, la entropía basada en norma, la energía logarítmica de la señal y la energía estándar, las cuales se han calculado de acuerdo a la literatura [114].

Como en los anteriores experimentos, para el desarrollo de éste también se llevó a cabo el preprocesamiento descrito en la Sección 4.2. A partir de la matriz de datos obtenida después del pre-clustering, se ha aplicado una búsqueda secuencial flotante hacia adelante empleando un clasificador bayesiano como función de evaluación. Esta es la matriz reducida que se evalúa con los tres clasificadores utilizados en las demás pruebas.

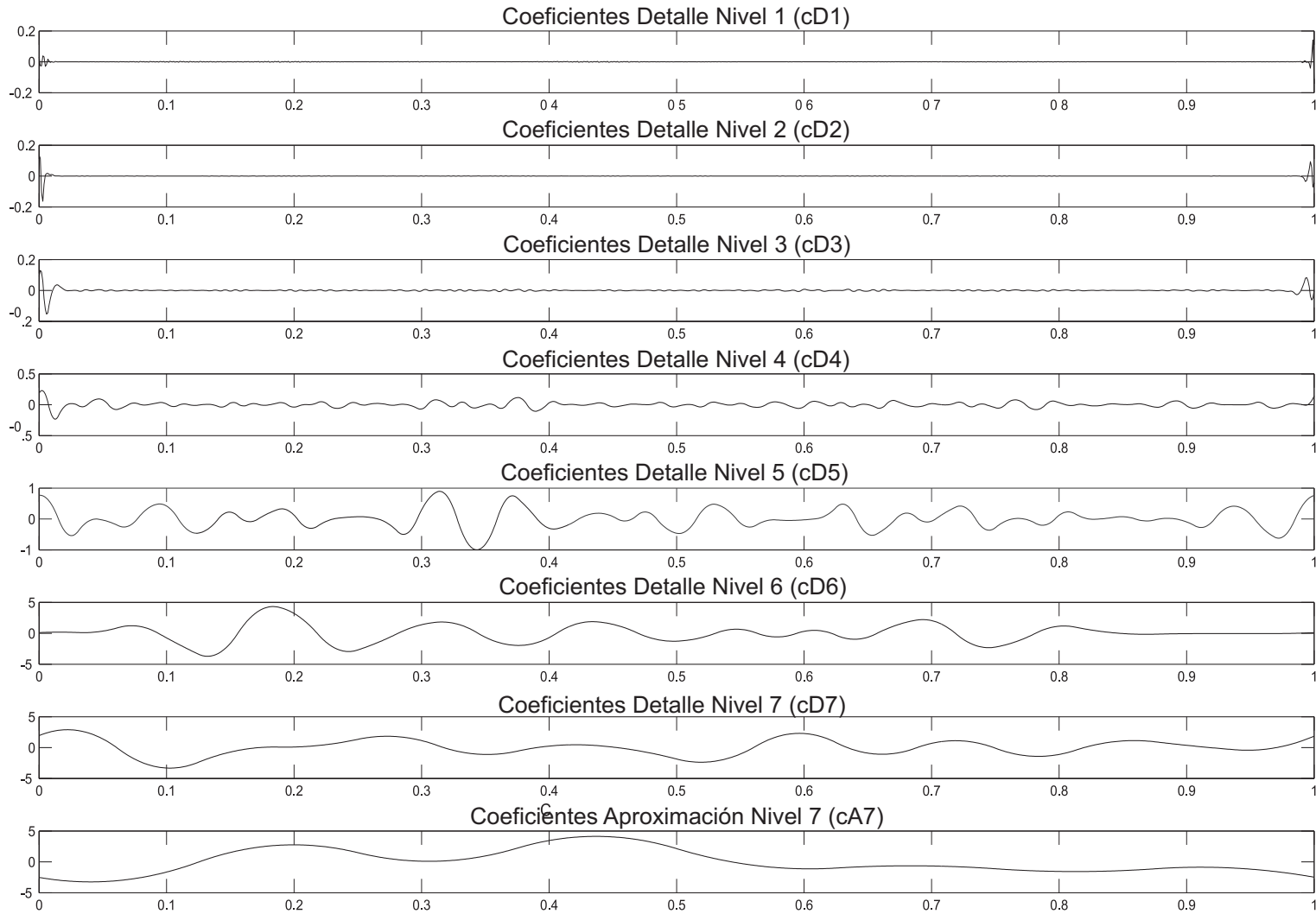


Figura 7-6.: Reconstrucción por niveles del análisis multiresolución

8. Resultados y Discusión

Este capítulo contiene los resultados obtenidos en los diferentes experimentos descritos en el capítulo 7, en los cuales puede apreciarse la capacidad discriminante de las características propuestas en el capítulo 4. Los resultados presentados son evaluados mediante las medidas de desempeño descritas en la Sección 6.2.2; además, se plantean hipótesis sobre la posible interpretación fisiológica de tales resultados.

8.1. Experimento 1

Resultados y discusión

Como se expuso en la Sección 7.4.1, el objetivo principal de este experimento es evaluar la calidad de los datos con los que se trabajará en posteriores experimentos, además de aplicar técnicas de preprocesamiento que garanticen la homogeneidad de los datos.

Dada la matriz de datos \mathbf{X} , obtenida en la etapa de caracterización, constituida por $n = 80$ observaciones y $p = 83$ variables, se procede a aplicar el respectivo preprocesamiento, según se describe en la Sección 4.2. Según el esquema de trabajo planteado en la Sección 7.4.1, en primera instancia se hizo la normalización y centralización de los datos de acuerdo a la Ecuación (4-28). Después, se implementó un algoritmo para la identificación de datos faltantes, donde se contabiliza (de forma univariada) la cantidad de campos en blanco y el número de datos no convergentes. Como no se hallaron datos de esta naturaleza, en esta instancia no fue eliminada ninguna variable ni excluida ninguna observación.

Posteriormente, se aplicó un algoritmo para la detección de datos atípicos basado en el valor de kurtosis, desarrollado como se explicó en la Sección 4.2.2. Debido al número limitado de observaciones, se llevó a cabo el método expuesto en la Sección 4.2.2, y no se eliminaron los datos atípicos sino que se excluyeron las variables que superaban el 10 % de valores anómalos; de esta manera ha quedado conformado un subconjunto de características que en adelante se denominará *subconjunto 1*. La etapa de preprocesamiento finaliza con la prueba de verificación de gaussividad univariada, la cual se desarrolló a partir del *subconjunto 1* con base en el test de Kolmogorov-Smirnov a un nivel de significación de $\alpha = 0,5$. Esta prueba se aplica sobre cada una de las clases, y las variables identificadas como no normales, son eliminadas de ambas clases; el conjunto resultante de esta reducción del número de variables se denominará *subconjunto 2*. En la tabla 8-1 se relacionan los subconjuntos de variables

obtenidos en cada etapa del preprocesamiento.

Tabla 8-1.: Remoción de variables en el preproceso

Etapa del preproceso	Tamaño del subconjunto	N° de variables excluidas	Variables eliminadas	Naturaleza de las variables eliminadas
Inicio del preproceso - Conjunto inicial	83	0	0	- - - -
Identificación de datos faltantes	83	0	0	- - - -
Detección de datos atípicos - <i>Subconjunto 1</i>	81	2	3, 17	Morfológica, espectral
Verificación de gaussividad - <i>Subconjunto 2</i>	76	5	14, 15, 16, 18, 53	Morfológicas, espectral, wavelet

Por otra parte, como una etapa extra en el preprocesamiento de datos se hace una depuración de la misma mediante el procedimiento explicado en la sección 4.2.4. En esta ocasión no se hace reducción de variables como ocurre en las etapas previamente explicadas, sino que se hace exclusión de las observaciones que, después de la formación de *clusters* mediante el algoritmo de *k-medias*, no coincida con la etiqueta asignada por los especialistas médicos.

De la tabla 8-1 puede observarse que al finalizar la etapa de preprocesamiento fueron eliminadas 7 variables, que corresponden a cuatro característica morfológica, dos características espectrales y una wavelet. Esta etapa de preprocesamiento garantiza una correcta aplicación de los métodos de reducción de dimensión al cumplir los requerimientos de normalidad, como es el caso del análisis multivariado de varianza como función de evaluación.

Después de realizada la etapa de preprocesamiento, se procede a realizar la reducción de dimensión mediante la extracción y selección de características, siguiendo los lineamientos explicados en el capítulo 5. La figura 8-1 muestra la gráfica de la matriz de covarianza por clase de \mathbf{X} reducida a partir del preprocesamiento. Esta imagen nos permite continuar con la fase de reducción de dimensión, pues en ella se pueden identificar dependencias lineales que permiten suponer la efectividad de una técnica de extracción y/o selección de características. Aunque no fue posible satisfacer uno de los dos requerimientos del análisis multivariado de varianza (tener distribución normal multivariada y matriz de covarianza por clases iguales), se implementó esta técnica con el fin de comprobar su desempeño, pero teniendo en mente que esta condición podría disminuir la capacidad discriminativa de la prueba. Sin embargo, dado que el algoritmo de MANOVA se implementó para clases balanceadas, es posible que los inconvenientes no se reflejen con severidad.

Con el fin de analizar la efectividad del preproceso aplicado a la base de datos en las etapas posteriores de extracción/selección de características, las pruebas fueron realizadas en la base de datos con y sin preprocesamiento y depuración. Para llevar a cabo la clasificación se emplearon clasificadores básicos de uso común en esta tarea, como los son el *k*-NN y el LDC,

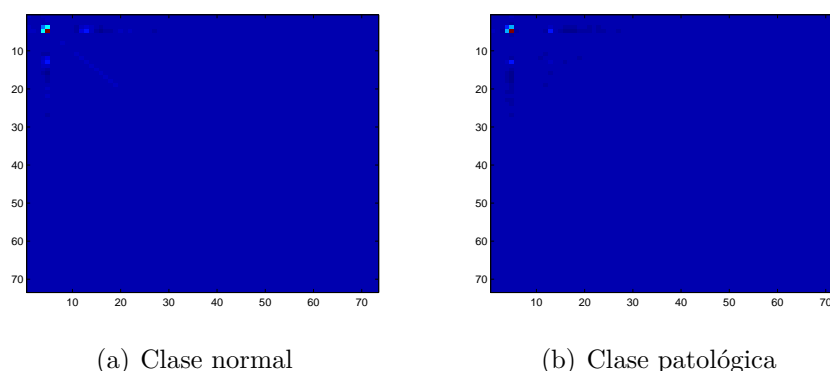


Figura 8-1.: Comparaciones de matrices de covarianza por clases

y uno más robusto que los anteriores, como el basado en SVM. Para la validación de las pruebas se utilizó una partición de 70 % de la base de datos para el grupo de entrenamiento y 30 % para el grupo de verificación. En las siguientes tablas se muestran los resultados obtenidos.

		Clasificador								
		<i>k</i> -NN			LDC			SVM		
		Prec (%)	Sen (%)	Esp (%)	Prec (%)	Sens (%)	Esp (%)	Prec (%)	Sens (%)	Esp (%)
Prueba	BD original	51.87 ± 5.6	58.33	45.41	48.75 ± 6.5	50.00	47.50	51.87 ± 8.7	54.58	49.16
	BD con preproceso	54.64 ± 2.4	24.16	77.5	37.50 ± 7.9	42.50	33.75	35.00 ± 5.1	30.86	32.45
	BD con preproceso	56.12 ± 6.3	62.40	51.76	55.43 ± 4.9	51.12	49.88	55.23 ± 6.1	55.82	48.38
	+ depuración									

Tabla 8-2.: Comparación del preprocesamiento y depuración de la base de datos (BD) en los resultados de clasificación

En la tabla 8-2 se hace una comparación de los resultados obtenidos al someter a clasificación: 1. la base de datos original (sin ningún preproceso), 2. la base de datos obtenida después del preprocesamiento anteriormente descrito, 3. la base de datos obtenida después del preprocesamiento y la depuración mencionada. La tabla muestra un leve pero progresivo incremento en las medidas de rendimiento después de cada una de las etapas evaluadas. Los anteriores resultados indican que efectivamente el preprocesamiento empleado sí tiene una incidencia positiva en la calidad de los datos; además, se puede inferir que la selección de las observaciones (llevada a cabo en la depuración) afecta directamente el rendimiento de clasificación, especialmente en el *k*-NN, que fue el clasificador que mostró mejor desempeño.

La tabla 8-3 relaciona los resultados obtenidos de aplicar la técnica de extracción PCA sobre el subconjunto generado al reducir la matriz **X** mediante la búsqueda heurística SFFS. La semejanza en los valores contenidos en la tabla indica que la reducción de dimensión mediante PCA no surtió un efecto relevante en el rendimiento de clasificación, por tanto, mediante este procedimiento no se alcanzó discriminancia alguna entre las clases evaluadas. No se reportan resultados de la reducción de dimensión usando MANOVA como función

		Clasificador								
		k -NN			LDC			SVM		
		Prec (%)	Sen (%)	Esp (%)	Prec (%)	Sens (%)	Esp (%)	Prec (%)	Sens (%)	Esp (%)
Prueba	Sin PCA	50.00 \pm 5.6	54.58	45.42	45.00 \pm 4.7	48.33	41.67	42.08 \pm 4.9	43.75	40.41
	Con PCA	51.25 \pm 3.8	50.00	52.50	47.92 \pm 4.1	49.16	46.66	44.79 \pm 6.5	45.83	43.75

Tabla 8-3.: Comparación de la extracción de características de la base de datos depurada, seguida de SFFS.

de costo sobre la base de datos depurada, dado que el método de análisis multivariado de varianza se implementó para clases balanceadas (la depuración puede provocar que las clases no contengan igual número de observaciones).

		Clasificador								
		k -NN			LDC			SVM		
		Prec (%)	Sen (%)	Esp (%)	Prec (%)	Sens (%)	Esp (%)	Prec (%)	Sens (%)	Esp (%)
Prueba	BDSD, sin PCA	43.96 \pm 6.4	52.91	35.00	32.29 \pm 5.9	34.58	30.00	34.37 \pm 4.0	36.66	32.08
	BDSD, con PCA	46.25 \pm 9.6	53.75	38.75	33.33 \pm 5.2	35.00	31.66	34.16 \pm 5.9	35.41	32.91
	BDD, sin PCA	77.85 \pm 7.2	65.83	86.87	45.71 \pm 7.1	50.00	76.25	51.07 \pm 9.5	58.33	85.00
	BDD, con PCA	68.57 \pm 7.3	63.33	72.50	45.00 \pm 8.4	10.83	70.63	53.93 \pm 5.1	3.33	91.97

Tabla 8-4.: Comparación del preprocesamiento y depuración, seguida de la extracción de características de la base de datos en los resultados de clasificación

En la tabla 8-4 se hace una comparación de los resultados obtenidos después de aplicar SFFS con un clasificador bayesiano como función de costo y seguido de la reducción de dimensión sobre la base de datos depurada (BDD) y sin depurar (BDSD). En la tabla puede observarse que el mejor rendimiento de clasificación fue alcanzado con la base de datos depurada; además, de las dos pruebas realizadas sobre la BDD, el clasificador k -NN ofrece mayor precisión cuando no se ha aplicado la técnica de reducción. Lo anterior indica que la proyección de los datos sobre los componentes principales no aporta mayor discriminancia entre clases que con la caracterización original.

La figura 8-2 muestra la sintonización del clasificador k -NN utilizado en la prueba que alcanzó mejor rendimiento (SFFS con clasificador bayesiano como función de evaluación y sin PCA, sobre base de datos depurada). Esta misma tendencia fue mostrada por el clasificador en las pruebas restantes, por tal motivo no hubo inconvenientes con la selección del número de vecinos, ya que desde $k = 1$ se alcanza el porcentaje de clasificación en el que el clasificador se estabiliza.

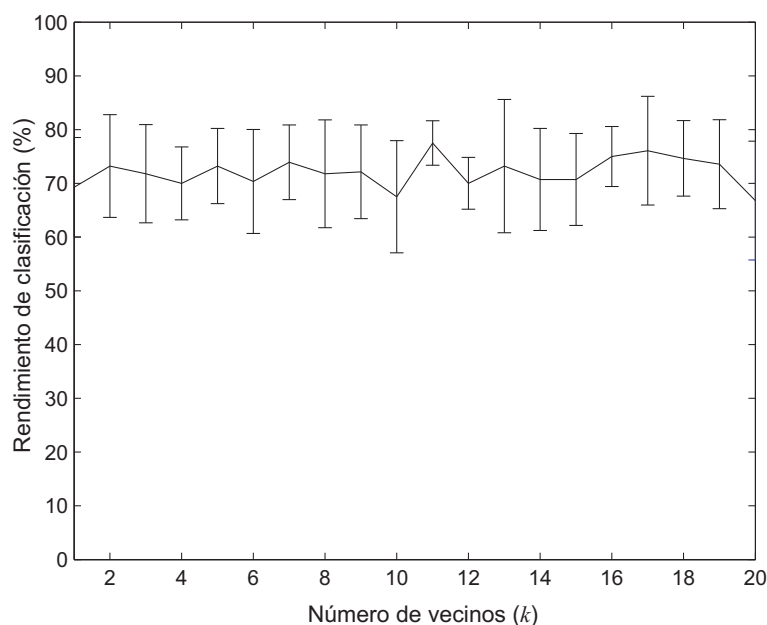


Figura 8-2.: Sintonización de clasificador k -nn usado sobre base de datos depurada, después de aplicar SFFS con clasificador bayesiano como función de evaluación y sin PCA

8.2. Experimento 2

Resultados y discusión

En este experimento se pone a prueba la metodología propuesta en este trabajo (descrita en la Sección 7.4.2), la cual consiste en la re-caracterización de la matriz \mathbf{X} a través de las distancias entre los datos y los respectivos agrupamientos formados mediante el algoritmo de k -medias (empleando la distancia euclidiana como medida de disimilitud). Para este fin, se parte del subconjunto de características que se genera después del procesamiento y del pre-clustering (como se hizo en el experimento anterior - Sección 8.1), y luego de la proyección lineal truncada de los datos con PCA (usando un criterio de varianza acumulada mayor al 90 %). En la figura 8-3 se muestra el esquema general de la metodología aplicada para el desarrollo de este experimento.

Adicionalmente, en esta prueba se desea explorar la capacidad discriminante de los tres grupos de características (morfológicas, espectrales y wavelet) como patrones descriptivos de las señales ERP, a fin de discriminar sujetos normales de pacientes con TDAH. Para llevar a cabo esta prueba, la matriz de datos \mathbf{X} es dividida en tres subconjuntos correspondientes a los tres grupos de características mencionados. Cada uno de estos subconjuntos es sometido a la metodología propuesta para definir el grupo más relevante, comparando el rendimiento de clasificación alcanzado por cada uno de ellos.

Cabe resaltar que en la etapa clasificatoria, al igual que en el anterior experimento, se usaron los clasificadores k -NN, LDC, y SVM, con particiones del 70 % de los datos para el entrenamiento del clasificador y 30 % para la validación.

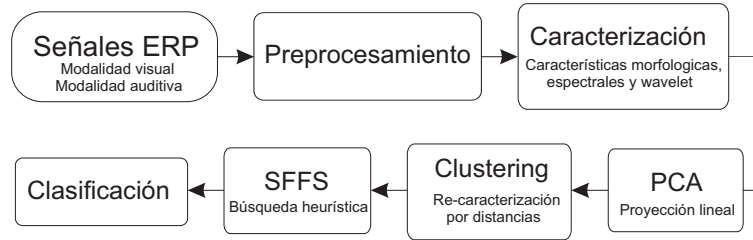


Figura 8-3.: Esquema de la metodología aplicada en experimento 2

En la figura 8-4 se muestra el desempeño de un clasificador k -NN en continuas repeticiones para comprobar la estabilidad de la metodología propuesta. En esta gráfica se observa que todos los valores del rendimiento de clasificación superan el 80 % y mantiene una desviación estándar no muy alta.

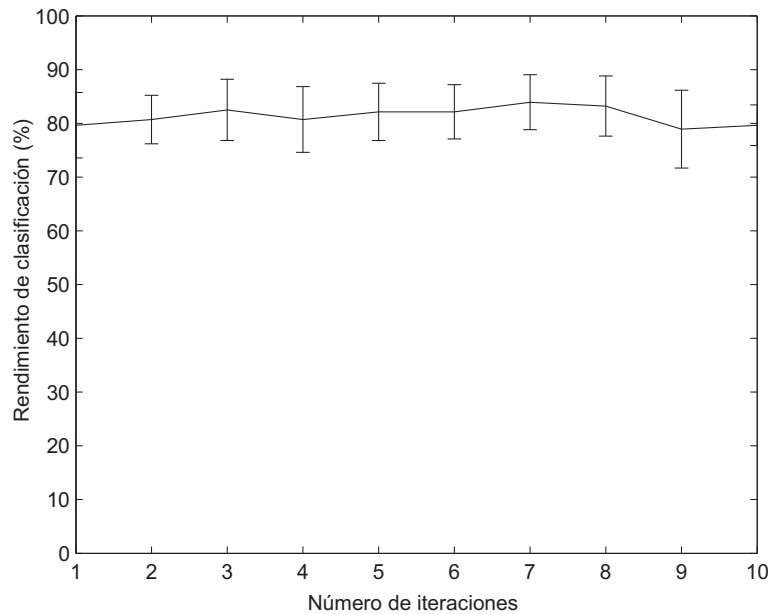


Figura 8-4.: Estabilidad de la metodología 2

La gráfica 8-5 muestra el rendimiento de clasificación alcanzado por los tres clasificadores al evaluar los subconjuntos obtenidos después de la búsqueda heurística. Tales subconjuntos están conformados de la siguiente manera: el primer subconjunto consta de la primera característica seleccionada; el segundo, está constituido por el anterior subconjunto sumado a la segunda característica seleccionada, y así sucesivamente. En esta figura se puede apreciar el efecto de la etapa de reducción de dimensión en la precisión del clasificador, dado que el

porcentaje de rendimiento va aumentando progresivamente, al tiempo que se incrementa el número de características seleccionadas.

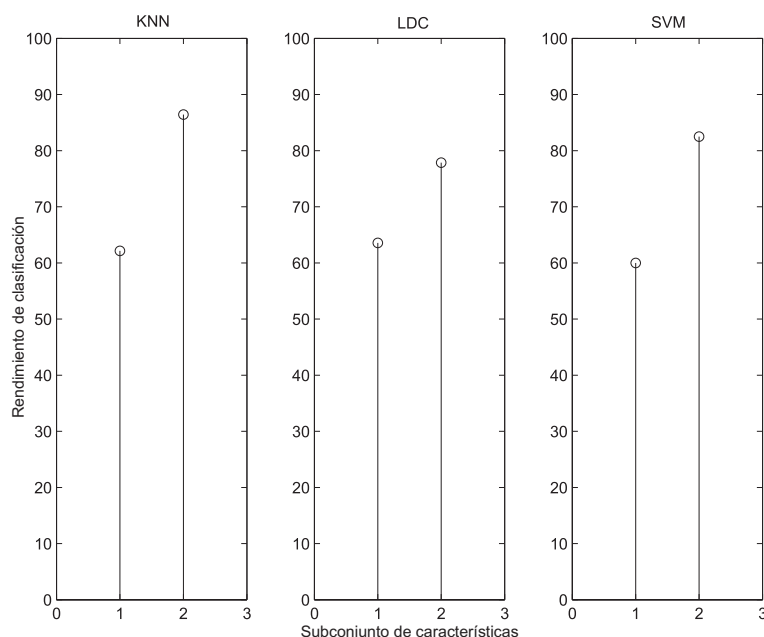


Figura 8-5.: Rendimiento de clasificación por subconjuntos de características

La tabla 8-5 muestra la precisión del clasificador, la sensibilidad y especificidad alcanzadas por cada grupo de características. Se puede observar que de los tres grupos de características que conforman la matriz **X**, el conjunto de características morfológicas es el que obtiene mayor porcentaje en todas las medidas de rendimiento. Lo anterior nos lleva a inferir que probablemente la mayoría de características (trasformadas ahora en distancias tras la re-caracterización) seleccionadas después de la búsqueda heurística y posteriormente llevadas al clasificador, están directamente asociadas a parámetros que en el espacio original de características se relacionan con la morfología de la señal.

Feature	Accuracy (%)	Specificity (%)	Sensitivity(%)
Morphological	85,35 ± 3,9	85,00	85,83
Spectral	63,92 ± 8,6	73,12	51,66
Wavelet	67,85 ± 8,5	73,12	60,83

Tabla 8-5.: Rendimiento de clasificación por grupos de características

La tabla 8-6 muestra los resultados de clasificación obtenidos con una matriz de datos **X** sometida al proceso de depuración (por pre-clustering). Comparando estos valores con los obtenidos empleando la base de datos depurada (ver tabla 8-7), se puede apreciar que la tasa de acierto de los clasificadores es mucho menor, lo cual puede atribuirse a la presencia de datos atípicos y heterogeneidad de los datos. De esta manera se comprueba la necesidad

de hacer un riguroso preprocesamiento a los datos; además, se pone en evidencia la baja confiabilidad de las etiquetación realizada por los especialistas.

Classifier	Accuracy (%)	Specificity (%)	Sensitivity(%)
k -NN	$56,04 \pm 5,6$	62,91	49,16
LDC	$48,54 \pm 6,5$	42,08	55,00
SVM	$48,33 \pm 4,2$	24,58	72,08

Tabla 8-6.: Medidas de rendimiento de clasificación para base de datos sin depurar

La tabla 8-7 contiene los porcentajes de precisión de los tres clasificadores considerados, que se obtuvieron empleado la matriz de datos original (la cual integra los tres grupos de características) al inicio de la aplicación de la metodología. En esta tabla se aprecia el alto desempeño que alcanzan los clasificadores tras la ejecución de la metodología propuesta, especialmente el k -NN, con el cual se obtiene un óptimo y estable porcentaje tanto en la precisión de clasificación como en la sensibilidad y especificidad. En comparación con los resultados obtenidos en el experimento anterior, es posible atribuir esta mejora en el rendimiento de clasificación a la re-caracterización mediante distancias entre datos y centroides, dado que hasta la etapa de proyección lineal, en ambos experimentos se consideraron los mismos procedimientos.

Classifier	Accuracy (%)	Specificity (%)	Sensitivity(%)
k -NN	$86,07 \pm 3,5$	85,00	87,50
LDC	$73,92 \pm 7,1$	82,5	62,50
SVM	$78,57 \pm 4,7$	81,25	75,00

Tabla 8-7.: Medidas de rendimiento de clasificación para base de datos depurada

8.3. Experimento 3

Resultados y discusión

Este experimento está orientado a probar la contribución del subconjunto de características morfológicas (en particular de la latencia) en la formación de agrupamientos compactos y bien definidos que puedan asociarse a las dos clases evaluadas: casos y controles. Para alcanzar este objetivo, se consideraron dos matrices de datos: la primera de ellas corresponde a \mathbf{X} (matriz de características extraídas de los registros originales); la segunda, es una matriz de datos estimada a partir de registros alineados de acuerdo a la ubicación del componente P300 (para llevar a cabo el alineamiento de los registros se siguió el procedimiento descrito en la Sección 7.4.3).

Para probar la influencia de la latencia se llevaron a cabo varias pruebas, cada una con una diferente configuración de \mathbf{X} . Para la formación de los agrupación se hizo uso de una técnica no supervisada tradicional como lo es el algoritmo de k -medias, para calcular las distancias

entre los datos y los centroides de los grupos estimados. Para evaluar la calidad de los grupos, se ha empleado una medida, introducida en este trabajo, denominada *Vector de valores relativos*, que calcula una distancia relativa entre los *clusters* y evalúa la compacidad de los mismos.

Como se se explicó en la Sección 7.4.3, la matriz de datos \mathbf{X} se definió de acuerdo a las cuatro diferentes pruebas que se llevaron a cabo:

1. En la primera prueba, los *clusters* fueron formados usando únicamente la latencia como vector de características \mathbf{X}_1 .
2. La segunda prueba consistió en aplicar el algoritmo de agrupamiento sobre el subconjunto formado por las 16 características morfológicas, incluyendo la latencia; así se forma la matriz \mathbf{X}_2 .
3. En la tercera prueba, para la formación de los grupos, se empleó un subconjunto conformado por 15 características morfológicas, el cual no incluye la latencia (\mathbf{X}_3).
4. En la última prueba, para la formación de la matriz \mathbf{X}_4 , se repite el segundo procedimiento previamente descrito, pero en esta ocasión los agrupamientos se forman a partir de una matriz de datos extraída de las señales ERPs alineadas mediante la técnica expuesta en la Sección 7.4.3

La tabla 8-8 muestra los valores relativos de las distancias calculadas en cada una de las pruebas llevadas a cabo. En ésta se puede observar que el mayor valor relativo se obtuvo en la cuarta prueba, en la cual los grupos fueron formados usando la matriz de datos estimada a partir de las señales ERP alineadas. El hecho de alinear la onda P300 en el mismo punto de dominio del tiempo, implica que se asignará el mismo valor de latencia a todos los registros por clase. Por tanto, si la distancia que separa a un cluster de otro es mayor que las obtenidas en las pruebas anteriores, entonces puede deducirse que la latencia no es una característica fuertemente relevante a la hora de clasificar, y el rendimiento alcanzado podría atribuirse a otras características morfológicas.

Vector de valores relativos	Prueba 1	Prueba 2	Prueba 3	Prueba 4
Valor relativo \mathbf{v}_1	0,0431	0,3731	0,3883	0,5124
Valor relativo \mathbf{v}_2	0,0525	0,3711	0,6171	1,3546

Tabla 8-8.: Vectores de valores relativos.

La anterior afirmación se puede también evidenciar en las figuras 8-6 y 8-7 (que fueron obtenidas con las matrices \mathbf{X}_2 y \mathbf{X}_3 respectivamente), en las cuales se muestrna los diagramas de dispersión bi-dimensionales de las características de interés. Para apreciar el efecto de la

latencia con respecto a la separabilidad y compacidad de los grupos formados, en la figura 8-6 se demuestra la interacción entre la latencia y las características morfológicas relacionadas al área de las señales ERP; se muestran estas parejas de características dado que, entre todas las características morfológicas, fueron aquellas con las cuales la latencia formó los grupos menos solapados y más compactos. En general, se puede observar que tal interacción entre las características mencionadas se expresa gráficamente como un solapamiento relativamente grande de los grupos formados; además, el tamaño de las circunferencias que contienen los agrupamientos indica que la variabilidad de los datos pertenecientes a cada grupo es mayor que aquella mostrada en las figuras 8-7 y 8-8.

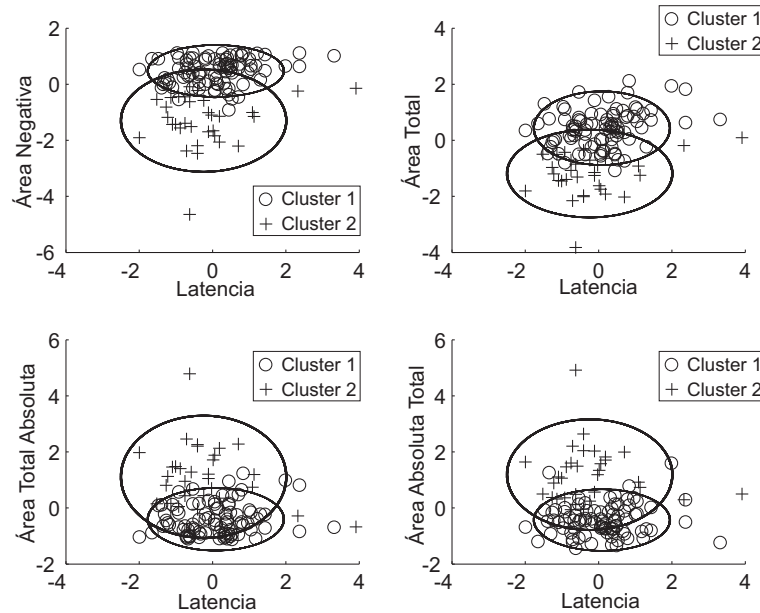


Figura 8-6.: Agrupamientos formados en el espacio de características original: latencia vs otras características morfológicas

Para propósitos de una mejor visualización del método de agrupamiento empleado sobre la base de datos, se escogieron sólo aquellas parejas de características cuyo diagrama de dispersión muestra los *clusters* mejor definidos. Estos grupos fueron seleccionados bajo un criterio de máxima distancia entre los centroides de cada grupo. La figura 8-7, que fue obtenida aplicando *clustering* a la matriz \mathbf{X}_3 , muestra unos diagramas de dispersión donde se forman grupos más compactos y más separados que los exhibidos en la gráfica 8-6. Sin embargo, es evidente que aún continúa existiendo algún solapamiento entre los grupos, y además, los datos están más dispersos en comparación con los *clusters* de la figura 8-8 (que se obtuvo con las señales alineadas).

El diagrama de dispersión de la figura 8-8 fue obtenido a partir de la matriz \mathbf{X}_4 , calculada para la señales alineadas. En esta gráfica, al igual que en las anteriores, de entre todas las posibles parejas de características morfológicas, se muestran aquellas que contienen los gru-

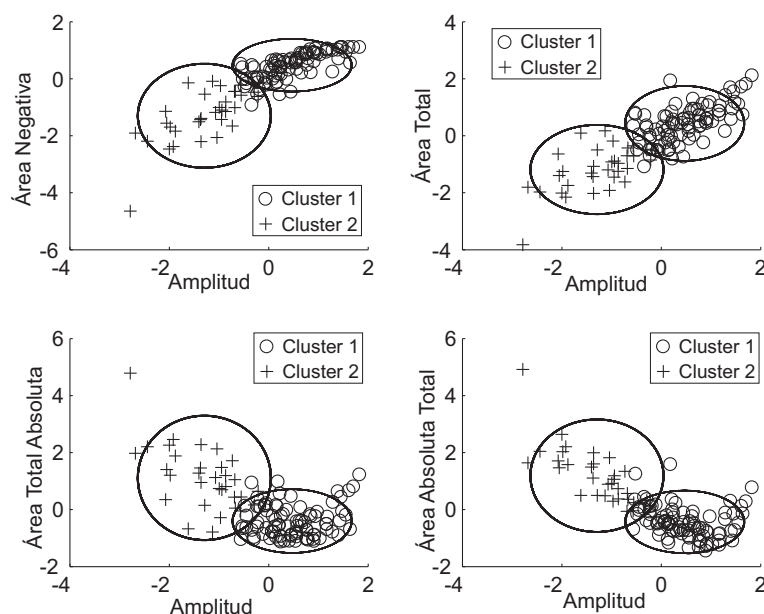


Figura 8-7.: Agrupamientos formados en el espacio de características que excluye la latencia

pos mejor definidos, que fueron generados al llevar a cabo la prueba N° 4. En concordancia con la quinta columna de la tabla 8-8, en la figura 8-8 se pueden observar agrupaciones mucho más separadas entre sí, y sin solapamiento alguno, además de mostrar grupos bien definidos y con mayor compacidad que en el resto de figuras. En conclusión, puede apreciarse que la interpretación de las figuras 8-6, 8-7 y 8-8 corresponden en su orden a los valores relativos exhibidos en la tabla 8-8; de esta manera, se da una mayor separabilidad entre los agrupamientos cuando el valor relativo va también en aumento.

Finalmente, es importante mencionar que la literatura ha documentado alteraciones (incrementos o decrementos) de la latencia en niños con TDAH, con respecto a un grupo control, mostrando éste como un parámetro discriminante entre las clases evaluadas. Sin embargo, con este experimento se ha demostrado que la latencia puede no tener una influencia determinante en la formación de grupos definidos asociados a las clases normal y patológica.

8.4. Experimento 4

Resultados y discusión

Como se explicó en la Sección 7.4.4, el cuarto experimento consiste en hacer la transformada wavelet discreta de las señales ERP usando funciones wavelet diseñadas exclusivamente para determinada aplicación. Para el desarrollo de este experimento, se parte de la base de datos resultante del preproceso y la depuración realizada en el experimento 1. A continuación se definen los parámetros de sintonización empleados en la personalización de la función wavelet

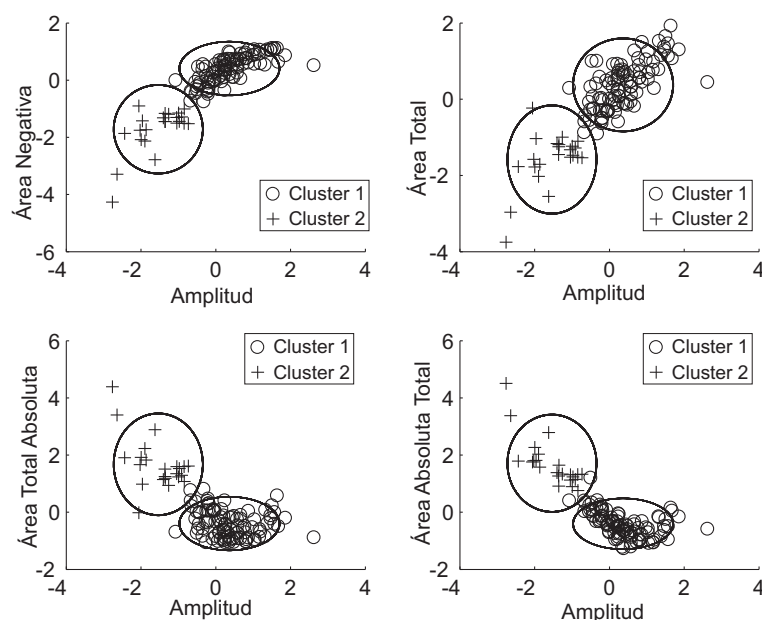


Figura 8-8.: Agrupamientos formados en espacio de características generado por las señales alineadas

y la optimización del espacio tiempo-frecuencia.

8.4.1. Sintonización de la función wavelet

Los parámetros de los algoritmos genéticos (GA) utilizados para la optimización de la función wavelet, fueron los siguientes: cruce aritmético y mutación no uniforme; la escala de la población es de 20, y el número de generaciones es de 50. Además para evitar la variabilidad de los resultados debido a la naturaleza iterativa de los GA, el procedimiento previamente planteado es repetido 10 veces. Como resultado, 10 diferentes actualizaciones y operadores predictivos fueron obtenidos, de los cuales la dupla en el valor medio se selecciona para los siguientes pasos. En la tabla 8-9 se presentan los valores de los operadores calculados.

Operadores	Coeficientes			
p	0,1207	0,3792	0,3792	0,1207
u	0,0869	0,1630	0,1630	1,0869

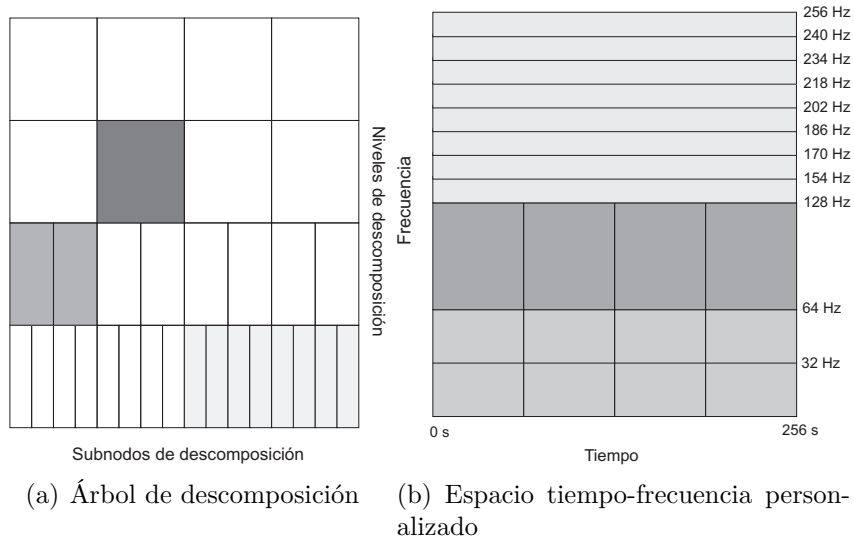
Tabla 8-9.: Coeficientes de los operadores predictivos

Aunque el proceso de sintonización muestra estabilidad sobre las señales de entrada y los resultados, es importante notar que empleando una base de datos no muy numerosa como la nuestra, los patrones en el conjunto de entrenamiento no deben incluir incertidumbre en los GA. De hecho, en bases de datos pequeñas se recomienda incluir más del 40 % de los datos para la sintonización de los GA.

8.4.2. Optimización del espacio tiempo-frecuencia

La optimización de la resolución del espacio tiempo-frecuencia es una tarea crucial cuando se emplea la transformada *wavelet packet* (WPT), dado que la adecuada selección de los subespacio podría revelar información importante. En este experimento, el conjunto de entrenamiento es descompuesto con WPT a 4 niveles, empleando los operadores u y p mencionados. Los registros del conjunto de entrenamiento son previamente filtrados con un filtro pasa bajos de rizado uniforme con banda de paso en 30 Hz y 256 ventanas. Se escogió una descomposición a 4 niveles porque la banda de frecuencia de interés (0 – 16 Hz) es obtenida en tal nivel. Una vez el árbol de descomposición binaria es computado, entonces $2^{j+1} - 1 = 31$ (con $j = 4$) subespacios son embebidos. Los subespacios óptimos, en este caso, son aquellos que maximicen el índice de separación de grupos (*cluster separation index- CSI*).

El espacio optimizado se muestra en la figura 8.9(b). Es importante mencionar que para obtener una representación completa de la señal de entrada, el árbol óptimo debe cubrir todo el dominio del tiempo, como se muestra en esa figura. Adicionalmente, se puede ver que la resolución temporal más alta se obtiene en la banda de frecuencia donde se han reportado los eventos neuronales (menor a 32 Hz). Además, se puede ver el efecto de la etapa de pre-filtrado en las frecuencias por encima de los 60 Hz, donde la resolución es muy alta debido al ruido que queda. Tal configuración del espacio tiempo-frecuencia ha sido posible únicamente por la optimización de la selección de los subespacios.



Con respecto a la extracción de características después de la sintonización, el resultado es un conjunto óptimo de coeficientes para cada registro ERP. Este conjunto de coeficientes es re-distribuido en el correspondiente espacio tiempo-frecuencia para cubrir tanto el dominio del tiempo completo (0 – 256 s) como todo el espectro de frecuencias (0 – 256 Hz). Entonces, las métricas mencionadas en la Sección 7.4.4 son extraídas.

8.4.3. Clasificación

En este experimento se emplearon clasificadores básicos para determinar cuáles registros son normales y cuáles patológicos. Se utilizó una partición del 70 % de la base de datos para entrenar los clasificadores y 30 % para la etapa de verificación. Además, el experimento se sometió a 10 *folds* para reducir la variabilidad de los resultados. Los resultados se dan en términos de precisión de clasificación, sensibilidad y especificidad.

Classifier	Accuracy (%)	Specificity (%)	Sensitivity(%)
<i>k</i> -NN	73,33 \pm 7,5	63,40	85,25
LDC	83,81 \pm 7,1	81,12	83,78
SVM	81,26 \pm 4,7	78,68	88,21

Tabla 8-10.: Medidas de rendimiento de clasificación usando *customized wavelet transform*

Según la tabla 8-10, con la metodología aplicada se ha alcanzado un rendimiento óptimo y equiparable a los resultados obtenidos en el experimento 2. Sin embargo, aunque se ha comprobado que esta metodología es capaz de discriminar entre las clases normal y patológica, en los bajos porcentajes de la especificidad se hace evidente que aún continúan presentándose falsas alarmas que no permiten un mejor desempeño de los clasificadores.

8.5. Experimento 5

Resultados y discusión

En el quinto experimento se desea poner a prueba la capacidad discriminante de un pequeño conjunto de características de naturaleza diferente a la de los atributos definidos en la Sección 4.1, que fueron empleados en los tres primeros experimentos. Se trata de estimaciones de energía basadas en entropía, que de igual manera están directamente relacionadas a la morfología de la señal. Para cuantificar la capacidad de generalización de dichas características y compararlas en términos de rendimiento con aquellas inicialmente planteadas, se empleará el método basado en proyección lineal y agrupamiento, aplicado en el experimento 2 (ver Sección 7.4.2). Al igual que en los casos anteriores, en este experimento se tuvieron en cuenta tres clasificadores, las mismas tres medidas de desempeño, y una partición de la base de datos del 70 % y 30 % para el entrenamiento y la validación, respectivamente.

En la figura 8-9 se observa el efecto de la etapa de selección de características, dado que la precisión del clasificador va aumentando a medida que se incrementa el número de características evaluadas. Comparando esta figura con la figura. 8-5, se aprecia que el rendimiento alcanzado con un subconjunto óptimo de características no es equiparable al logrado en el experimento 2. Debido a que se aplicó la misma metodología en ambas pruebas, el bajo rendimiento se puede atribuir a la calidad de los parámetros utilizados en el presente expe-

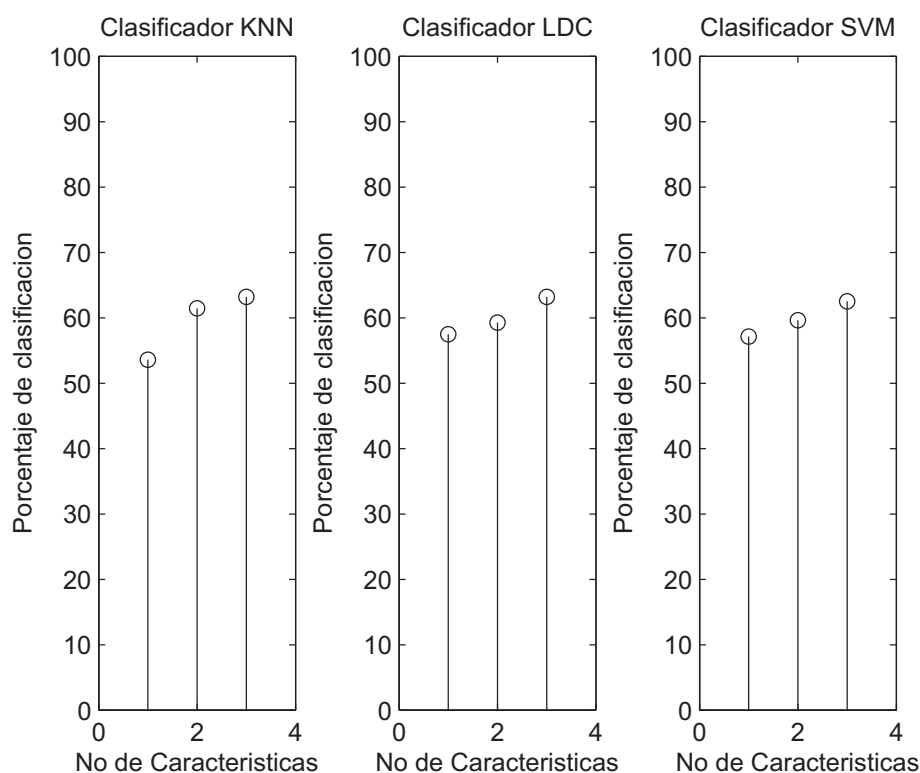


Figura 8-9.: Rendimiento de clasificación por subconjuntos de características

rimento. De esta manera se puede inferir que el conjunto de datos \mathbf{X} tiene una capacidad de generalización mucho mayor que las características basadas en entropía.

Classifier	Accuracy (%)	Specificity (%)	Sensitivity(%)
k -NN	67,50 \pm 8,1	68,33	66,87
LDC	62,14 \pm 7,5	33,33	83,75
SVM	62,14 \pm 5,8	23,33	91,25

Tabla 8-11.: Medidas de rendimiento de clasificación usando características informativas

La anterior afirmación es sustentada también en los resultados mostrados en la tabla 8-11, en la cual se aprecian los bajos porcentajes alcanzados por los tres clasificadores tanto en precisión como en sensibilidad y especificidad. Las diferencias entre los valores de estas dos últimas medidas de desempeño sugieren que se está clasificando correctamente la clase patológica, pero están ocurriendo numerosos falsos positivos al momento de clasificar la clase normal.

Es importante mencionar que los resultados obtenidos señalan que en el presente experimento no se escogieron las características adecuadas para extraer información relevante y discriminante de los ritmos a la hora de clasificar, más no indican que los ritmos generados sean señales inapropiadas para caracterizar los diferentes eventos neurológicos que ocurren

en el cerebro.

Parte V.

Comentarios finales

9. Conclusiones y Trabajo Futuro

Este capítulo presenta las conclusiones obtenidas durante el desarrollo de este trabajo de investigación y se proponen las posibles líneas de trabajo futuro a seguir; además, se desatacan de forma explícita los aportes hechos con esta tesis.

9.1. Conclusiones

Un conjunto óptimo de patrones descriptivos de la señal ERP puede lograrse a partir de características morfológicas, espectrales y coeficientes wavelet . En este trabajo se ha definido un grupo de patrones que han mostrado tener una aceptable capacidad discriminante, logrando la correcta clasificación de las clases patológica y normal en un 86 %. Para obtener un buen desempeño de los clasificadores es indispensable contar con una etapa de preprocesamiento que identifique los valores atípicos y las variables no gaussianas, adicional a una depuración de la base de datos que la haga más homogénea.

Debido a la naturaleza de las señales ERP y a la baja fiabilidad del etiquetado emitido por los especialistas, la identificación del TDAH representa una tarea compleja tanto para el campo de la medicina como para el reconocimiento de patrones. El diseño de un sistema de clasificación para la discriminación entre sujetos normales y patológicos requiere de una nueva representación de datos dado que las muestras de las señales o ciertos atributos morfológicos y espectrales de la misma, no son suficientes para obtener una buena separabilidad entre clases. En respuesta a este inconveniente, en este trabajo se desarrolló un enfoque que contempla una proyección lineal óptima en términos de mínimos cuadrados, seguida de una re-caracterización basada en técnicas de agrupamiento. Esta metodología mejora el rendimiento de clasificación en comparación con la clasificación llevada a cabo sobre el conjunto original de características. Tal hecho se puede atribuir a que la proyección lineal obtiene una mejor representación de datos en términos de compacidad, y entonces un proceso de extracción de características basado en distancias entre los datos y los centroides de los respectivos grupos, tendería a mejorar el rendimiento de clasificación; de esta manera estas dos etapas resultan coherentes y complementarias.

La correcta ubicación del componente P300 sobre la forma de onda resulta ser fundamental para hacer un análisis apropiado de los registros ERP, ya sea de forma automática o simplemente visual, dado que de ese procedimiento depende la estimación de parámetros como la latencia, amplitud, valor pico a pico, entre otras. En respuesta a este requerimiento, en este

trabajo se ha propuesto un algoritmo basado en la señal derivada y la distancia euclidiana de sus pendientes, que no sólo considera las definiciones teóricas de la latencia sino que además tiene en cuenta otros criterios médicos reales que visualmente son evaluados en la señal ERP.

En términos de reconocimiento de patrones, el diagnóstico automático de patologías está relacionado al agrupamiento de patrones homogéneos de tal forma que las clases de interés puedan ser identificadas. Por tanto, las características a ser analizadas deben representar apropiadamente las señales consideradas con el fin de generar una buena separabilidad. Características diagnósticas tales como la latencia y otras morfológicas basadas en la ubicación del P300, son tenidas en cuenta en el diseño de un sistema automático para la detección de TDAH, en particular, la latencia, que intuitivamente puede ser importante en la inspección visual. Sin embargo, clínicamente aún no se determina un consenso que defina el tipo de alteración que sufre la latencia de acuerdo a una determinada patología. Con el desarrollo de este trabajo se puede concluir que, en comparación con otras características morfológicas, la latencia podría no ser una característica fuertemente relevante para llevar a cabo un proceso de agrupamiento de alta calidad cuantificado en términos de separabilidad y compacidad.

La cuantificación del desempeño del *clustering* ha sido siempre un problema abierto en el contexto de los métodos no supervisados. Dado que en estos métodos no es coherente usar medidas supervisadas para medir la calidad del clustering, se han desarrollado medidas basadas en el valor de la función objetivo, coherencia de *clusters* y, en general, la naturaleza de los datos. En este trabajo se comprobó que los valores relativos obtenidos de las distancias entre las observaciones y los centroides pueden dar un indicio de la calidad del agrupamiento porque cuantifican la varianza intra y entre clases cuando se considera la norma euclidiana.

9.2. Trabajo futuro

Como trabajo futuro de esta tesis se propone:

1. Continuar con el estudio y evaluación del conjunto de características propuesto en esta tesis, aplicando técnicas de extracción y selección más robustas.
2. Emplear otro tipo de patrones que describan los cambios generados de forma continua en el dominio del tiempo (características dinámicas).
3. Estudiar la posible no linealidad de los fenómenos neurológicos manifestados en la señal ERP, mediante el uso de técnicas de dinámica no lineal.
4. Considerar el uso de técnicas de agrupamiento no supervisado más complejas que las particionales, y usando un valor de k mayor a dos, de manera que se pueda jerarquizar para identificar subtipos de TDAH y, en general, subgrupos en ambas clases.

5. Realizar el análisis de los potenciales evocados, partiendo del filtrado de la señal EEG mediante técnicas como el análisis de componentes principales (ICA).

10. Discusión académica

- *“Evaluation of Feature Extraction Techniques on Event-Related Potentials for Detection of Attention-Deficit/Hyperactivity Disorder”*. P. Castro-Cabrera, J. Gómez-García, F. Restrepo, O. Moscoso, G. Castellanos-Domínguez. EMBC 2010 - 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Buenos Aires - Argentina, August 31 - September 4, 2010.
- *“Detection of Attention-Deficit/Hyperactivity Disorder based on Customized Wavelet”*. R.D. Pinzon-Morales, F. Restrepo, O. Moscoso, P.A. Castro-Cabrera, A.A. Orozco-Gutiérrez, C.G. Castellanos-Domínguez. CASEIB 2010 - XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica. Madrid - España, 24, 25 y 26 de noviembre de 2010.
- *“Methodology for attention deficit/hyperactivity disorder detection by means of Event-related potentials”*. P.A. Castro-Cabrera, J.A. Gómez-García, F. Restrepo, O. Moscoso, G. Castellanos-Domínguez. IWINAC 2011 - International Work-conference on the Interplay between Natural and Artificial Computation. Canary Islands - Spain, May 30 - June 3, 2011.
- *“ADHD identification based on a linear projection and clustering”*. P.A. Castro-Cabrera, D.H. Peluffo-Ordoñez, F. Restrepo, G. Castellanos-Domínguez. BMEI 2011 - 4th International Conference on BioMedical Engineering and Informatics. Shanghai, China, 15-17 October 2011. En revisión
- *“Latency as discriminant parameter between P300 recordings of ADHD and healthy control children”*. P.A. Castro-Cabrera, D.H. Peluffo-Ordoñez, F. Restrepo, G. Castellanos-Domínguez. BIOSTEC 2012. En revisión.

A. Anexo: Criterios Diagnósticos del TDAH, según DSM-IV

El diagnóstico de TDAH se da si [120]:

- Se cumplen todos los criterios esenciales, y
- 6 ó más de los 9 ítems de déficit de atención, y/o
- 6 ó más de los 9 ítems de hiperactividad-impulsividad

Criterios Esenciales

- *Duración:* Los criterios sintomatológicos de inatención y/o hiperactividad-impulsividad deben haber persistido al menos los últimos 6 meses.
- *Edad de comienzo:* Algunos síntomas de hiperactividad-impulsividad o de inatención causantes de deterioro deben haber estado presentes antes de los 7 años.
- *Ubicuidad:* Algún grado de disfunción debida a los síntomas deben haber estado presentes en dos situaciones o más (escuela, trabajo, casa, etc.).
- *Disfunción:* Debe existir evidencia clara de alteración clínica significativa en el funcionamiento social, académico, familiar, etc. Un niño puede cumplir los 18 criterios del TDAH pero si no le afectan su vida diaria no es un niño hiperactivo.
- *Discrepancia:* Los síntomas son excesivos comparando con otros niños de la misma edad y CI.
- *Exclusión:* Los síntomas no se explican mejor por la presencia de otro trastorno mental.

Criterios de Inatención

1. A menudo no presta atención suficiente a los detalles o incurre en errores por descuido en las tareas escolares, en el trabajo o en otras actividades.
2. Frecuentemente tiene dificultades para mantener la atención en tareas o en actividades lúdicas.
3. A menudo parece no escuchar cuando se le habla directamente.

4. A menudo no sigue las instrucciones y no termina sus tareas, oficios o responsabilidades en el lugar de trabajo (no explicable por presencia de comportamiento desafiante o por dificultades para comprender las instrucciones).
5. Frecuentemente tiene dificultades para organizar tareas y otras actividades.
6. A menudo evita, le disgusta, o rechaza el comprometerse en tareas que requieran concentración o esfuerzo mental sostenido (como actividades académicas en la escuela o en casa).
7. Con frecuencia extravía los materiales necesarios para realizar una tarea o actividad (ej., juguetes, tareas escolares, lápices, libros o herramientas).
8. A menudo se distrae fácilmente por estímulos irrelevantes.
9. Es frecuentemente olvidadizo en las actividades de la vida diaria.

Criterios de Hiperactividad

1. A menudo mueve en exceso las manos y los pies mientras está sentado.
2. Se levanta frecuentemente de su asiento en la clase o en otras situaciones en que se espera que permanezca sentado.
3. A menudo corre o salta excesivamente en situaciones en las que es inapropiado hacerlo.
4. Con frecuencia tiene dificultades para relajarse o practicar juegos donde deba permanecer quieto.
5. Está permanentemente en marcha, como si tuviera un motor por su interior.
6. A menudo habla excesivamente.

Criterios de Impulsividad

1. A menudo precipita respuestas actuando o contestando antes de haber sido completadas las preguntas.
2. Frecuentemente tiene dificultades para hacer filas o esperar turnos en los juegos.
3. Interrumpe las conversaciones o los juegos de los demás, perturbando a otros.

B. Anexo: Algoritmo K-medias

El algoritmo K-medias, es el método clásico del agrupamiento particional. El algoritmo comienza con una partición inicial asociada a un conjunto de centroides iniciales, y el movimiento de los mismos, que se da para generar nuevas particiones, se evalúa por cada centroide; por tanto, una vez se haya movido un centroide, se realizan las asignaciones de todos los elementos y se calcula el cambio de la función objetivo [107].

Asumiendo que un elemento \mathbf{x}_i perteneciente a un grupo \mathbf{C}_l en la solución actual es reasignado a algún otro grupo \mathbf{C}_j , el cambio de los centroides puede realizarse así:

$$\mathbf{q}_l \leftarrow \frac{n_l \mathbf{q}_l - \mathbf{x}_i}{n_l - 1}, \quad \mathbf{q}_j \leftarrow \frac{n_j \mathbf{q}_j + \mathbf{x}_i}{n_j + 1} \quad (\text{B-1})$$

donde $n_i = n_e(\mathbf{C}_i)$ y $l \neq j$.

El cambio de la función objetivo generado por cada reasignación se puede medir mediante con:

$$v_{ij} = \frac{n_j}{n_j + 1} \|\mathbf{q}_j - \mathbf{x}_i\|^2 - \frac{n_l}{n_l - 1} \|\mathbf{q}_l - \mathbf{x}_i\|^2, \quad \mathbf{x}_i \in \mathbf{C}_l \quad (\text{B-2})$$

La anterior ecuación aplica en caso de que la función objetivo sea la de MSSC, en el caso general debe tenerse en cuenta la naturaleza de la función objetivo.

Los cambios de la función objetivo son calculados para todas las posibles reasignaciones, de forma que si todos los valores son no negativos ($v_{ij} \geq 0$), el algoritmo converge y la solución lograda corresponde a una partición asociada a un mínimo local de la función objetivo [107].

En el Algoritmo 5 se reúnen los pasos del método K-medias.

Es de mencionar que en el algoritmo se numera únicamente los pasos significativos.

Algoritmo 5 Algoritmo de K-medias

1. Inicialización: escoger un valor de k y una partición inicial $\mathbf{C}^{(0)}$ con centroides $\mathbf{Q}^{(0)}$, fijar número máximo de iteraciones N_{iter} , inicializar el contador: $r = 1$

Mientras $r < N_{iter}$

Desde $j = 1$ **hasta** k **hacer**

2. Mover los centroides: $\mathbf{q}_l^{(r)} \leftarrow \frac{n_l \mathbf{q}_l^{(r-1)} - \mathbf{x}_i}{n_l - 1}$, $\mathbf{q}_j^{(r)} \leftarrow \frac{n_j \mathbf{q}_j^{(r-1)} + \mathbf{x}_i}{n_j + 1}$

3. Calcular el cambio de la función objetivo:

$$v_{ij} = \frac{n_j}{n_j + 1} \|\mathbf{q}_j^{(r)} - \mathbf{x}_i\|^2 - \frac{n_l}{n_l - 1} \|\mathbf{q}_l^{(r)} - \mathbf{x}_i\|^2, \quad \mathbf{x}_i \in \mathbf{C}_l^{(r)}$$

Si $v_{ij} \geq 0$ ($i = 1, \dots, n$ y $j = 1, \dots, k$)

4. El proceso termina y la solución es $\mathbf{C}^{(r)}$

en caso contrario

$r \leftarrow r + 1$

Termina Si

Termina Desde

Termina Mientras

C. Anexo: Métodos de Selección de Características

Como se mencionó en la Sección 5.2, la selección de características contempla dos etapas básicas: la estrategia de generación y la función de evaluación. En este sentido, se presentan a continuación, diferentes algoritmos de búsqueda para la generación de subconjuntos, y algunas funciones de evaluación, los cuales son ampliamente referenciados en la literatura de reconocimiento de patrones.

C.1. Algoritmos de búsqueda para la generación de subconjuntos

C.1.1. Búsqueda exhaustiva

Es la aproximación básica al proceso de selección de características. Consiste en conformar todos los subconjuntos posibles a partir de las p variables originales, la cantidad de subconjuntos posibles está dada por,

$$\text{Nº de subconjuntos} = \frac{p!}{(p-m)!m!} \quad (\text{C-1})$$

donde $m \leq p$. Cada conjunto es evaluado por medio de la función de costo f_{ξ_j} determinada, la cual en adelante se llamará \mathcal{J} . El subconjunto que optimice (maximice o minimice, según sea el caso) la función de costo es el seleccionado. La posibilidad de llevar a cabo este procedimiento, está reducida al caso en que la dimensión del espacio de características es relativamente pequeño, por cuestiones prácticas $p \leq 15$ ó 20 . Así, si la dimensión del espacio inicial de características es $p = 20$ el número de subconjuntos a evaluar es 1048575. Lo cual representa un proceso que puede ser muy costoso computacionalmente.

C.1.2. Búsqueda heurística

Mejores m variables individuales

Se evalúa cada una de las p variables de forma individual a través del criterio \mathcal{J} de relevancia asociado a alguna métrica. Posteriormente, se seleccionan las m características que reporten mayores valores de evaluación.

Selección secuencial hacia adelante

La selección secuencial hacia adelante (SFS - *Sequential Forward Selection*) es una técnica de búsqueda abajo-arriba (*bottom-up*). Selecciona primero la mejor variable según el criterio \mathcal{J} , luego se combina la variable original con cada una de las variables restantes, entonces, se busca la pareja que aporta el mayor valor de evaluación y se escoge como nuevo conjunto de partida. A continuación se combina esta pareja con cada una de las variables restantes, formando ternas, se selecciona la terna que dé un mayor valor en el criterio de evaluación. El proceso se repite una y otra vez en la misma forma. La búsqueda se detiene cuando un conjunto de más variables no mejore los resultados de la función de costo para un conjunto de menos variables, es decir cuando

$\mathcal{J}(\Omega_{\hat{k}}) > \mathcal{J}(\Omega_{\hat{k}+1})$, donde la dimensión del subconjunto $\Omega_{\hat{k}}$ es menor que la dimensión del subconjunto $\Omega_{\hat{k}+1}$.

Selección secuencial hacia atrás

La selección secuencial hacia atrás (SBS - *Sequential Backward Selection*), al igual que la técnica anterior, la idea es construir conjuntos diferentes, iteración tras iteración, con la diferencia que ahora se inicia con el conjunto completo de características de dimensión p , y en cada iteración se remueve una variable. La variable que se elimina es aquella que al no estar presente en el subconjunto a evaluar, hace que la función de costo reporte el mayor valor entre todos los subconjuntos evaluados en la misma iteración. El algoritmo se detiene cuando el valor de \mathcal{J} no supera cierta cota preestablecida, es decir, cuando en una iteración ninguno de los subconjuntos formados al ser evaluados supera dicho umbral. Entonces, se selecciona el último subconjunto que al ser evaluado haya superado el umbral. Otro criterio de parada puede ser simplemente seleccionar el tamaño m del subconjunto final.

Adicionar l y quitar r

Este procedimiento permite de alguna forma dar marcha atrás en el proceso de selección de características. Si $l > r$, el procedimiento es abajo-arriba. Se agregan l variables usando selección secuencial hacia adelante y luego las peores r características son removidas empleando selección secuencial hacia atrás. Si $l < r$, entonces el procedimiento es arriba-abajo,

iniciando con el conjunto completo de características, removiendo r , luego agregando l y repetir el proceso sucesivamente.

Selección secuencial flotante

La selección secuencial flotante (SFFS - *Sequential Floating Forward Selection*) se considera como la generalización de la técnica: adicionar l y quitar r . Asumiendo que en la iteración \hat{k} se tiene un grupo de subconjuntos $\Omega_1, \dots, \Omega_{\hat{k}}$ de tamaños 1 a \hat{k} respectivamente. Los correspondientes valores del criterio de selección de características van desde $\mathcal{J}(\Omega_1)$ hasta $\mathcal{J}(\Omega_{\hat{k}})$. Además, el conjunto total de características es Ω . En la iteración \hat{k} , se procede de la siguiente forma,

1. Seleccionar la característica ξ_j de $\Omega - \Omega_{\hat{k}}$ que incremente el valor de \mathcal{J} lo más posible y agregarla al subconjunto actual, $\Omega_{\hat{k}+1} = \Omega_{\hat{k}} + \xi_j$.
2. Hallar la característica ξ_r , en el subconjunto actual, $\Omega_{\hat{k}+1}$, que más reduzca el valor de \mathcal{J} . Si esta característica es la misma que ξ_j (del paso anterior), entonces el conjunto actual no se modifica, se incrementa \hat{k} y se retorna al primer paso. En caso contrario, se remueve la característica del subconjunto actual para formar, $\hat{\Omega}_{\hat{k}} = \Omega_{\hat{k}+1} - \xi_r$.
3. Continuar removiendo las características del conjunto $\hat{\Omega}_{\hat{k}}$ para formar el subconjunto reducido $\hat{\Omega}_{\hat{k}-1}$, mientras que $J(\hat{\Omega}_{\hat{k}-1}) > J(\Omega_{\hat{k}-1})$, $\hat{k} = \hat{k} - 1$, ó $\hat{k} = 2$, entonces continuar al paso 1.

El algoritmo es inicializado con $\hat{k} = 0$ y $\Omega_0 = \text{vacío}$, luego se usa la selección secuencial hacia adelante hasta que el tamaño del subconjunto sea 2.

C.2. Funciones de evaluación

C.2.1. Análisis multivariado de varianza

En MANOVA el criterio de evaluación estadístico es la separabilidad entre clases, que se realiza mediante una prueba de hipótesis sobre la igualdad o desigualdad de los vectores de promedios entre las clases. Se asume que los datos están generados con base en el siguiente modelo:

$$\mathbf{x}_{k,j} = \mathbf{m}_k + \varepsilon_{k,j} \quad \mathbf{m}_k = \mathbf{m} + \alpha_k \quad (\text{C-2})$$

siendo j la observación y k la clase, \mathbf{m}_k es el vector de medias para cada clase y $\varepsilon_{k,j}$ es la respectiva perturbación del modelo, \mathbf{m} es la media global de las clases y α_k es la perturbación sobre esta media global.

La comparación de los vectores de medias de las k clases para encontrar diferencias significativas, se realiza mediante la prueba hipótesis:

$$\begin{aligned} H_0 &= \mathbf{m}_1 = \mathbf{m}_1 = \dots = \mathbf{m}_L \\ H_1 &= \exists \text{ al menos un par } \mathbf{m}_k \neq \mathbf{m}_i; \forall k, i \in \{1, \dots, L\} \end{aligned} \quad (\text{C-3})$$

La estadística de Wilks es comúnmente usada al interior de MANOVA, para probar la hipótesis H_0 , que corresponde a la relación de verosimilitud dada por:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \quad (\text{C-4})$$

la cual es conocida como Λ de Wilks. Siendo \mathbf{H} la matriz de hipótesis que puede entenderse como una medida de dispersion entre la media de las clases, mientras que la matriz de error \mathbf{E} se relaciona con la medida de dispersion entre las observaciones para cada clase. La hipótesis H_0 se rechaza si la dispersion entre los patrones es mayor que la dispersion de las observaciones dentro de los patrones, y así, $\Lambda \in [0, 1]$ tiende a cero. Por otro lado, el Λ de Wilks puede ser similar a un estadístico F , pero de manera inversa. Un valor grande del estadístico F rechaza H_0 [121].

Bibliografía

- [1] C. Capdevila-Brophy, J. Artigas-Pallarés, and J. Obiols-Llandrich, “Tempo cognitivo lento: ¿ síntomas del trastorno de déficit de atención/hiperactividad predominantemente desatento o una nueva entidad clínica?” *Revista de Neurología*, vol. 42, no. Supl2, pp. 127–134, 2006.
- [2] G. Overturf, “American academy of pediatrics. committee on infectious diseases,” Technical report: prevention of pneumococcal infections, including the use of pneumococcal conjugate and polysaccharide vaccines and antibiotic prophylaxis. Pediatrics 2000, Tech. Rep.
- [3] V. De Luca, P. Muglia, U. Jain, V. Basile, M. Sokolowski, and J. Kennedy, “A drosophila model for attention deficit hyperactivity disorder (adhd),” *NeuroMolecular Medicine*, vol. 2, no. 3, pp. 281–287, 2002.
- [4] K. H. Chiappa, *Evoked potentials in clinical medicine*, 3rd ed., K. H. Chiappa, Ed. New York: Lippincott - Raven, 1997.
- [5] R. J. Barry, S. J. Johnstone, and A. R. Clarke, “A review of electrophysiology in attentiondeficit/hyperactivity disorder: Ii. event-related potentials.” *Clin Neurophysiol*, vol. 114, pp. 184–198, 2003.
- [6] P. Robaey, F. Breton, M. Dugas, and B. Renault, “An event-related potential study of controlled and automatic processes in 6-8 year old boys with attention deficit hyperactivity disorder,” *Electroencephalography and Clinical Neurophysiology*, vol. 82, pp. 330–340, 1992.
- [7] J. Satterfield, A. Schell, and T. Nicholas, “Ontogeny of selective attention effects on event related potentials in attention deficit hyperactivity disorder and normal boys.” *Biological Psychiatry*, vol. 28, 1990.
- [8] K. Klorman, “Cognitive event- related potentials in attention deficit disorder,” *Journal of learning Disabilities*, vol. 24, no. 130-140, 1991.
- [9] R. Strandburg, J. Marsh, W. Brown, R. Asarnow, J. Higa, R. Harper, and D. Guthrie, “Continuous-processing-related event-related potentials in children with attention deficit hyperactivity disorder,” *Biological psychiatry*, vol. 40, no. 10, pp. 964–980, 1996.

- [10] M. Taylor, J. Voros, W. Logan, and M. Malone, "Changes in event-related potentials with stimulant medication in children with attention deficit hyperactivity disorder," *Biological psychology*, vol. 36, no. 3, pp. 139–156, 1993.
- [11] D. A. Pineda, F. J. Lopera, G. C. Henao, J. D. Palacio, and F. X. Castellanos, "Confirmación de la alta prevalencia del trastorno por déficit de atención en una comunidad colombiana," *Revista de Neurología*, vol. 32, no. 3, pp. 1–6, 2001.
- [12] A. Vera, M. I. Ruano, and L. P. Ramírez, "Características clínicas y neurobiológicas del trastorno por déficit de la atención e hiperactividad," *Colombia Médica*, vol. 38, no. 4, pp. 433–439, Octubre, 2007.
- [13] R. Q. Cracco, *Evoked potentials (Frontiers of clinical neuroscience)*, I. Bodis-Wollne, Ed. Wiley-Liss, 1986.
- [14] S. López, M. Ángel, and E. Amenedo Losada, *Manual de Psicofisiología Clínica*, 1st ed. Pirámide S.A., 2001.
- [15] C. Sellán-Paniceres, "Registros de los potenciales evocados: aproximación a los componentes p300 y n400," *Revista de Psicología General y Aplicaciones*, vol. 44, no. 3, pp. 317–322, 1991.
- [16] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neurociencia : explorando el cerebro*. Barcelona, España: Masson - Williams and Wilkins, 1998.
- [17] D. E. Haines, *Fundamental Neuroscience for Basic and Clinical Applications*, 3rd ed. McGraw-Hill Livingstone, 2005.
- [18] M. J. Presentación-Herrero and M. A. Martínez-Benedicto, "Trastorno por déficit de atención con hiperactividad y potenciales evocados cognitivos," Universitat Jaume, Tech. Rep., 1999.
- [19] J. G. Webster, *Encyclopedia of Medical Devices and Instrumentation*. John Wiley & Sons, 2006, vol. 6.
- [20] C. Muñoz-Gamboa and J. Jiménez-Cruz, "Potenciales evocados y cognición," *Revista mexicana de Bioingeniería*, vol. 29, no. 1, pp. 57–63, Junio 2008.
- [21] I. Hoonhorst, W. Serniclaes, G. Collet, C. Colin, E. Markessis, M. Radeau, and P. Deltenre, "N1b and Na subcomponents of the N100 long latency auditory evoked-potential: Neurophysiological correlates of voicing in french-speaking subjects," *Clinical Neurophysiology*, vol. 120, p. 897–903, 2009.

- [22] R. Näätänen and P. T. Michie, "Early selective attention effects on the evoked potential: a critical review and reinterpretation." *Biological Psychology*, vol. 8, pp. 81–136, 1979.
- [23] S. J. Luck, *An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience)*. Cambridge: MIT Press, 2005.
- [24] S. H. Patel and P. N. Azzam, "Characterization of N200 and P300: Selected studies of the event-related potential," *International Journal of Medical Sciences*, vol. 2, no. 4, pp. 147–154, 2005.
- [25] R. Näätänen and T. W. Picton, "N2 and automatic versus controlled processes," *Electroencephalography And Clinical Neurophysiology Supplement*, vol. 36, pp. 169–186, 1986.
- [26] T. W. Picton, "The p300 wave of the human event-related potential," *Clinical Neurophysiology*, vol. 9, p. 456–479, 1992.
- [27] J. Polich, C. Ladish, and F. E. Bloom, "P300 assessment of early alzheimer's disease," *Electroencephalography and Clinical Neurophysiology*, vol. 77, no. 3, pp. 179–189, May 1990.
- [28] S. A. H. M. Kutas, "Event-related brain potentials to semantically inappropriate and surprisingly large words," *Biological Psychology*, vol. 11, no. 2, p. 99–116, Sep 1980.
- [29] —, "Reading senseless sentences: Brain potentials reflect semantic incongruity," *Science*, vol. 207, no. 4427, p. 203–205, 1980.
- [30] J. Bogacz, *Los potenciales evocados en el hombre*, E. Ateneo, Ed., Buenos Aires, Argentina, 1985.
- [31] R. Spehlmann, *Evoked potential primer: Visual, auditory, and somatosensory evoked potentials in clinical diagnosis*. Boston-London: Butterworth-Heinemann Ltd, 1985.
- [32] C. C. Duncan, R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Näätänen, J. Polich, I. Reinvang, and C. V. Petten, "Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1883–1908, 2009.
- [33] A. Fernández-Jáen and B. Calleja-Pérez, "Trastorno por déficit de atención y/o hiperactividad (tdah). abordaje multidisciplinar," servicio de Neurología Infantil. Hospital "La Zarzuela".
- [34] T. Ollendick and R. J. Prinz, *Clinical Child and Family Psychology Review*, International Consensus Statement on ADHD, January 2002.

- [35] I. Menéndez-Benavente, "Trastorno de déficit de atención con hiperactividad: clínica y diagnóstico," *Revista de Psiquiatría y Psicología del Niño y del Adolescente*, vol. 5, no. 1, 2001.
- [36] R. J. Barry, A. R. Clarke, R. McCarthy, M. Selikowitz, C. R. Brown, and P. C. Heaven, "Event-related potentials in adults with attention-deficit/hyperactivity disorder: An investigation using an inter-modal auditory/visual oddball task," *International Journal of Psychophysiology*, vol. 71, p. 124–131, 2009.
- [37] J. Narbona-García and R. Sánchez-Carpintero, "Neurobiología del trastorno de la atención e hipercinesia en el niño," *Revista de Neurología*, vol. 28, no. 2, pp. 160–164, 1999.
- [38] D. Pineda, "Avances en la investigación del trastorno de atención con y sin hiperactividad," *Revista de Neuropsicología*, vol. 5, pp. 20–21, 2003.
- [39] J. R. Valdizan, M. A. Navascués, and M. V. Sebastián, "Cartografía cerebral y trastorno por deficit de atención con hiperactividad," *Revista de Neurología*, vol. 32, no. 2, pp. 127–132, 2001.
- [40] E. Niedermeyer and F. L.-D. Silva, *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*, 3rd ed. Williams and Williams, 1993.
- [41] T. Fuchs, N. Birbaumer, W. Lutzenberger, J. H. Gruzelier, and J. Kaiser, "Neuro-feedback treatment for attention-deficit/hyperactivity disorder in children: A comparison with methylphenidate," *Applied Psychophysiology and Biofeedback*, vol. 28, no. 1, March 2003.
- [42] K. Busby, P. Firestone, and R. T. Pivik, "Sleep patterns in hyperkinetic and normal children," *Sleep*, vol. 4, no. 4, pp. 366–383, 1981.
- [43] K. Busby and R. T. Pivik, "Auditory arousal thresholds during sleep in hyperkinetic children," *Sleep*, vol. 8, no. 4, pp. 332–341, Dec 1985.
- [44] D. L. Picchetti, S. J. England, A. S. Walters, K. Willis, and T. Verrico, "Periodic limb movement disorder and restless legs syndrome in children with attention-deficit hyperactivity disorder," *Child Neurology*, vol. 13, no. 12, pp. 588–594, Dec 1998.
- [45] J. H. Satterfield, D. P. Cantwell, L. I. Lesser, and R. L. Podosin, "Physiological studies of the hyperkinetic child," *American Journal of Psychiatry*, vol. 128, p. 102–108, 1972.
- [46] M. Buchsbaum and P. Wender, "Averages evoked responses in normal and minimally brain dysfunctioned children treated with amphetamine," *Archives of General Psychiatry*, vol. 29, p. 764–770, 1973.

- [47] R. J. Schachar, R. Tannock, and G. Logan, "Inhibitory control, impulsiveness, and attention deficit hyperactivity disorder," *Clinical Psychology Review*, vol. 13, pp. 721–739, 1993.
- [48] J. M. Halperin, J. H. Newcorn, V. Sharma, J. M. Healey, and L. E. Wolf, "Inattentive and noninattentive ADHD children: do they constitute a unitary group?" *Abnormal Child Psychology*, vol. 18, p. 437–449, 1990.
- [49] Y. Frank, J. A. Seiden, and B. Napolitano, "Event-related potentials to an "oddball" paradigm in children with learning disabilities with or without attention deficit hyperactivity disorder," *Clin Electroencephalogr*, vol. 25, pp. 136–141, 1994.
- [50] M. C. Etchepareborda, J. Etchepareborda, and S. A. González, "Relación entre el espectro de coherencia de la actividad bioeléctrica cerebral y los hallazgos neuropsicológicos en pacientes con síndrome disatencional e hiperactividad," *Revista de Neurología*, vol. 22, pp. 266–271, 1994.
- [51] M. C. Etchepareborda, "Subtipos neuropsicológicos del síndrome disatencional con hiperactividad," *Revista de Neurología*, vol. 28, pp. 165–173, 1999.
- [52] M. A. Idiazábal-Alecha, A. B. Palencia-Taboada, J. Sangorrín, and J. M. Espadaler-Gamissans, "Potenciales evocados cognitivos en el trastorno por déficit de atención con hiperactividad," *Revista de Neurología*, vol. 34, no. 4, pp. 301–305, 2002.
- [53] R. S. Ledley and L. B. Lusted, "Reasoning foundations of medical diagnosis," *Science*, vol. 130, p. 9–21, 1959.
- [54] H. R. Warner, A. F. Toronto, L. G. Veasey, and R. Stephenson, "A mathematical approach to medical diagnosis. application to congenital heart disease," *Journal of the American Medical Association*, vol. 22, p. 177–183, 1961.
- [55] L. S. Goggin, R. H. Eikelboom, and M. D. Atlas, "Clinical decision support systems and computer-aided diagnosis in otology," *Otolaryngology – Head and Neck Surgery*, vol. 136, p. 521–526, 2007.
- [56] G. Lindgaard, C. Pyper, M. Frize, and R. Walker, "Does bayes have it? decision support systems in diagnostic medicine," *International Journal of Industrial Ergonomics*, vol. 39, p. 524–532, 2009.
- [57] D. Kahneman and A. Tversky, "The simulation heuristic," Cambridge University, Technical report, 1982.
- [58] J. A. Kozielicki, "A model for diagnostic problem solving," *Acta Psychologica*, vol. 36, pp. 370–380, 1972.

- [59] J. Ridderikhoff and B. van Herk, "Who is afraid of the system? doctors' attitude towards diagnostic systems," *International Journal of Medical Informatics*, vol. 53, p. 91–100, 1999.
- [60] E. Delgado-Trejos, "Generación y extracción/selección de características en la detección automática de isquemia y deficiencias valvulares sobre registros de la actividad cardíaca," Ph.D. dissertation, Universidad Nacional de Colombia, Junio 2008.
- [61] C. Vasios, G. Matsopoulos, K. Nikita, N. Uzunoglu, and C. Papageorgiou, "A decision support system for the classification of event-related potentials," in *6th Seminar on Neural Network Applications in Electrical Engineering - NEUREL, 2002*. IEEE, 2002, pp. 159–164.
- [62] M. Delavarian, F. Towhidkhah, S. Gharibzadeh, and P. Dibajnia, "Automatic classification of hyperactive children: Comparing multiple artificial intelligence approaches," *Neuroscience letters*, vol. In press, 2011.
- [63] M. Delavarian, F. Towhidkhah, P. Dibajnia, and S. Gharibzadeh, "Designing a decision support system for distinguishing adhd from similar children behavioral disorders," *Journal of Medical Systems*, pp. 1–9, 2011.
- [64] A. Nait-Ali, *Advanced biosignal processing*. Springer Verlag, 2009.
- [65] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory - TIT*, vol. 41, no. 3, pp. 613–627, 1995.
- [66] R. Quian-Quiroga and E. van Luijtelaaar, "A wavelet denoising implementation for obtaining single-stimulus evoked potentials."
- [67] R. Quian-Quiroga, "Obtaining single stimulus evoked potentials with wavelet denoising," *Physica D*, vol. 145, p. 278–292, 2000.
- [68] T. Demiralp, Y. Istefanopulos, A. Ademoglu, J. Yordanova, and V. Kolev, "Analysis of functional componentes of p300 by wavelet transform," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 4, 1998.
- [69] T. Demiralp, A. Ademoglu, Y. Istefanopulos, C. Basar-Eroglu, and E. Basar, "Wavelet analysis of oddball p300," *International Journal of Psychophysiology*, vol. 39, pp. 221–227, 2001.
- [70] R. Quian-Quiroga, O. A. Rosso, E. Basar, and M. Schürmann, "Wavelet entropy in event-related potenciales: a new method shows ordering of eeg oscillations," *Biol. Cybern*, vol. 84, pp. 291–299, 2001.

-
- [71] S. Aviyente, L. A. W. Brakel, R. K. Kushwaha, M. Snodgrass, H. Shevrin, and W. J. Williams, "Characterization of event related potentials using information theoretic distance measures," in *IEEE Transactions on Biomedical Engineering*, May 2004.
- [72] S. Ayoubian-Markazi and L. S. Stergioulas, "Wavelet filtering of the p300 component in event-related potentials," in *Proceedings of the 28th IEEE EMBS Annual International Conference*, New York City, USA,, Aug 30-sept 3 2006.
- [73] I. Kalatzis, N. Piliouras, E. Ventouras, C. Papageorgiou, A. Rabavilas, and D. Cavouras, "Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals," *Computer Methods and Programs in Biomedicine*, vol. 75, pp. 11–22, 2004.
- [74] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," *Computer methods and programs in biomedicine*, vol. 94, p. 48–57, 2009.
- [75] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM Journal on Mathematical Analysis*, vol. 15, no. 4, pp. 723–736, 1984.
- [76] P. Addison, *The Illustrated Wavelet Transform Handbook*, T. . Francis, Ed., Jul 2002.
- [77] A. Cohen and J. Kovacevic, "Wavelets: The mathematical background," in *Proceedings of the IEEE*, vol. 84, no. 4, 1996, p. 514–522.
- [78] S. Qian, *Introduction to time-frequency and wavelet transforms*. Prentice Hall PTR, 2002.
- [79] I. Daubechies, *Ten lectures on wavelets*. Society for industrial and applied mathematics, 2004.
- [80] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.
- [81] D. Peña, *Análisis de datos multivariantes*, C. F. Madrid, Ed. Madrid, España: McGraw-Hill, 2002.
- [82] D. Peña and F. J. Prieto, "Multivariate outlier detection and robust covariance matrix estimation," *Technometrics*, vol. 43, no. 3, p. 286–310, August 2001.

-
- [83] S. Murillo-Rendón and G. castellanos Domínguez, “Construcción, limpieza y depuración previa al análisis estadístico de bases de datos.” in *XV SIMPOSIO DE TRATAMIENTO DE SE~NALES, IM´AGENES Y VISI ´ON ARTIFICIAL - STSI-VA 2010*, 2010.
- [84] A. Jain, M. Murty, and P. FLynn, “Data clustering: A review,” *ACM Computing Surveys*, vol. 31, no. 3, September 1999.
- [85] A. L. Blum and P. Langley, “Selection of relevant features and examples in machine learning,” *Artificial Intelligence*, vol. 97, no. 1, pp. 245–271, Dec 1997.
- [86] H. Brunzell and J. Eriksson, “Feature reduction for classification of multidimensional data,” *Pattern Recognition*, vol. 33, pp. 1741–1748, 2000.
- [87] H. Bast, “Dimension reduction: A powerful principle for automatically finding concepts in unstructured data,” in *Proceedings of the International Workshop on Self-Properties in Complex Information Systems (SELF-STAR’04)*, 2004, pp. 113–116.
- [88] A. Jain and D. Zongker, “Feature selection: Evaluation, application, and small sample performance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.
- [89] A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: A review,” *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [90] M. Dash and H. Liu, “Feature selection for classification,” *Intelligent Data Analysis. Elsevier*, vol. 1, p. 131–156, 1997.
- [91] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. Indianapolis, IN, USA: John Wiley & Sons, Ltd, 2002.
- [92] G. Brassard and P. Bratley, *Fundamentals of Algorithms*. New Jersey: Prentice Hall, 1996.
- [93] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 2nd ed. San Diego, CA, USA: ELSEVIER Academic Press, 2003.
- [94] D. Koller and M. Sahami, “Toward optimal feature selection,” in *Proceedings of the 13th International Conference on Machine Learning*, Italy, July 1996, pp. 284–292.
- [95] M. Ben-Bassat, “Pattern recognition and reduction of dimensionality,” in *Handbook of Statistics II*. Amsterdam, North-Holland: P. R. Krishnaiah and L. N. Kanal, eds., 1982, p. 773–791.

- [96] M. Last, A. Kandel, and O. Maimon, "Information-theoretic algorithm for feature selection," *Pattern Recognition Letters*, vol. 22, pp. 799–811, 2001.
- [97] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proceedings of the Twentieth International Conference on Machine Learning (ICML'03)*, Washington DC, 2003.
- [98] H. Liu and R. Setiono, "Feature selection and classification — a probabilistic wrapper approach," in *Proceedings of Ninth International Conference on Industrial and Engineering Applications of AI and ES*, 1996, p. 284–292.
- [99] L. Wolf and S. Bileschi, "Combining variable selection with dimensionality reduction," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, june 2005, pp. 801–806.
- [100] J. L. Rodríguez-Sotelo, E. Delgado-Trejos, D. H. Peluffo-Ordoñez, D. Cuesta-Frau, and C. G. Castellanos-Domínguez, 31.
- [101] J. Shlens, "A tutorial on principal component analysis," april 2009.
- [102] F. Jager, "Feature extraction and shape representation of ambulatory electrocardiogram using the Karhunen-Lòeve transform," *Electrotechnical Review*, vol. 69, no. 2, pp. 83–89, march 2002.
- [103] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, and M. Strintzis, "ECG analysis using nonlinear PCA neural networks for ischemia detection," *IEEE Transactions On Signal Processing*, vol. 46, no. 11, pp. 3058–3066, Noviembre 1998.
- [104] H. Y. Wang and X. J. Wu, "Weighted PCA space and its application in face recognition," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, vol. 7, 2005. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1527735
- [105] D. H. Peluffo-Ordoñez, J. L. Rodríguez-Sotelo, D. Cuesta-Frau, and C. G. Castellanos-Domínguez, "Estudio comparativo de métodos de selección de características de inferencia supervisada y no supervisada," *Tecno Lógicas*, no. 23, pp. 149–166, December 2009.
- [106] Y. S. X. and S. Jianbo, "Multiclass spectral clustering," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 313.
- [107] P. Hansen and N. Mladenovic, "J-means: a new local search heuristic for minimum sum of squares clustering," *Pattern Recognition*, pp. 405 – 413, 2001.

- [108] G. Hamerly and C. Elkan, “Alternatives to the k-means algorithm that find better clusterings,” in *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*. New York, NY, USA: ACM, 2002, pp. 600–607.
- [109] D. Cuesta, J. Pérez-Cortés, and G. Andreau-García, “Clustering of ECG signals in computer-aided holter analysis,” *Computer Methods and Programs in Biomedicine*, vol. 72, pp. 179–196, 2003.
- [110] J. L. R. Sotelo, D. Peluffo, D. C. Frau, D. P. Ordóñez, and G. C. Domínguez, “Non-parametric density-based clustering for cardiac arrhythmia analysis,” *Computers in cardiology. CINC*, 2009.
- [111] Y. S. X. and S. Jianbo, “Multiclass spectral clustering,” in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 313.
- [112] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 849–856.
- [113] K. B. E. Böcker, J. A. G. van Avermaete, and M. M. C. van den Berg-Lenssen, *Brain Topography*. Springer, 1994, ch. The international 10–20 system revisited: Cartesian and spherical co-ordinates, pp. 231–235.
- [114] R. Coifman and M. Wickerhauser, “Entropy-based algorithms for best basis selection,” *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 713–718, mar 1992.
- [115] E. Delgado-Trejos, A. Perera-Lluna, M. Vallverdú-Ferrer, P. Caminal-Magrans, and G. Castellanos-Domínguez, “Dimensionality reduction oriented toward the feature visualization for ischemia detection,” in *IEEE Transactions on Information Technology in Biomedicine VOL. 13, NO. 4, JULY 2009*, vol. 13, no. 4, July 2009.
- [116] D. Ververidis and C. Kotropoulos, “Fast and accurate sequential floating forward feature selection with the bayes classifier applied to speech emotion recognition,” *Signal Processing*, vol. 88, no. 12, pp. 2956–2970, 2008.
- [117] R. Pinzon-Morales, A. Orozco-Gutierrez, H. Carmona-Villada, and G. Castellanos-Dominguez, “Towards high accuracy classification of mer signals for target localization in parkinson’s disease,” in *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010.
- [118] R. Pinzon-Morales, F. Restrepo, O. Moscoso, P. C.-C. adn A.A. Orozco-Gutiérrez, and C. Castellanos-Dominguez, “Detection of attention-deficit/hyperactivity disorder

based on customized wavelet,” in *XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, Madrid, 24, 25 y 26 de noviembre 2010.

- [119] R. Pinzon-Morales, A. Orozco-Gutierrez, and G.Castellanos-Dominguez, “Feature selection using an ensemble of optimal wavelet packet and learning machine: Application to mer signals,” in *IET International Symposium on Communication Systems, Networks and Digital Signal Processing*. UK: IEEE, 2010.
- [120] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, cuarta ed., American Psychiatric Association, Washington, DC., 2000.
- [121] A. Rencher, *Methods of multivariate analysis*. Wiley-Interscience, 2002.