

UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN

EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

Desarrollo y evaluación de clasificadores neuro-genéticos basados en señales de EEG para la ayuda al diagnóstico de la esquizofrenia

Autor:

D. Lorenzo Santos Mayo

Tutores:

Dr. D. Luis Miguel San José Revuelta

Dr. D. Juan Ignacio Arribas Sánchez

Valladolid, 14 de Septiembre de 2015

TÍTULO: **Desarrollo y evaluación de clasificadores neuro-genéticos basados en señales de EEG para la ayuda al diagnóstico de la esquizofrenia**

AUTOR: **D. Lorenzo Santos Mayo**

TUTORES: **Dr. D. Luis Miguel San José Revuelta
Dr. D. Juan Ignacio Arribas Sánchez**

DEPARTAMENTO: **Teoría de la Señal y Comunicaciones e Ingeniería Telemática**

TRIBUNAL

PRESIDENTE: **Dr. D. Pablo de la Fuente Redondo**

SECRETARIO: **Dr. D. Rodrigo de Luis García**

VOCAL: **Dr. D. César Gutiérrez Vaquero**

FECHA: **14 de Septiembre de 2015**

CALIFICACIÓN:

Resumen de TFM

El presente Trabajo de Fin de Máster (TFM) constituye una ampliación del Proyecto Fin de Carrera (PFC) titulado "Diagnóstico automático de la esquizofrenia basado en el estudio de señales de EEG de actividad cortical en el cerebro". En este proyecto se abordaba el problema de la realización de una clasificación automática entre sujetos sanos y esquizofrénicos basada en la onda P3b, evocada por el cerebro cuando es sometido a una tarea *oddball* auditiva, y en un esquema de reconocimiento de patrones. Para ello, en primer lugar, se realizaba el pre-procesado de las señales de EEG proporcionadas en un banco de datos. Posteriormente, se extraían diferentes características de dichas señales que eran seleccionadas mediante algoritmos de selección de características basados en la teoría de la información y en la distancia entre las clases (MIFS, DISR y J5). Finalmente, estas características seleccionadas eran introducidas en unos clasificadores altamente no lineales (MLP y SVM) llegando a obtener, en el mejor caso, una tasa de acierto del 93.4%, una sensibilidad del 87.3% y una especificidad de un 96.7%. Además, el PFC relataba diversas conclusiones sobre el origen espacial de la onda P3b y sobre la comparación de los distintos algoritmos utilizados. Este TFM ha continuado sobre esta línea de investigación con el fin de tratar algunos de los aspectos que se habían mencionado como líneas futuras en el PFC anterior. En primer lugar, se ha revisado la etapa de selección de características añadiendo los Algoritmos Genéticos (AG) como un nuevo método que pretende mejorar los resultados finales de las diversas clasificaciones realizadas. Además, los AG se implementan con el fin de lograr mantener o superar las tasas de acierto mediante un menor número de características seleccionadas por lo que pretenden reducir la carga computacional de la clasificación final. En segundo lugar, en este TFM se han realizado diversas agrupaciones espaciales de las características y un análisis estadístico de todas ellas individualmente con el fin de lograr reforzar las diversas conclusiones espaciales del PFC. Por último, este TFM realiza una optimización de los clasificadores utilizados ya que éstos carecían de la misma en el PFC. Los nuevos resultados de este TFM han mostrado que los AG se alzan como un método de selección de características muy interesante a la hora de abordar el problema de la detección de la esquizofrenia ya que, aunque las tasas de acierto para la mejor clasificación se siguen logrando mediante el algoritmo J5, las tasas de aciertos mediante una extracción menor de características crece para los AG y, además, se consigue una carga computacional mucho menor en los clasificadores finales. Además, los AG, al contrario que los otros métodos de selección de características utilizados, poseen una gran cantidad de parámetros a configurar para su implementación por lo que se puede pensar que los resultados obtenidos mediante los AG pueden ser mejorados mediante una optimización más profunda. Por otro lado, los resultados del TFM evidencian una gran importancia de la zona parietal-temporal a la hora de la detección

de la onda P3b concordando con otras investigaciones similares. Finalmente, la optimización de los clasificadores automáticos proporciona un incremento generalizado de todas las clasificaciones y, en concreto, aunque la clasificación más eficiente sigue proporcionando una tasa de acierto de un 93.4%, la sensibilidad, que es el parámetro más crítico para aplicaciones médicas, logra subir hasta el 87.7% a costa de un descenso en la especificidad obteniendo un 96.3%.

Palabras clave

Aprendizaje Automático, Curvas ROC, EEG, Esquizofrenia, P3b, Reconocimiento de Patrones.

Abstract

This Master Thesis (TFM) is an extension of the Final Project (PFC) called "Automatic schizophrenia diagnosis based on the EEG signals study of cortical activity in the brain." In this project an automatic classification between healthy and schizophrenics subjects, based on P3b wave evoked by the brain when it is subjected to an auditory oddball task and a pattern recognition scheme, was done. In order to do this, firstly, it was performed a pre-processing of the EEG signals provided in a database. Subsequently, some features of the signals which were selected by features selection algorithms based on information theory and the distance between groups (MIFS, DISR and J5) were extracted. Finally, these selected features were introduced in a highly non-linear classifiers (MLP and SVM) reaching, in the best case, a success rate of 93.4%, a sensitivity of 87.3% and a specificity of 96.7%. In addition, the PFC recounted several conclusions about the spatial origin of the P3b wave and comparing the different algorithms used. This TFM has continued this line of research in order to solve some of the issues that were mentioned as future lines in the previous PFC. Firstly, it has revised the selection stage of adding features Genetic Algorithms (GA) as a new method that aims to improve the final results of the various classifications made. In addition, the AG are implemented in order to improve the performances rates with a smaller number of features with the aim to reduce the computational cost of the final classification. Secondly, in this TFM some spatial groups of features and a statistical analysis of all of them individually have been made in order to underline some conclusions of the PFC. Finally, this TFM performs an optimization of classifiers used because they lacked it in PFC. The new results of this TFM show that the AG are a really important method of selecting in order to detect the schizophrenia, because, although success rates for achieving better rankings continue using the algorithm J5, success rates with less features grows for AG and also much lower computational cost on the final classifier is achieved. In addition, AG, unlike other selection methods used, have a lot of parameters to be configured for deployment, so it can be think that the results get by the AG can be improved through further

optimization. On the other hand, the results of TFM show a great importance in the parietal-temporal area when detecting the wave P3b agreeing with similar researches. Finally, optimization of automatic classifiers provides a general increase in all classifications and in particular, although the more efficient classification is providing a success rate of 93.4%, sensitivity, which is the most critical parameter for medical applications, rises to 87.7% with a decline of 96.3% in specificity.

Keywords

Machine Learning, ROC Curves, EEG, Schizophrenia, P3b, Pattern Recognition.

It's better to burn out than to fade away.
-Kurt Cobain-

Agradecimientos

En primer lugar, quiero dar las gracias a toda mi familia por el apoyo que siempre he recibido de ellos.

También quiero mencionar a mis amigos y compañeros por todos los momentos vividos este curso.

Por último, quiero agradecer a mis tutores, Luis Miguel y Juan Ignacio, toda la ayuda que me han proporcionado y la paciencia que han tenido conmigo.

Gracias a todos.

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	3
1.3. Metodología	4
1.4. Medios	5
1.5. Estructura	6
2. Conocimientos previos	9
2.1. La esquizofrenia	9
2.1.1. ¿Qué es la esquizofrenia?	9
2.1.2. Síntomas	10
2.1.3. Variantes y tratamientos	11
2.2. Electroencefalografía	12
2.2.1. El encéfalo	13
2.2.2. Captación del electroencefalograma	13
2.3. Potenciales evocados	17
2.4. Detección de la esquizofrenia	19
2.4.1. Onda P300	20
2.4.2. Reconocimiento de patrones	20
3. Extracción y agrupación de características de la señal de EEG	25
3.1. Adquisición de señales de EEG	25
3.1.1. Análisis de los sujetos del banco de datos	25
3.1.2. Captación de las señales de EEG	27
3.2. Pre-procesado	28
3.2.1. Filtrado	29
3.2.2. Segmentación	31
3.2.3. Análisis ICA	34
3.2.4. Grand Average	37
3.3. Extracción de características	38

3.4. Agrupación de características	41
4. Selección de características para las señales de EEG	43
4.1. Algoritmos basados en la teoría de la información	44
4.1.1. MIFS	44
4.1.2. DISR	48
4.2. Criterio J5	50
4.3. Algoritmos genéticos	53
5. Análisis estadístico y clasificación de las señales de EEG	61
5.1. Análisis estadístico	61
5.2. Clasificación	65
5.2.1. Máquinas de Vectores Soporte	65
5.2.2. Redes Neuronales	72
6. Resultados	83
6.1. Técnicas para el análisis de resultados	83
6.1.1. Análisis de los resultados estadísticos	83
6.1.2. Análisis de los resultados de la clasificación	84
6.2. Resultados finales	89
6.2.1. Comparación de métodos de selección de características	89
6.2.2. Análisis de resultados para las agrupaciones de características de electrodos individuales	90
6.2.3. Curvas ROC para las agrupaciones de características	94
6.2.4. Análisis de las características más discriminantes	96
6.2.5. Procesado óptimo para la clasificación	97
7. Conclusiones y líneas futuras	101
7.1. Conclusiones	101
7.2. Líneas Futuras	104
A. Áreas bajo las curvas ROC para todos los procesados estudiados	117
B. Resultados completos del análisis estadístico	123
C. Parámetros de las señales de EEG incluidos en cada matriz de características	133

Índice de figuras

2.1. CórteX cerebral	14
2.2. Gorro de electrodos	15
2.3. Sistema diez-veinte de posicionamiento de electrodos según la versión europea	17
2.4. Esquema básico de un método de reconocimiento de patrones	21
2.5. Esquema del método de reconocimiento de patrones implementado en el TFM	23
3.1. Representación gráfica de una señal continua de EEG	29
3.2. Representación gráfica de una señal continua de EEG filtrada	31
3.3. Representación gráfica de las épocas de una señal de EEG filtrada	32
3.4. Ejemplo de señales fuente de audio	35
3.5. Ejemplo de señales mezcla de audio	36
4.1. Ilustración de varios conceptos relativos a la teoría de la información	45
4.2. Esquema básico de un algoritmo genético	55
4.3. Esquema del algoritmo genético implementado	60
5.1. Representación gráfica de las distintas partes de un clasificador SVM de margen máximo	66
5.2. Representación gráfica de las distintas partes de un clasificador SVM de margen blando	70
5.3. Transformación $\Phi(\cdot)$ para hacer los datos linealmente separables en las SVM	71
5.4. Neurona biológica	73
5.5. Neurona artificial	74
5.6. Tipos de redes neuronales según su estructura	75
5.7. Perceptrón monocapa	77
5.8. Posibles funciones de activación para una neurona artificial	78
5.9. Ejemplo de una red MLP con dos capas ocultas	79
6.1. Ejemplo de matriz de confusión	85
6.2. Ejemplo de curva ROC real.	86

6.3.	AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las diferentes asociaciones de electrodos. Las etiquetas del eje x describen los procesos utilizados en las diferentes etapas del método de reconocimiento de patrones para los valores de AUC del eje y: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (J5, AG, MIFS o DISR) y el tercero al tipo de clasificador utilizado (MLP o SVM)	90
6.4.	Número de características discriminantes para cada electrodo individual mediante el test U: (a) Características extraídas de la señal de EEG filtrada a 15Hz; (b) Características extraídas de la señal de EEG filtrada a 35Hz	92
6.5.	AUC para los diferentes electrodos individuales según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, criterio J5 y clasificador MLP; (b) filtrado de 35Hz, criterio J5 y clasificador MLP; (c) filtrado de 15Hz, criterio J5 y clasificador SVM; y (d) filtrado de 35Hz, criterio J5 y clasificador SVM	92
6.6.	AUC para los diferentes electrodos individuales según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, AG y clasificador MLP; (b) filtrado de 35Hz, AG y clasificador MLP; (c) filtrado de 15Hz, AG y clasificador SVM; y (d) filtrado de 35Hz, AG y clasificador SVM	93
6.7.	Curvas ROC para las diferentes agrupaciones de electrodos según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, criterio J5 y clasificador MLP; (b) filtrado de 35Hz, criterio J5 y clasificador MLP; (c) filtrado de 15Hz, criterio J5 y clasificador SVM; y (d) filtrado de 35Hz, criterio J5 y clasificador SVM	94
6.8.	Curvas ROC para las diferentes agrupaciones de electrodos según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, AG y clasificador MLP; (b) filtrado de 35Hz, AG y clasificador MLP; (c) filtrado de 15Hz, AG y clasificador SVM; y (d) filtrado de 35Hz, AG y clasificador SVM	95
6.9.	Curvas ROC para los procesados más eficientes. La leyenda indica el tipo de procesado implementado para generar cada ROC: el primer término indica el filtrado realizado en el pre-procesado (15Hz o 35Hz), el segundo el tipo de agrupación seleccionada (total), el tercero el algoritmo de selección de características (J5 o AG) y el cuarto el clasificador (MLP o SVM)	99

6.10. Matrices de confusión para la clasificación llevada a cabo cuando el umbral del clasificador se sitúa en el punto óptimo de la curva ROC para los procesados de: (a)15Hz-Total-J5-MLP, (b)35Hz-Total-J5-MLP, (c)15Hz-Total-J5-SVM, (d)35Hz-Total-J5-SVM, (e)15Hz-Total-AG-MLP y (f)15Hz-Total-AG-SVM	100
--	-----

Índice de tablas

3.1. Parámetros biológicos y su influencia en la onda P300.	26
5.1. Tipos de errores cometidos en el contraste de hipótesis	63
6.1. Número de características discriminantes para cada electrodo individual mediante el test U. La primera columna describe el electrodo al que se refiere. El resto de columnas hacen referencia a los filtrados llevados a cabo en el pre-procesado (15Hz o 35Hz) anterior a la extracción de las características	91
6.2. Número de electrodos discriminantes para cada característica mediante el test U. La primera columna describe la característica a la que se refiere. El resto de columnas hacen referencia a los filtrados llevados a cabo en el pre-procesado (15Hz o 35Hz) anterior a la extracción de las características	97
A.1. AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (J5) y el tercero al tipo de clasificador utilizado (MLP o SVM)	118
A.2. AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (AG) y el tercero al tipo de clasificador utilizado (MLP o SVM)	119

A.3.	AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (MIFS) y el tercero al tipo de clasificador utilizado (MLP o SVM)	120
A.4.	AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (DISR) y el tercero al tipo de clasificador utilizado (MLP o SVM)	121
B.1.	Resultados del análisis estadístico obtenidos para los diferentes filtrados realizados en el pre-procesado. La primera columna describe la característica que se utiliza para realizar el test U: el primer término hace referencia a la característica y el segundo al electrodo del que se extrae la mismas. El resto de columnas hacen referencia a los distintos p-valores y h-valores para el filtrado de 15Hz y el de 35Hz	131
C.1.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, algoritmo J5).	133
C.2.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, algoritmo J5).	134
C.3.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, algoritmo MIFS).	134
C.4.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, algoritmo MIFS).	134
C.5.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, algoritmo DISR).	135
C.6.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, algoritmo DISR).	135
C.7.	Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, AG).	135

C.8. Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, AG).	136
--	-----

Capítulo 1

Introducción

La señal de electroencefalograma (EEG) permite monitorizar la actividad eléctrica generada por las neuronas cerebrales. En concreto, la onda P3b, evocada por el cerebro cuando es sometido a una tarea *oddball* auditiva, ha sido ampliamente estudiada y considerada como un importante marcador de la psicosis de la esquizofrenia. Pero a pesar de existir numerosos estudios sobre la relación entre la esquizofrenia y la onda P3b, aún no se ha realizado ninguna investigación que proponga un método para el diagnóstico asistido por ordenador de dicha enfermedad basado en dicha onda con unos resultados lo suficientemente completos. La mayoría de las investigaciones realizadas sobre este tema se conforman con un análisis estadístico de las características extraídas de la señal de EEG o no llegan a completar los resultados finales de la clasificación de las mismas mediante parámetros de gran importancia como son la sensibilidad y la especificidad. En este TFM se pretende implementar un método de reconocimiento de patrones basado en algoritmos neuro-genéticos, optimizando el proporcionado en [?], con el fin de lograr una clasificación entre sujetos esquizofrénicos y sujetos de control con unas altas tasas de acierto. Para ello se propondrán diferentes algoritmos de procesamiento no lineal para las distintas etapas del método de reconocimiento de patrones con el fin de compararlos y definir, finalmente, las etapas óptimas de dicho método en cuanto a tasas de acierto a la hora de la clasificación entre sujetos sanos y esquizofrénicos se refiere. En este capítulo se expone la motivación del presente TFM, los objetivos que se perseguirán durante su realización, la metodología seguida y los medios empleados. Finalmente, se presentará también la estructura que seguirá el presente documento.

1.1. Motivación

La esquizofrenia es un trastorno mental que afecta aproximadamente a un 1% de la población mundial [?]. De acuerdo con el manual diagnóstico y estadístico de los trastornos mentales (DSM-V) [?] los sujetos esquizofrénicos presentan diferentes

síntomas como ilusiones, alucinaciones o un lenguaje desorganizado. La esquizofrenia es un trastorno crónico por lo que el único tratamiento posible es el control de sus efectos con el fin de que los enfermos puedan llevar una vida lo más normal posible.

La detección de la esquizofrenia ha sido, y es, un tema de gran importancia y está sobradamente documentado en numerosos trabajos médicos [?, ?, ?, ?]. La gran variedad de síntomas producidos por dicha psicosis hacen muy complejo su diagnóstico. Por ejemplo, la detección de la esquizofrenia entre adolescentes es muy complicada ya que los primeros síntomas como pueden ser cambios de amigos, bajo rendimiento escolar o problemas de insomnio, son comportamientos comunes entre este colectivo. Existe una combinación de síntomas que hacen posible predecir este trastorno de forma subjetiva en los jóvenes con riesgo a padecerlo. Estos factores engloban: el aislamiento y retraimiento, el aumento de pensamientos y sospechas inusuales, y los antecedentes familiares de esta psicosis [?].

Sin embargo, las causas que propician esta enfermedad son desconocidas en la actualidad. Numerosos estudios han intentado relacionar diferentes factores como el genético o el ambiental con la esquizofrenia [?] pero, a día de hoy, se desconoce con precisión cuál es su origen. La importancia de la detección de la esquizofrenia radica en que un rápido diagnóstico hace que el enfermo pueda ser tratado y sus síntomas puedan ser controlados con mayor eficacia. Por el contrario, si la enfermedad no se diagnostica a tiempo, los síntomas pueden volverse crónicos y difíciles de tratar.

Por este motivo, un diagnóstico objetivo de la enfermedad mediante el análisis de las señales de EEG puede aportar un gran valor a la hora de lograr realizar una detección temprana con el fin de conseguir obtener un mayor control de la enfermedad y sus síntomas. En concreto, la subcomponente P3b del potencial P300 evocado por el cerebro ante un paradigma *oddball* de estímulos auditivos ha sido establecida históricamente como un marcador fiable para el diagnóstico de la esquizofrenia en diferentes investigaciones [?, ?, ?, ?]. Pero la mayoría de estos estudios solo llegan a realizar un análisis estadístico de los parámetros ofrecidos por el potencial P3b sin llegar a realizar una clasificación entre los distintos grupos de sujetos. Además, estas investigaciones solo tienen en cuenta las características de la onda P3b en el dominio temporal.

Con el avance de las distintas técnicas de procesado de señal, las investigaciones para lograr un diagnóstico de la esquizofrenia asistido por ordenador estudiaron las señales de EEG evocadas por diferentes tareas en diversos dominios transformados utilizando, por ejemplo, la transformada de Fourier o la transformada Wavelet. De esta forma, en [?] se utilizaba la transformada Wavelet para la clasificación entre sujetos sanos y esquizofrénicos logrando una sensibilidad de un 71 % y una especificidad de un 66 %. Por otro lado, en [?] se hacía uso de la transformada de Fourier de los EEGs para diferenciar

entre sujetos sanos y esquizofrénicos, logrando una tasa de aciertos del 62.2%. Además, en [?] se combinaban características temporales y frecuenciales del EEG obteniendo una tasa de aciertos del 87.51% en la clasificación entre ambos grupos.

Además, existen investigaciones más específicas que no solo pretenden clasificar entre sujetos esquizofrénicos y sanos. En este sentido, en [?] se realizó un estudio estadístico para evaluar las diferencias producidas en los EEG en reposo entre sujetos esquizofrénicos, con trastorno bipolar y sanos. Por otro lado, en [?] se pretendía investigar las señales de EEG para observar la respuesta de los diferentes sujetos a la terapia de clozapina.

Pero quizá la carencia más importante en la mayoría de las investigaciones actuales es la utilidad de los resultados de las mismas. En [?] se realizó un estado del arte estudiando 53 artículos publicados relacionados con las anormalidades producidas en el espectro de la señal de EEG para diagnosticar la esquizofrenia. Este resumen advirtió que la mayoría de los artículos concluían únicamente con la demostración de que una variable biológica era potencialmente útil para el diagnóstico de la enfermedad pero muy pocos llegaban a realizar un método de clasificación suficientemente completo para su validación, describiendo en profundidad las características de rendimiento del mismo tales como la sensibilidad, la especificidad o un análisis de las curvas ROC.

Este trabajo pretende realizar una clasificación basada en algoritmos neuro-genéticos entre sujetos sanos y esquizofrénicos que obtenga un rendimiento lo suficientemente bueno para poder ser implementado como herramienta de ayuda al diagnóstico real aportando unos resultados de clasificación lo suficientemente completos. Para ello se partirá de la hipótesis de que la onda P3b posee características diferentes para ambos grupos, que ha sido validada a lo largo de diferentes investigaciones anteriores. Una vez tomada esta hipótesis se pretende llevar a cabo un método de reconocimiento de patrones optimizando el que ya se realizó en [?] con el fin de conseguir el mayor rendimiento posible a la hora de diagnosticar la enfermedad. De esta forma, se utilizarán diferentes algoritmos en las distintas fases del esquema de reconocimiento de patrones con el fin de encontrar la combinación que maximice la tasa de aciertos en la clasificación final. Si se consiguen unas tasas de acierto lo suficientemente altas, este TFM podría solucionar, en gran medida, la problemática actual del difícil diagnóstico de la esquizofrenia, sirviendo de gran ayuda para poder tratar la enfermedad de forma adecuada.

1.2. Objetivos

De acuerdo con lo descrito en la sección anterior, el objetivo global de este trabajo puede descomponerse en los siguientes objetivos específicos:

- Analizar el estado del arte de los diferentes métodos de reconocimiento de patrones

utilizados hasta la fecha con el fin de proponer diferentes optimizaciones para el método implementado en [?].

- Seleccionar nuevos algoritmos para cada etapa del método de reconocimiento de patrones que parezcan encajar mejor en el mismo teniendo en cuenta el tipo de datos a tratar en el problema de la detección de la esquizofrenia.
- Desarrollar un *software* en MATLAB que lleve a cabo las novedades propuestas para el método de reconocimiento de patrones de [?]: la agrupación de características, la implementación del AG, el análisis estadístico y la optimización de los clasificadores. Se pretenden incorporar estas mejoras dentro del método con el fin de mejorar los resultados proporcionados y lograr extraer unas conclusiones más contrastadas.
- Utilizar los EEGs del banco de datos real proporcionado por un equipo médico como datos de entrada para el procesamiento finalmente utilizado con el fin de estudiar los resultados generados por los mismos.
- Analizar y comparar los diferentes resultados obtenidos para las diferentes combinaciones de algoritmos implementados con el fin de seleccionar la combinación que mejor se ajuste al problema. De esta forma, se pretende conseguir un método de reconocimiento de patrones que pueda ser implementado en un entorno real como herramienta de ayuda al diagnóstico. Además, se pretenden extraer otras conclusiones de interés como la zona cerebral en la que mejor se detecta el potencial P3b.
- Obtener líneas futuras de estudio que permitan ampliar el sistema desarrollado.

1.3. Metodología

Como se ha mencionado en la sección anterior, el trabajo presentado pretende mejorar el esquema de reconocimiento de patrones llevado a cabo en [?]. Por lo tanto, la consecución de este objetivo implica el siguiente programa de tareas:

1. Revisión bibliográfica de todos los contenidos implicados en este TFM, desde la enfermedad de la esquizofrenia hasta el estado del arte de las máquinas de clasificación de patrones y el análisis de sus resultados. De esta forma, se pretende verificar la originalidad del trabajo a realizar en el presente TFM.
2. Realización de diversas agrupaciones de las características extraídas en [?] en diferentes grupos basados en la localización espacial de las mismas. De esta forma se pretende lograr extraer diferentes conclusiones sobre el origen espacial

de la señal que permite discriminar la psicosis de la esquizofrenia dentro de la limitada resolución espacial proporcionada por la señal de EEG complementando los resultados expuestos en [?].

3. Utilización de los Algoritmos Genéticos como criterio de selección de características discriminantes con el fin de comparar su rendimiento al de los criterios ya utilizados en [?].
4. Implementación de un análisis estadístico de las diferentes características seleccionadas para realizar un primer análisis de resultados basado en la separabilidad estadística que existe entre los sujetos sanos y esquizofrénicos utilizando dichas características por separado.
5. Optimización de máquinas automáticas implementadas en [?] para la clasificación entre sujetos sanos y esquizofrénicos. Mediante esta optimización se pretenden mejorar, en la medida de lo posible, los resultados del PFC del que parte este trabajo.
6. Análisis de los resultados en los que se compara la precisión obtenida empleando los distintos algoritmos en cada etapa del método de reconocimiento de patrones desarrollado. Los resultados se emplearán para estudiar la viabilidad de implantar esta técnica en aplicaciones reales y para obtener otras conclusiones de interés.
7. Obtención de líneas futuras de estudio que permitan evolucionar y mejorar el sistema propuesto.

1.4. Medios

Los medios que se emplearán para llevar a cabo los objetivos del TFM son los siguientes:

- Ordenador portátil conectado a Internet con *Windows 8.1* como sistema operativo.
- Entorno de programación MATLAB 8.1 (R2013a). Dentro de este entorno cabe destacar el uso de los siguientes *toolbox*: EEGLAB, FEAST (*Feature Selection Toolbox*), GLOBALOPTIM (*Global Optimization Toolbox*), NNET (*Neural Network Toolbox*) y STATS (*Statistics Toolbox*).
- Paquete *Microsoft Office*.
- Sistema de composición de textos *LaTex*.

- Banco de datos de electroencefalogramas¹ con un total de 47 sujetos válidos para su posterior procesado y clasificación. De estos 47 sujetos, 16 de ellos pertenecen a EEGs de pacientes esquizofrénicos y 31 a controles.

1.5. Estructura

El resto de este Trabajo de Fin de Máster se estructura de la siguiente forma:

- **Capítulo 2. Conocimientos previos.** En este capítulo se expondrán diferentes conocimientos básicos sobre la psicosis de la esquizofrenia y su detección mediante la señal de EEG con el fin de introducir al lector en la problemática de este trabajo. En primer lugar, se describirá la psicosis desde el punto de vista médico describiendo la enfermedad y sus distintos síntomas. Posteriormente, se detallarán los conceptos básicos sobre la electroencefalografía y los potenciales relacionados con eventos que se manifiestan en las señales de EEG durante la realización de diversas tareas. Por último, se relacionará la esquizofrenia con el potencial evocado P300 y se expondrán las bases del reconocimiento de patrones como una posible herramienta para la diferenciación de dicho potencial entre los diferentes grupos de sujetos (sanos y esquizofrénicos). Finalmente, se describirá el esquema de reconocimiento de patrones utilizado en el presente TFM.
- **Capítulo 3. Extracción y agrupación de las características de la señal de EEG.** Esta sección pretenden describir las primeras etapas del método de reconocimiento de patrones utilizado en la presente investigación. En primer lugar, se comenzará describiendo el proceso seguido por los psiquiatras que proporcionaron el banco de datos utilizado en este trabajo para captar las señales de EEG de los sujetos. En segundo lugar, se detallará el tratamiento que se deberá realizar a los EEGs proporcionados con el fin de obtener los potenciales evocados con los que se podrá estudiar la existencia de la esquizofrenia en los diferentes individuos. Posteriormente, se expondrá el proceso de extracción de características de la señal de EEG pre-procesada llevada a cabo en el estudio. Finalmente, se tratará sobre la realización de diferentes agrupaciones de dichas características.
- **Capítulo 4. Selección de características para las señales de EEG.** En este epígrafe se expondrán los distintos métodos implementados en el presente estudio para la realización de la selección de las características más discriminantes de las extraídas previamente. En primer lugar, se detallarán algunos algoritmos basados

¹Agradecemos al Prof. Vicente Molina, del Depto. de Psiquiatría de la Univ. Valladolid por habernos permitido el acceso al banco de datos de EEG.

en la teoría de la información. En segundo lugar, se detallará el criterio de selección J5 basado en las distancias intraclase e interclase en el espacio de características. Finalmente, se describirán los algoritmos genéticos como una solución altamente no lineal para la selección de características.

- **Capítulo 5. Análisis estadístico y clasificación de las señales de EEG.** En este capítulo se afrontará la problemática sobre la diferenciación de los diferentes individuos entre esquizofrénicos y sanos. Para ello se relatarán dos tipos de análisis a realizar en los datos extraídos de las señales de EEG con el fin de cuantificar el grado en que es posible separar ambas clases de sujetos. Por un lado, se expondrán las bases del análisis estadístico de todas las características extraídas de las señales de EEG con el fin de cuantificar el grado en que cada una de ellas es capaz de diferenciar los sujetos. Finalmente, se describirán algunas de las máquinas de clasificación automática más importantes en la actualidad como son las Máquinas de Vectores Soporte y las Redes Neuronales Artificiales.
- **Capítulo 6. Resultados.** En este epígrafe se pretenden exponer los diferentes resultados extraídos referentes a la aplicación del método de reconocimiento de patrones para la ayuda al diagnóstico de la esquizofrenia. En primer lugar, se realizará un resumen del algoritmo de reconocimiento de patrones utilizado con el fin de recordar al lector el procesamiento realizado a las señales de EEG. En segundo lugar, se describirán los métodos de análisis de resultados mediante los que se generarán los resultados a estudiar en este trabajo. Para finalizar, se hará uso de estos métodos para extraer los distintos resultados de interés del estudio.
- **Capítulo 7. Conclusiones y líneas futuras.** En este capítulo se extraerán las conclusiones más relevantes del TFM y de los resultados obtenidos. Asimismo, se plantearán algunas líneas futuras que contribuirían a mejorar y ampliar el trabajo realizado.

Capítulo 2

Conocimientos previos

En este capítulo se introducirán diferentes conceptos básicos fundamentales para la comprensión del resto del trabajo. En primer lugar se tratará sobre la problemática de la esquizofrenia desde el punto de vista médico describiendo la enfermedad y sus distintos síntomas. En segundo lugar, se describirán los conceptos básicos sobre la electroencefalografía como método para cuantificar los potenciales emitidos por las neuronas cerebrales. Posteriormente, se expondrán algunos de los potenciales evocados relacionados con eventos que se manifiestan en las señales de EEG cuando se realizan diversas tareas cognitivas. Para finalizar, se pondrá de manifiesto la relación entre el potencial P300 y la psicosis de la esquizofrenia y se expondrán los conceptos básicos del reconocimiento de patrones como una herramienta para lograr diferenciar los distintos grupos de sujetos (sanos y esquizofrénicos) mediante las características extraídas de dicho potencial evocado. Finalmente, se describirá el esquema de reconocimiento de patrones utilizado en el presente TFM.

2.1. La esquizofrenia

En este apartado se abordarán, desde un punto de vista médico, aspectos tales como la definición de la psicosis de la esquizofrenia, sus síntomas y otros datos de interés. El fin de esta descripción será introducir el concepto de esquizofrenia para asentar las bases biológicas de este trabajo.

2.1.1. ¿Qué es la esquizofrenia?

La esquizofrenia es un trastorno cerebral crónico, severo e incapacitante que ha afectado a las personas a lo largo de la historia [?]. Se considera que el origen de este término se encuentra en la descripción de la llamada *dementia praecox* por parte del psiquiatra alemán Emil Kraepelin (1856-1926), aunque ya se encontraban alusiones a

este término en algunos textos franceses de mediados del siglo XIX. Kraepelin propuso el término *dementia praecox* para referirse a este trastorno, y así diferenciarlo del resto [?].

El origen de la esquizofrenia es actualmente desconocido y aún no existen pruebas diagnósticas definitivas. Hay investigaciones que están a favor de la existencia de una predisposición genética, incluso que podría ser provocada por algunos virus que afectan al feto durante el embarazo, pero no hay ninguna conclusión clara sobre su origen. Aún así, los pacientes con esquizofrenia tienen una serie de rasgos comunes que ayudan a su diagnóstico. Algunos de estos rasgos, presentados en mayor o menor medida según el paciente, son: delirios, alucinaciones, trastornos del pensamiento, alteración de la sensación sobre sí mismo, deterioro de las emociones o aislamiento.

2.1.2. Síntomas

Como ya se ha mencionado anteriormente, en la actualidad, el diagnóstico de la esquizofrenia no es realizado de forma objetiva o cuantitativa sino que se trata de un cómputo de síntomas que se clasifican en positivos, negativos y cognitivos [?].

Síntomas positivos

Los síntomas positivos engloban todos los comportamientos que no se observan en sujetos sanos pero sí en esquizofrénicos. Pueden aparecer y desaparecer, ser intensos o casi imperceptibles. Estos síntomas incluyen:

- Alucinaciones. El tipo más común son las “voces”, pero los enfermos también pueden ver personas u objetos que no están presentes, oler aromas que nadie más percibe o sentir dedos invisibles acariciándole entre otras.
- Delirios. Engloban creencias falsas que no forman parte de la cultura de la persona y se mantienen firmes incluso después de que otras personas le demuestren que no son reales ni lógicas. También abarcan delirios paranoicos y de persecución.
- Trastornos del pensamiento. Son maneras inusuales o disfuncionales de pensar. Una manifestación de ello es el “pensamiento desorganizado” que se da cuando un individuo tiene dificultad para organizar sus pensamientos o conectarlos en una secuencia lógica. Otra manifestación es el “bloqueo del pensamiento”, es decir, que el sujeto deja de hablar repentinamente en medio de una idea.
- Trastornos del movimiento. Estos síntomas se pueden manifestar en movimientos agitados del cuerpo o, por el contrario, en catatonia (el sujeto no se mueve ni habla con los demás).

Síntomas negativos

Los síntomas negativos se asocian con la interrupción de las emociones y los comportamientos normales, es decir, son comportamientos que poseen los individuos sanos que no se manifiestan en esquizofrénicos. Son más difíciles de reconocer ya que pueden confundirse con estados depresivos. Estos síntomas incluyen:

- Afecto plano. Es decir, voz monótona sin cambio de rostro.
- Falta de satisfacción en la vida diaria.
- Falta de habilidad para iniciar y mantener actividades planificadas.
- Falta de comunicación. Incluso cuando la persona se ve forzada a interactuar.

Síntomas cognitivos

Los síntomas cognitivos son los más sutiles y, al igual que los negativos, son difíciles de reconocer como tales. Por lo general, estos síntomas hacen que sea más difícil llevar una vida normal y ganarse la vida y eso puede conllevar una gran angustia emocional. Incluyen:

- Funcionamiento ejecutivo deficiente. Es decir, incapacidad para comprender información y utilizarla para tomar decisiones.
- Problemas de concentración y de atención.
- Problemas con la “memoria de trabajo”. Es decir, dificultad a la hora de utilizar información inmediatamente después de haberla aprendido.

2.1.3. Variantes y tratamientos

Como se viene destacando, la esquizofrenia es una enfermedad compleja cuyo diagnóstico se determina mediante la suma de síntomas discriminatorios. Esta complejidad hace que la enfermedad pueda ser dividida en multitud de subtipos para facilitar su tratamiento. Los cinco tipos principales de esquizofrenia son: paranoide, desorganizada, catatónica, indiferenciada y residual [?].

Los tratamientos para la esquizofrenia se centran en eliminar los síntomas ya que, actualmente, su causa es desconocida y no tiene cura. Normalmente, el protocolo a seguir es el uso de medicamentos antipsicóticos combinado con tratamientos psicosociales [?].

En la década de 1950 se creó una generación de antipsicóticos, los llamados “típicos”. Entre ellos se encuentran la clorpromazina, el haloperidol, la perfenazina y la flufenazina.

Más adelante, en la década de 1990, se fabricaron nuevos compuestos, denominados de segunda generación o “atípicos”. El primero fue la clozapina, pero al observarse que uno de sus efectos adversos, la agranulocitosis, era demasiado agresivo, se fabricaron nuevos como la risperidona, la olanzapina, la quetiapina, la ziprasidona, el aripiprazol y la paliperidona.

Por otra parte, junto con los medicamentos, se suele utilizar una terapia psicosocial. Este tratamiento consiste en enseñar al paciente cómo convivir con su enfermedad, rehabilitarlo, educar a sus familiares para que sepan afrontar el problema, realizar terapias cognitivo-conductuales y en la creación de grupos de autoayuda.

Ninguno de estos tratamientos consigue curar la esquizofrenia, e incluso son poco efectivos en el caso de que la enfermedad esté en un alto grado de desarrollo. Por ello, es fundamental realizar un diagnóstico adecuado en las primeras etapas de la enfermedad con el fin de poder controlar la esquizofrenia a tiempo.

2.2. Electroencefalografía

La electroencefalografía es una técnica de exploración neurofisiológica que registra la actividad bioeléctrica cerebral. Un electroencefalograma (EEG) es el registro obtenido mediante esta técnica. Con el EEG se pueden observar los distintos tipos de ondas cerebrales en estado de vigilia, durante el sueño, en reposo o bajo ciertas condiciones de activación. Estas señales de EEG pueden ser utilizadas para diagnosticar multitud de enfermedades cerebrales o trastornos del sueño.

El electroencefalograma fue inventado por Hans Berger en 1924. Sin embargo su historia se remonta a mucho antes, en concreto a 1770, cuando el italiano Luigi Galvani publicó sus observaciones sobre la electricidad animal que no tuvieron demasiado éxito [?].

El 6 de julio de 1924 Hans Berger fue capaz de registrar la actividad eléctrica cerebral de su hijo Klaus, y estas señales formaron el primer electroencefalograma de la historia. La comunidad científica acogió su hallazgo con escepticismo, pero él decidió insistir. De esta forma, en la década siguiente, Hans Berger publicó una quincena de artículos sobre los distintos aspectos de su técnica, y fue en el Congreso Internacional de Psicología de París en 1937 cuando recibió el reconocimiento de los demás científicos. Sin embargo, hasta la década de 1950 no se generalizó la electroencefalografía ya que no se disponía de un perfeccionamiento tecnológico suficiente.

En esta sección, se comenzará describiendo la fisiología del encéfalo en el que se aloja la corteza cerebral que emite la mayor parte de las señales cerebrales. Finalmente, se describirán los métodos de captación de dichas señales para formar los EEGs.

2.2.1. El encéfalo

El sistema nervioso permite al organismo la comunicación y adaptación al medio en el que vive. Junto con el sistema endocrino, son los sistemas que coordinan todas las funciones corporales. En concreto, el sistema nervioso se encarga de integrar la información exterior, desechar la irrelevante y generar las respuestas adecuadas.

El encéfalo, contenido en el cráneo, es la parte más voluminosa del sistema nervioso central y es la parte que interesa describir en el presente TFM ya que es el que genera las señales de EEG. Este órgano puede dividirse en tres partes:

- *Tronco cerebral.* Conecta la corteza cerebral (córtex cerebral), la médula espinal y el cerebro. Este sistema se encarga de controlar el ritmo cardiaco y respiratorio, y algunos reflejos motores.
- *Cerebelo.* Se encarga de coordinar movimientos voluntarios y mantener el equilibrio.
- *Cerebro.* Se encarga de manejar todas las funciones conscientes del sistema nervioso. La superficie externa es la que se conoce como córtex cerebral y es la que recibe la información sensorial y que contiene la mayor parte de las neuronas del cuerpo humano.

El córtex cerebral está formado por una especie de fisuras denominadas surcos que son empleados como límites para dividir la corteza en ciertos lóbulos. Se cree que cada una de estas regiones está relacionada específicamente con ciertas modalidades de información sensitiva y motora [?]. En la Figura ?? se pueden observar las distintas regiones del córtex cerebral con sus diferentes funciones asociadas. Las señales eléctricas generadas por las neuronas del córtex, mediante distintos procesos electro-químicos, serán las que se midan posteriormente mediante una electroencefalografía [?]. Por lo tanto, sería interesante analizar las señales de cada lóbulo con el fin de detectar en qué lugares se hacen más evidentes las diferencias entre sujetos sanos y esquizofrénicos para el presente estudio.

2.2.2. Captación del electroencefalograma

Para poder registrar la actividad cerebral por medio de un electroencefalograma, es necesario colocar en la cabeza del sujeto una serie de electrodos [?]. Estos electrodos son una especie de piezas metálicas, buenas conductoras, que se insertan sobre el cuero cabelludo o el cerebro con el fin de registrar los potenciales eléctricos neuronales. Existen dos tipos de electrodos según su localización: superficiales o extracraneales (situados en el cuero cabelludo), y neuroquirúrgicos o intracraneales (situados en el córtex cerebral).

Los electrodos sobre el cuero cabelludo, que son los más utilizados, pueden ser de muchos tipos y materiales, pero los más empleados para la captación de EEG son los de

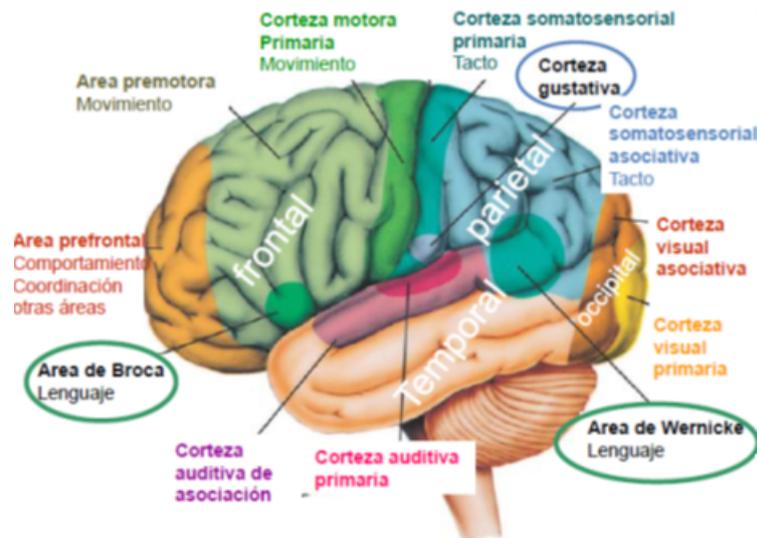


Figura 2.1: Córtex cerebral: lóbulos cerebrales y funciones asociadas a cada uno.

plata fina. Dentro de esta clase se incluyen los electrodos de almohadilla y los gorros de electrodos.

- Los electrodos de almohadilla han sido por mucho tiempo los más utilizados. Suelen ser de plata fina y su contacto se recubre con una almohadilla de espuma y gasa para posteriormente introducirse en una solución salina con el fin de mejorar su conducción. Estos electrodos se colocan en la cabeza del paciente y se sujetan mediante un gorro de gomas entrecruzadas.
- Los gorros de electrodos están constituidos por una malla elástica, mucho más cómoda que los almohadillados, en la que se posicionan los electrodos. Además, en cada posición se aplica un gel que facilita el contacto con el cuero cabelludo. Su gran ventaja frente al otro tipo es su facilidad de uso y colocación. En la Figura ?? se puede observar un gorro real de este tipo.

Para poder registrar la actividad cerebral, es necesario que los electrodos se dispongan en un orden determinado, y para ello existen una gran cantidad de sistemas que definen esta colocación, como el de Illinois, Montreal, etc. Sin embargo, el sistema más utilizado en la actualidad es el conocido como “sistema diez-veinte” [?]. Según este sistema, los electrodos han de situarse siguiendo las siguientes pautas hasta disponerlos por toda la cabeza:

1. Se mide la distancia entre el nasion y el inion pasando por el vertex. El 10% de dicha distancia sobre el nasion señala la posición del punto Fp (Frontal Polar) y, la misma distancia sobre el inion el punto O (Occipital).



Figura 2.2: Gorro con electrodos incorporados para la realización de EEG.

2. Entre los puntos Fp y O calculados en el primer paso, se sitúan otros tres puntos equiespaciados. Estos tres puntos serán, de adelante hacia atrás, el Fz (Frontal), el Cz (Central) y el Pz (Parietal).
3. A continuación, se mide la distancia entre los puntos preauriculares (situados delante del pabellón auditivo) pasando por el punto (Cz). El 10 % de esta distancia marcará la posición de los puntos temporales mediales, T3 (izquierdo) y T4 (derecho).
4. A un 20 % por encima de los puntos temporales anteriores se colocan los electrodos C3 (izquierda) y C4 (derecha).
5. Se sitúan los electrodos F3 (izquierda) y F4 (derecha) de forma equidistante entre el punto Fz y la línea que une los electrodos temporales.
6. Los electrodos P3 (izquierda) y P4 (derecha) se posicionarán equidistantes entre el punto Pz y la línea de electrodos temporales.
7. Se calcula la distancia entre Fp y O a través de T3. El 10 % de dicha distancia a través de Fp corresponde con los electrodos Fp1 y Fp2, y a través de O con los O1 y O2.
8. Los electrodos F7 y F8 se colocan equidistantes entre los puntos Fp1 y T3, y Fp2 y T4 respectivamente.

9. Los electrodos T5 y T6 se sitúan en la línea media entre los puntos T3 y O1, y T4 y O2 respectivamente.
10. A un 10% de los electrodos T3 y T4 se sitúan los electrodos auriculares A1 y A2 respectivamente.

Como regla general, se puede comprobar cómo los electrodos del lado izquierdo llevan asociado una numeración impar y viceversa. Además, los electrodos pertenecientes a la línea media reciben el subíndice “z”. En la versión europea del sistema 10-20 se presentan ligeras variaciones a la notación descrita:

- Los electrodos temporales mediales (T3 y T4) se denotan como Tm (temporal medial).
- Los electrodos F7, F8, T5 y T6 son denominados como temporales anteriores, Ta (F7 y F8), y temporales posteriores Tp (T5 y T6). Además existe una ligera diferencia en la posición.
- Los electrodos F3 y F4 son denotados como Fs (frontales superiores) y los P3 y P4 como P (parietales).
- Los electrodos C3 y C4 son representados como CI y CD.
- Se consideran dos nuevos electrodos adicionales, los mastoideos, colocados junto a las apófisis mastoides (M).

Para verificar gráficamente la posición final de los electrodos en el sistema europeo puede observarse la Figura ???. Mediante este estándar en el posicionamiento de los electrodos es posible la reproducibilidad de las diferentes investigaciones relacionadas con la señal de EEG y, por lo tanto, que éstas puedan avanzar con mayor rapidez.

La actividad eléctrica que recogen los electrodos consiste en ondas cuya amplitud oscila entre los $0,5\mu V$ y los $10mV$, y cuya frecuencia se mueve entre los $0,5Hz$ y los $100Hz$ dependiendo de la actividad que realice el cerebro. Estas señales poseen una amplitud muy baja por lo que será necesario un posterior proceso de amplificación. Además, es necesario realizar otros procesos de acondicionamiento como la referenciación, el filtrado y su posterior digitalización en caso de precisar un estudio de los mismos mediante un ordenador como se pretende en este trabajo [?].

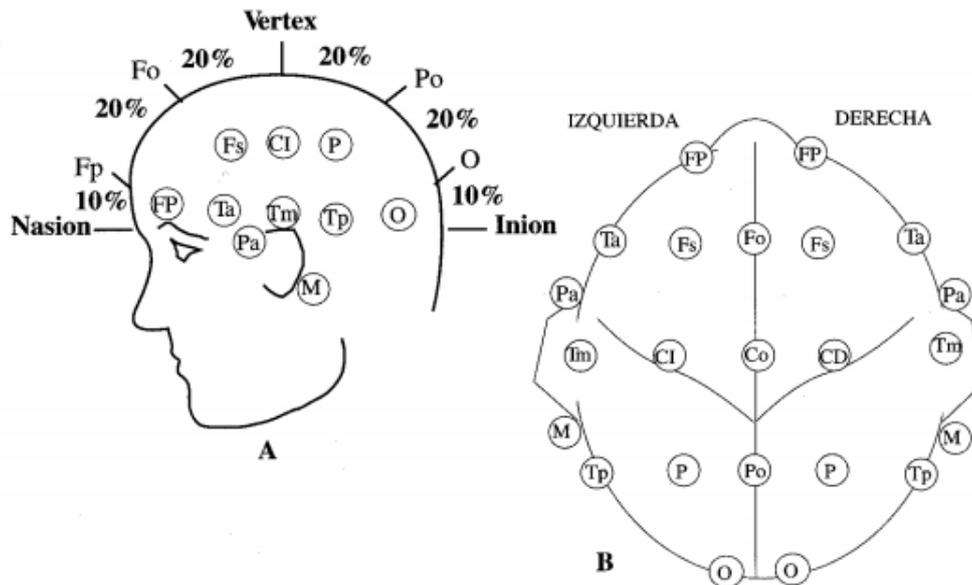


Figura 2.3: Sistema diez-veinte de posicionamiento de electrodos según la versión europea [?].

2.3. Potenciales evocados

Los potenciales evocados (PEs) son indicadores de procesos o subprocesos cognitivos o perceptivos. Su utilización se está extendiendo cada vez más en varios campos de la psicología cognitiva.

Coles y Rugg (1995) [?] definen los PEs como fluctuaciones en el voltaje del EEG provocadas por sucesos sensoriales, motores o cognitivos. Estos cambios adoptan forma de picos o valles, y son capaces de informar sobre los procesos cognitivos subyacentes asociados a los diferentes sucesos.

Pero para que estos potenciales se manifiesten en una señal de EEG, el sujeto que produce la señal debe de estar realizando una tarea que los evoque. Una de las tareas más típicas a la hora de detectar PEs es el denominado paradigma *oddball*. Este paradigma consiste en proponer al sujeto la discriminación en dos categorías: “probables” e “improbables” de una serie de elementos (auditivos, visuales, etc.) [?].

La diferencia entre un PE y un EEG radica en dos aspectos:

- Su generación. Un PE precisa de la exposición a un estímulo para aparecer. Si no se presenta un suceso, no se puede provocar un potencial evocado.
- Su voltaje. La amplitud de un PE oscila entre $0,5\mu V$ y $20\mu V$, mientras que la de un EEG es de entre $50\mu V$ y $10mV$. Esto tiene como consecuencia que las fluctuaciones registradas por un PE sean imperceptibles en un EEG.

Para solucionar la imperceptibilidad de los potenciales evocados frente a la señal de electroencefalograma existen varios métodos que permiten aislarlos de la misma. El procedimiento más utilizado consiste en el cálculo de un potencial promedio [?]. Este potencial se calcula a partir de varias exposiciones a un mismo estímulo. La actividad que se encuentre al margen de dicha excitación aparecerá aleatoriamente, mientras que la de ésta seguirá constante. Por ello, será sencillo eliminar todas las ondas residuales y obtener el registro correcto promediando varios ensayos o épocas.

Características de los PEs

Donchin, Ritter y McCallum (1978) [?] indican que una componente (potencial evocado) ha de definirse por una combinación de su polaridad, latencia, topografía y sensibilidad a las características de la manipulación experimental.

Respecto a su polaridad, las componentes pueden ser positivos o negativos (se denominan P y N, respectivamente). En lo que se refiere a su latencia, ésta suele medirse tomando el tiempo en milisegundos desde la presentación del estímulo hasta la aparición del pico o valle (punto de máxima amplitud). Así, por ejemplo, la componente N400 es una onda negativa que presenta un pico aproximadamente a los 400ms después de la presentación del estímulo.

La tercera característica es su topografía, es decir, su distribución en el cuero cabelludo. Para ello, se debe tener en cuenta que las fluctuaciones en el voltaje recogidas con un electrodo sobre el cuero cabelludo no han de tomarse como actividad originada por el tejido cerebral directamente subyacente a esta localización, es decir, que la topografía sobre el cuero cabelludo no suministra un mapa de la localización neuronal.

Como última característica se debe tener en cuenta su sensibilidad respecto a una determinada manipulación experimental. Este es el criterio más difícil de establecer, puesto que existen componentes bastante inespecíficos. A pesar de esto, sigue siendo un variable fundamental para separar componentes.

Clasificación de los PEs

Existen varias clasificaciones de los PEs [?], la más común agrupa los componentes en dos clases: los que son previos a la ocurrencia del suceso y los posteriores a ella. Entre los primeros destacan el potencial de respuesta (LRP) y la variación negativa contingente (CNV). En cuanto a los segundos, se distingue entre componentes sensoriales, las ondas ND, el potencial de disparidad (MMN), las ondas P300s y la onda N400. A continuación se describirá cada uno de ellos.

- **LRP.** Se trata de un potencial negativo que aparece cuando el sujeto se prepara para realizar un movimiento con una mano. Además, se ha observado que presenta

máxima amplitud en las localizaciones precentrales y es contralateral a la mano de respuesta escogida.

- **CNV.** Aparece cuando se presentan dos estímulos separados por un intervalo de tiempo y se establece una contingencia entre ambos. Se trata de una onda negativa lenta que alcanza su máxima amplitud cuanto más cerca se está del segundo estímulo, y parece que tiene que ver con los procesos de preparación de respuesta.
- **Componentes sensoriales.** En esta clase se recogen todos los denominados exógenos. Se caracterizan por ser respuestas obligatorias del sistema nervioso a estímulos externos.
- **Nd o N100.** Onda negativa que aparece aproximadamente a los 100ms después del estímulo. Surge al comparar la actividad evocada por estímulos atendidos con aquellos no atendidos, por lo que parece estar asociada a la atención.
- **P300, P3, P3a y P3b.** Se trata de una serie de picos positivos, los cuales, a pesar de presentar múltiples diferencias entre ellos, también presentan varios elementos comunes. En primer lugar, suelen aparecer ante estímulos inesperados que son relevantes para la tarea que ha de realizar el sujeto. En segundo, su latencia varía dependiendo de la dificultad de la tarea. En tercero, su amplitud que se ve modulada por el grado de expectación que el sujeto tiene de un determinado estímulo.
- **Potencial de disparidad (MMN).** Se obtiene al presentar un estímulo discrepante dentro de una secuencia de estímulos auditivos. Este estímulo discrepante presenta una baja probabilidad de aparición. Este componente se ha registrado incluso en niños recién nacidos y en pacientes en coma.
- **N400.** Se trata de una onda negativa, con topografía centroparietal y latencia en torno a los 400ms, que se considera un indicador del procesamiento semántico.

2.4. Detección de la esquizofrenia

En esta sección, se comenzarán describiendo las características del potencial evocado P300 como medida cuantitativa para poder distinguir la psicosis de la esquizofrenia. Finalmente, se exponen los conceptos básicos relativos al reconocimiento de patrones con el fin de describir una herramienta de gran utilidad para la clasificación entre sujetos sanos y esquizofrénicos basada en las características cuantitativas de la onda P300.

2.4.1. Onda P300

La onda P300 es un tipo de potencial evocado que puede ser registrado mediante electroencefalografía como una deflexión positiva de voltaje con una latencia de unos 300ms. Se piensa que esta componente está compuesta por dos ondas secundarias: la P3a y la P3b. Estas subcomponentes responden individualmente a diferentes estímulos: la onda P3a se origina en los mecanismos de atención frontal dirigidos por estímulo durante el procesamiento de tareas, mientras que la P3b se origina en la actividad parietal-temporal asociada con la atención y parece estar relacionada con el consiguiente procesamiento de la memoria. Se suele denominar potencial evocado diana (*target*) a la P3b, y potencial no diana (*non-target*) a la P3a [?].

En la práctica, la forma de onda P300 se debe evocar utilizando un estímulo provisto por alguna de las modalidades sensoriales. El procedimiento más típico es el paradigma *oddball*, en el que un estímulo objetivo se presenta entre estímulos con mayor frecuencia de aparición. También puede utilizarse un estímulo distractor para asegurarse de que la respuesta se debe al objetivo en lugar de a un cambio de patrón.

La onda P300, y más en concreto su componente P3b, es el potencial evocado más importante para el estudio de la psicosis de la esquizofrenia. Numerosos estudios han demostrado la variación, tanto en amplitud como en latencia, de la onda P300, mostrando una menor amplitud y mayor latencia en los individuos que poseen la enfermedad [?]. Una posible justificación a estas diferencias podría ser la definición de la onda P3b como una medida de la atención que un sujeto pone al desarrollo de una tarea y, por tanto, los individuos esquizofrénicos estarán más dispersos en la ejecución del trabajo por lo que su amplitud será menor.

2.4.2. Reconocimiento de patrones

Esquema general de reconocimiento de patrones

El reconocimiento de patrones es una disciplina científica cuyo objetivo es la clasificación de objetos en un determinado número de categorías o clases [?, ?]. Dependiendo de la aplicación, estos objetos pueden ser imágenes, señales o cualquier tipo de medida que pueda ser clasificada. En este sentido, los objetos a clasificar serán denominados “patrones”.

Teniendo en cuenta esta definición, parece obvio que este tipo de procesado puede ser aplicado al reconocimiento del patrón de la onda P300 descrita anteriormente con el fin de distinguir entre sujetos sanos y esquizofrénicos. Las etapas básicas de un esquema de reconocimiento de patrones pueden observarse en la Figura ?? y la descripción de las mismas es la siguiente:

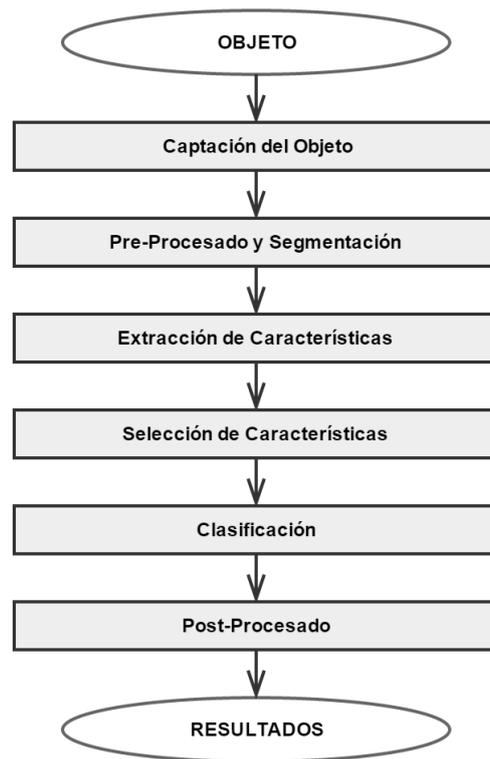


Figura 2.4: Esquema básico de un método de reconocimiento de patrones.

1. *Captación del objeto a reconocer.* Esta etapa consiste en realizar alguna medida u observación sobre el objeto o patrón que se pretende clasificar. Por ejemplo, los objetos pueden ser sonidos o imágenes y el equipamiento de medida para los mismos podría ser un micrófono o una cámara respectivamente.
2. *Pre-procesado y segmentación.* El pre-procesado consiste en la realización de un filtrado de los datos medidos del objeto con el fin de suprimir su ruido y otras operaciones que se puedan realizar en los mismos para mejorar su calidad. En la segmentación, los datos medidos son divididos de tal forma que cada división represente exactamente un objeto a clasificar.
3. *Extracción de características.* Normalmente, los objetos medidos pueden poseer una gran cantidad de componentes. Por ejemplo, la medición de una señal de EEG puede contener una cantidad enorme de valores de voltaje para cada instante de tiempo, que aumentan proporcionalmente con la tasa de muestreo utilizado y con el intervalo de tiempo medido. La extracción de características pretende extraer las características que mejor caracterizan estos datos para la clasificación ya que la mayor parte de los datos medidos no son útiles para la misma.
4. *Selección de características.* En algunas ocasiones, las características extraídas en la

etapa anterior son demasiado numerosas e incluso algunas siguen siendo inservibles para la clasificación. La selección de características permite realizar un filtrado de las características extraídas del objeto o patrón con el fin de que la clasificación de los mismos posea mayores tasas de acierto.

5. *Clasificación.* En esta etapa se toma el vector con las características seleccionadas del objeto para ser clasificadas. Existen una gran cantidad de clasificadores para realizar esta tarea.
6. *Post-procesado.* La tarea final en el reconocimiento de patrones consiste en decidir la forma en la que exponer los resultados de la clasificación realizada con el fin de cuantificar los resultados finales del reconocimiento. De esta forma, por ejemplo, se podría definir como resultado a analizar, la tasa de aciertos producida para el reconocimiento de un solo tipo de objeto en lugar de la tasa de aciertos total si la detección de un objeto concreto fuera crítico para el reconocimiento.

Esquema del método de reconocimiento de patrones del TFM

Las etapas expuestas en la Figura ?? no son más que una guía general para llevar a cabo el reconocimiento de patrones. Algunas fases pueden ser omitidas o, incluso, modificadas dependiendo del problema a tratar.

Tal y como se expuso anteriormente, el presente estudio pretende llevar a cabo un método de reconocimiento de patrones optimizado, basado en algoritmos neuro-genéticos, con el fin de lograr una clasificación entre sujetos esquizofrénicos y sujetos de control con unas altas tasas de acierto. Para ello se han propuesto diferentes algoritmos de procesado no lineal para las distintas etapas a lo largo de los diversos capítulos del presente documento. La Figura ?? resume esquemáticamente el método de reconocimiento de patrones utilizado finalmente junto con los algoritmos implementados y otros parámetros de interés de las diferentes etapas.

Puede comprobarse que existen diversos algoritmos para la realización de las mismas etapas del método de reconocimiento de patrones. Estos algoritmos pretenden ser comparados mediante el análisis de los resultados finales con el fin de obtener la combinación óptima de los mismos en cuanto a mejores resultados a la hora de la clasificación entre esquizofrénicos y controles. Además, se puede observar como tanto la etapa de clasificación como la de análisis estadístico pretenden realizar un análisis final de resultados relativos a la separabilidad entre las clases. De esta forma, se pretende que ambos análisis puedan ser complementarios para reforzar algunas conclusiones extraídas del procesado de las señales de EEG a la hora de realizar la ayuda al diagnóstico de la esquizofrenia.

Para una mayor comprensión del método final de reconocimiento de patrones empleado, los capítulos ??, ?? y ?? expondrán la teoría e implementación de los distintos algoritmos utilizados en este TFM para cada una de las etapas de dicho método.

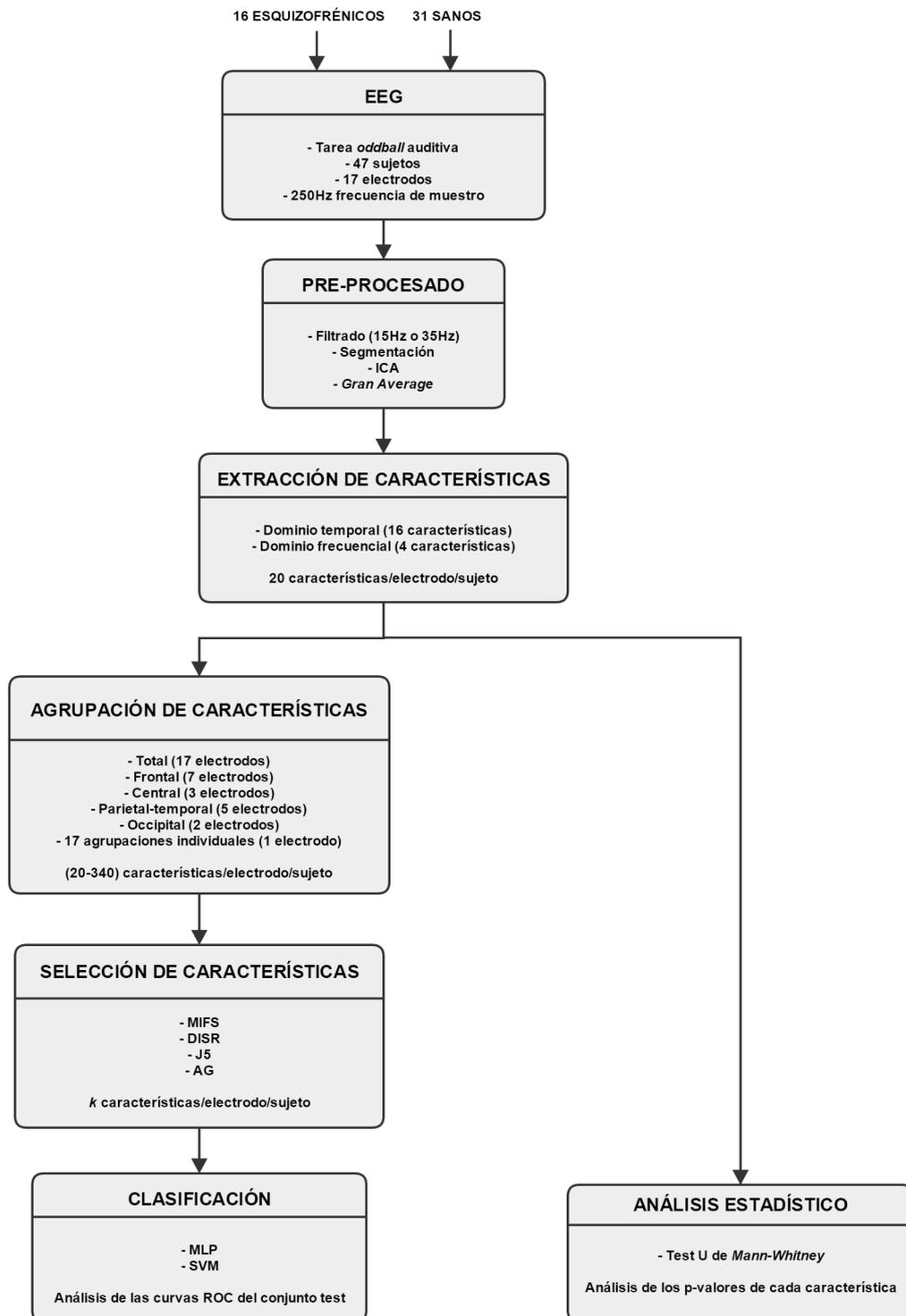


Figura 2.5: Esquema del método de reconocimiento de patrones implementado en el TFM.

Capítulo 3

Extracción y agrupación de características de la señal de EEG

En este capítulo se describirá el método llevado a cabo en la presente investigación para realizar las etapas 1, 2 y 3 del esquema de reconocimiento de patrones general descrito en el Apartado ?? además de las posibles bases teóricas subyacentes. Cabe destacar que los métodos implementados en el presente TFM para realizar estas etapas serán los mismos que se seleccionaron en [?]. En primer lugar, se describirá el método de captación de señales de EEG utilizado. En segundo lugar, se especificarán las diferentes operaciones a realizar en estas señales con el fin de adecuarlas a su posterior estudio, es decir, su pre-procesado. Posteriormente, se extraerán ciertos parámetros de interés de las señales pre-procesadas. Finalmente, se tratará sobre la realización de diferentes agrupaciones de dichos parámetros incluyendo, de esta forma, una “nueva” etapa al método general descrito anteriormente.

3.1. Adquisición de señales de EEG

En el Capítulo ?? se expusieron las diferentes técnicas empleadas para la captación de las señales de EEG y más concretamente de los potenciales evocados relacionados con la esquizofrenia. En esta sección, se procede a la concretización del método utilizado para la extracción de las señales en este trabajo que es similar al realizado en [?]. Cabe destacar que en primer lugar se realizará un análisis de la base de datos utilizada con el fin de validarla para el estudio que acontece en el presente trabajo.

3.1.1. Análisis de los sujetos del banco de datos

La onda P300, y los potenciales evocados en general, varían, tanto en amplitud como en latencia, dependiendo de diversas características de los sujetos y de su ambiente [?, ?]. Esto quiere decir que para la evaluación cognitiva no influyen exclusivamente los sistemas

cognitivos de los sujetos sino que también hay que tener en cuenta sus sistemas biológicos. Estos últimos sistemas están compuestos por los factores que afectan a la psicología del organismo y, por ende, a su capacidad para participar en acciones cognitivas.

Por lo tanto, es evidente que para evaluar la capacidad cognitiva de un sujeto, atendiendo a la forma de su onda P300, hay que tener en cuenta todos estos parámetros. Como ya se ha mencionada, el estudio que se pretende realizar en este trabajo es una clasificación automática que diferencie entre sujetos sanos y esquizofrénicos. Para realizar esta tarea es importante que los individuos a clasificar sólo se diferencien en su sistema cognitivo, siendo constante o similar entre ellos el sistema biológico. En la Tabla ?? se muestran los diferentes factores biológicos y su influencia en la onda P300, tanto en su amplitud como en su latencia.

FACTOR	AMPLITUD	LATENCIA	COMENTARIO
<i>Natural</i>			
Circadiano	Indirecto	Indirecto	Cambios en el cuerpo circadiano afectan a la P300
Temperatura del cuerpo	No	Si	Si crece la temperatura, decrece latencia
Frecuencia cardiaca	No	Si	Mayor frecuencia cardiaca, menor latencia
Ingesta de alimentos	Si	Alguno	Mayor amplitud, menor latencia
Tiempo de actividad	Si	Alguno	La comida interacciona con el tiempo de preferencia de actividad
Ultradiano	Alguno	Si	Ciclos de 90 minutos de latencia
Estacional	Si	No	Estaciones con luz aumentan la amplitud
Menstrual	No	No	Estimulo neutral, sin efectos
<i>Inducido</i>			
Ejercicio	Indirecto	Directo	Afecta sobretodo al nivel de excitación
Tónico	Si	Si	Aumenta amplitud, decrece latencia
Crónico	No	Si	Latencia decrece, resultados variables en diversos estudios
Fatiga	Si	Si	Amplitud decrece, latencia aumenta
Drogas (Comunes)	Alguno	Si	Drogas específicas, nivel de excitación, uso tónico/crónico
Cafeína	Alguno	Si	Crece la amplitud si hay fatiga, decrece latencia
Nicotina	Pequeño	Si	Efectos débiles en amplitud, decrece latencia
Alcohol			
Agudo	Si	Si	Decrece amplitud, aumenta latencia
Crónico	No	No	Beber en sociedad: No tiene efectos permanentes a largo plazo
Riesgo de alcoholismo	Si	No	Tasa de riesgo: amplitudes mas pequeñas con tareas visuales
<i>Constitucional</i>			
Edad	Si	Si	Modalidad, tarea, parámetros importantes de respuesta
Niños	Si	Si	Crece amplitud, decrece latencia
Adultos	Si	Si	Desciende amplitud, crece latencia
Inteligencia	Si	Si	Cuanta más inteligencia, menor amplitud en tareas complejas y menor latencia en tareas de clasificación perceptuales
Lateralidad	Si	Si	Amplitud: izquierda > derecha para lóbulos frontales/centrales Latencia: izquierda < derecha para lóbulos frontales/centrales
Género	Pequeño	Pequeño	Amplitud: hombre < mujer Latencia: hombre > mujer
Personalidad	Si	No	Amplitud: Introversos < Extroversos
Genetic	Si	Si	

Tabla 3.1: Parámetros biológicos y su influencia en la onda P300 [?].

Como puede observarse existen una gran cantidad de parámetros biológicos influyentes en la forma de la onda P300. Se puede hablar de factores naturales como por ejemplo la hora del día en que se realice la captación de datos de EEG. También, de factores inducidos por los propios sujetos como la cantidad de ejercicio que realicen o el consumo

de drogas (tabaco, alcohol, cannabis ...). Por último, existen parámetros constitutivos de cada sujeto como pueden ser la edad, el sexo, la personalidad o si son zurdos o diestros.

Como se mencionó anteriormente, es importante que los sujetos incluidos en el estudio posean un sistema biológico similar para que su única diferencia sea el sistema cognitivo y, al clasificarlos, lo único que los distinga sea la posesión o no de la psicosis de la esquizofrenia. Es importante advertir que no se dispone de información de todos los parámetros biológicos en la base de datos, por ejemplo, no se proporcionan detalles sobre la hora del día en que se realizó la prueba, la personalidad de cada paciente o el ejercicio que realiza. Aún así, mediante los parámetros biológicos proporcionados como la edad, sexo, nivel de educación, consumo de tabaco, alcohol o cannabis, se han estudiado los sujetos de la base de datos y, se han seleccionado 47 para este estudio. Estos individuos seleccionados poseen unas características biológicas lo suficientemente semejantes como para poder afirmar que, en gran medida, la comparación entre ellos se basa en sus características cognitivas.

Siguiendo estas directrices, para el presente estudio, se dispondrá de 47 sujetos a estudiar. De estos individuos, 16 pertenecen al grupo de esquizofrénicos con medicación (edad = 36.33 ± 10.47 , media \pm desviación estándar DE) y 31 a controles (edad = 29.85 ± 9.75), es decir, individuos que no padecen la enfermedad. Entre estos sujetos existe una paridad entre hombres y mujeres para que este factor no influya en el posterior análisis de los datos. Además, se trata de individuos relativamente jóvenes por lo que la edad no será determinante a la hora de evaluar la onda P300. La finalidad de este trabajo será construir un método de reconocimiento de patrones que diferencie las señales de EEG de ambos grupos. Por lo tanto, una vez seleccionados los individuos a caracterizar, ya se está a disposición de comenzar la tarea propuesta en este estudio.

3.1.2. Captación de las señales de EEG

Las señales de EEG del banco de datos proporcionado para el presente estudio fueron adquiridas mientras los participantes se encontraban realizando una tarea *oddball* auditiva. En este ejercicio, se pedía a los sujetos que diferenciaban entre 3 estímulos presentados de forma aleatoria: un tono objetivo (*target*) a $500Hz$ con una probabilidad de aparición de 0,20, para estudiar la onda P3b (que es la que permite diagnosticar la esquizofrenia); un tono distractor (*distracter*) a $1000Hz$ con una probabilidad de aparición de 0,20, que permitiría detectar la onda P3a (aunque no interesa en este estudio); y un tono referencia (*standard*) a $2000Hz$ con una probabilidad de 0,60.

En este proceso, se pedía a los participantes que presionaran un botón cuando detectaran el tono objetivo. Además, se les solicitaba que mantuvieran una actitud de reposo, realizando los mínimos movimientos y con los ojos cerrados. Estos consejos se

realizaban con el fin de obtener una señal de EEG en la que la mayor parte de la potencia procediera de señales cerebrales y no de otras áreas, evitando en la medida de lo posible los artefactos oculares y musculares.

Los EEGs fueron grabados usando un equipo *BrainVision* (*Brain Products GmbH; Munich, Germany*) de 17 sensores de estaño montados en un casco (*Electro-Cap International, Inc.; Eaton, Ohio, USA*) de acuerdo al sistema internacional 10-20. Los 17 electrodos seleccionados fueron: C3, C4, Cz, F3, F4, F7, F8, Fp1, Fp2, Fz, O1, O2, P3, P4, Pz, T5 y T6. La impedancia de los electrodos se mantuvo siempre por debajo de los $5k\Omega$. Además, estas grabaciones de señales cerebrales fueron referenciadas al electrodo Cz, la tasa de muestreo fue de $250Hz$, y el EEG fue adquirido de manera continua en todo el tiempo que duró la tarea.

Por lo tanto, finalmente se obtuvo una señal de EEG para cada sujeto, grabada en todo el periodo que duró la tarea *oddball* auditiva, que consta de los diferentes voltajes en el tiempo para cada uno de los 17 electrodos. Además, dentro de los datos obtenidos se incluyó un fichero de *markers* que contenía los instantes de tiempo en los que se producían los eventos S1, S2, S3 y S4 que se identificaban con la aparición del tono de referencia, del distractor, del objetivo y de la pulsación del botón por parte del sujeto respectivamente. Por último, el *software* de grabación proporcionó un archivo con la localización de los electrodos en el casco para poder realizar un análisis espacial de las señales.

La Figura ?? muestra un ejemplo de EEG grabado en un intervalo de tiempo aleatorio durante la realización de la tarea *oddball* auditiva (entre los 10 y los 15 primeros segundos) para un sujeto. Puede verificarse cómo se representa el voltaje para cada uno de los 17 electrodos descritos anteriormente. Además, se puede comprobar cómo los eventos S1, S2, S3 y S4 se marcan con una línea vertical en su momento de ocurrencia. Estos datos en bruto serán los que habrá que procesar para poder extraer diferentes conclusiones sobre el tema a tratar en la presente investigación.

3.2. Pre-procesado

Una vez adquiridas las señales de EEG es necesario adaptarlas de forma que puedan proporcionar información suficientemente relevante para realizar una clasificación adecuada a través de las mismas. Las señales de EEG grabadas en primera instancia (datos en bruto) suelen contener gran cantidad de ruido y diferentes artefactos que ocultan en mayor o menor medida la información cognitiva que se pretende analizar.

Por lo tanto, es evidente que las señales adquiridas, descritas en la sección anterior, precisan de un pre-procesado antes de poder ser utilizadas para la detección de la esquizofrenia. Para realizar este pre-procesado se dispone del *toolbox* EEGLAB [?]

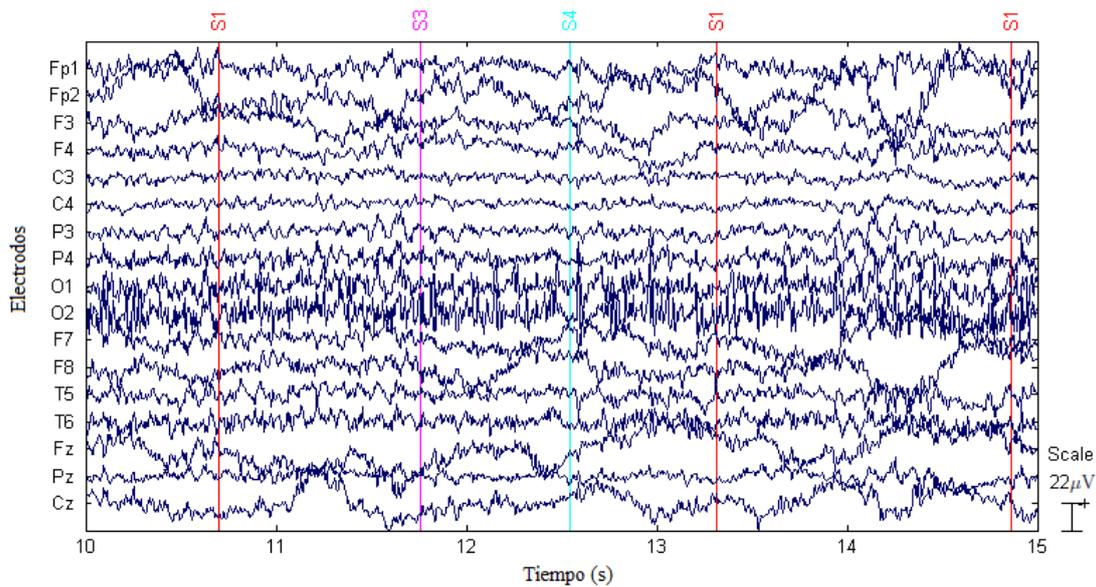


Figura 3.1: Señal en bruto de la señal de EEG continua.

de MATLAB. Este *toolbox* es de código abierto e incluye el análisis de componentes independientes (ICA) que es una de las herramientas más potentes para la supresión de artefactos en la actualidad. Además, el EEGLAB presenta una interfaz gráfica muy intuitiva que facilita en gran medida su uso.

En los siguientes sub-apartados de esta sección se detallarán las distintas operaciones que ya se llevaron a cabo en [?] para obtener una señal libre de ruidos y artefactos.

3.2.1. Filtrado

Para el estudio que pretende este trabajo, sólo es necesario el análisis de los potenciales evocados relacionados con el evento “*pulsación del botón por parte del sujeto al escuchar el tono objetivo*”. Como ya se ha mencionado, el potencial evocado más importante es la onda P3b que surge como respuesta del cerebro a la realización de una actividad en la que los sujetos tienen que prestar atención. De esta forma, midiendo distintas características del potencial P3b, y por lo tanto, cuantificando el nivel de atención de los diferentes sujetos, se podrían estudiar las posibles diferencias entre ambos grupos de sujetos (sanos y esquizofrénicos) que inferirían una capacidad de atención diferente para cada uno como se ha venido afirmando en investigaciones anteriores.

Los potenciales evocados poseen unas frecuencias, por regla general, menores de 35Hz. Esta circunstancia hace que sea conveniente un filtrado paso bajo de las señales de EEG ya que, de esta forma, se podrá eliminar gran cantidad de ruido y artefactos que suelen

aparecer a altas frecuencias [?] sin eliminar la componente de señal relativo a los procesos cognitivos que se pretenden analizar. Por otro lado, también es importante la supresión de las frecuencias cercanas a 0Hz con el fin de eliminar las componentes de continua en los EEGs.

Por todo lo comentado anteriormente, la gran mayoría de los trabajos vinculados con la detección de la P300 realizan este tipo de filtrado como paso fundamental en el pre-procesado de la señal de EEG. Por ejemplo en [?], que ha sido referencia para multitud de investigaciones actuales, se usa un filtrado paso bajo a 35Hz y paso alto para suprimir la componente en continua con el fin de detectar la onda P300. El tema del filtrado es tan crítico a la hora de poder alcanzar un buen pre-procesado que en [?] se realiza un análisis para investigar cuál es el filtro más eficiente en la detección de la onda P300. Además, se llega a la conclusión de que el filtrado que mejor detectaría el potencial evocado y más cantidad de artefactos y ruido eliminaría sería un filtrado paso bajo a 15Hz sumado a otro paso alto que elimine la componente de continua de la señal de EEG.

En este trabajo se han llevado a cabo 4 tipos de filtrado para evaluar su eficacia a la hora de detectar de la onda P3b y suprimir el ruido y los artefactos de los EEGs del banco de datos utilizado. Los filtros implementados han sido los siguientes:

- **Filtro paso alto a 1Hz.** Este filtrado se realiza para eliminar exclusivamente la componente de continua tal y como se recomienda el *toolbox* EEGLAB [?] con el fin de minimizar los artefactos en los bordes de las épocas.
- **Filtro paso alto a 1Hz + filtro de banda eliminada a 50Hz.** Este filtro sumaría al anterior la exclusión de la componente ruidosa de la señal eléctrica.
- **Filtro paso banda entre 0.02Hz y 35Hz.** Este filtrado es implementado en [?] y es propuesto para verificar su rendimiento en el problema que acontece a este trabajo.
- **Filtro paso banda entre 0.1Hz y 15Hz.** Esta opción es la que [?] considera óptima a la hora de abordar el problema de la detección de la onda P300 y la supresión de los artefactos en la señal de EEG. En la Figura ?? puede verse la señal de EEG mostrada en la Figura ?? después de haber sido filtrada mediante este filtro.

Una vez realizados los 4 filtrados se evaluaron sus rendimientos. Para ello se utilizó el análisis ICA (que será detallado en las secciones posteriores de este trabajo) de las señales resultantes. Se comprobó cómo los dos primeros filtrados constaban de gran cantidad de componentes ruidosos y, por consiguiente, estos artefactos ocultaban en gran medida las

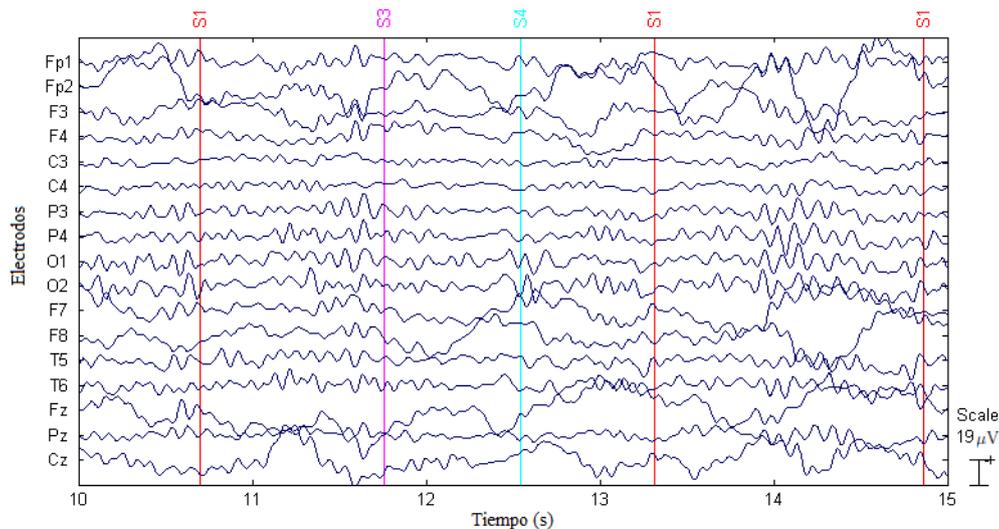


Figura 3.2: Señal continua de EEG tras el filtrado paso banda entre 0.1 y 15Hz.

potencias evocadas por el cerebro. Por tanto, se decidió continuar con el pre-procesado utilizando exclusivamente el filtro paso banda entre $0,02Hz$ y $35Hz$ y el filtro paso banda entre $0,1Hz$ y $15Hz$. Se pretende comparar ambos filtrados posteriormente en lo que respecta a su bondad a la hora de clasificar los sujetos mediante el esquema de reconocimiento de patrones, y verificar cuál de los dos proporciona unas tasas de acierto mayores en la clasificación.

3.2.2. Segmentación

Una vez realizado el filtrado de la señal para suprimir la mayor cantidad de ruido posible de los EEGs, llega el momento de seleccionar los instantes temporales de señal cerebral que proporcionan información sobre la onda P3b. Para ello, es preciso segmentar la señal de EEG en bruto en diferentes épocas que recojan los eventos que originan el potencial evocado de interés. Estas épocas serán instantes temporales de cierta duración dentro del total del EEG continuo. Además, una vez identificadas, no todas ellas serán válidas para el estudio que acontece en este trabajo. A continuación, se procede a la descripción del proceso realizado para recoger estos instantes temporales y seleccionar cuáles de ellos son legítimos para el objetivo de esta investigación.

Extracción de épocas

La onda P3b aparece en un intervalo entre 250 y 500ms después de escuchar el tono objetivo y reconocerlo. Por lo tanto, las épocas extraídas para el estudio de estos

potenciales se extenderán desde los 200ms antes del estímulo *target*, con el fin de lograr eliminar la línea base, hasta los 800ms después del estímulo, con el objetivo de detectar sin problemas la onda P3b y verificar que existe un evento S4 posteriormente, es decir, que el usuario haya pulsado el botón frente al tono objetivo.

Por lo tanto, se creará una época por cada estímulo objetivo que se muestre a los sujetos que efectúen la tarea *oddball*. Pero no todos los instantes temporales seleccionados serán válidos para el estudio de este trabajo. En la Figura ?? pueden verse cinco épocas de un EEG filtrado en las que es inmediato comprobar que no todas están compuestas por un estímulo S3 seguido de un S4 por lo que no serían válidas. La selección de las épocas es otra de las etapas más críticas a la hora del pre-procesado de las señales de EEG.

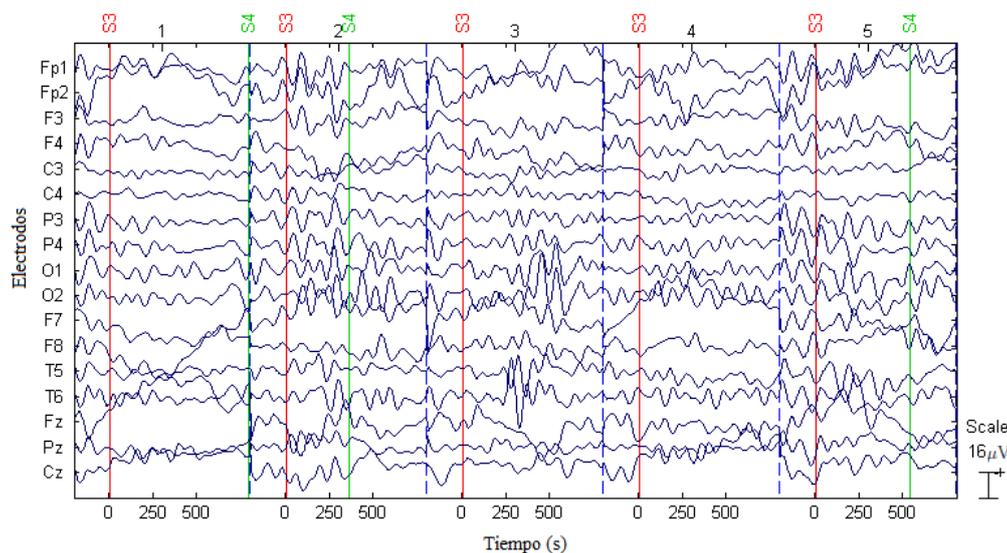


Figura 3.3: Cinco épocas de una señal de EEG filtrada seleccionada al azar.

Selección de épocas

Una vez segmentado el EEG continuo en diferentes épocas para obtener sus potenciales evocados, el siguiente paso es seleccionar cuáles de éstas son válidas para su posterior estudio. La primera forma de “limpiar” el EEG consiste en seleccionar las épocas que contengan únicamente un evento S3 seguido de otro S4, es decir, los instantes temporales en los que se haya presentado el tono objetivo (S3) y, posteriormente, el sujeto haya pulsado el botón detectándolo (S4). Es importante advertir, que tanto si el sujeto no ha pulsado el botón o, incluso si lo ha hecho a partir de los 800ms después del tono *target*, se considerará que el individuo no ha realizado correctamente la tarea y la época

será eliminada. Una vez realizado este paso, las épocas se acortarán y, únicamente, abarcarán hasta los 600ms con el objetivo de eliminar ventanas temporales en las que no se manifieste la onda P300.

Por otro lado, también se eliminarán las épocas que contengan artefactos fácilmente detectables. Se definirán una serie de características para marcar una señal perteneciente a una época como artefacto y, por consiguiente, suprimirla. Las señales que se descartarán serán las siguientes [?]:

- Las señales de las épocas que no se encuentren en el intervalo de valores entre $+30\mu\text{V}$ y $-30\mu\text{V}$. La explicación para el uso de este criterio es que los potenciales evocados, y más en concreto la onda P3b, no exceden nunca de estos valores tanto para individuos sanos como esquizofrénicos (a excepción de los esquizofrénicos con psicosis cíploide que no se contemplan en el banco de datos de este estudio) [?,?].
- Las señales que posean tendencias anormales. En este método se ajustan los datos a una línea recta para detectar alguna tendencia, eliminando las épocas donde la pendiente exceda los $40\mu\text{V}$, basándose en que los artefactos causan tendencias lineales en algunos electrodos.
- Las señales que posean datos improbables. Para ello se crea una función que determine la distribución de probabilidad de las épocas, calculando la probabilidad de ocurrencia de cada una. Posteriormente, se eliminan las épocas con eventos improbables y se espera que éstas sean debidas a artefactos. Se usó un umbral de rechazo de valor 3 (expresado en término de las desviaciones estándar de la distribución de probabilidad). Para más detalles sobre el uso de este valor se puede consultar [?].
- Las señales que tengan datos anormalmente distribuidos. Para el uso de este procedimiento se emplea el cálculo del coeficiente de *curtosis*, cuarto momento estadístico, y se utiliza para detectar el comportamiento de la distribución de las señales. Por lo tanto, si el valor de *curtosis* es alto y negativo, la distribución será plana o *platicúrtica*; si el valor es cercano a 0, la distribución será parecida a una distribución normal o *mesocúrtica*; pero si el valor es alto y positivo, la distribución será puntiaguda o *leptocúrtica*. Este método es de gran utilidad debido a que las épocas de datos con artefactos a menudo poseen distribuciones de la actividad muy picudas o muy planas por algún canal suelto. Los umbrales son definidos en términos de las desviaciones estándar del valor y también se utilizó un valor de 3 para rechazar las épocas. Para comprender mejor este valor se puede revisar [?].

Una vez realizado esta etapa, se habrán suprimido una gran cantidad de artefactos de las señales de EEG. Es importante advertir que el filtrado paso banda entre 0.1 y 15Hz es el que menor número de épocas elimina en este paso por lo que se puede pensar que es el que suprime mayor número de artefactos.

3.2.3. Análisis ICA

El Análisis de Componentes Independientes (*Independent Component Analysis*, ICA) [?] es un método de descomposición lineal que tiene como objetivo la recuperación de fuentes de señal estadísticamente independientes a partir de mezclas de dichas señales. Esta representación es de gran utilidad para una gran variedad de campos que incluyen la extracción de características y la separación de señales.

Fundamento

El fundamento del análisis ICA es la separación de fuentes. Para entender este problema se plantea una habitación en la que dos individuos están hablando simultáneamente. Además, en dicho recinto se encuentran dos micrófonos situados en distintas localizaciones que están recibiendo ambas señales mezcladas. Estos micrófonos grabarán, por tanto, dos señales que se denotarán como $x_1(t)$ y $x_2(t)$ donde x_1 y x_2 son las amplitudes y t es el índice de tiempos. Cada una de estas señales puede ser expresada como combinación lineal de las señales emitidas por cada uno de los individuos que se denominan $s_1(t)$ y $s_2(t)$ y, por tanto, se puede escribir que:

$$\begin{aligned}x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t)\end{aligned}\tag{3.1}$$

Donde a_{ij} son parámetros que dependen de las distancias entre los individuos y los micrófonos. En este punto, sería de gran utilidad conseguir estimar las señales originales $s_1(t)$ y $s_2(t)$ usando solamente las señales grabadas $x_1(t)$ y $x_2(t)$. Este problema es denominado *Cocktail-Party Problem*.

Si se conocieran los parámetros a_{ij} , se podría resolver sin problema la ecuación lineal anterior mediante métodos clásicos. Pero el desconocimiento de los valores de a_{ij} hace el problema considerablemente más complicado. Un posible enfoque para la resolución de este problema podría ser la estimación de los parámetros a_{ij} basada en las propiedades estadísticas de las señales $s_i(t)$. En la mayoría de los casos, basta con asumir que las señales $s_1(t)$ y $s_2(t)$ son estadísticamente independientes aunque en la práctica muchas no lo sean.

De forma ilustrativa, a modo de ejemplo, se pueden considerar las dos señales mostradas en las Figuras ?? y ?. Estas señales de voz no son reales pero son bastante representativas para comprender el objetivo del análisis ICA. Las señales de la Figura ?? representan los datos originales correspondientes a $s_1(t)$ y $s_2(t)$. Una vez las señales originales son emitidas a la habitación, los micrófonos reciben las de la Figura ?? que se corresponden con las señales mezcladas $x_1(t)$ y $x_2(t)$. Para las dos señales combinadas no se conocen los emisores ni la forma en la que han sido mezcladas dichas fuentes de señal. Lo que pretende el análisis ICA es reconstruir la información original a partir de la mezclada, es decir, recuperar las señales de la Figura ?? a partir de las de la Figura ??.

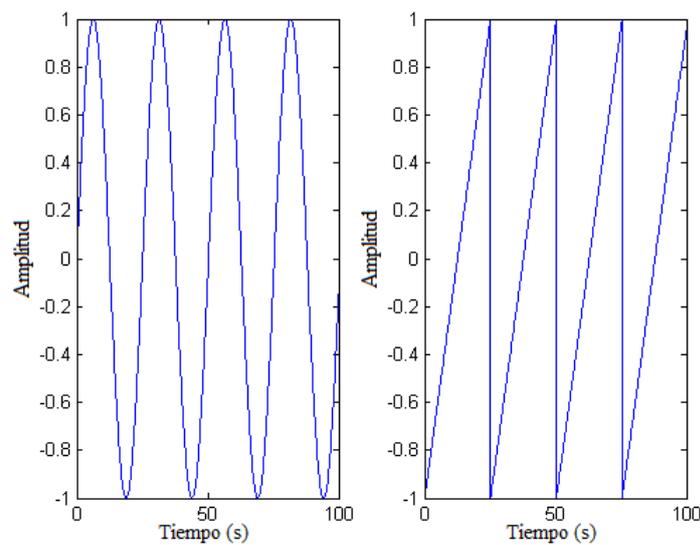


Figura 3.4: Ejemplo de señales fuente $s_1(t)$ y $s_2(t)$.

Definición

Para definir rigurosamente el análisis ICA [?] se emplea el modelo estadístico de variables latentes, es decir, de variables que no pueden ser directamente observadas. Asumiendo que se recogen n combinaciones lineales x_1, \dots, x_n de n componentes independientes, se puede expresar que:

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n \quad \forall j \quad (3.2)$$

Donde cada x_j representa una mezcla y cada s_n una componente independiente. Estos parámetros serán variables aleatorias de media cero.

Es conveniente expresar la ecuación (??) de forma vectorial y matricial. Se denota \mathbf{x} al vector aleatorio cuyos elementos son las mezclas x_1, \dots, x_n y de la misma forma se define el vector \mathbf{s} con elementos s_1, \dots, s_n . Además, se definirá \mathbf{A} como la matriz cuyos elementos

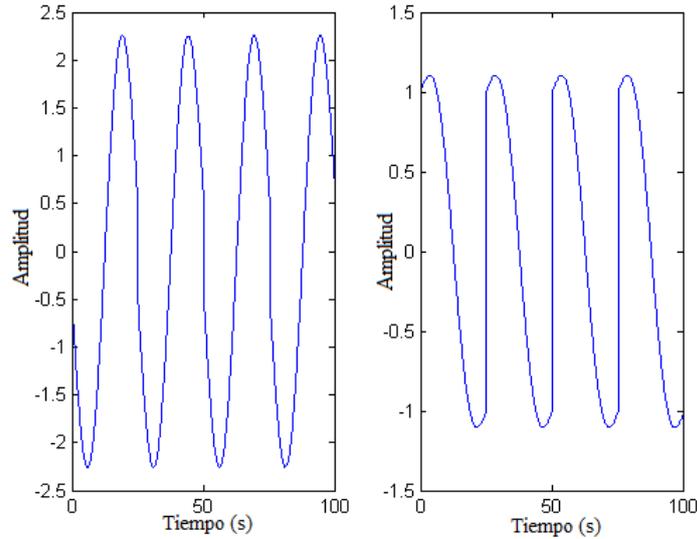


Figura 3.5: Ejemplo de señales mezcla $x_1(t)$ y $x_2(t)$.

son los parámetros a_{ij} . Es importante advertir que se entiende que todos los vectores se definen como vectores columna. Usando esta notación la ecuación (??) se escribe:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (3.3)$$

En algunas ocasiones se precisará de las columnas de \mathbf{A} que serán denotadas como \mathbf{a}_j . De esta forma, el modelo podría ser descrito matemáticamente como:

$$\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i \quad (3.4)$$

El modelo estadístico de la ecuación (??) es denominado análisis de componentes independientes o ICA. El modelo de ICA es un modelo generativo, lo que significa que los datos observados son generados por un proceso de mezcla de las componentes originales (s_i). Las componentes independientes son variables latentes, es decir, no pueden ser directamente observadas. Incluso la matriz de mezclado se supone desconocida. Todo lo que se observa es el vector aleatorio \mathbf{x} mediante el cual se deberán de estimar tanto \mathbf{A} como \mathbf{s} .

En primer lugar, se debe asumir que las componentes s_i son estadísticamente independientes. Además, se admite que las fuentes tienen distribuciones *no-gaussianas* y el modelo de mezcla es conocido (lineal, convolutivo o no lineal). Por simplicidad, se asume que la matriz desconocida de mezclado es cuadrada. Entonces, después de estimar \mathbf{A} , es posible calcular su inversa, \mathbf{W} , y obtener las componentes independientes como:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (3.5)$$

Supresión de artefactos

Una vez descrito el funcionamiento del análisis de componentes independientes parece evidente que una aplicación inmediata de éste puede ser la supresión de artefactos y ruidos. Para realizar esta tarea bastaría con separar las señales de EEG y eliminar las componentes ruidosas no deseadas. De esta forma el EEG sólo se compondría de las señales originadas por la actividad neuronal.

Este proceso es exactamente el que se ha realizado sobre las señales filtradas y segmentadas del banco de datos en los pasos anteriores del pre-procesado. Haciendo uso del *toolbox* EEGLAB de MATLAB, se han dividido dichas señales en sus componentes independientes y se han suprimido las que no estaban originadas por el cerebro como, por ejemplo, los movimientos oculares o musculares. Además, se ha hecho uso del *plugin* ADJUST [?] del EEGLAB que es capaz de reconocer automáticamente la mayor parte de los artefactos oculares, aunque la supresión final de artefactos ha sido supervisada. El algoritmo utilizado para realizar el análisis ICA ha sido el *Infomax* [?] que se basa en el principio de transmisión de información óptima a partir de neuronas. La finalidad de este algoritmo es el cálculo de la matriz de pesos \mathbf{W} con el fin de, a partir de ella, obtener las componentes independientes y eliminar los que pertenezcan a artefactos.

3.2.4. Grand Average

Una vez suprimido ruido de la señal en la medida de lo posible y, por lo tanto, teniendo una señal de EEG perteneciente únicamente a los procesos cognitivos, el último paso en el pre-procesado será la extracción del potencial evocado P3b del mismo. Como se mencionó en el Capítulo ??, el procedimiento más utilizado consiste en el cálculo de un potencial promedio [?] ya que la actividad que se encuentre al margen de la excitación producida por el estímulo auditivo aparecerá aleatoriamente y la perteneciente al mismo permanecerá constante. De esta forma, será sencillo eliminar todas las ondas residuales y obtener la onda P3b promediando varias épocas.

Por lo tanto, como última etapa en el pre-procesado se realizará un promediado de todas las épocas libres de artefactos para cada sujeto con el fin de obtener sus potenciales evocados. Esta señal promediada será el objeto de análisis en los siguientes puntos. De esta manera, finaliza el pre-procesado de la señal que ya estará dispuesta para poder ser utilizada en las etapas posteriores del reconocimiento de patrones implementado.

3.3. Extracción de características

La extracción de características de la señal EEG se define como el cálculo de algunos parámetros de dicha señal con el fin de poder caracterizarla mediante los mismos. Cabe destacar que la segmentación realizada en la señal de EEG inicial hace que la *grand average* calculada finalmente pueda ser considerada como una señal estacionaria lo que será de gran utilidad a la hora de su posible transformación a otros dominios.

Actualmente, existen multitud de características utilizadas para caracterizar los diferentes potenciales cerebrales tanto en el dominio temporal como frecuencial [?, ?, ?]. En este estudio se han seleccionado las 20 características sencillas que ya fueron extraídas en [?] y que permiten caracterizar el potencial evocado P3b en cada canal y sujeto con el fin de seleccionar posteriormente las que mejor caractericen las diferencias entre el grupo de sujetos sanos y esquizofrénicos. A continuación se mostrarán dichas características divididas en dos grandes grupos: las calculadas a partir la señal en el dominio temporal y las calculadas a través de su densidad espectral de potencia en el dominio frecuencial.

Medidas en el dominio temporal

La señal exportada del pre-procesado anterior ya se encuentra en el dominio del tiempo. Cabe destacar que el potencial P3b se evoca aproximadamente entre los 300ms y los 600ms después del estímulo objetivo por lo que la mayoría de las características temporales se refieren a este intervalo. Además, es importante recordar que la señal que ofrecía finalmente el pre-procesado consistía en una época promediada que iba desde los 200ms antes del estímulo objetivo, hasta los 600ms post-estímulo. Por tanto, basta con estudiar dicha señal de voltaje $x(t)$ ¹ para extraer los siguientes parámetros:

1. **Media de la señal.** Se define como la suma de todos los valores que toma la señal dividido entre el número de puntos de los que consta dicha señal (N puntos).

$$media = \frac{\int_{-200}^{600} x(t)dt}{800} \quad (3.6)$$

2. **Varianza de la señal.** Se define como la media de las diferencias cuadráticas en los N puntos de señal con respecto a su media aritmética.

$$varianza = \frac{1}{800} \int_{-200}^{600} (x(t) - media)^2 dt \quad (3.7)$$

¹En realidad la señal a analizar será discreta ya que se está realizando un procesado digital.

3. **Energía de la señal.** Energía que posee la señal en toda su extensión.

$$energia = \int_{-200}^{600} (x(t))^2 dt \quad (3.8)$$

4. **Onda P300 como valor de pico positivo.** Este parámetro intenta captar la onda P300 definida como el valor de pico positivo del EEG en el intervalo de tiempos entre 300 y 600ms que es donde aparece dicha onda.

$$P300ValorPico = \max(x(t)) \quad \forall t \in (300, 600)ms \quad (3.9)$$

5. **Onda P300 como valor de pico en valor absoluto.** Define la P300 como el valor absoluto máximo de señal encontrada en el intervalo 300-600ms.

$$P300ValorPicoAbs = \max(|x(t)|) \quad \forall t \in (300, 600)ms \quad (3.10)$$

6. **Onda P300 como amplitud media.** Define la P300 como la amplitud media de la señal en el intervalo 300-600ms.

$$P300Media = \frac{\int_{300}^{600} x(t)}{600 - 300} \quad (3.11)$$

7. **Latencia de la onda P300 como valor de pico positivo.** Se define como el instante temporal en que se encuentra la onda P300 si es definida como el valor de pico positivo en el intervalo 300-600ms.

8. **Latencia de la onda P300 como valor de pico en valor absoluto.** Se define como el instante temporal en que se encuentra la onda P300 si es definida como el valor de pico en valor absoluto en el intervalo 300-600ms.

9. **Razón latencia/amplitud, LAR.** Se define como la división entre la latencia y la amplitud de la onda P300 definida como el valor de pico positivo en el intervalo 300-600ms.

$$LAR = \frac{P300Latencia}{P300ValorPico} \quad (3.12)$$

10. **Razón latencia/amplitud en valor absoluto.** Se define como la división entre la latencia y la amplitud de la onda P300 definida como el valor absoluto del pico en el intervalo 300-600ms.

$$LARAbs = \frac{P300LatenciaAbs}{P300ValorPicoAbs} \quad (3.13)$$

11. **Área positiva en el intervalo 300-600ms.** Se define como la suma de los valores positivos de la señal en el intervalo 300-600ms.

$$A_{positiva} = \frac{1}{2} \int_{300}^{600} (x(t) + |x(t)|) dt \quad (3.14)$$

12. **Área negativa en el intervalo 300-600ms.** Se define como la suma de los valores negativos de la señal en el intervalo 300-600ms.

$$A_{negativa} = \frac{1}{2} \int_{300}^{600} (x(t) - |x(t)|) dt \quad (3.15)$$

13. **Área total en el intervalo 300-600ms.** Se define como la suma de las áreas positiva y negativa en el intervalo 300-600ms.

$$A_{total} = A_{positiva} + A_{negativa} \quad (3.16)$$

14. **Área total absoluta, ATAR, en el intervalo 300-600ms.** Se define como el valor absoluto del área total.

$$ATAR = |A_{total}| \quad (3.17)$$

15. **Área absoluta total, TAAR, en el intervalo 300-600ms.** Se define como la suma del área positiva con la negativa en valor absoluto, ambas en el intervalo 300-600ms.

$$TAAR = A_{positiva} + |A_{negativa}| \quad (3.18)$$

16. **Cruces por cero en el intervalo 300-600ms.** Se define como el número de veces que la señal pasa por cero en el intervalo 300-600ms.

$$ZC = \int_{300}^{600} \delta_x dt \quad (3.19)$$

Donde δ_x se refiere a $\delta(x(t))$ y, por lo tanto, toma valor 1 cuando $x(t) = 0$.

Medidas en el dominio frecuencial

Para calcular las características de la señal en el dominio frecuencia, el primer paso es la transformación de $x(t)$ a $X(f)$ por medio de la Transformada de *Fourier* considerando la señal estacionaria. Posteriormente, se calcula la densidad espectral de potencia [?], $S_x(f)$, de la que se extraerán finalmente las siguientes características frecuenciales:

1. **Entropía espectral.** Entropía de la densidad espectral de potencia [?].

$$PSE = - \int_0^{f_n} S_x(f) \log S_x(f) df \quad (3.20)$$

Donde f_n es la frecuencia de *Nyquist* (es decir, la mitad de la frecuencia de muestreo) y $\log S_x(f)$ se refiere al logaritmo en base 2.

2. **Frecuencia mediana.** Definida como la frecuencia que separa el espectro de potencia en dos áreas iguales bajo la curva.

$$\int_0^{f_{mediana}} S_x(f) df = \int_{f_{mediana}}^{\infty} S_x(f) df \quad (3.21)$$

3. **Frecuencia media.** Se calcula a partir del promedio ponderado de las frecuencias en la densidad espectral de potencia.

$$f_{media} = \frac{\int_0^{\infty} f S_x(f) df}{\int_0^{\infty} S_x(f) df} \quad (3.22)$$

4. **Frecuencia moda.** Se define como la frecuencia con mayor concentración de energía en el espectro de la señal, es decir, donde se encuentra la máxima amplitud de la densidad del espectro de potencia de la señal.

$$S_x(f_{moda}) = \max(S_x(f)) \quad (3.23)$$

3.4. Agrupación de características

El estudio de las características extraídas anteriormente para cada electrodo de cada sujeto (20 características para cada uno de los 17 electrodos) será completado utilizando también ciertas combinaciones de los parámetros de diferentes electrodos. Estas agrupaciones tendrán la finalidad de combinar características de diferentes canales con alguna relación con el fin de permitir extraer ciertas conclusiones espaciales posteriormente. De esta forma, se realizarán las agrupaciones de características descritas a continuación:

1. **Agrupación total.** Agrupa todas las características de los 17 canales.
2. **Agrupación frontal.** Agrupa todas las características de los 7 canales frontales (Fp1, Fp2, F3, F4, F7, F8 y Fz).
3. **Agrupación central.** Agrupa todas las características de los 3 canales centrales (C3, C4 y Cz).

4. **Agrupación parieto-temporal.** Agrupa todas las características de los 5 canales parieto-temporales (P3, P4, Pz, T5 y T6).
5. **Agrupación occipital.** Agrupa todas las características de los 2 canales occipitales (O1 y O2).
6. **17 agrupaciones individuales.** Toman las 20 características de cada electrodo por separado.

En este punto del método de reconocimiento de patrones ya se han calculado todas las características de las señales de EEG que podrían ser de interés a la hora de diferenciar entre sujetos sanos y esquizofrénicos. El inconveniente radica en que el número de parámetros extraídos es demasiado alto para poder ser introducidos en un clasificador por lo que se precisará de una selección de los mismos posteriormente.

Capítulo 4

Selección de características para las señales de EEG

Las características extraídas de las señales de EEG en el punto anterior son muy numerosas. Además, muchos de estos parámetros pueden ser irrelevantes o redundantes en el contexto de diferenciabilidad entre las dos clases de individuos. El aprendizaje de los clasificadores en este contexto plantea cuestiones importantes como el sobre-aprendizaje de datos irrelevantes de las señales¹ y la gran carga computacional de procesamiento que se requiere. Por estos motivos, es importante poder identificar automáticamente subconjuntos más pequeños dentro de las características iniciales, es decir, realizar una selección de características (*Feature Selection*) [?].

El objetivo de esta sección es describir la etapa de selección de características que se implementará en la presente investigación junto con sus bases teóricas para lograr una mayor comprensión de la misma. Para realizar la selección de características de este estudio se han implementado cuatro algoritmos distintos con el fin de evaluar la bondad de los mismos relativa a las tasas de acierto de la clasificación final del método de reconocimiento de patrones.

En los siguientes puntos se procede a la descripción de los diferentes algoritmos de *feature selection* a incluir en la etapa 4 del método de reconocimiento de patrones general descrito en el Apartado ???. En primer lugar se detallarán dos de los algoritmos, que ya se utilizaron en [?], basados en la teoría de la información (MIFS y DISR). En segundo lugar, se describirá el algoritmo J5, también utilizado en [?], para la selección de características que está fundamentado en la suposición de que las diferentes clases de individuos ocupan distintas regiones del espacio de las mismas. Finalmente, se abordarán los Algoritmos Genéticos como un nuevo método de selección de características basado en la teoría de la evolución de Darwin.

¹Error por el que un clasificador aprende a diferenciar un determinado conjunto de datos pero no tiene capacidad de generalizar para nuevos datos.

4.1. Algoritmos basados en la teoría de la información

4.1.1. MIFS

La Selección de Características mediante Información Mutua (*Mutual Information Feature Selection*, MIFS) [?] es un algoritmo para la selección de características que considera el uso de la información mutua para evaluar el “contenido en información” de cada característica con el fin de cuantificar la separabilidad que proporciona cada una de ellas. A continuación se procede a la descripción de las bases teóricas de este algoritmo y de su implementación en el presente trabajo.

Entropía e Información Mutua

La entropía de una variable aleatoria es la unidad fundamental de información en este análisis [?]. Este parámetro se denota como $H(X)$ y cuantifica la incertidumbre presente en la distribución de X . Es definida como:

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (4.1)$$

Donde x denota un posible valor que la variable X puede adoptar. Además, el logaritmo utilizado es en base 2. Por tanto, para calcular la entropía se precisa de una estimación de la distribución $p(X)$ y de la resolución de la ecuación (??). Si el valor de la entropía es bajo significa que la incertidumbre para conocer los valores de la variable X es pequeña. Por el contrario, si todos los valores de X son igual de probables, la incertidumbre para conocer su valor será máxima y, por ende, la entropía también. Siguiendo las reglas de la teoría de la probabilidad, la entropía puede estar condicionada a otros eventos. De esta forma, la entropía de X condicionada a Y se denota como:

$$H(X|Y) = - \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) \quad (4.2)$$

Este concepto puede definirse como la cantidad de incertidumbre que posee la variable X cuando se conoce el valor de Y . Una vez comprendido el concepto de entropía, se puede definir la información mutua entre las variables X e Y ($I(X;Y)$) como la cantidad de información compartida por dichas variables calculada como:

$$I(X;Y) = H(X) - H(X|Y) = \sum_{x \in X} \sum_{y \in Y} p(xy) \log \frac{p(xy)}{p(x)p(y)} \quad (4.3)$$

Por lo tanto, teniendo en cuenta esta definición, se puede verificar como la información mutua no es más que la diferencia entre dos entropías. Esta definición también puede ser

interpretada como la cantidad de incertidumbre de X que es eliminada al conocer Y , es decir, la cantidad de información que una variable proporciona sobre la otra. Cabe destacar que, según esta definición, la información mutua es simétrica, en otras palabras, $I(X; Y) = I(Y; X)$ y es cero si y solo si las variables son estadísticamente independientes, es decir, $p(x, y) = p(x)p(y)$. Todos estos conceptos son mostrados gráficamente en la Figura ?? para una mayor comprensión de las mismas.

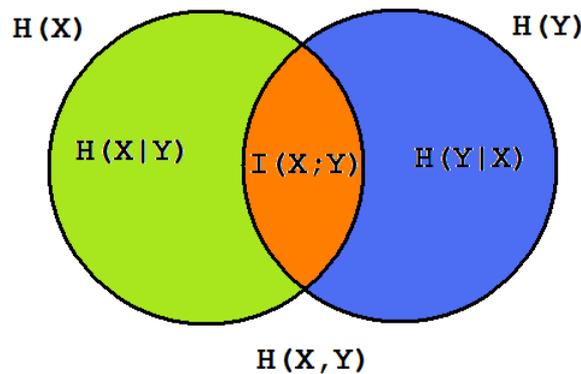


Figura 4.1: Ilustración de varios conceptos relativos a la teoría de la información.

Por último, es importante mencionar que la información mutua también puede ser condicionada. La información condicionada se define como:

$$I(X; Y|Z) = H(X|Z) - H(X|YZ) = \sum_{z \in Z} p(z) \sum_{x \in X} \sum_{y \in Y} p(xy|z) \log \frac{p(xy|z)}{p(x|z)p(y|z)} \quad (4.4)$$

La información condicionada se define como la información compartida por X e Y después de conocer el valor de la variable Z . Con todos estos conceptos, se han descrito brevemente los principios básicos de la teoría de la información necesarios para comprender el algoritmo MIFS.

Selección de características mediante la información mutua

El desarrollo de un clasificador a menudo se enfrenta con restricciones prácticas como el *hardware* disponible o el tiempo máximo permitido para la realización de la tarea. Existen muchas características que pueden ser extraídas de los datos a analizar y la información contenida en ellas debe ser suficiente para determinar su clase con poca ambigüedad. Pero el número de estas características puede ser una restricción para el posterior clasificador y debe de ser reducido.

Los algoritmos basados en la teoría de la información parten del siguiente problema de *Feature Selection* a solucionar: “*Dado un conjunto inicial de n características, encontrar*

un subconjunto de $k < n$ características que maximice la información sobre cada clase”. Si aplicamos la ecuación (??) y el hecho de que la incertidumbre para cada clase es fija, el anterior problema se puede reformular como: “Dado un conjunto inicial F con n características, encontrar los subconjuntos $S \subset F$ de k características que minimicen $H(F|S)$, es decir, que maximicen la información mutua $I(F; S)$ ”.

Desafortunadamente, en la práctica, la solución al problema propuesto para un clasificador complejo que requiera un gran número de características está limitada por problemas computacionales. Por ejemplo, la cantidad de operaciones y tiempo requeridos por la CPU para calcular la información mutua se vuelve excesiva cuando las dimensiones del vector de características \mathbf{F} son muy elevadas [?]. Por lo tanto, para utilizar estos conceptos en la práctica, es imprescindible considerar soluciones aproximadas al problema.

Algoritmo MIFS

El algoritmo MIFS implementa una solución aproximada al problema descrito para resolver las limitaciones de cálculo mencionadas anteriormente. Considerando que la información mutua entre vectores variables se aproxima usando la información mutua entre las componentes individuales de dichos vectores. De esta forma, en lugar de computar la información mutua $I(F; C)$ entre un vector de características \mathbf{F} y la variable clase c , se calculan solo $I(f; C)$ e $I(f; f')$ donde f y f' son características individuales.

Por lo tanto, el cálculo teórico de la información mutua exacta es sustituido por una serie de cálculos factibles. De esta forma, el análisis de todos los posibles subconjuntos es remplazado por el algoritmo MIFS precisando de un menor coste computacional. Dado un conjunto de características ya seleccionadas, este algoritmo elige la siguiente característica como la que maximiza la información de la clase correcta mediante la resta de una cantidad proporcional a la media de información mutua con la característica seleccionada. Por tanto, para ser seleccionada, una característica debe proporcionar información sobre la clase que no es predicha por las características ya seleccionadas. Por ejemplo, si dos características f y f' son altamente dependientes, $I(f, f')$ es grande, y la mejor de ellas ha sido seleccionada para el subconjunto final, la selección de la otra será poco probable.

El algoritmo MIFS es descrito por un procedimiento que consta de las siguientes etapas:

1. **Iniciación.** Consiste en establecer un conjunto F con las n características iniciales, y un conjunto S vacío por el momento.
2. **Cálculo de la información mutua con la clase de salida.** En la que para cada característica $f \subset F$ de calcula $I(C, f)$.

3. **Selección de la primera característica.** En este paso se busca la característica f que maximiza $I(C; f)$. La característica cumple el criterio se elimina del conjunto F y se coloca en S .
4. **Selección del resto de características.** Consiste en repetir el siguiente proceso hasta que el conjunto S posea el número de parámetros k que se definió previamente.
 - a) Cálculo de la información mutua entre variables. Donde para todas las parejas de variables (f, s) con $f \in F$, $s \in S$, se calcula $I(f; s)$.
 - b) Selección de la siguiente característica. Se elige la característica f como la que maximice $I(C; f) - \beta \sum_{s \in S} I(f; s)$. Esta característica se incluye en el conjunto S y se suprime del F .
5. **Fin.** Finalmente, se muestra el conjunto S que contiene los parámetros seleccionados.

El parámetro β regula la importancia relativa de la información mutua de la característica candidata con las características ya seleccionadas con respecto a la información mutua con la clase de salida. Si β es cero, solo la información mutua con la clase de salida es considerada para la selección. El aumento de β implica el incremento de la importancia de la otra información mutua.

Implementación en el TFM

Para la implementación en esta investigación del algoritmo MIFS se ha hecho uso del *toolbox* FEAST [?] de MATLAB. Los parámetros necesarios para su definición, como se ha podido observar, son β y en número de parámetros k a seleccionar.

En la práctica, se ha verificado que un valor de β entre 0.5 y 1 es apropiada para la mayoría de las tareas de clasificación. El valor seleccionado en el presente trabajo para realizar este algoritmo es $\beta = 1$ ya que, según el creador del algoritmo, es el más adecuado en la mayoría de los casos [?].

Por otro lado, el número k de características a seleccionar será un factor crítico a la hora de la implementación de los clasificadores posteriores. Un alto número de las mismas podría acarrear problemas como el sobre-entrenamiento de dichos clasificadores. Teniendo en cuenta el reducido número de individuos del banco de datos a estudiar y mediante la optimización heurística del parámetro, se ha decidido que un $k = 10$ será adecuado para este algoritmo MIFS.

4.1.2. DISR

La Relevancia Simétrica de Entrada Doble (*Double Input Symmetrical Relevance*, DISR) [?] es otro algoritmo basado en la teoría de la información para llevar a cabo la selección de características. Este método puede ser usado para elegir entre un número finito de subconjuntos alternativos el que maximice la cantidad de información con respecto a la clase de salida. A continuación se procede a la descripción de las bases teóricas de este algoritmo y de su implementación.

Relevancia y complementariedad

Uno de los conceptos más importantes en *feature selection* es el de eliminar la redundancia de las variables iniciales. Esta noción de relevancia y redundancia puede ser realizada en términos más formales mediante la teoría de la información. En este sentido, a partir de los conceptos de entropía condicional e información mutua condicional, definidos anteriormente en (??) y (??), se está a disposición de introducir la siguiente definición de relevancia propuesta en [?]: “*Considerando tres variables X , Y y Z y su distribución de probabilidad conjunta $p_{X,Y,Z}(x, y, z)$. Si $H(Y|Z) = 0$, entonces la relevancia de X para Y dada Z y denotada por $r(X; Y|Z)$, es cero. En cambio, si $H(Y|Z) \neq 0$, entonces la variable de relevancia de X para Y dada Z es definida como:*”

$$r(X; Y|Z) = \frac{I(X; Y|Z)}{H(Y|Z)} \quad (4.5)$$

De acuerdo a esta definición, el valor numérico de la relevancia se encuentra en el intervalo $[0,1]$ y mide la reducción relativa de incertidumbre de Y proporcionada por X cuando se conoce el valor de Z .

Otro concepto importante a la hora de definir el algoritmo DISR es la complementariedad entre variables. En relación a esto se introduce la medida de complementariedad entre dos variables X_i y X_j con respecto a una salida Y como:

$$C_Y(X_i, X_j) = I(X_{i,j}; Y) - I(X_i; Y) - I(X_j; Y) \quad (4.6)$$

donde $X_{i,j} = \{X_i, X_j\}$.

Se dice que dos variables son complementarias si su medida de complementariedad con respecto a Y es positiva. Es importante advertir que un valor negativo es sinónimo de redundancia del par de variables en la tarea de predecir Y .

Algoritmo DISR

El algoritmo DISR se basa en los conceptos de relevancia y complementariedad para realizar la selección de características. El valor de separabilidad de una característica X_k

viene dada por:

$$J_{disr}(X_j) = \sum_{X_j \in S} \frac{I(X_k X_j; Y)}{H(X_k X_j; Y)} \quad (4.7)$$

Con esta medida ya se está a disposición de realizar el algoritmo DISR que poseerá las siguientes etapas:

1. **Iniciación.** Establecer el conjunto F con las n características iniciales, y el conjunto S vacío por el momento.
2. **Cálculo del valor de separabilidad para las características de F .** En la que para cada característica $f \in F$ se calcula $J_{disr}(f)$.
3. **Selección de la primera característica.** En este paso se busca la característica f que maximiza $J_{disr}(f)$. La característica cumple el criterio se elimina del conjunto F y se coloca en S .
4. **Selección del resto de características.** Consiste en repetir el siguiente proceso hasta que el conjunto S posea el número de parámetros k que se definió previamente.
 - a) Cálculo de la separabilidad de las variables. Donde para todas las variables se calcula $J_{disr}(f)$.
 - b) Selección de la siguiente característica. Se elige la característica f como la que maximice $J_{disr}(f)$. Esta característica se incluye en el conjunto S y se suprime del F .
5. **Fin.** Finalmente, se muestra el conjunto S que contiene los parámetros seleccionados.

Implementación en el TFM

Al igual que para el MÉTODO MIFS, Para la implementación en esta investigación del algoritmo DISR se ha hecho uso del *toolbox* FEAST [?] de MATLAB. Además, como se ha podido observar, el único parámetro necesario para su definición es el número de parámetros k a seleccionar.

En la práctica, el número k de características a seleccionar será un factor crítico a la hora de la implementación de los clasificadores posteriores. Un alto número de las mismas podría acarrear problemas como el sobre-entrenamiento de dichos clasificadores. Teniendo en cuenta el reducido número de individuos del banco de datos a estudiar y mediante la optimización heurística del parámetro, se ha decidido que un $k = 10$ será adecuado para el algoritmo DISR.

4.2. Criterio J5

El criterio J5 para la selección de características, a diferencia de los anteriores de los basados en la teoría de la información, se fundamenta en la asunción de que las diferentes clases ocupan regiones distintas aunque posiblemente solapadas del espacio de características [?]. De esta forma, la situación ideal será aquella en la que vectores de distintas clases se encuentren muy separados en el espacio, mientras que vectores de una misma clase estén muy próximos entre sí. A continuación, se procede a la descripción de las bases teóricas de este algoritmo y de su implementación.

Fundamento

Los fundamentos del algoritmo J5 sobre la proximidad y la lejanía se concretan mediante los conceptos de distancia interclase y distancias intraclase [?]. Se denotará ξ_{ik} al k -ésimo vector de características de la clase i -ésima; \mathbf{m}_i al promedio de los vectores de características del conjunto de entrenamiento pertenecientes a la clase i -ésima, y n_i al número de vectores de la clase i -ésima existentes en el conjunto de entrenamiento. Por lo tanto:

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \xi_{ik} \quad (4.8)$$

Por otro lado, si se denota la probabilidad *a priori* de que un determinado vector ξ pertenezca a la i -ésima clase por P_i , se llamará \mathbf{m} al promedio global de los vectores de características, definido como:

$$\mathbf{m} = \sum_{i=1}^C P_i \mathbf{m}_i \quad (4.9)$$

donde C es el número de clases. De esta forma, se puede dar una expresión cuantitativa para la distancia intraclase Δ_W e interclase Δ_b :

$$\begin{aligned} \Delta_W &= \sum_{i=1}^C \frac{P_i}{n_i} \sum_{k=1}^{n_i} \|\xi_{ik} - \mathbf{m}_i\|^2 \\ \Delta_b &= \sum_{i=1}^C P_i \|\mathbf{m}_i - \mathbf{m}\|^2 \end{aligned} \quad (4.10)$$

En la ecuación (4.10) se miden las distancias en términos de la distancia euclídea. Para la distancia intraclase se promedian las distancias de cada vector del conjunto de entrenamiento al centroide de la clase a la que pertenece éste. Evidentemente, cuanto más compactadas estén las clases menor será esta distancia y más sencilla la clasificación.

Por el contrario, la distancia interclase mide lo dispersos que están los centroides de las diferentes clases. Cuanto mayor sea ésta, mayor será la separación y más sencilla la clasificación. Cabe destacar que, normalmente, no se conocerán las probabilidades *a priori* P_i y, por tanto, habrá que conformarse con los estimadores \tilde{P}_i definidos como:

$$\tilde{P}_i = \frac{n_i}{n}, \text{ con } n = \sum_{i=1}^C n_i \quad (4.11)$$

El objetivo a conseguir es la obtención de un criterio $J(\boldsymbol{\xi})$ que mida la separabilidad entre las diferentes clases, es decir, cómo de cerca están los vectores de una misma clase y cómo de lejos están las clases entre ellas. Para la obtención de este criterio se llevarán a cabo una serie de operaciones matemáticas que ya fueron definidas en [?] llegando a definición del criterio $J(\boldsymbol{\xi}_5)$ que cuantifica la separabilidad entre clases para un vector de características de una característica concreta.

Pero puede ocurrir que ciertas combinaciones lineales de características sean más discriminantes que cualquiera de las características de forma aislada. Por lo tanto, el problema final será encontrar una transformación del espacio de entrada, dada por una matriz W de tamaño $d' \times d$, con d el número inicial de características y d' el número deseado, que maximice la separabilidad, es decir encontrar [?]:

$$W_{\text{óptima}} = \max_W J_5(W\boldsymbol{\xi}) \quad (4.12)$$

Aunque esta maximización se puede llevar a cabo analíticamente a través del gradiente de la función criterio J_5 con respecto a la matriz W , hay dos razones para no hacer esto. En primer lugar, la gran complejidad de este proceso. En segundo lugar, que las posteriores transformaciones del espacio de características, junto con la capacidad de los clasificadores automáticos para adaptar sus pesos sinápticos a las posibles combinaciones lineales más discriminantes, harán que este proceso sea menos crítico.

Por tanto, el objetivo final será buscar, dentro del conjunto global de características, aquel subconjunto que contenga las más discriminantes. Este proceso se puede entender como una particularización del caso general en el que se restringe la matriz W a una matriz binaria que contenga como mucho un uno por cada fila y columna.

Suponiendo que W es una matriz $d' \times d$ cuyas columnas son los d vectores de tamaño d' que forman el sistema $\{\mathbf{w}_i\}_{i=1}^d$, y denotando la k -ésima componente del vector \mathbf{w}_i por $w_{i,k}$. Entonces el vector \mathbf{w}_i será nulo si la i -ésima característica no está entre las d' más discriminantes, y si lo está tendrá todas sus componentes nulas salvo la componente $w_{i,l}$, con l el orden de la característica en el nuevo espacio.

Una vez restringido el espacio de las posibles matrices de transformación, la optimización de la separabilidad en función de éstas es más sencilla. En lugar de emplear

un método de gradiente, las restricciones sobre W hacen posible probar todas las matrices que se pueden considerar, calcular el valor del criterio de separabilidad y elegir aquella que lo maximice.

Algoritmo J5

El principal problema es que *a priori* no se conoce el número final de características que se emplearán. Si se supone que el número inicial del que se dispone es d , habrá que considerar todos los posibles subconjuntos contenidos en el inicial, es decir, habrá que calcular el criterio de separabilidad para:

$$N = \sum_{i=1}^d \binom{d}{i} = \sum_{i=0}^d \binom{d}{i} \cdot 1^i \cdot 1^{d-i} - \binom{d}{0} = (1+1)^d - 1 = 2^d - 1 \quad (4.13)$$

subconjuntos, por lo que con un gran número de características el problema será intratable. Es evidente que si no fuese por la enorme carga computacional este planteamiento llevaría a la solución óptima, ya que se lleva a cabo una búsqueda exhaustiva de todas las posibles combinaciones. En lugar de ello, hay que conformarse con una solución subóptima utilizando el siguiente algoritmo:

1. Primero se busca la característica más discriminante, lo que requerirá un total de d cálculos del criterio de separabilidad. No obstante, en este caso las matrices S_W y S_b tendrán tamaño 1×1 , con lo que su inversión será trivial y el cálculo muy rápido.
2. Una vez encontrada la mejor característica, quedan $d - 1$. Entre ellas, se busca una segunda tal que el conjunto formado con la primera característica sea el que muestre una mayor separabilidad de entre todos los conjuntos de dos características que incluyen la primera. Esto requiere manejar $d - 1$ matrices de tamaño 2×2 .
3. El mismo proceso se repite hasta agotar todas las características. En cada paso se tendrá un conjunto de d' características al que hay que añadir una nueva que sea la que maximice el criterio. Esto requiere $d - d'$ cálculos, cada uno de los cuales involucra la inversión de una matriz de tamaño $(d' + 1) \times (d' + 1)$; las matrices son cada vez mayores, pero también son cada vez menos numerosas.

En este caso el proceso no tiene por qué conducir a la solución óptima, pero sí a una solución subóptima que en principio será válida para su clasificación en los posteriores clasificadores automáticos. Finalmente, para incluir las características en un clasificador automático, se puede decidir seleccionar las k primeras características más separables con el fin de lograr un número óptimo para el funcionamiento de estas máquinas.

Implementación en el TFM

Para utilizar el criterio de selección J5 se ha dispuesto de un conjunto de *scripts* de MATLAB creados en el trabajo [?]. Estos *scripts* generan de forma automática un vector de todas las características introducidas a la entrada en orden descendente según su grado de discriminación. Es decir, la primera característica del vector será la característica más discriminante y la última la que peor discrimina ambas clases.

Una vez calculado el vector, el último paso será la definición del número k de características a seleccionar que será un factor crítico a la hora de la implementación de los clasificadores posteriores. Como ya se mencionó para los anteriores algoritmos, un alto número de las mismas podría acarrear problemas como el sobre-entrenamiento de dichos clasificadores. Por lo tanto, teniendo en cuenta el reducido número de individuos del banco de datos a estudiar y mediante la optimización heurística del parámetro, se ha decidido que un $k = 10$ será adecuado para este algoritmo J5.

4.3. Algoritmos genéticos

Los Algoritmos Genéticos (AGs) son métodos adaptativos de búsqueda de soluciones (estimación de parámetros) en problemas no lineales con espacio de soluciones multimodal. Los AG pertenecen a los denominados a los algoritmos de computación evolutiva y están basados en conceptos biológicos derivados de la teoría Darwiniana de la evolución natural [?]. De esta forma, para lograr obtener la solución de este tipo de problemas, se parte de un conjunto inicial de individuos (posibles soluciones) denominado población. Estos individuos evolucionarán, tomando como base los esquemas propuestos por Darwin sobre la selección natural, y se adaptarán en mayor medida tras el paso de cada generación a la solución requerida. A continuación, se procede a la descripción de las bases teóricas de este tipo de algoritmos y de su esquema general de implementación.

Introducción

Los orígenes de los AGs se remontan a finales de la década de los 60 en la que John Holland desarrolló una técnica que imitaba el funcionamiento de la selección natural [?]. En general, se definió que un AG consistía en una población de soluciones codificadas de forma similar a los cromosomas biológicos. A su vez, cada uno de estos cromosomas poseía un valor de aptitud o *fitness* que cuantificaba su validez como solución al problema que se pretendía resolver. Además, estos valores de aptitud eran claves a la hora de aplicar las diferentes operaciones definidas en la teoría de la evolución de Darwin como el cruce o la mutación de los individuos.

En la actualidad, los Algoritmos Genéticos son ampliamente utilizados para la solución de diversos problemas de búsqueda y optimización de parámetros. En concreto, los AGs son una herramienta común para la realización de la selección de características en las aplicaciones biomédicas de reconocimiento de patrones como la que acontece en este trabajo, logrando mejorar los resultados proporcionados por otras técnicas [?, ?, ?, ?].

Esquema general de implementación de un AG

En general, un AG sigue el diagrama de flujo expuesto en la Figura ?? [?]. El algoritmo genético comienza seleccionando de alguna forma, normalmente aleatoriamente, los cromosomas o posibles soluciones iniciales que constituirán la denominada población inicial. En segundo lugar, se calcula la aptitud de cada sujeto de la población inicial, es decir, lo válido que es dicho cromosoma como solución del problema a resolver. Posteriormente, se seleccionarán los cromosomas que formarán la siguiente generación por diversos operadores genéticos como son el cruce, la mutación o el elitismo. El número de cromosomas generado por cada operador genético es un parámetro de gran importancia a definir por cada AG concreto. Una vez creada la nueva población o generación de individuos, se calcula la aptitud de los mismos. Finalmente, si se cumple el criterio de parada establecido por el algoritmo, se escoge el cromosoma más apto de la población y éste será la solución final del problema. En caso de no cumplirse el criterio de parada, se vuelve a realizar otra nueva selección de cromosomas con la nueva generación y se repite el proceso descrito desde dicha etapa.

Parámetros típicos de un AG

En base a la teoría Darwiniana, y como puede observarse en la Figura ??, los AGs constan de una gran cantidad de factores a tener en cuenta a la hora de su implementación. Por lo tanto, la construcción de un AG conlleva una elevada cantidad de parámetros que deben de optimizarse según el problema que se pretenda resolver. Los factores más importantes a la hora de construir un AG suelen ser [?]:

1. **La codificación de los cromosomas.** Cualquier solución potencial a un problema puede ser descrita mediante una serie de valores o genes. El conjunto de todos estos valores se codifica en una cadena de parámetros dando lugar a un cromosoma. Los primeros estudios solían codificar estos parámetros como cadenas binarias asignando un número de bits concreto para cada parámetro, pero también existen representaciones que codifican cada parámetro con un valor entero o real. Cabe destacar que la codificación binaria hace la creación de los operadores genéticos posteriores más cercana a la teoría Darwiniana.

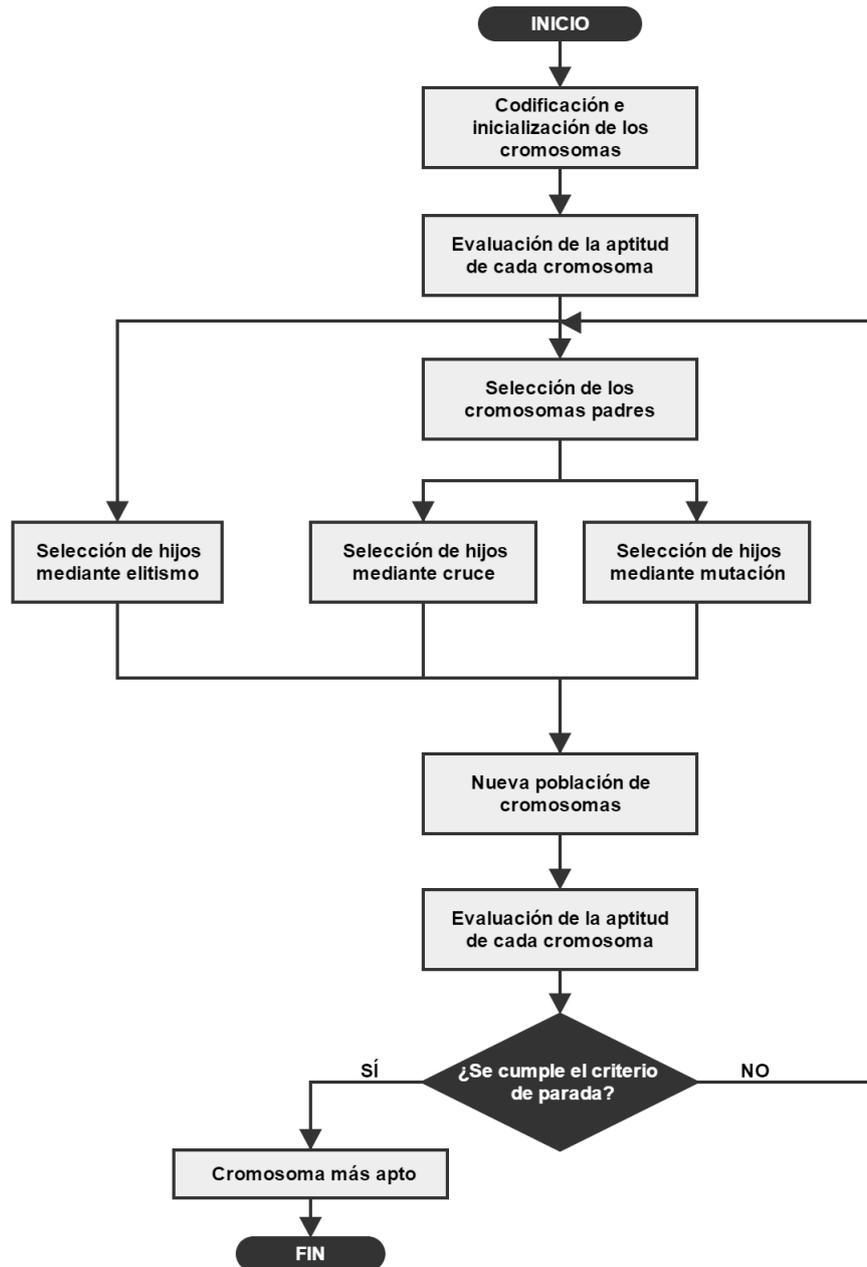


Figura 4.2: Esquema básico de un algoritmo genético.

2. **El cálculo de la aptitud de los cromosomas.** Para el correcto funcionamiento del AG es de vital importancia poseer una métrica que determine la calidad de las distintas soluciones o cromosomas a la hora de resolver el problema propuesto. De esta tarea se encarga la función de aptitud que establece una medida numérica de la bondad de cada solución o cromosoma. Esta función deberá de ser definida según el problema que se pretenda resolver de forma que calcule de alguna manera la distancia entre el cromosoma y la solución óptima a dicho problema.

3. **Los mecanismos de selección.** Los algoritmos de selección son los encargados de escoger los individuos que van a reproducirse dentro de una población. Teniendo en cuenta la teoría Darwiniana, los algoritmos de selección han de otorgar mayor probabilidad de reproducción a los individuos más aptos pero sin eliminar por completo las opciones de reproducción de los otros para que la población no se vuelva homogénea. Existen una gran cantidad de métodos para realizar este procedimiento, los más comunes son la selección por ruleta o la selección por torneo pero existen muchos otros como la selección por jerarquías, el muestreo determinístico o la brecha generacional.
4. **Los operadores genéticos.** Estos operadores definen los distintos cambios que se producen al pasar de una generación de cromosomas a otra. Los operadores genéticos más comunes son:
 - a) *El cruce.* Este operador parte de los individuos elegidos mediante el mecanismo de selección para ser recombinados simulando la reproducción sexual biológica. La idea principal de la implementación de este operador se basa en que, si se toman dos individuos correctamente adaptados al medio y se obtiene una descendencia que comparta genes de ambos, cabe la posibilidad de que los genes heredados sean precisamente los causantes de la bondad de los padres y se mejore la misma. Existen una gran cantidad de operadores de cruce como el cruce de 1 punto, el cruce de 2 puntos o el cruce uniforme. Estos operadores serán definidos dependiendo de la codificación de los cromosomas en el AG y del problema a resolver.
 - b) *La mutación.* Este operador provoca que algunos de los genes de un cromosoma varíen de alguna forma determinada. Al igual que para el cruce, es común que los cromosomas a mutar sean elegidos por el mecanismo de selección implementado. Existen también diferentes operadores de mutación aunque la más habitual es el reemplazo aleatorio de algunos de los genes.
 - c) *El elitismo.* Este operador consiste en realizar la copia del mejor, o en su caso los mejores, individuos de cada generación en la generación siguiente. De esta manera se garantiza que el proceso de búsqueda nunca empeorará la calidad de la mejor solución obtenida, sino que un cambio en ésta siempre implicará una mejora.
5. **El criterio de parada.** Este criterio define la condición que ha de producirse en el AG para que finalice la búsqueda del mejor cromosoma. Existen gran cantidad de criterios de parada dependiendo de distintas características de la aplicación del

AG como son la carga computacional o el tiempo de procesamiento. Algunas de las condiciones de parada más comunes son la realización de un número máximo de generaciones o la consecución de un individuo con una aptitud lo suficientemente próxima a un valor fijado.

Como puede observarse, existe un gran abanico de posibilidades a la hora de la creación de un AG. La selección del esquema final de implementación, así como de los diferentes métodos utilizados en cada etapa, serán críticos a la hora de obtener la solución más óptima del problema a resolver. Por lo tanto, es necesaria una profunda optimización de todos estos factores para lograr un AG que se adapte en la mayor medida de lo posible a la aplicación concreta.

Implementación en el TFM

Para la implementación del AG en el presente estudio se dispuso del *toolbox* GLOBALOPTIM (*Global Optimization Toolbox*) [?] del entorno de programación MATLAB. Este *toolbox* permite configurar todos los parámetros de un AG básico por lo que permite adaptar el mismo a la aplicación de reconocimiento de patrones que se pretende realizar.

Como se advirtió anteriormente, la implementación del AG precisa de la definición de una gran cantidad de parámetros que deben de adaptarse a las necesidades del problema que se está llevando a cabo. Partiendo del esquema general que se describió en la Figura ??, y basándose en [?], se han configurado las operaciones de dicho esquema de la siguiente forma:

1. *Codificación de los cromosomas.* Los diferentes individuos del AG implementado serán codificados como cadenas binarias de tamaño igual al número de características total de cada agrupación introducida. De esta forma, un 1 en un gen del cromosoma quiere decir que la característica situada en la posición de dicho gen es seleccionada, y un 0 lo contrario.
2. *Inicialización de los cromosomas.* Se llevará a cabo de manera totalmente aleatoria. Se generará un número de individuos para cada generación igual a cinco veces el tamaño del cromosoma.
3. *Función de aptitud.* Lo ideal sería que la aptitud a ser minimizada fuera la tasa de error en la clasificación de los clasificadores posteriores del método de reconocimiento de patrones. Pero debido a la complejidad y al alto coste computacional de los clasificadores implementados en este trabajo, esta medida es poco factible computacionalmente. De todas maneras, existen otros clasificadores

más sencillos, y al igual que en [?], se ha decidido tomar como función de aptitud la que retorna el error en la clasificación (α) mediante un método de K-vecinos más cercanos (*K nearest neighbors*, KNN) con $K = 3$. De esta forma, se espera que las tasas de acierto mediante clasificadores más complejos sean proporcionales a ésta aunque posiblemente más altas. Por último, cabe destacar que el valor final de aptitud es una corrección de este error en la clasificación con el fin de volver más aptos a los cromosomas con un menor número de características seleccionadas ya que un alto número de ellas repercute en diversos problemas para los clasificadores posteriores como puede ser el sobre-entrenamiento. Esta corrección será distinta según sea la longitud de los cromosomas, es decir, el número de características (N) de la agrupación inicial. De esta forma²:

- a) Si $N \leq 100$, la función de aptitud a minimizar será la tasa de error en la clasificación dividida entre el número de características no seleccionadas. De esta forma, un menor número de características seleccionadas proporcionará al individuo menor aptitud y, por lo tanto, como se pretende minimizar el valor de la misma, será más apto. De esta forma:

$$\text{aptitud} = \frac{\alpha}{N - k} \quad (4.14)$$

- b) Si $N > 100$, el valor de la aptitud será el mismo que para el otro caso pero con una corrección extra basada en una función sigmoideal para minimizar la aptitud de los cromosomas con un menor número de características. De esta forma:

$$\text{aptitud} = \frac{\alpha}{N - k} + \frac{1}{(N - k)(1 + \exp(k - 70))} \quad (4.15)$$

4. *Mecanismo de selección.* Se utilizará la selección por torneo de tamaño 2 debido a la simplicidad, rapidez y eficiencia del mismo. Este mecanismo de selección toma dos individuos al azar y selecciona el más apto como padre de la nueva generación. Este proceso será repetido hasta que se hayan seleccionado todos los padres necesarios para llenar la siguiente generación.

5. *Operadores genéticos.* Se llevará a cabo:

- a) Un elitismo de tamaño 2. Es decir, los dos cromosomas más aptos de cada generación serán copiados directamente en la siguiente.
- b) Un cruce aleatorio. Este cruce consistirá en la realización de un vector aleatorio binario del mismo tamaño que los cromosomas del AG (cromosoma máscara).

²Denominando el número de características seleccionadas por cada cromosoma, es decir, el número de unos del mismo como a k .

Si el gen del cromosoma máscara es un 0, el hijo poseerá el valor del primer padre en dicho gen, y si es un 1 del segundo. Cabe destacar que se tomará una fracción de 0.8 de los cromosomas seleccionados para realizar el cruce de los mismos.

- c) Una mutación uniforme para los cromosomas que no han sido cruzados ni copiados mediante elitismo y que consta de dos pasos. El primer paso consiste en la generación de un cromosoma máscara aleatorio con valores reales entre 0 y 1. Esta máscara se traducirá a un vector binario de tal forma que si el gen es menor que la probabilidad de mutación (p_m) será un 1, y si es mayor un 0. El segundo paso será la comparación del cromosoma máscara con el cromosoma a mutar. Si el cromosoma máscara posee un 1 en un gen, el gen en la misma posición del cromosoma a mutar será invertido, es decir, si era un 0 pasa a ser un 1 y viceversa. Cabe destacar que la p_m será distinta según el tamaño de los cromosomas con el fin de que el número de genes mutados sea similar. De esta forma:

- 1) Si $N \leq 100$, $p_m = 0,1$.
- 2) Si $N > 100$, $p_m = 0,01$.

6. *Criterio de parada.* Existirán dos criterios de parada para finalizar el algoritmo genético. El primero consiste en el paso de un número de generaciones igual al doble de la longitud de los cromosomas. El segundo detiene el AG si el cambio en la media de la función de aptitud de todos los cromosomas es menor que 0,000001 una vez pasado un número de generaciones igual a la longitud de los cromosomas.

Pero el esquema del AG general no se llevará a cabo de forma exacta. Este esquema básico será repetido 10 veces de forma que se seleccionarán 10 individuos óptimos (uno por repetición). Estos 10 individuos serán clasificados mediante un perceptrón multicapa³ (MLP) con 10 nodos intermedios que pretende generar una frontera de decisión más completa que el clasificador utilizado en la función de aptitud del AG básico. De esta forma, se seleccionará el cromosoma que mayor área bajo la curva ROC⁴ obtenga que será la solución final del AG implementado. Con esta última etapa se pretende incluir una clasificación con una frontera más compleja para la selección de los cromosomas y evitar algunas ejecuciones en las que el AG pueda converger a algún mínimo local. Por lo tanto, el AG implementado en este trabajo seguirá el diagrama de flujo expuesto en la Figura ??.

³Cuyo fundamento será descrito en el siguiente Capítulo.

⁴Medida que se explicará en el siguiente Capítulo.

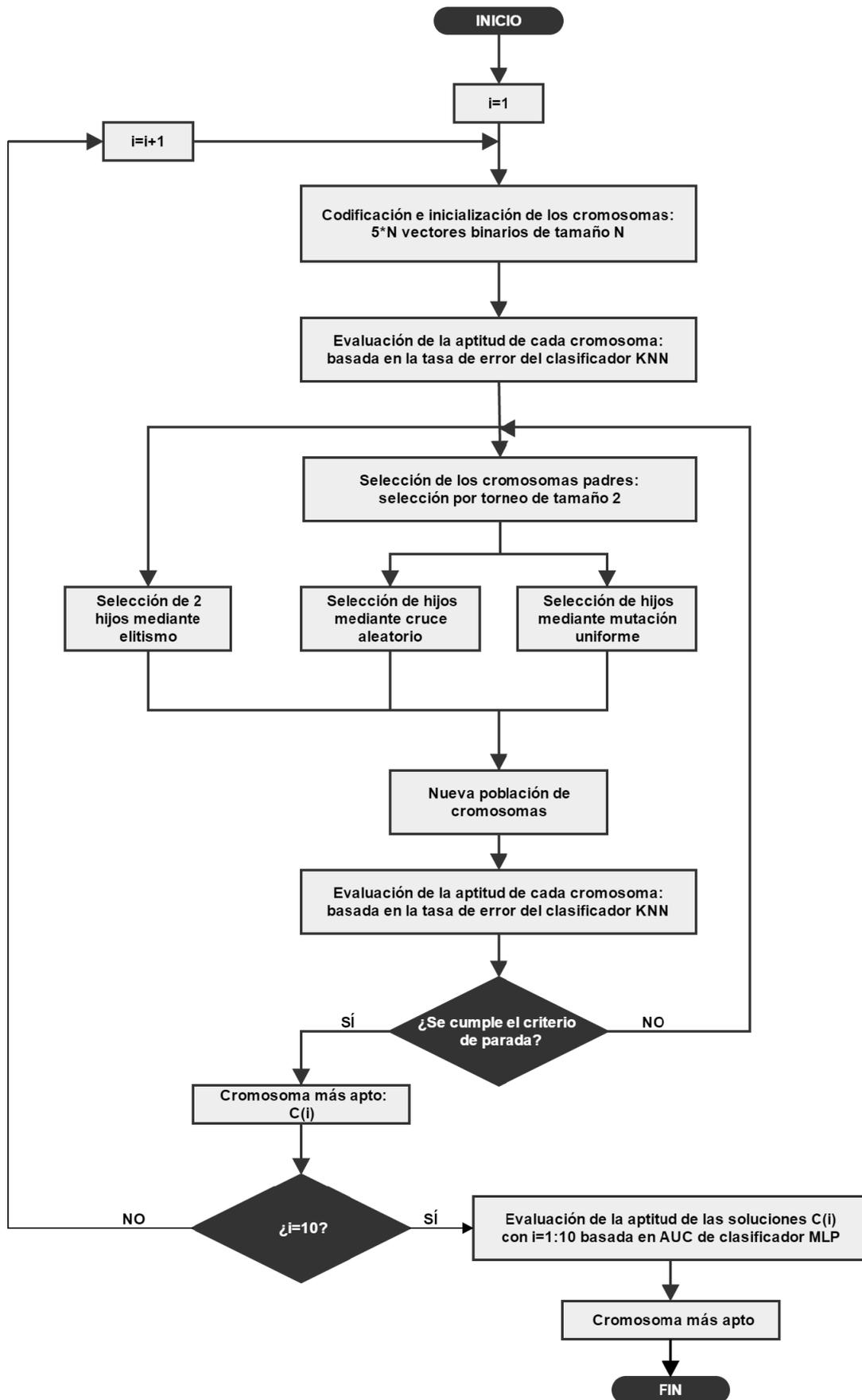


Figura 4.3: Esquema del algoritmo genético implementado.

Capítulo 5

Análisis estadístico y clasificación de las señales de EEG

En este epígrafe se afrontará la problemática sobre la diferenciación de los diferentes individuos entre esquizofrénicos y sanos. Para ello se relatarán dos tipos de análisis a realizar en los datos extraídos de las señales de EEG con el fin de cuantificar el grado en que es posible separar ambas clases de sujetos y que se incluirán en la etapa 5 del método de reconocimiento de patrones general descrito en el Apartado ???. Por un lado, se implementará un análisis estadístico de todas las características extraídas de las señales de EEG con el fin de cuantificar el grado en que cada una de ellas es capaz de diferenciar los sujetos. Finalmente, se implementarán algunas de las máquinas de clasificación automática más importantes en la actualidad como son las Máquinas de Vectores Soporte y las Redes Neuronales. Mediante estos clasificadores se pretende obtener una medida de las tasas de aciertos en la clasificación de las distintas agrupaciones de características, tomando como entradas las características seleccionadas en la etapa de selección de características. Estas dos medidas de separabilidad entre clases pretenden ser comparadas en el capítulo relativo a los resultados del presente estudio con el fin de verificar si realmente existen coincidencias entre ellas.

5.1. Análisis estadístico

La estadística es una disciplina científica que proporciona una gran cantidad de herramientas necesarias para el análisis de datos [?]. Dado que, frecuentemente, no es posible realizar estudios contando con una población completa, la estadística permite cuantificar el error que se comete al aceptar los resultados proporcionados por solo una muestra de la misma.

El análisis estadístico mediante el cual se pretende cuantificar el grado de diferenciación entre las dos poblaciones de sujetos en este estudio es el denominado

método de contraste de hipótesis o test de hipótesis [?]. Los contrastes de hipótesis pertenecen a una rama de la estadística, denominada estadística inferencial, que pretende hacer deducciones con respecto a una o varias poblaciones basándose en la información contenida solamente en muestras de las mismas. A continuación se presentan los fundamentos estadísticos de los contrastes de hipótesis y la descripción de algunos de los algoritmos concretos más comunes para su implementación en este trabajo.

Hipótesis nula e hipótesis alternativa

Un método de contraste de hipótesis consiste en estudiar dos hipótesis: una hipótesis nula (H_0) y una hipótesis alternativa (H_1). En primer lugar, se parte de la aceptación de la hipótesis nula sobre una población y, posteriormente, se pretende demostrar la veracidad de la hipótesis alternativa, H_1 , demostrando que la hipótesis nula debe de ser rechazada en base a los datos de una muestra de la población. En base a esta definición, un test de hipótesis no consiste en decidir cuál de las dos hipótesis es más creíble, sino en decidir si la muestra de población proporciona o no suficiente evidencia para descartar H_0 . De esta forma:

- Se acepta H_0 si los resultados proporcionados por la muestra no contradicen las suposiciones de H_0 .
- Se rechaza H_0 si los resultados proporcionados por la muestra son “poco probables” bajo la suposición de H_0 .

Teniendo en cuenta todo lo anterior, es inmediato deducir que existen dos decisiones posibles a realizar en un contraste de hipótesis. Estas decisiones son:

- Aceptar H_0 y, por tanto, rechazar H_1 .
- Rechazar H_0 y, por tanto, aceptar H_1 .

Tipos de error

Como ya se mencionó, la estadística inferencial, y más en concreto los contrastes de hipótesis, toman las decisiones sobre una población en función de los datos de una muestra de la misma. Por lo tanto, al tomar una decisión sobre solo una muestra, pueden cometerse los siguientes errores (mostrados gráficamente en la Tabla ??):

- **Errores de Tipo I.** Producidos cuando se rechaza H_0 y realmente es cierta.
- **Errores de Tipo II.** Producidos cuando se acepta H_0 y realmente es falsa.

	H_0 es cierta	H_1 es cierta
Se acepta H_0	Sin error	Error de Tipo II
Se acepta H_1	Error de Tipo I	Sin error

Tabla 5.1: Tipos de errores cometidos en el contraste de hipótesis.

Obviamente, la situación ideal se produciría cuando ambos errores fueran nulos o, por lo menos, muy pequeños. Sin embargo, no se pueden disminuir ambos de forma simultánea así que se suele recurrir a la reducción de los errores de Tipo I que se consideran los más importantes. Para ello, se suele fijar un nivel de significación (α) que representa la probabilidad máxima de error de Tipo I permitida. Este valor deberá de ser muy pequeño para reducir al máximo este tipo de errores. Los valores más típicos para α suelen tomarse entre 0,1 y 0,01 siendo 0,05 el más común en la mayoría de los casos.

Etapas para la construcción de un test de contraste

Teniendo en cuenta todas las definiciones anteriores, para la creación de un contraste estadístico hay que seguir los siguientes pasos:

1. Identificar el parámetro de interés de la población que se pretende validar.
2. Establecer las hipótesis H_0 y H_1 .
3. Fijar el nivel de significación α .
4. Determinar el estadístico del contraste que cuantifica la probabilidad de que se cumpla la hipótesis nula.
5. Calcular el valor que toma el estadístico de contraste para la muestra seleccionada.
6. Decidir si debe rechazarse H_0 e interpretar la decisión tomada.

Test de hipótesis

Existen múltiples tests estadísticos en función de la distribución de los datos de las poblaciones analizadas. De esta forma existen: los tests paramétricos, que realizan una serie de suposiciones sobre la distribución de los datos de las poblaciones; y los tests no paramétricos, que no realizan ninguna suposición sobre dichas distribuciones.

El test estadístico más utilizado en aplicaciones biomédicas, como la acontecida en este trabajo, es el análisis de la varianza (*Analysis of variance*, ANOVA). ANOVA es una colección de modelos estadísticos paramétricos que permiten inferir si un conjunto de datos numérico es significativamente diferente a otro. Este análisis retorna un p-valor (estadístico del contraste) que si posee un valor pequeño (típicamente por debajo de

0,05) determina que las diferencias entre las poblaciones son importantes frente a las diferencias entre los sujetos de una misma población y, por tanto, los grupos presentan diferencias significativas entre ellos. El problema que presenta ANOVA es que supone que las distribuciones de ambas poblaciones son normales y poseen varianzas similares (homocedásticas).

Pero si las hipótesis necesarias para la aplicación del ANOVA no se cumplieran habría que recurrir a algún test no paramétrico para estudiar las diferencias significativas entre las poblaciones. En aplicaciones biomédicas se suele recurrir al test U de *Mann-Whitney* que parte de la hipótesis nula de que ambas distribuciones poseen medianas similares. Además, de manera similar a ANOVA, la existencia de diferencias significativas, es decir, de que no se cumpla la hipótesis nula, se cuantifica en términos de p-valores.

Por otro lado, cabe destacar que existen otros tests de hipótesis que permiten verificar las hipótesis de gaussianidad y homocedasticidad con el fin de determinar la posibilidad de uso de tests paramétricos como el ANOVA. En este sentido, el test de *Lilliefors* [?] es comúnmente utilizado para la verificación de la hipótesis de normalidad de los datos de cada población. Del mismo modo, el test de *Ansari-Bradley* [?] es ampliamente utilizado para verificar la homocedasticidad entre dos poblaciones.

Por último, cabe destacar que existen diferentes métodos de ajuste en los p-valores cuando se realizan comparaciones múltiples como sucede en el presente estudio. Estas correcciones tienen como finalidad reducir los errores de Tipo I generados por la realización de varias comparaciones. Este tipo de corrección está ampliamente apoyada por los profesionales de la estadística, pero existen muchos epidemiólogos que indican que esta corrección debe de ser omitida en investigaciones biomédicas con el fin de no entorpecer la inferencia estadística producida inicialmente aumentando, de manera innecesaria, los errores de Tipo II [?, ?].

Por lo tanto, a la hora de realizar el análisis estadístico en el presente trabajo habrá que contemplar tres etapas. La primera etapa consistirá en la ejecución de diversos tests de hipótesis que permitan verificar las diferentes suposiciones precisadas para el uso de los tests paramétricos. En la segunda etapa, teniendo en cuenta las distribuciones estadísticas de las poblaciones a analizar, habrá que seleccionar el test un test de hipótesis válido para cuantificar la separabilidad de las poblaciones. Finalmente, tomando como hipótesis alternativa el hecho de que las poblaciones de sujetos (sanos y esquizofrénicos) posean un parámetro estadístico similar, como puede ser la media o la mediana, se podría inferir que ambas poblaciones no son diferenciables y viceversa.

Implementación en el TFM

Para realizar en análisis estadístico en el presente estudio, se hizo uso del *toolbox* STATS (*Statistics Toolbox*) [?] de MATLAB. Este *toolbox* contiene programados los diferentes tests estadísticos citados anteriormente por lo que su ejecución es inmediata.

En primer lugar, es importante definir los datos sobre los que se pretende realizar el análisis estadístico. Estos datos serán las 20 características extraídas en el Capítulo ?? para los 17 electrodos, es decir, las 340 características. De esta forma, se pretenderán realizar 340 tests estadísticos con el fin de cuantificar el grado de diferenciabilidad que aporta cada característica entre las dos poblaciones a estudiar (individuos sanos y esquizofrénicos).

El segundo paso a realizar en el análisis estadístico consistirá en la verificación de la hipótesis de normalidad de los datos mediante el test de *Lilliefors*. Este test se aplicará a la distribución de los datos de las características de cada población. De esta forma, se podrá verificar que la mayoría de estas poblaciones no cumplen la hipótesis de gaussianidad. Por lo tanto, los test no paramétricos serán la opción más recomendable para verificar la separabilidad entre las poblaciones.

Finalmente, se seleccionará el test U de *Mann-Whitney* como medida de separabilidad de las dos poblaciones cuantificada mediante los p-valores que retorna. De esta forma, un p-valor pequeño, por debajo de 0.05, determinará que las diferencias entre ambas poblaciones serán significativas, y viceversa.

5.2. Clasificación

La clasificación entre ambas poblaciones es una de las etapas del método básico de reconocimiento de patrones que se definió en la Figura ?. En este estudio, se pretenden utilizar dos clasificadores altamente no lineales y capaces de separar las poblaciones mediante fronteras de decisión muy complejas que han sido ampliamente utilizados en las investigaciones actuales proporcionando unas altas tasas de acierto. A continuación, se procede a la descripción de las bases teóricas relativas a los clasificadores y de su implementación en este trabajo.

5.2.1. Máquinas de Vectores Soporte

Las Máquinas de Soporte de Vectores (*Support Vector Machine*, SVM) son unos sistemas de aprendizaje que pertenecen a la rama de clasificadores lineales generalizados, aplicables tanto en problemas de clasificación como de regresión [?]. En el presente estudio se pretende usar este sistema para realizar la clasificación entre individuos sanos y esquizofrénicos por lo que a partir de ahora sólo se describirá esta utilidad. Las SVM

basan su estructura en una red estática de núcleos o *kernels* que operan sobre vectores de características que han sido transformados a un espacio de dimensión mayor a la del espacio de características original.

La mayor ventaja de este tipo de clasificadores es su buena capacidad de generalización, basada en la determinación del hiperplano que da lugar a la máxima distancia de separación entre los vectores transformados. Esta distancia de separación se obtiene mediante la construcción de dos hiperplanos paralelos al de separación óptima, localizados a ambos lados del mismo y que contengan al menos a uno de los vectores transformados, denominado vector soporte. Cuanto mayor sea esta distancia, mejor será la capacidad de generalización del clasificador. Este proceso puede observarse en la Figura ?? para una SVM y unos datos aleatorios en dos dimensiones. Además, cabe destacar que otra gran ventaja de las SVM frente a otros métodos de clasificación, como por ejemplo las redes neuronales, es su capacidad de minimizar el efecto del sobre-aprendizaje del conjunto de entrenamiento ya que en el entrenamiento sólo se aprenden los vectores soporte.

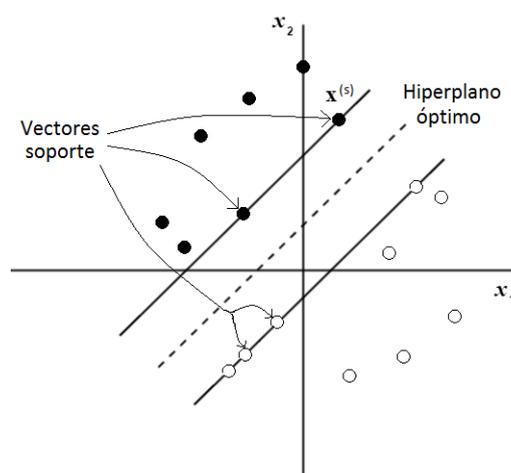


Figura 5.1: Representación del hiperplano óptimo junto con sus vectores soporte asociados para un SVM de margen máximo [?].

Las etapas empleadas por las SVM para realizar una clasificación de patrones son las siguientes:

1. Transformación de los datos o vectores de características de entrada a un espacio de mayor dimensión a través de una función Φ . El objetivo es formular el problema de modo que el procedimiento de clasificación transforme implícitamente los datos operando con funciones *kernel*, K .
2. Cálculo del hiperplano óptimo que maximiza la distancia entre las clases consideradas.

Clasificador de margen máximo

El modelo más simple de SVM es el denominado clasificador de margen máximo. El margen se define como la distancia mínima entre el hiperplano de separación de las clases y los datos de entrenamiento más cercanos. Desde un punto de vista intuitivo, se puede deducir que cuanto mayor sea el margen, mayor poder de generalización poseerá el clasificador implementado.

La clasificación de margen máximo sólo tiene validez para datos linealmente separables en el espacio de características y, por tanto, no puede ser usada en la mayor parte de los problemas reales. Sin embargo, es un modelo muy sencillo de entender y sirve como punto de partida para introducir la forma de procesamiento usada por las SVM. Cabe destacar que, aunque la explicación de la implementación de este modelo se detalla sólo para la clasificación de los datos en dos clases, las conclusiones serán extrapolables para el caso de múltiples clases, por ejemplo, generando tantos clasificadores como clases se pretendan distinguir. La Figura ?? ilustra este problema de clasificación de margen máximo.

El problema que se plantea es el siguiente [?]: dado un conjunto de datos o vectores de entrada $\mathbf{x} \in \mathbb{R}^n$ de las diferentes clases $\mathbf{y} \in \mathbb{N}$, se deseará encontrar el valor de los parámetros del hiperplano óptimo: $\{\mathbf{w}, b\}$, que da lugar al margen máximo. Esto es:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b \geq y_i \quad / \quad &= \begin{cases} y_i = 1 & \forall \mathbf{x}_i \in C_i = A \\ y_i = -1 & \forall \mathbf{x}_i \in C_i = B \end{cases} \\ \Rightarrow \exists(\mathbf{w}^*, b^*) \quad / \quad &\mathbf{w}^{T*} \mathbf{x} + b^* = 0 \quad \text{Hiperplano \textit{optimo}.} \end{aligned} \quad (5.1)$$

Paralelos al hiperplano óptimo, que maximiza el margen, se encuentran otros dos hiperplanos, H_1 y H_2 , que contienen al menos un vector de cada clase, denominados vectores soporte. La expresión matemática de los hiperplanos que contienen los vectores soporte es la siguiente:

$$H_1 \quad : \quad \mathbf{w}^T \mathbf{x} + b = 1 \quad (5.2)$$

$$H_2 \quad : \quad \mathbf{w}^T \mathbf{x} + b = -1 \quad (5.3)$$

Como el vector \mathbf{w} es perpendicular al hiperplano que define, y la distancia del hiperplano al origen del sistema de coordenadas viene dada por:

$$d = \frac{b}{\|\mathbf{w}\|} \quad (5.4)$$

donde $\|\mathbf{w}\|$ se refiere a la norma euclídea de \mathbf{w} . Utilizando las ecuaciones (??) y (??), se obtiene que la distancia entre H_1 y H_2 es:

$$d = \frac{2}{\|\mathbf{w}\|} \quad (5.5)$$

Por lo tanto, el objetivo de maximizar la distancia de separación entre ambos hiperplanos es equivalente a minimizar $\|\mathbf{w}\|$. Además, la solución propuesta debe evitar que haya vectores de entrada dentro de la región entre los hiperplanos H_1 y H_2 , es decir:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad 1 \leq i \leq N \quad (5.6)$$

donde N es el número de datos o vectores de entrada proporcionados para el entrenamiento de la SVM. Todas estas restricciones dan lugar al siguiente problema de optimización:

$$\text{minimizar: } \quad \Omega(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\| \quad (5.7)$$

$$\text{sujeto a: } \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad 1 \leq i \leq N \quad (5.8)$$

Para resolver el problema anterior se utiliza el método de los multiplicadores de *Lagrange*, dando lugar a la siguiente función lagrangiana:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i [y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (5.9)$$

donde $\alpha_i \geq 0$ son los multiplicadores de *Lagrange*. Derivando esta ecuación en función de \mathbf{w} y b , e igualando a cero, se consiguen las dos condiciones de optimización:

$$\frac{\delta L(\mathbf{w}, b, \boldsymbol{\alpha})}{\delta \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (5.10)$$

$$\frac{\delta L(\mathbf{w}, b, \boldsymbol{\alpha})}{\delta b} = 0 \rightarrow 0 = \sum_{i=1}^N \alpha_i y_i \quad (5.11)$$

Por lo tanto, se puede decir que:

$$\mathbf{w}^T \mathbf{w} = \mathbf{w}^T \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i = \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j \quad (5.12)$$

Y, sustituyendo en la ecuación (??), se obtiene la función a maximizar $Q(\boldsymbol{\alpha})$:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = Q(\boldsymbol{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j \quad (5.13)$$

sujeta a las siguientes restricciones:

$$\begin{aligned} \sum_{i=1}^N \alpha_i y_i &= 0 \\ \alpha_i &\geq 0 \quad \forall i = 1, \dots, N \end{aligned} \quad (5.14)$$

donde se puede demostrar que sólo los α_i de los vectores soporte son distintos de cero.

Por lo tanto, el hiperplano de separación óptima, H_0 , depende únicamente de los vectores soporte calculados en el entrenamiento. Entonces, una vez calculados los valores óptimos α_i^* se obtendrán finalmente los coeficientes del hiperplano:

$$\alpha_i^* \Rightarrow \mathbf{w}^* = \sum_{i=1}^N \alpha_i^* y_i \mathbf{x}_i \rightarrow b^* = 1 - \mathbf{w}^{*T} \mathbf{x}_s \quad (5.15)$$

Clasificador de margen blando

El principal problema del clasificador de margen máximo es que considera que la región comprendida entre los hiperplanos H_1 y H_2 está libre de vectores ya que no contempla errores en el entrenamiento. Pero esta consideración no se cumple prácticamente nunca en aplicaciones reales y por ello se creó la idea de margen blando [?] para los casos en que no sea posible encontrar un hiperplano de separación óptimo H_0 .

En la clasificación de margen blando se busca el hiperplano óptimo en el sentido de que sea capaz de separar los datos con el menor número de errores y, además, el margen sea máximo entre los datos clasificados correctamente. Esta nueva variante introduce un nuevo conjunto de variables, denominadas variables de holgura: $\xi_i \geq 0, i = 1, \dots, N$. Estas nuevas variables permiten ponderar el grado de desclasificación de cada vector tal que:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N \quad (5.16)$$

donde:

$$\xi_i \geq 0 \quad \forall i \Rightarrow \begin{cases} 0 \leq \xi_i \leq 1 & \rightarrow \text{clasificación correcta} \\ \xi_i > 1 & \rightarrow \text{clasificación incorrecta} \end{cases} \quad (5.17)$$

Este problema del margen blando puede observarse en la Figura ???. Además, de esta forma se ha originado una nueva función objetivo, $\Omega(\mathbf{w}, \xi)$ que contiene un término que penaliza a las variables de holgura, convirtiendo la optimización en un balance entre el mayor margen y el error de penalización. En el caso de que la función de penalización sea lineal, la ecuación objetivo (??) se transforma en:

$$\text{minimizar: } \Omega(\mathbf{w}, h) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (5.18)$$

$$\text{sujeto a: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad 1 \leq i \leq N \quad (5.19)$$

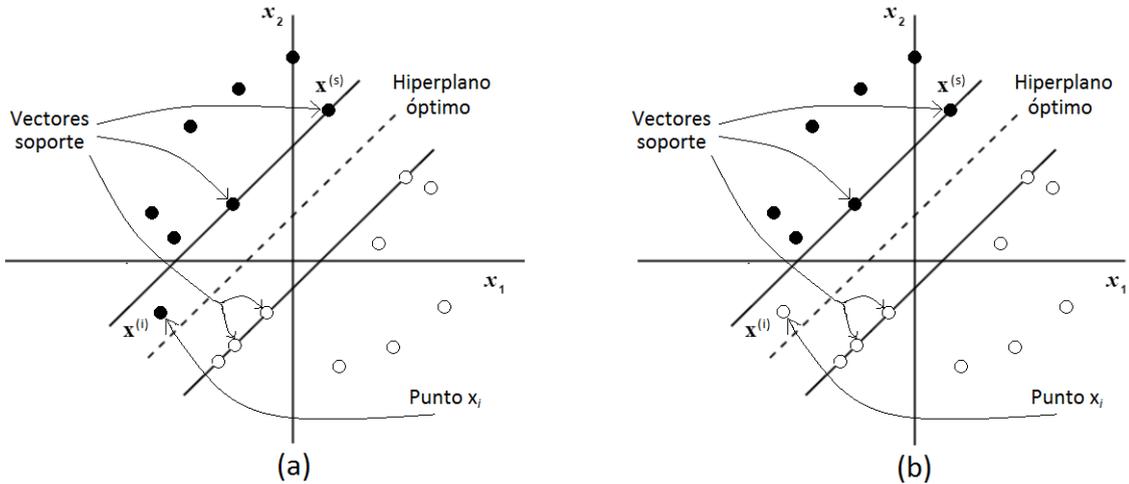


Figura 5.2: Clasificaciones de margen blando: (a) Punto \mathbf{x}_i clasificado correctamente [?]. (b) Punto \mathbf{x}_i clasificado incorrectamente [?].

Por lo tanto, siguiendo un proceso de resolución similar al del clasificador de margen máximo se lograría implementar esta nueva restricción [?,?]. Las funciones de penalización no lineales también se utilizan, sobre todo en los casos donde es preciso minimizar el efecto de muestras anómalas. Aunque estas funciones presentan el inconveniente de aumentar la complejidad de la resolución de los problemas.

Clasificador no lineal

Para realizar la clasificación en la mayoría de las aplicaciones reales, los anteriores procedimientos lineales de clasificación no serían válidos. Esta limitación motivó que se creara un nuevo tipo de SVM basada en una función de transformación no lineal $\Phi(\cdot)$ para transformar el conjunto de datos originales a otro de mayor dimensión en el que las características transformadas fueran separables linealmente [?,?]. De esta forma, mediante el conjunto transformado, se emplearía el algoritmo lineal de determinación del hiperplano óptimo, H_0 , para la separación de los conjuntos. Por lo tanto, es posible realizar una de las clasificaciones anteriores ya que la separación de las clases en el espacio transformado, donde se realiza la clasificación, es lineal, aunque en el espacio original no tiene por qué serlo. Esta nueva situación se ilustra en la Figura ??.

El algoritmo resultante para la clasificación no lineal será similar al descrito para la clasificación lineal con la única diferencia de que el vector \mathbf{x} tendrá que ser sustituido por

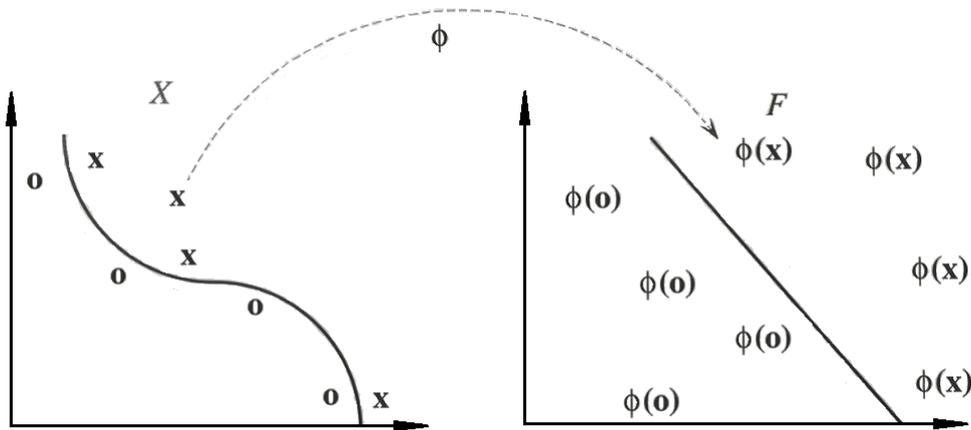


Figura 5.3: Transformación $\Phi(\cdot)$ para hacer los datos linealmente separables en las SVM [?].

$\Phi(\mathbf{x})$, que poseerá mayor dimensión. Además, el producto escalar de las transformaciones de los vectores originales será sustituido por una función *kernel*, $K(\mathbf{x}, \mathbf{x}_k)$, no lineal que solo dependerá de los vectores originales. De esta forma, no sería preciso transformar directamente los datos de entrada al entrenar la SVM y, por lo tanto, sería mucho menos costoso computacionalmente. Existen multitud de funciones *kernel*, las más comunes son las funciones de base radial y las polinómicas de grado n . Para más detalles sobre las funciones *kernel* y, en general sobre las SVM, pueden consultarse [?, ?, ?]. Este análisis no lineal será del que se valga este trabajo para implementar la SVM necesaria para la clasificación entre sujetos sanos y esquizofrénicos que se detallará posteriormente.

Implementación en el TFM

Para poner en funcionamiento las máquinas de vectores soporte se ha hecho uso del *toolbox* LIBSVM (*A Library for Support Vector Machines*) [?] de MATLAB. LIBSVM es un *software* integrado para realizar la clasificación, la regresión y la estimación de distribuciones mediante SVM.

Las muestras de los sujetos proporcionadas en la base de datos del presente estudio han sido divididas de forma aleatoria: un 75% de ellas (36 individuos) se han usado para entrenar la máquina y el 25% restante (11 sujetos) ha servido para testear su funcionamiento. Cabe destacar que, debido al reducido número de sujetos del banco de datos, y con el fin de reducir los efectos del ruido en los datos y obtener resultados estadísticamente significativos, las clasificaciones se repetirán $n_{cic} = 1000$ veces de forma similar a [?]. Este proceso es una especie de validación cruzada que permite aumentar artificialmente el número de sujetos a utilizar. Por lo tanto, este procedimiento consigue que el conjunto de test que se evaluará finalmente para estudiar la bondad del clasificador implementado aumente de forma artificial a razón de 1000 veces.

Por último, para la implementación final de este tipo de clasificador, se ha realizado una exploración con la finalidad de seleccionar el tipo de *kernel*, junto con sus parámetros óptimos, para cada agrupación de electrodos, de forma que se maximicen las tasas de acierto proporcionadas por la SVM. Los *kernels* permitidos para las SVM han sido: el lineal, la función de base radial (RBF), el polinómico de orden 3 y el sigmoideal. Por lo tanto, finalmente, se obtendrán los parámetros óptimos para la implementación de una SVM para cada agrupación de características mediante la que se realizará la clasificación final de los sujetos.

5.2.2. Redes Neuronales

Las Redes Neuronales Artificiales (*Artificial Neural Networks*, ANN) son modelos matemáticos inspirados en sistemas biológicos, adaptados y simulados en computadoras convencionales [?]. Estas redes, son procesadores masivamente paralelos y distribuidos compuestos por unidades procesadoras simples, las cuales son de forma natural propensas a almacenar conocimiento adquirido de la experiencia y hacerlo útil [?]. Estas unidades procesadoras que las conforman se asemejan a las neuronas biológicas. Las ANN se asimilan al cerebro humano en dos aspectos:

1. La red neuronal adquiere el conocimiento del ambiente, mediante un sistema de aprendizaje denominado algoritmo de entrenamiento.
2. Las fuerzas de conexión entre las neuronas, conocidas como los pesos sinápticos, se usan para almacenar el conocimiento adquirido.

A continuación, se procederá a la descripción de las bases de este modelo. Finalmente, el estudio se centrará en las redes neuronales MLP ya que son las que se han decidido implementar en el presente trabajo.

Introducción a las ANN

La neurona es la unidad fundamental del sistema nervioso y en particular del cerebro. El cerebro humano contiene más de 100 billones de neuronas. En la Figura ?? pueden diferenciarse las tres partes fundamentales de una neurona biológica descritas a continuación:

- **Dentritas.** Las cuales reciben las señales de otras células en sus puntos de conexión llamados sinapsis. Estas señales recibidas son enviadas al núcleo de la neurona donde se promediarán con otras señales.

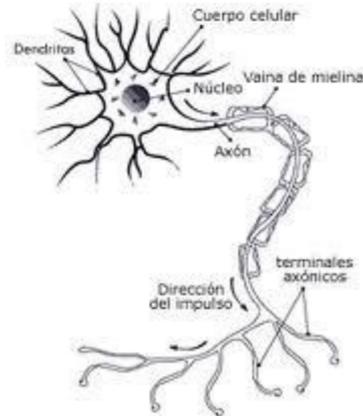


Figura 5.4: Neurona biológica con sus partes fundamentales.

- **Núcleo o Soma.** Es una fábrica química capaz de procesar y emitir señales, controlar la energía de la neurona, etc.
- **Axón.** Se encarga de enviar pulsos a otras células en caso de que el promedio obtenido en el núcleo de la neurona sea suficientemente grande a lo largo de un tiempo (la neurona se active). La punta del axón posee ramificaciones que terminan en botones sinápticos (sinapsis) para transmitir la información de una neurona a otra.

Todas las neuronas del cerebro poseen un alto grado de interconectividad entre ellas, teniendo alrededor de 10^4 botones sinápticos conectados a las dendritas de otras neuronas. Esta conectividad es la que proporciona al cerebro su gran capacidad de procesamiento.

Análogamente a la neurona biológica, una neurona artificial es un elemento que posee varias entradas y las combina, normalmente con una suma ponderada a la que se puede añadir otro sumando llamado sesgo o *bias*. Esta suma es modificada por una función de transferencia y su valor de salida se pasa directamente a la salida de la neurona. En la Figura ?? se puede observar la estructura básica de una neurona artificial.

Esta analogía de las ANN con los sistemas biológicos hace que éstas reciban sus propiedades. Es importante destacar las siguientes características heredadas por estas redes neuronales:

- **Aprendizaje.** Pueden adquirir conocimiento de una cosa por medio del estudio, ejercicio o experiencia. Además pueden cambiar su comportamiento en función del entorno.
- **Generalización.** Una vez entrenada, la red neuronal puede ofrecer respuesta correcta a una entrada no introducida anteriormente si presenta pequeñas variaciones con otras ya estudiadas.

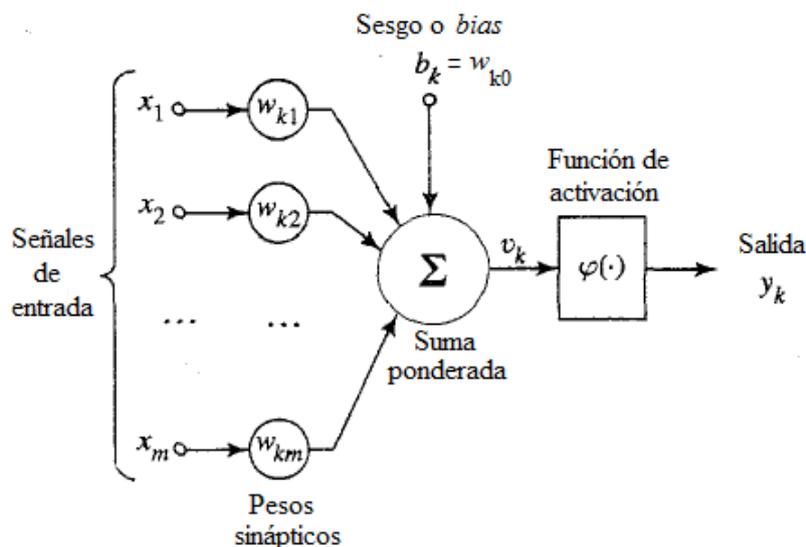


Figura 5.5: Neurona artificial con sus partes fundamentales [?].

- **Abstracción.** Pueden determinar la esencia o características principales de un conjunto de datos.

Estructura de las ANN

Una red neuronal consiste en un conjunto de neuronas artificiales conectadas de forma concreta. El interés de las ANN no reside solo en el modelo de las neuronas sino en las formas en que éstas se conectan. Generalmente, las neuronas están organizadas en grupos denominados niveles o capas. Una red típica consiste en una secuencia de capas con conexiones entre capas adyacentes consecutivas.

Existen dos capas con conexiones al mundo exterior. Una capa de entrada, *buffer* de entrada, donde se presentan datos a la red, y una capa *buffer* de salida que mantiene la respuesta de la red a una entrada determinada. El resto de capas reciben el nombre de capas ocultas.

La estructuración de estas capas de neuronas depende del enfoque que se quiera dar al problema a estudiar. Se pueden diferenciar 3 tipos de ANN en función de su estructura:

- Dos tipos de redes de propagación hacia delante (*feedforward*) cuyas conexiones siempre se encuentran en el mismo sentido: **Monocapa** como el perceptrón y **Multicapa** como el perceptrón multicapa.
- Redes **Recurrentes** que presentan al menos un ciclo cerrado de activación neuronal (realimentación). Por ejemplo la máquina de *Bolzman*.

En la Figura ?? puede encontrarse un ejemplo de cada tipo de red según su estructura. Atendiendo a estas configuraciones, se puede observar cómo las ANN poseen una carga

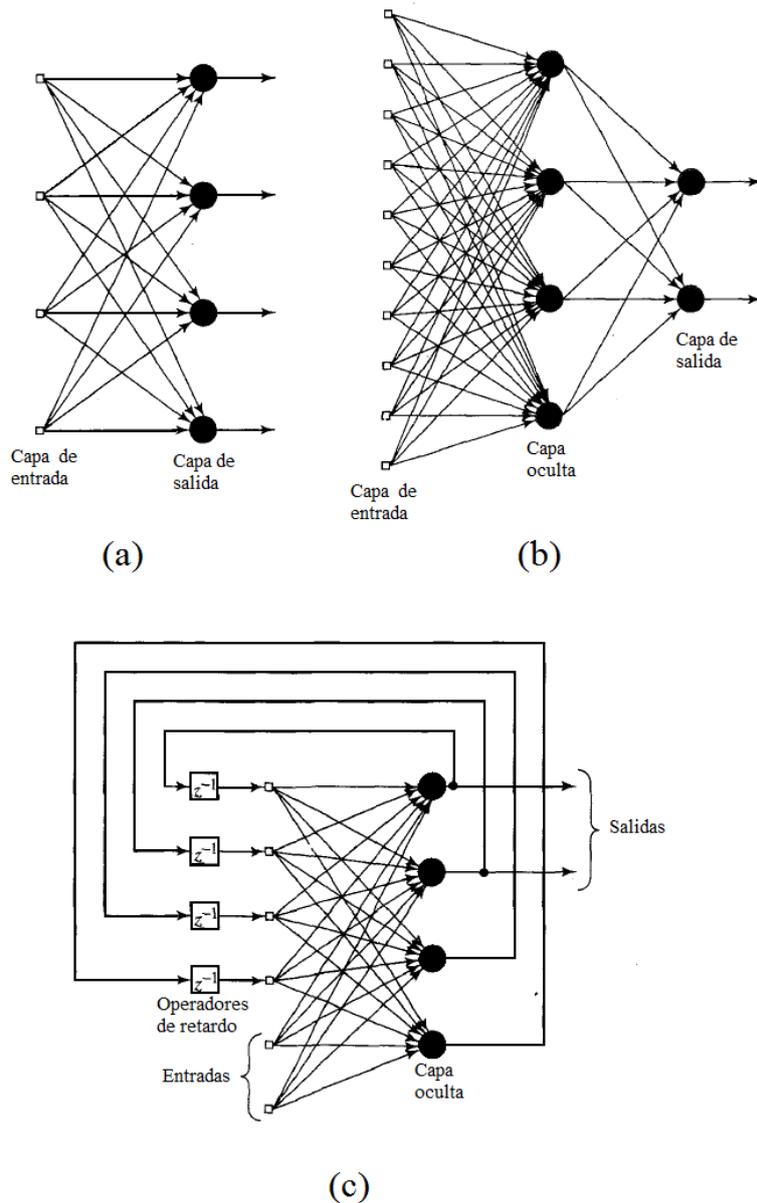


Figura 5.6: Tipos de ANN según su estructura [?]: (a) Red *feedforwards* monocapa; (b) Red *feedforwards* multicapa; (c) Red recurrente con una capa oculta.

computacional distribuida ya que cada neurona realiza un pequeño porcentaje de la carga computacional total. Además existen varias conexiones en paralelo entre cada par de capas y, por tanto, existen muchos caminos diferentes entre la entrada y la salida. Todo esto ayuda a que el funcionamiento incorrecto de alguna neurona no afecte en exceso al funcionamiento global de la red y, por ende, aporta robustez a estos clasificadores.

El funcionamiento de las ANN está determinado únicamente por las fuerzas de conexión entre cada par de neuronas ya que las operaciones realizadas en cada neurona son idénticas típicamente. Estas fuerzas entre neuronas son denominadas como pesos (w_{ij}),

siendo i el índice de la neurona, y j el de la entrada a dicho elemento. La variación de dichos pesos permite realizar distintas funciones a una misma red neuronal. Para adaptar los pesos a una tarea determinada se usa un algoritmo de entrenamiento.

Aparte de la clasificación atendiendo a su estructura, las ANN se clasifican comúnmente en dos grandes grupos atendiendo a la posesión de memoria en la clasificación. De esta forma existen:

- **Redes estáticas.** Son las más utilizadas. Carecen de memoria, por lo que, una vez entrenadas, la salida es sólo función de la entrada actual.
- **Redes dinámicas.** Se diferencian de las estáticas en que poseen memoria, es decir, su salida puede depender de sus entradas pasadas o futuras, y no sólo de las actuales.

Las ANN utilizadas en este trabajo (MLP) son redes estáticas por lo que es importante hacer hincapié en este tipo de redes con el fin de asentar las bases para la comprensión de los clasificadores que se implementarán posteriormente.

ANN estáticas

Las redes neuronales estáticas pueden implementar transformaciones no lineales de la forma $d = G(x)$ donde típicamente $x \in \mathbb{R}^n$ y $d \in [0, 1]^m$. Los coeficientes n y m representan las dimensiones de x y d respectivamente. Los tres modelos más representativos de ANN estáticas son:

- **Redes supervisadas.** Que poseen aprendizaje supervisado, es decir, se presentan pares de entrada y sus correspondientes salidas (x_i, d_i) y la red intenta adaptarse a ellos.
- **Redes auto-organizativas.** Que poseen aprendizaje no supervisado, es decir, sólo se presentan las entradas (x_i) y la red las agrupa en grupos de entradas similares.
- **Redes híbridas.** Combinación de dos o más redes básicas para juntar los atributos positivos de las redes individuales que las forman. Lo más típico es la combinación de una red auto-organizativa con otra supervisada.

El perceptrón

La ANN estática más sencilla es el perceptrón. El perceptrón pertenece a las redes supervisadas y su implementación puede observarse en la Figura ???. La entrada de esta red es un vector x_n , de dimensión n . Una vez recogidas las entradas se realiza la suma ponderada de éstas, aumentada en un valor w_0 (sesgo o *bias*). La salida de la suma

es introducida en la función de activación f que puede ser de muy diversos tipos, las más típicas son la *Hard-limiting nonlinearity* (no derivable) y la *Sigmoid nonlinearity* (derivable y, por ende, permite algoritmos de aprendizaje basados en búsqueda de gradiente). Estas funciones de activación pueden observarse en la Figura ??.

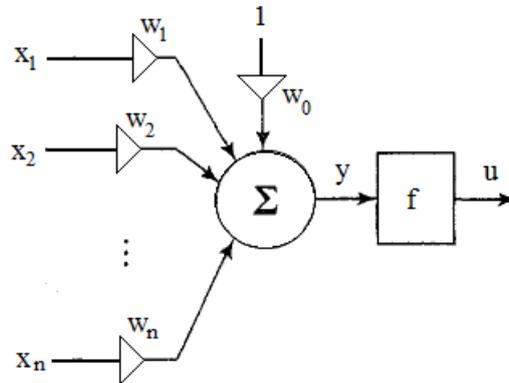


Figura 5.7: Perceptrón monocapa.

Matemáticamente, la suma ponderada de los componentes del vector de entrada se puede representar como el producto interno entre el vector de entradas x y el vector de pesos w . De esta forma, la función realizada por el perceptrón es:

$$\begin{aligned} y &= w^T x \\ u &= f(y) \end{aligned} \quad (5.20)$$

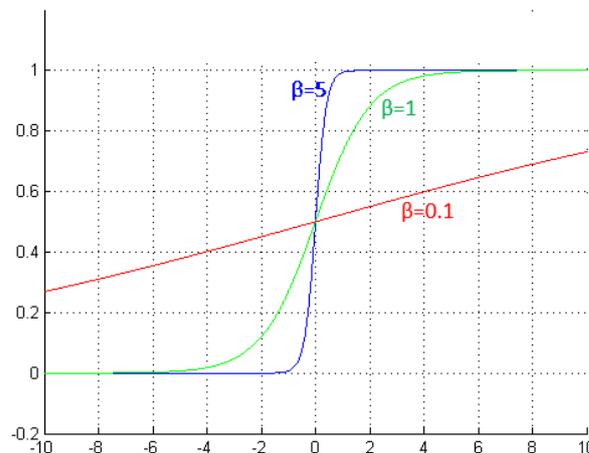
donde el valor de x_0 será 1 para introducir el sesgo en la suma.

Es posible considerar la funcionalidad del perceptrón desde dos puntos de vista:

1. *Función discriminatoria en los problemas de reconocimiento de patrones de dos clases.* Con esta red es obligatorio un límite de decisión lineal para dividir el espacio de entradas en dos regiones. Por lo tanto, no es posible resolver problemas no lineales de clasificación como el presentado en este TFM.
2. *Unidad lógica básica.* Es capaz de implementar algunas operaciones lógicas básicas: AND, OR o complemento.

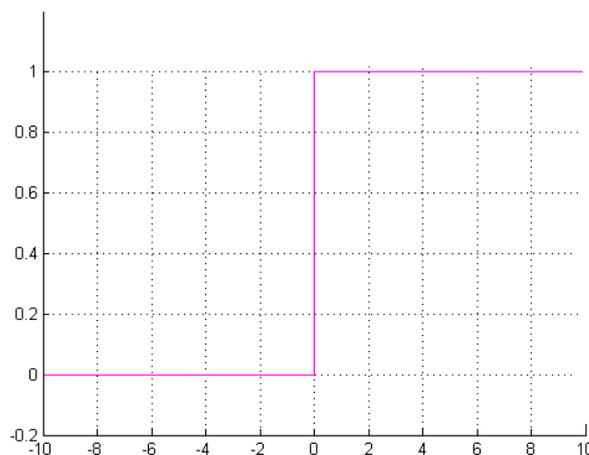
El algoritmo de entrenamiento típico de esta red es el LMS (*Least Mean Squares*) que intenta minimizar el error cuadrático entre la salida deseada, $d(x_i)$, y la salida que da el perceptrón con la suma ponderada de las entradas y_i . Por lo tanto, la función a minimizar es:

$$f_s(y) = (1 + e^{-\beta y})^{-1}$$



(a)

$$f_{HL}(y) = \begin{cases} 1 & \text{si } y > 0 \\ 0 & \text{si } y < 0 \end{cases}$$



(b)

Figura 5.8: Ejemplos de funciones de activación: (a) Función *sigmoid nonlinearity* ($\beta = 5$ en azul, $\beta = 1$ en verde, $\beta = 0,1$ en rojo); (b) Función *hard-limiting nonlinearity*.

$$J(w, x) = \frac{1}{2} \sum_{i=1}^M (y_i - d(x_i))^2 = \frac{1}{2} \sum_{i=1}^M (w^T x_i - d(x_i))^2 \quad (5.21)$$

La ecuación (??) describe una superficie por lo que se puede encontrar el mínimo mediante la técnica de descenso por gradiente. Esta técnica se fundamenta en seleccionar inicialmente un conjunto de pesos aleatorios y, a partir del valor inicial de $J(w, x)$ ir dando pequeños pasos en la superficie del error en dirección hacia el mínimo. Por lo tanto, en la iteración k , se sumaría al vector de pesos, el gradiente negativo de $J(w, x)$, multiplicado por un valor positivo μ (tasa de aprendizaje):

$$w(k + 1) = w(k) - \mu \nabla_w J(w, x)|_{w(k)} \tag{5.22}$$

donde el valor del gradiente es $\nabla_w J(w, x) = \sum_{i=1}^M (y_i(k) - d(x_i)) x_i$.

Con la descripción de esta red tan sencilla y de los anteriores conceptos de ANN, se está a disposición de describir e implementar las redes MLP que se pretenden utilizar en este trabajo.

MLP

El perceptrón multicapa (*MultiLayer Perceptron*, MLP) es una red neuronal estática de aprendizaje supervisado formado por varias capas de perceptrones denominados neuronas o nodos: una capa de entrada, posiblemente una o más capas ocultas y una capa de salida. A continuación se procede a dar una somera descripción de este tipo de redes, pero para una mayor profundización puede consultarse multitud de bibliografía que trata este tema como [?, ?].

La característica fundamental de la red MLP es que entre las capas que la conforman existe una conectividad total, esto significa que la entrada a una neurona está conectada con la salida de todas las neuronas de la capa anterior. En la Figura ?? se puede observar un ejemplo de este tipo red formada por dos capas ocultas.

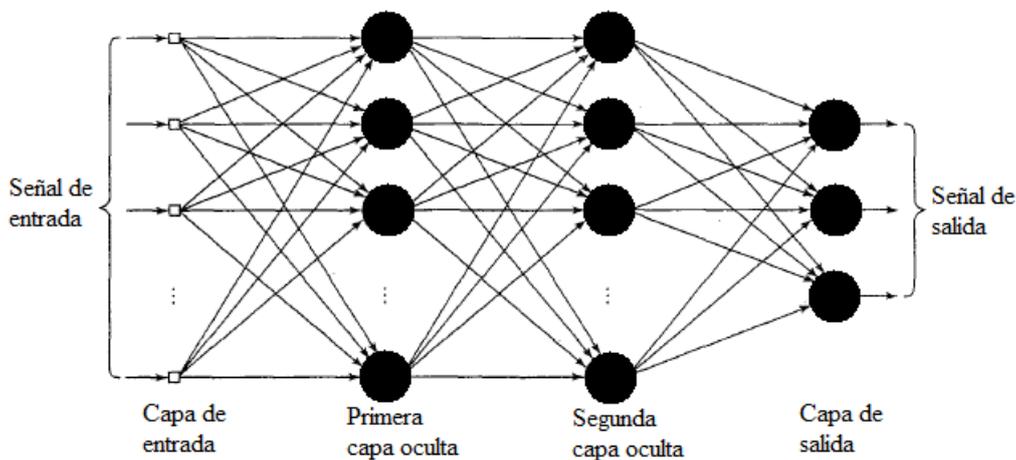


Figura 5.9: Red MLP con dos capas ocultas [?].

La función de activación típica del perceptrón multicapa es la *sigmoid nonlinearity* en lugar del *hard-limiting nonlinearity* ya que es una función derivable y, por ende, permite implementar un algoritmo de aprendizaje basado en el gradiente en todas las capas.

El algoritmo de aprendizaje más característico de MLP se denomina *backpropagation* (propagación hacia atrás) [?, ?] que es un método de entrenamiento supervisado. A la ANN

se le presentan parejas de patrones: un patrón de entrada con su correspondiente patrón de salida deseada. En cada presentación, los pesos de la red neuronal son ajustados de forma que disminuya el error entre la salida deseada y la respuesta de la red. El algoritmo *backpropagation* conlleva dos fases:

1. **Fase de propagación hacia adelante.** Se inicia cuando se presenta un patrón en la capa de entrada de la red. Posteriormente, este patrón de entrada es procesado por las distintas neuronas de la red siguiendo el proceso que se describió con la ecuación (??) en cada neurona.
2. **Fase de propagación hacia atrás o de corrección.** Se inicia una vez finalizada la fase de propagación hacia adelante. Es una etapa de modificación de los valores de los pesos, comenzando por los de la capa de salida y continuando hacia atrás hasta la de entrada. Esta modificación de los pesos se realiza con el fin de reducir el error entre la salida deseada para una determinada entrada y la respuesta que ofreció la red a dicha entrada. La búsqueda del mínimo error se basa en la técnica del descenso del gradiente, detallada para el perceptrón simple con la ecuación (??), para todos los perceptrones que conforman la red.

El clasificador MLP supera muchas de las limitaciones del perceptrón simple, por ejemplo permite implementar fronteras de decisión complejas y expresiones arbitrarias. Estas características hacen que este tipo de red sea idónea para el problema de clasificación que acontece a este trabajo ya que se prevé que la frontera de decisión del problema posea una gran complejidad. Además, el uso de esta red no requiere el conocimiento matemático de la función que relaciona las entradas con las salidas, sino sólo ejemplos de éstas.

El mayor inconveniente de MLP es su tiempo de convergencia. Las aplicaciones reales pueden tener millones de muestras de clasificación que requerirían días de cálculo de estas máquinas. De todas formas, la clasificación a realizar en esta investigación no posee restricciones de tiempo y, además, no requiere del procesamiento de un gran número de datos. Por lo tanto, esta red parece a simple vista una buena opción para la diferenciación de los patrones del presente estudio.

Implementación en el TFM

Para la puesta a punto de las redes neuronales MLP se hará uso del *toolbox* NNET (*Neural Network Toolbox*) [?] de MATLAB. Este programa permite la creación, el entrenamiento y la simulación de la mayor parte de las redes neuronales empleadas en la actualidad permitiendo variar todo tipo de parámetros de las mismas.

Los datos de sujetos proporcionados en la base de datos del presente estudio han sido divididos de manera aleatoria: un 60 % de ellos (28 individuos) se han usado para entrenar la máquina, un 10 % (5 muestras) para la validación del entrenamiento, y el 30 % restante (14 sujetos) ha servido para testear su funcionamiento. Al igual que para las SVM, debido al reducido número de sujetos del banco de datos, y con el fin de reducir los efectos del ruido en los datos y obtener resultados estadísticamente significativos, las clasificaciones se repetirán $n_{cic} = 1000$ veces de forma similar a [?]. Este proceso es una especie de validación cruzada que permite aumentar artificialmente el número de sujetos a utilizar. Por lo tanto, este procedimiento consigue que el conjunto de test que se evaluará finalmente para estudiar la bondad del clasificador implementado aumente de forma artificial a razón de 1000 veces.

La estructura de la red MLP estará compuesta por: una capa de entrada con un número de nodos igual al número de características a introducir en la red, una capa oculta formada por l perceptrones y una capa de salida compuesta por una neurona que definirá finalmente si los individuos son esquizofrénicos (la neurona se podría a 1) o sanos (al contrario). Además, los pesos de la red serán inicializados aleatoriamente y el algoritmo de aprendizaje de la red utilizará el método del gradiente conjugado escalado [?] tomando como valor de rendimiento de la red el error cuadrático medio.

Por último, para la implementación final de este tipo de clasificador, se ha realizado una exploración con la finalidad de seleccionar el número óptimo, l , de nodos de la capa oculta, para cada agrupación de electrodos, de forma que se maximicen las tasas de acierto proporcionadas por la red MLP. El número de nodos podrá oscilar entre 5 y 30 en su optimización. Por lo tanto, finalmente, se obtendrán los parámetros óptimos para la implementación de una red MLP para cada agrupación de características mediante la cual se realizará la clasificación final de los sujetos.

Capítulo 6

Resultados

En este capítulo se pretenden exponer los diferentes resultados extraídos referentes a la aplicación del método de reconocimiento de patrones para la ayuda al diagnóstico de la esquizofrenia. Este análisis de los resultados se correspondería con el realizado en la etapa 6 del método de reconocimiento general de patrones descrito en el Apartado ???. En primer lugar, se describirán los métodos de análisis de resultados mediante los que se generarán los resultados a estudiar en este TFM. Para finalizar, se hará uso de estos métodos para extraer los distintos resultados de interés del estudio.

6.1. Técnicas para el análisis de resultados

Como ya se ha mencionado, el método de reconocimiento de patrones implementado en este trabajo genera dos tipos de resultados. Por un lado, se generan los resultados relativos al análisis estadístico de cada una de las características extraídas de las señales de EEG y, por otro, los generados a partir de la clasificación de los datos. En este apartado se procede a la descripción de las técnicas utilizadas para la comprensión de ambos tipos de resultados.

6.1.1. Análisis de los resultados estadísticos

El esquema de reconocimiento de patrones implementado en el presente estudio utiliza al test U de *Mann-Whitney* para realizar el análisis estadístico. Este método parte de la hipótesis nula de que ambas distribuciones poseen medianas similares y la probabilidad de veracidad de la misma se cuantifica en términos de p-valores [?]. Un p-valor representa la probabilidad de error de rechazo de la hipótesis nula.

Para el análisis de los p-valores, es imprescindible establecer un nivel de significación tal que si el p-valor que compara las dos poblaciones está por debajo de este nivel, se pueda rechazar la hipótesis nula. Como ya se mencionó, el nivel de significación establecido para este estudio es de 0,05 que es comúnmente utilizado en las investigaciones actuales. Esto

quiere decir que si el p-valor del test U fuera inferior al nivel de significación establecido, la hipótesis de que las poblaciones poseen medianas similares sería falsa y se podría concluir infiriendo que ambas son estadísticamente independientes.

Por lo tanto, el análisis de los resultados de la ejecución de los tests de contraste de hipótesis realizados serán cuantificados en términos de p-valores. Cuanto menor sean estos p-valores, mayor nivel de diferenciación se podría atribuir a las poblaciones por lo que se, para el análisis estadístico, esta será la medida de la bondad de la separabilidad entre los sujetos sanos y los esquizofrénicos.

6.1.2. Análisis de los resultados de la clasificación

Existen gran variedad de métodos para presentar los resultados de clasificación finales de una máquina automática. Los más utilizados e intuitivos a la hora de evaluar el funcionamiento de un clasificador son las matrices de confusión y las curvas ROC. A continuación se procede a la descripción de estos métodos de análisis con el fin de que sean utilizados para la evaluación de las máquinas automáticas implementadas en el Capítulo ??.

Matrices de confusión

Las matrices de confusión son una forma simple de resumir los resultados de un clasificador de patrones que establece un valor umbral t para distinguir entre dos poblaciones. Los resultados del clasificador con valor superior al umbral son considerados de la población positiva, 1. Por el contrario, si los valores de salida son menores o iguales a t , se asignará al patrón una pertenencia a la clase negativa, 0 [?].

En la matriz de confusión se representa el tipo de los datos que se proporcionaron a la entrada de un clasificador frente a la clase que el clasificador predijo para dichos datos. En la Figura ?? puede observarse una matriz de confusión como las que se representarán posteriormente. La matriz de confusión en esta figura está formada exclusivamente por las celdas de color rojo y verde (una matriz de 2×2). El resto de celdas albergan otra información complementaria.

La diagonal principal de la matriz de confusión representa el número de datos que fueron correctamente clasificados, es decir, los verdaderos positivos (datos que se clasificaron como positivos y realmente lo eran) y los verdaderos negativos (datos clasificados como negativos que realmente lo son). La diagonal secundaria muestra los casos de clasificación errónea, es decir, los falsos positivos (datos que se clasificaron como positivos y realmente no lo eran) y los falsos negativos (datos clasificados como negativos que realmente no lo son). La celda azul muestra el porcentaje total de acierto en la clasificación en verde, y el porcentaje total de casos erróneamente clasificados en rojo.

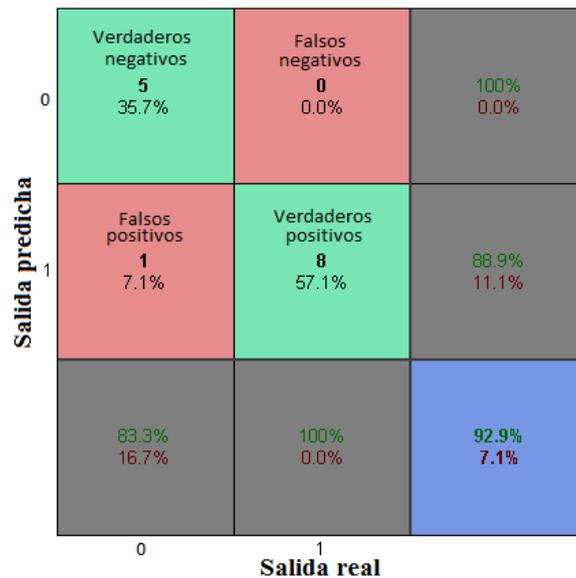


Figura 6.1: Matriz de confusión aleatoria.

Por último, las celdas grises representan el porcentaje de aciertos (en verde) y de fallos (en rojo) para las filas o columnas que llevan asociadas.

Es importante destacar que conociendo a *posteriori* los resultados de un clasificador es posible asignar una serie de *ratios* relativos. Los más importantes son:

- **La sensibilidad.** También denominada ratio de verdaderos positivos (*True Positive Rate*). Expresada como:

$$Sensibilidad = \frac{n^0 \text{ de verdaderos positivos}}{n^0 \text{ de positivos real total}} \tag{6.1}$$

- **La especificidad.** También denominada ratio de verdaderos negativos (*True Negative Rate*). Expresada como:

$$Especificidad = \frac{n^0 \text{ de verdaderos negativos}}{n^0 \text{ de negativos real total}} \tag{6.2}$$

Por lo tanto, las matrices de confusión son una forma sencilla de analizar los resultados de clasificación de una máquina automática para un determinado umbral de detección entre las clases. Pero, el hecho de que se analicen los resultados para un solo umbral de detección entre clases, produce que no se pueda verificar a ciencia cierta si una máquina posee un mejor rendimiento a otra en todos los casos, es decir, para otros umbrales. Con el fin de superar esta limitación se usan las curvas ROC.

Curvas ROC

Las gráficas de características de operadores recibidos (*Receiver Operating Characteristic*, ROC) son una técnica útil para analizar los resultados proporcionados por los

clasificadores visualizando su comportamiento para diferentes umbrales de clasificación. Para esbozar estas gráficas se parte de un clasificador al que se le varía el umbral de detección para todos sus posibles valores con el fin de obtener los resultados de clasificación producidos por cada uno de ellos. A partir de los resultados para cada umbral, se marca la sensibilidad (tasa de verdaderos positivos) en el eje de ordenadas y la tasa de falsos positivos (1-especificidad) en el eje de abscisas formando finalmente lo que se conoce como curva ROC. Por lo tanto, el eje de ordenadas de la curva ROC de relacionará con la tasa de aciertos ya que es el *ratio* de verdaderos positivos, y el eje de abscisas con la de cometer errores al ser el *ratio* de falsos positivos [?, ?].

Una restricción importante para la representación de la curva ROC es que existe un número infinito de umbrales. Como consecuencia de esto, será imposible plantear todos los umbrales posibles para un clasificador concreto. Para solucionar esta limitación se elige normalizar el clasificador previamente. De esta forma, se consigue que el umbral solamente deba de variar entre 0 y 1.

Las curvas ROC cumplen las siguientes propiedades [?]:

- Son una función monótona creciente en el primer cuadrante.
- No se ven afectadas si experimentan una transformación estrictamente creciente en los resultados de clasificación. Por tanto, no varía su resultado en el proceso de normalización de clasificadores.
- Dado un umbral $T = t$, la pendiente de la curva ROC para dicho umbral será $\frac{dy}{dx} = \frac{P(t|positivo)}{P(t|negativo)}$.

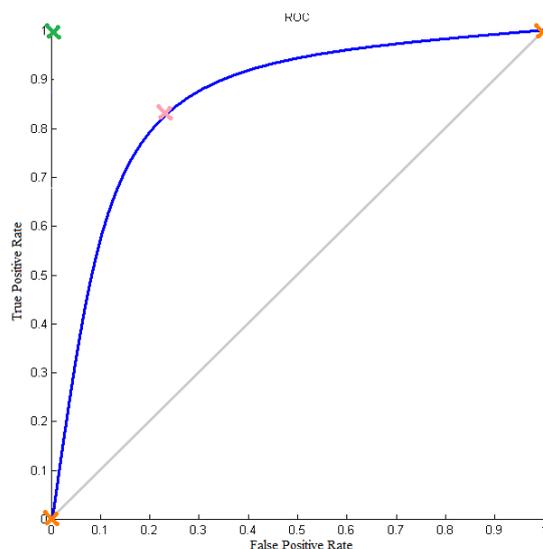


Figura 6.2: Ejemplo de curva ROC real.

En la Figura ?? se puede observar una curva ROC real aleatoria. En este tipo de curvas es muy sencillo discriminar entre algunas regiones de interés:

- Al relacionar el eje de ordenadas con los aciertos y el de abscisas con los errores, cuanto más cerca esté un punto del $(0, 1)$ mayor eficiencia de clasificación poseerá. Por lo tanto, el punto $(0, 1)$ se considera como óptimo ya que es el que permite la clasificación perfecta del 100 % de los datos sin cometer errores. Este punto aparece marcado en verde en la Figura ?. Siguiendo esta hipótesis, el punto marcado en rosa en la misma figura será el punto más óptimo del clasificador que representa esa curva ROC en concreto.
- Los puntos extremos de la gráfica, $(0, 0)$ y $(1, 1)$, marcan las situaciones límite fijadas por los umbrales. Por ejemplo, para el punto $(0, 0)$, el umbral se ha situado en un punto muy elevado y siempre considerará todas las muestras como negativas aunque no lo sean, por lo tanto clasifica correctamente todos los datos de la clase negativa y erróneamente los de la positiva. Para el otro punto extremo pasa justo al contrario, clasificará todos los datos como si fueran de la clase positiva. Estos puntos están marcados en color naranja en la curva ROC de la Figura ?.
- Para todos los puntos de la línea que une los puntos $(0, 0)$ y $(1, 1)$ (recta de comportamiento aleatorio) se considera que se posee la misma cantidad de acierto que de error. La proximidad de un punto a esta región se entiende como negativa ya que indica que las probabilidades de pertenecer a cualquiera de las dos poblaciones son idénticas y, por tanto, la clasificación no aporta nada de información de separabilidad. Esta recta está marcada en color gris en la Figura ?, y se correspondería con el hecho de tirar una moneda al aire para decidir si la muestra es de la clase 0 o de la 1.
- Si el clasificador se encuentran por debajo de la recta de comportamiento aleatorio quiere decir que la probabilidad de acierto es menor del 50 %. Por tanto, esta es una situación absurda ya que si se cambiara el criterio y se clasificara justamente con la hipótesis contraria se lograría una tasa de aciertos mayor.

Si se realizara una comparación entre dos o más clasificadores basada en las matrices de confusión, se estarían confrontando ambas máquinas únicamente en base a un umbral determinado, es decir, a un punto de la curva ROC. Para superar esta limitación, se suele evaluar la calidad de un clasificador mediante el área bajo la curva ROC (*Area Under Curve*, AUC). Este área tendrá un valor comprendido entre 0 y 1:

- Un valor próximo a 0 indica que el clasificador tiene un número mayor de errores que de aciertos. Esto quiere decir que si el criterio de elección fuera justo al contrario, se obtendrían mejores resultados por lo que infiere un error en la implementación de dicho criterio.
- Un valor cercano a 0.5 sugiere que la máquina se mantiene cerca de la región de probabilidad aleatoria y, por lo tanto, carece de utilidad.
- Un valor próximo a 1 indica que el clasificador es bueno y proporciona una alta tasa de aciertos.

Por último, otra medida interesante extraída de la curva ROC es el punto óptimo (*Best Operation Point*. BOP) [?] de la misma. El BOP es el punto que mayores tasas de acierto proporciona al clasificador que suele ser el más cercano al punto (0,1) que, como se recalcó anteriormente, pertenecería al clasificador de mayor eficiencia. Para el cálculo del punto óptimo de la curva ROC lo primero que hay que hacer es encontrar la pendiente de dicho BOP (S):

$$S = \frac{\text{coste}(P|N) - \text{coste}(N|N) \frac{N}{P}}{\text{coste}(N|P) - \text{coste}(P|P)} \quad (6.3)$$

donde P y N hacen referencia al número total de sujetos clasificados finalmente como positivos y negativos respectivamente, y la función coste calcula el coste de error que producido. Por ejemplo, $\text{coste}(P|N)$ devuelve el coste de clasificar la muestra como positiva siendo ésta negativa. Los costes de $N|N$ y $P|P$ suelen ser cero ya que se estaría realizando la clasificación correctamente, y los otros se suelen fijar como 0,5. Una vez calculada la pendiente S, se transportará una línea recta con dicha pendiente desde el punto (0,1) de la curva ROC hasta el (1,0). El primer lugar en que esta recta corte la curva ROC se definirá, finalmente, como el BOP.

Es importante destacar que la detección precoz de una enfermedad es un factor primordial a la hora de realizar cualquier diagnóstico clínico y en concreto el de la esquizofrenia ya que si no se trata a tiempo puede volverse crónica. Por este motivo, es fundamental que la cantidad de falsos sanos, es decir, de individuos esquizofrénicos diagnosticados como sanos, sea mínima. En el presente trabajo se ha elegido a los sujetos sanos como la clase negativa (“clase 0”) y a los individuos esquizofrénicos como la positiva (“clase 1”). Por lo tanto, en la presente investigación, para realizar una detección precoz de la esquizofrenia es fundamental que la tasa de falsos negativos sea mínima. Este parámetro puede ser estudiado cómodamente mediante la creación de matrices de confusión que muestran este *ratio* de la misma forma que en la matriz descrita en la Figura ??.

Por último, es importante mencionar que para la implementación, tanto de las curvas ROC como de las matrices de confusión descritas anteriormente, se ha hecho uso del

entorno de programación MATLAB el cual posee una gran variedad de funciones que facilitan en gran medida su construcción en sus *toolbox* STATS y NNET. Con estas herramientas ya se está a disposición de analizar el comportamiento de las máquinas automáticas implementadas mediante el banco de datos del presente TFM.

6.2. Resultados finales

El propósito de este apartado será la realización de un análisis profundo de los resultados generados con el fin de extraer diferentes conclusiones. A continuación, se procede a la descripción de estos resultados y a la discusión de los mismos.

6.2.1. Comparación de métodos de selección de características

El primer resultado a analizar se compone de la comparación de los resultados obtenidos por los diferentes algoritmos de selección de características. La Figura ?? muestra las distintas AUC para los diferentes filtrados utilizados en el pre-procesado (de 15 o 35Hz), los distintos criterios de selección de características (J5, AG, MIFS o DISR) y los distintos clasificadores (SVM o MLP) únicamente para las agrupaciones de más de un electrodo utilizadas¹. Puede verificarse a simple vista como, por regla general, los resultados de clasificación obtenidos mediante el uso del criterio J5 y del AG mejoran en gran medida a los resultados obtenidos utilizando los otros algoritmos de selección implementados. Por lo tanto, parece obvio que estos criterios aportan una selección de características más óptima que los basados en la teoría de la información por lo que los resultados a analizar se centrarán en los generados por estos algoritmos.

Centrándose en el criterio J5 y en el AG, puede observarse como las agrupaciones que integran un mayor número de electrodos (Total y Frontal) poseen, por regla general, una mayor AUC para el criterio J5 en comparación con el AG. Pero, por otro lado, para las agrupaciones más pequeñas (Occipitales y Parietal-Temporales), el AG parece proporcionar resultados superiores. De todas formas, ambos algoritmos presentan AUC similares por lo que se presentan como los dos métodos para la selección de características más óptimos para implementar en el esquema final de reconocimiento de patrones.

Es importante advertir que se verifica que las mayores AUC se producen, normalmente, para la agrupación de todos los canales. Pero, por otro lado, las AUC producidas para la zona parietal-temporal se acercan en gran medida a las producidas por la agrupación total a pesar de poseer un número mucho menor de características para la selección. Por lo tanto, se podría inferir que la onda P3b podría detectarse con mayor precisión en

¹Para verificar los valores concretos de las AUC mostradas en la Figura ?? puede consultarse las tablas proporcionadas en el Apéndice ??.

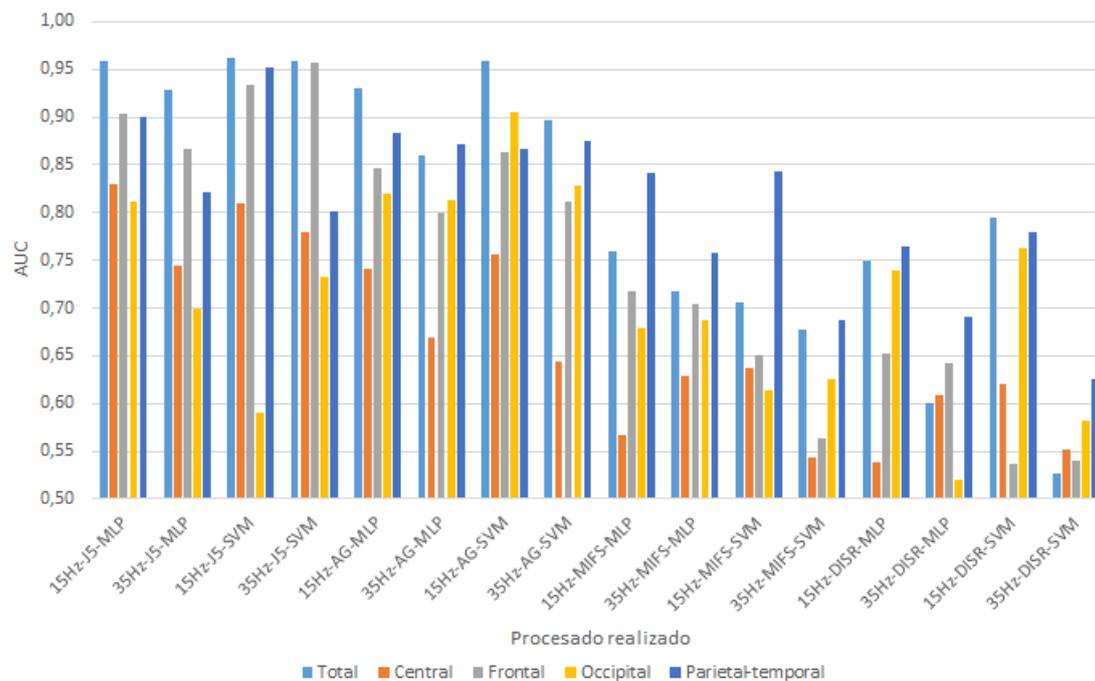


Figura 6.3: AUC obtenidas para los diferentes algoritmos de procesamiento de la señal de EEG en las diferentes asociaciones de electrodos. Las etiquetas del eje x describen los procesos utilizados en las diferentes etapas del método de reconocimiento de patrones para los valores de AUC del eje y: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (J5, AG, MIFS o DISR) y el tercero al tipo de clasificador utilizado (MLP o SVM).

la zona parietal-temporal por lo que esta región podría ser en la que se generara dicho potencial.

Por último, es importante profundizar en las posibles mejoras que podría proporcionar una optimización profunda del criterio J5 y del AG. Por un lado, el criterio J5 es un algoritmo cerrado por lo que no cabe margen de mejora en el mismo. Pero, por otro lado, el AG posee una gran cantidad de parámetros y estructuras a analizar por lo que llevando a cabo un estudio mucho más exhaustivo cabe la posibilidad de que sus resultados puedan tener todavía un margen de mejora.

6.2.2. Análisis de resultados para las agrupaciones de características de electrodos individuales

El siguiente resultado pretende analizar las agrupaciones formadas por electrodos individuales con el fin de encontrar los electrodos que mejor discriminen entre sujetos sanos y esquizofrénicos por sí mismos. Para ello, se tomarán los datos del análisis estadístico de las características² y de las AUC obtenidas a partir de la clasificación

²Para más detalles sobre los resultados estadísticos puede consultarse el Apéndice ??.

de las características seleccionadas para los electrodos individuales³. De esta forma, se pretende realizar también la extracción de diferentes conclusiones espaciales mediante ambos métodos con el fin de que éstas sean validadas por los dos análisis de resultados.

Los resultados aportados por el análisis estadístico para cuantificar la bondad de cada electrodo se calculan como el número de características estadísticamente diferenciables que posee cada uno. De esta forma, la Tabla ?? muestra el número de características discriminantes que posee cada electrodo para las señales de EEG filtradas a 15Hz y a 35Hz. Además, estos resultados son mostrados de forma gráfica en la Figura ??.

Electrodo	Características discriminantes	
	<i>Filtrado de 15Hz</i>	<i>Filtrado de 35Hz</i>
<i>C3</i>	0	0
<i>C4</i>	0	0
<i>Cz</i>	1	1
<i>F3</i>	4	4
<i>F4</i>	1	0
<i>F7</i>	6	3
<i>F8</i>	0	5
<i>Fp1</i>	3	1
<i>Fp2</i>	0	1
<i>Fz</i>	0	0
<i>O1</i>	0	0
<i>O2</i>	0	0
<i>P3</i>	0	1
<i>P4</i>	10	5
<i>Pz</i>	9	2
<i>T5</i>	2	3
<i>T6</i>	1	1

Tabla 6.1: Número de características discriminantes para cada electrodo individual mediante el test U. La primera columna describe el electrodo al que se refiere. El resto de columnas hacen referencia a los filtrados llevados a cabo en el pre-procesado (15Hz o 35Hz) anterior a la extracción de las características.

Por otro lado, para analizar los resultados aportados por los clasificadores implementados, solo se tendrán en cuenta aquellos que hayan sido obtenidos mediante el criterio J5 y el AG por ser los métodos para la selección de características más óptimos de entre los implementados en el presente estudio. Las AUC obtenidas por los electrodos individuales mediante el uso del criterio J5 pueden ser observadas en la Figura ?. Finalmente, las AUC obtenidas por los electrodos individuales utilizando el AG se muestran de igual manera en la Figura ?.

³Para más detalles sobre los resultados del AUC de la clasificación puede consultarse el Apéndice ?.

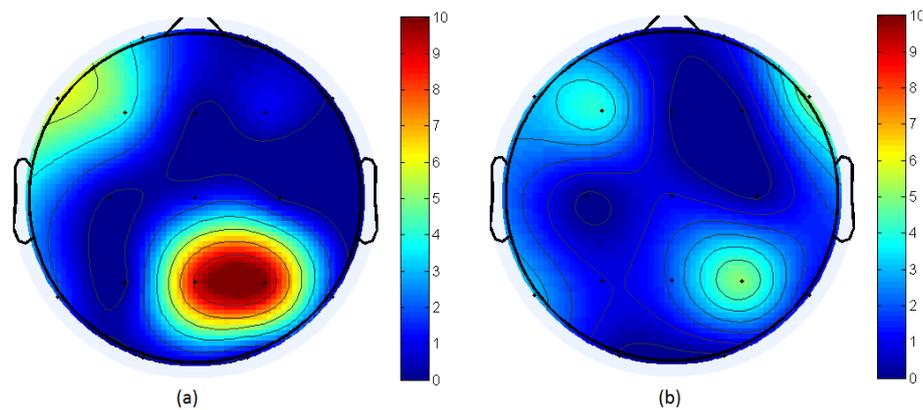


Figura 6.4: Número de características discriminantes para cada electrodo individual mediante el test U: (a) Características extraídas de la señal de EEG filtrada a 15Hz; (b) Características extraídas de la señal de EEG filtrada a 35Hz.

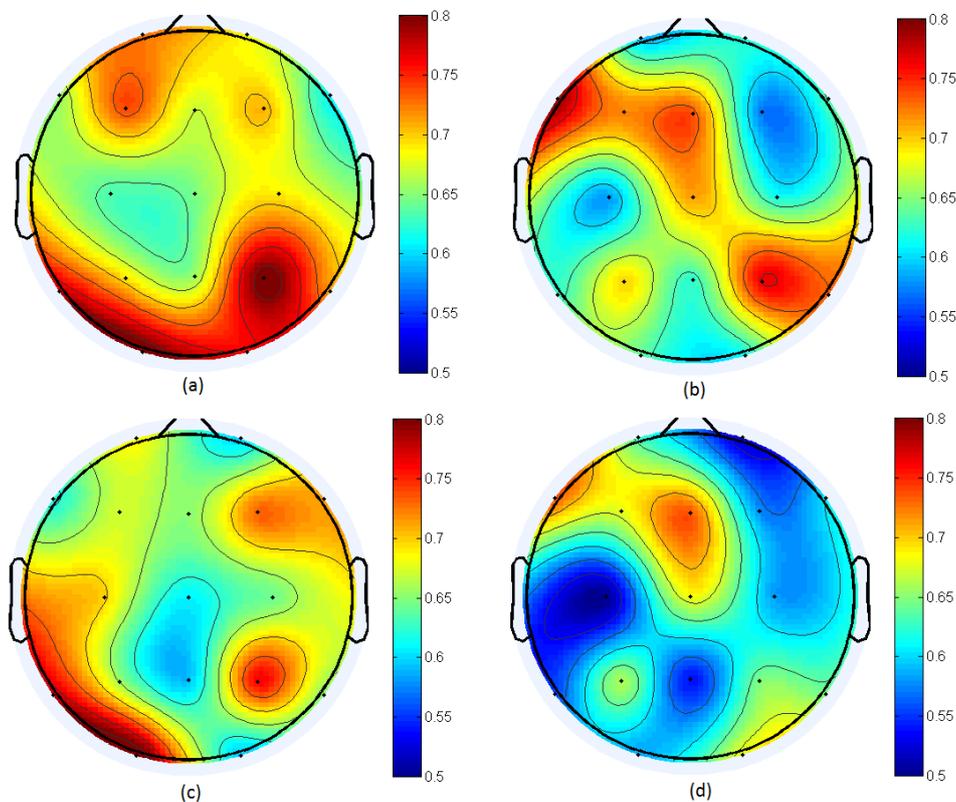


Figura 6.5: AUC para los diferentes electrodos individuales según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, criterio J5 y clasificador MLP; (b) filtrado de 35Hz, criterio J5 y clasificador MLP; (c) filtrado de 15Hz, criterio J5 y clasificador SVM; y (d) filtrado de 35Hz, criterio J5 y clasificador SVM.

Mediante el análisis de estos resultados, se puede observar como las señales que fueron filtradas a 15Hz poseen, en general, un mayor número de características significativas y

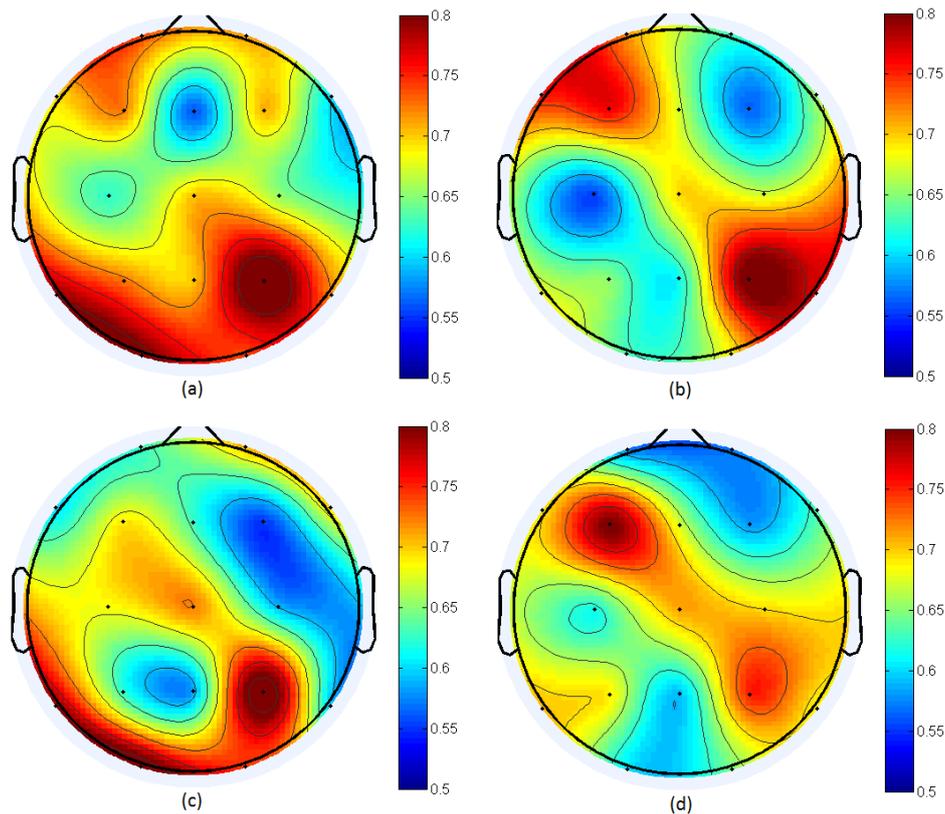


Figura 6.6: AUC para los diferentes electrodos individuales según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, AG y clasificador MLP; (b) filtrado de 35Hz, AG y clasificador MLP; (c) filtrado de 15Hz, AG y clasificador SVM; y (d) filtrado de 35Hz, AG y clasificador SVM.

una mayor precisión en la clasificación frente a las que fueron filtradas a 35Hz. Además, se puede verificar como los electrodos de la zona parietal-temporal proporcionan unas tasas de acierto ligeramente superiores y un mayor número de características discriminantes por lo que, unido a los resultados de la sección ??, hace pensar que esta zona posee una gran importancia a la hora de percibir la onda P3b. En concreto, el electrodo P4 parece ser el más importante de la zona parietal-temporal siendo el que mejor discrimina la enfermedad por sí solo.

Por último, es importante analizar los resultados para la región occipital en la que los resultados del análisis estadístico y de clasificación no son coincidentes. Por un lado, el electrodo O1 no proporciona ninguna característica significativamente discriminante mediante el análisis estadístico. Pero, por otro lado, este electrodo posee altas AUC para las clasificaciones realizadas. Por lo tanto, cabe pensar que las características extraídas por el electrodo O1 no son discriminantes por sí solas, pero la combinación de ellas sí lo es. De esta forma, se puede pensar que la zona occipital es también una región en la que la onda P3b se detecta con claridad.

6.2.3. Curvas ROC para las agrupaciones de características

Para realizar un análisis en profundidad de los resultados de clasificación de las diferentes agrupaciones de electrodos es preciso recurrir al análisis de las curvas ROC. Las Figuras ?? y ?? muestran las curvas ROC de las mismas para los distintos filtrados y clasificadores utilizados, manteniendo constante el uso de los criterios de selección de características J5 y el AG respectivamente.

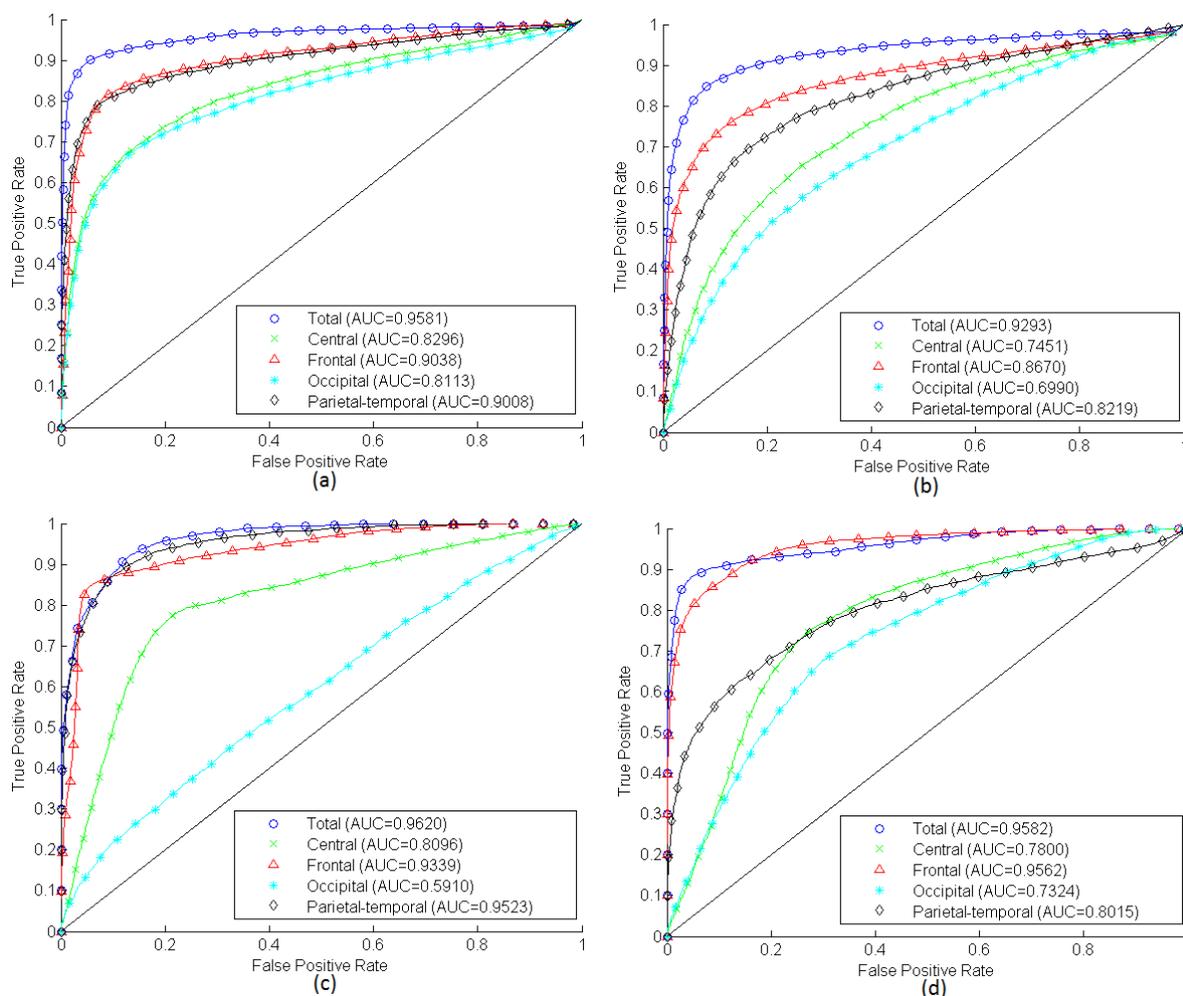


Figura 6.7: Curvas ROC para las diferentes agrupaciones de electrodos según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, criterio J5 y clasificador MLP; (b) filtrado de 35Hz, criterio J5 y clasificador MLP; (c) filtrado de 15Hz, criterio J5 y clasificador SVM; y (d) filtrado de 35Hz, criterio J5 y clasificador SVM.

En primer lugar, como se esperaba, se observa como la agrupación de todos los electrodos es, por regla general, la que mejores resultados de clasificación proporciona. Este resultado es debido a que la agrupación de todos los electrodos posee todas las características incluidas en el resto de agrupaciones. Por tanto, si la selección de

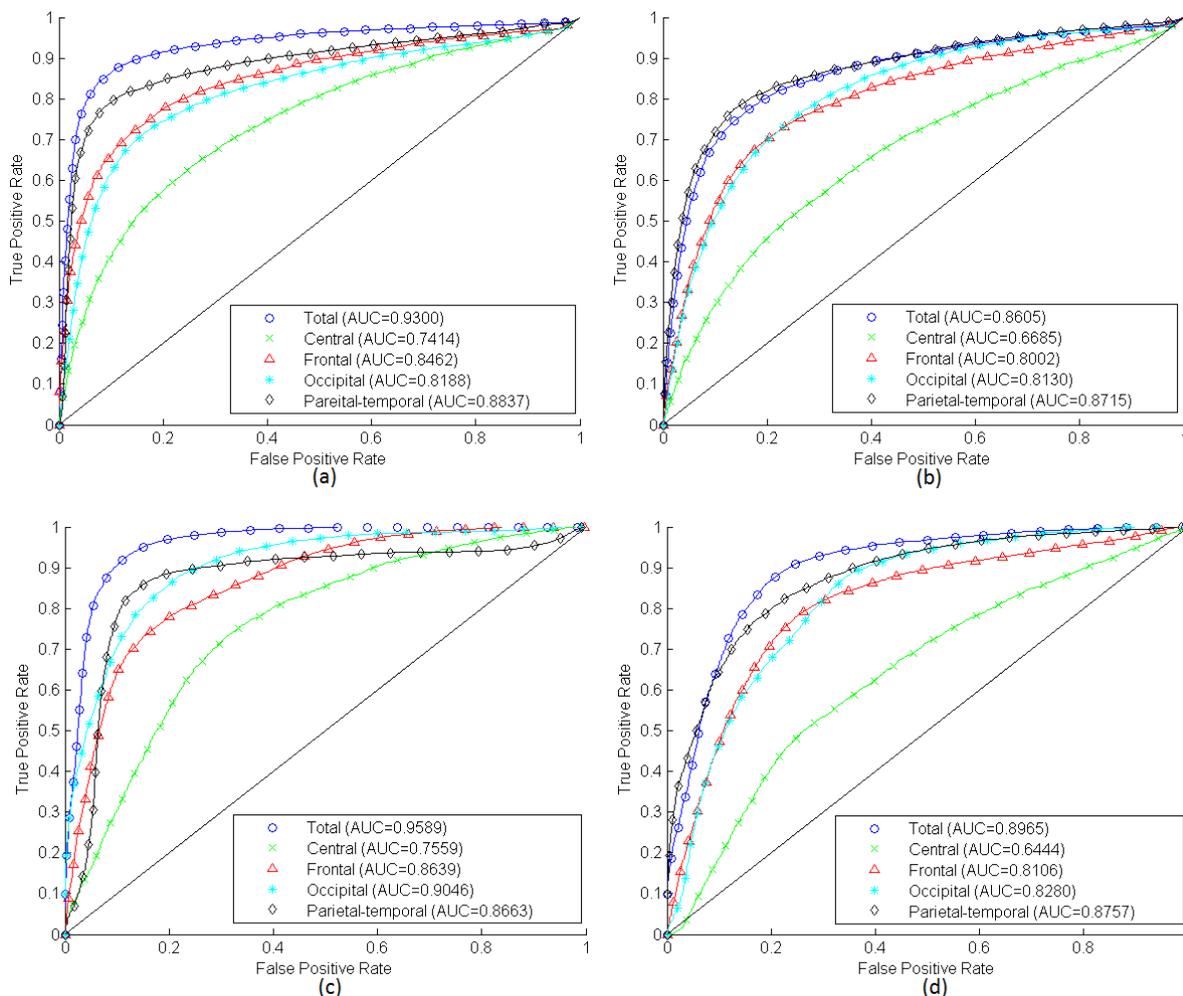


Figura 6.8: Curvas ROC para las diferentes agrupaciones de electrodos según los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: (a) filtrado de 15Hz, AG y clasificador MLP; (b) filtrado de 35Hz, AG y clasificador MLP; (c) filtrado de 15Hz, AG y clasificador SVM; y (d) filtrado de 35Hz, AG y clasificador SVM.

características ha sido eficiente, las características seleccionadas para esta agrupación serán las que mejor discriminen entre sujetos sanos y enfermos. Una peor prestación de esta agrupación solo evidenciaría una selección de características deficiente para la misma.

En cuanto a las agrupaciones espaciales realizadas, es importante advertir que la agrupación frontal, que posee 7 electrodos, posee unas curvas ROC con mejores prestaciones normalmente lo que puede ser debido al mayor número de características que posee en comparación con el resto. Pero, por otro lado, cabe destacar que los resultados obtenidos por la agrupación de electrodos parietal-temporal igual e incluso, en algunos casos, supera a la frontal con 2 electrodos menos, es decir, 40 características menos. Por lo tanto, y teniendo en cuenta los resultados extraídos anteriormente, parece evidente que esta zona es de gran importancia para la detección de la esquizofrenia.

Por último, se puede observar como las curvas ROC pertenecientes al filtrado más estricto, de 15Hz, se acercan en mayor medida al punto óptimo de la clasificación que las generadas por señales de EEG filtradas a 35Hz. Por lo tanto, parece evidente también que el uso de este filtrado en el pre-procesado mejora los resultados finales de la clasificación.

6.2.4. Análisis de las características más discriminantes

El análisis de las características que mejores tasas de acierto proporcionan en la clasificación es un resultado de gran importancia a la hora de entender en mayor profundidad el problema a resolver. Para realizar este análisis se partirá de los resultados proporcionados mediante el análisis estadístico realizado⁴.

Se ha decidido cuantificar la discriminación de una característica a la hora de su bondad a la hora de discriminar entre la población de sujetos sanos y la de esquizofrénicos como el número de electrodos que poseen dicha característica como discriminante. De esta forma, la Tabla ?? muestra el número de electrodos en los que cada característica es discriminante según si éstas han sido extraídas por un filtrado de 15Hz o de 35Hz.

Los resultados estadísticos parecen evidenciar una gran importancia de la PSE en cuanto a su valor discriminante entre la población de sujetos sanos y la de esquizofrénicos ya que es la que mayor número de electrodos discriminantes posee para ambos filtrados (la PSE es una característica discriminante en 5 de los 17 electrodos para el filtrado de 15Hz y en 7 para el de 35Hz). Además, teniendo en cuenta que la agrupación de características que mayores AUC obtiene es la agrupación llevada a cabo mediante las 10 características de todos los electrodos seleccionadas mediante el criterio J5 y filtradas mediante el filtro de 15Hz, y que 2 de las 10 características seleccionadas en ella son las PSE en diferentes electrodos⁵, queda validada la gran importancia de la PSE a la hora de la discriminación.

Por último, estudiando en mayor profundidad los valores de los p-valores reportados por los test estadísticos, puede observarse cómo la PSE en el electrodo P4 parece una característica de gran importancia a la hora de lograr diferenciar entre sujetos sanos y esquizofrénicos ya que es la que menor p-valor posee de todas. Además, la PSE en el electrodo P4 es primera característica que se selecciona tanto en la matriz de todos los electrodos como en la agrupación parietal-temporal que son las que mayores AUC proporcionan para el criterio J5 y el filtrado de 15Hz por lo que su importancia queda validada mediante ambos análisis.

⁴Para más detalles sobre los resultados estadísticos puede consultarse el Apéndice ??.

⁵Ver Apéndice ??.

Característica	Electrodos discriminantes	
	<i>Filtrado de 15Hz</i>	<i>Filtrado de 35Hz</i>
<i>varianza</i>	0	0
<i>media</i>	2	1
<i>P300ValorPico</i>	3	0
<i>P300ValorPicoAbs</i>	1	0
<i>P300Media</i>	3	0
<i>energia</i>	0	1
<i>P300Latencia</i>	0	1
<i>P300LatenciaAbs</i>	2	2
<i>LAR</i>	1	0
<i>LARAbs</i>	1	4
<i>Apositiva</i>	4	0
<i>Anegativa</i>	0	0
<i>Atotal</i>	3	0
<i>ATAR</i>	2	0
<i>TAAR</i>	2	0
<i>ZC</i>	3	2
<i>PSE</i>	5	7
<i>fmediana</i>	1	2
<i>fmedia</i>	3	4
<i>fmoda</i>	1	3

Tabla 6.2: Número de electrodos discriminantes para cada característica mediante el test U. La primera columna describe la característica a la que se refiere. El resto de columnas hacen referencia a los filtrados llevados a cabo en el pre-procesado (15Hz o 35Hz) anterior a la extracción de las características.

6.2.5. Procesado óptimo para la clasificación

Finalmente, después de todos los análisis realizados, es importante establecer el método más óptimo para llevar a cabo la clasificación entre individuos sanos y esquizofrénicos de entre los implementados en el presente estudio. Para realizar esta tarea se pretenden llevar a cabo todos los análisis de clasificación mencionados anteriormente con el fin de verificar que la combinación de algoritmos más óptima mejora al resto en todos sus aspectos.

Según los resultados anteriores, es evidente que la agrupación de todos los electrodos junto con los métodos de selección de características J5 y AG proporcionan mejores prestaciones que los otros algoritmos implementados en sus mismas etapas del método de reconocimiento de patrones. Por este motivo, se prescindirá del análisis de los resultados proporcionados por el resto de agrupaciones y de los criterios de selección MIFS y DISR. Pero, por otro lado, aunque ya se infería una ligera mejora del filtrado de

15Hz, se ha decidido tener en cuenta el filtrado de 35Hz cuando se utiliza el criterio de selección de características J5 al proporcionar curvas ROC similares al de 15Hz como se pudo observar en la Figura ???. Además, también se han decidido estudiar los resultados proporcionados por los dos clasificadores implementados al proporcionar resultados similares si se mantienen constantes el resto de algoritmos del método de reconocimiento de patrones.

Teniendo en cuenta el filtro realizado para la selección del procesado que mejores prestaciones proporcione a la hora de la clasificación entre sujetos sanos y esquizofrénicos, los procesados a tener en cuenta para la búsqueda del más óptimo son:

1. **El procesado 15Hz-Total-J5-MLP.** Con: un filtro de 15Hz en el pre-procesado, la agrupación total de electrodos, el criterio J5 en la selección de características y el clasificador MLP.
2. **El procesado 35Hz-Total-J5-MLP.** Con: un filtro de 35Hz en el pre-procesado, la agrupación total de electrodos, el criterio J5 en la selección de características y el clasificador MLP.
3. **El procesado 15Hz-Total-J5-SVM.** Con: un filtro de 15Hz en el pre-procesado, la agrupación total de electrodos, el criterio J5 en la selección de características y el clasificador SVM.
4. **El procesado 35Hz-Total-J5-SVM.** Con: un filtro de 35Hz en el pre-procesado, la agrupación total de electrodos, el criterio J5 en la selección de características y el clasificador SVM.
5. **El procesado 15Hz-Total-AG-MLP.** Con: un filtro de 15Hz en el pre-procesado, la agrupación total de electrodos, el AG en la selección de características y el clasificador MLP.
6. **El procesado 15Hz-Total-AG-SVM.** Con: un filtro de 15Hz en el pre-procesado, la agrupación total de electrodos, el AG en la selección de características y el clasificador SVM.

La Figura ?? muestra las curvas ROC para los procesados anteriores. En ella se puede observar como las curvas ROC para las combinaciones 15Hz-Total-J5-MLP y 35Hz-Total-J5-SVM parecen ser las que poseen un punto óptimo más cercano al óptimo total. De todas formas, la mayoría de las curvas ROC son muy similares y habrá que recurrir a un análisis más específico para decidirse por una de ellas.

Debido a la similitud de las curvas ROC comparadas en la Figura ??, se han decidido comparar las matrices de confusión generadas en el punto óptimo de las mismas con el fin

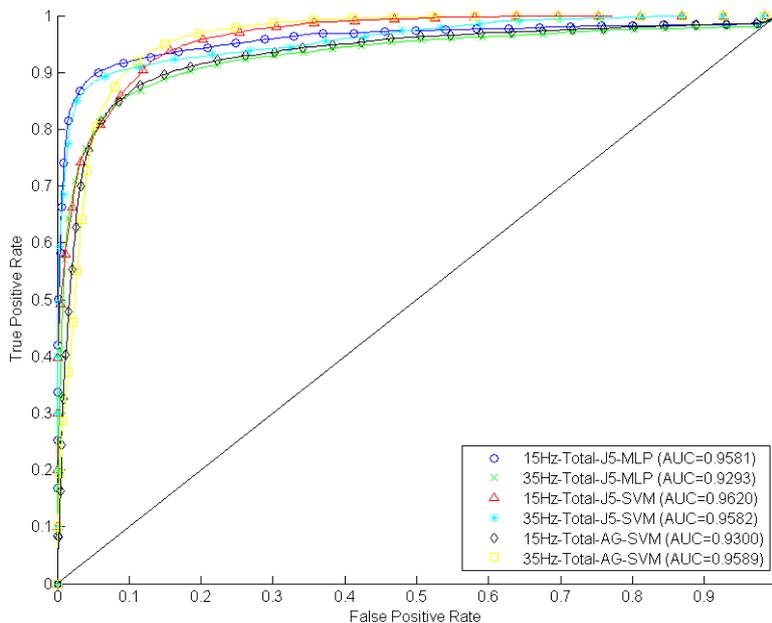


Figura 6.9: Curvas ROC para los procesados más eficientes. La leyenda indica el tipo de procesado implementado para generar cada ROC: el primer término indica el filtrado realizado en el pre-procesado (15Hz o 35Hz), el segundo el tipo de agrupación seleccionada (total), el tercero el algoritmo de selección de características (J5 o AG) y el cuarto el clasificador (MLP o SVM).

de poder seleccionar el esquema de procesado que mejores tasas de acierto proporcione en las mismas. La Figura ?? muestra las matrices de confusión cuando el umbral del clasificador se sitúa en el punto óptimo de las curvas ROC para los esquemas de procesado definidos anteriormente.

Mediante el análisis de las matrices de confusión, puede observarse como el procesado de 15Hz-Total-J5-MLP es el que logra obtener mayores tasas de acierto con un 93,4%, de forma que iguala los resultados que ya se obtenían en el PFC del que parte este trabajo. Pero, cabe destacar que la sensibilidad lograda en este TFM es de un 87,7% por lo que logra superar la sensibilidad obtenida en el PFC (87,3%) a costa de empeorar la especificidad que pasa de un 96,7% del PFC a un 96,3% en este TFM. De todas maneras, como ya se mencionó anteriormente, la sensibilidad es el factor más crítico para aplicaciones médicas por lo que se pueden concluir con que, realmente, se han mejorado los resultados finales del PFC del que partía este TFM. La consecuencia de esta mejora ha tenido que ser la optimización de los clasificadores llevada a cabo ya que el resto de parámetros de este procesado más óptimo han permanecido constantes entre ambos.

Por último, cabe destacar que estas elevadas tasas de acierto consiguen mejorar las proporcionadas por los diversos artículos revisados del estado del arte y mencionados en el apartado ?? de este trabajo. Por lo tanto, el procesado más óptimo encontrado en

este estudio podría proponerse como una herramienta bastante precisa para la ayuda al diagnóstico real de la esquizofrenia mediante las señales de EEG de los sujetos.

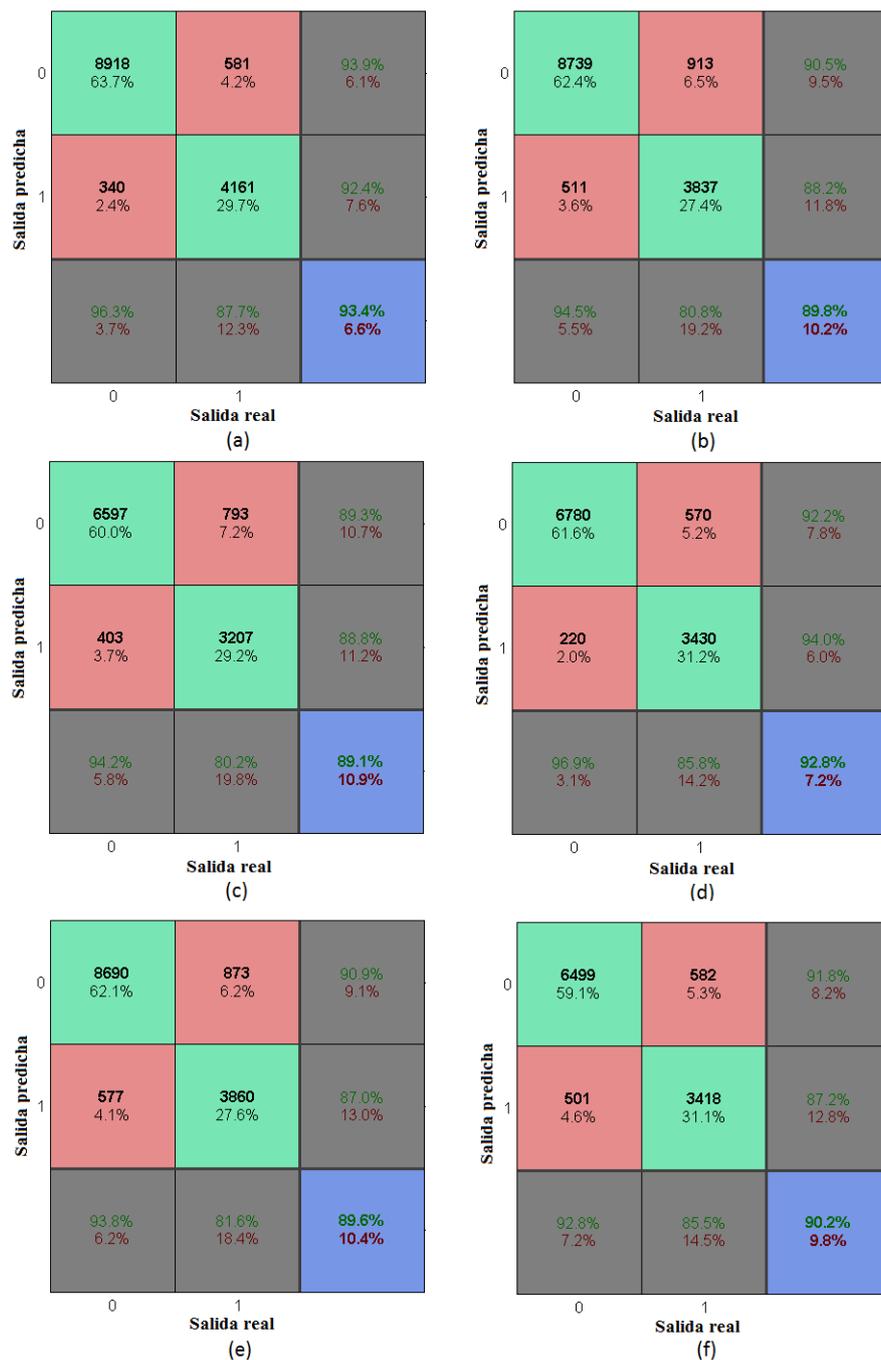


Figura 6.10: Matrices de confusión para la clasificación llevada a cabo cuando el umbral del clasificador se sitúa en el punto óptimo de la curva ROC para los procesados de: (a)15Hz-Total-J5-MLP, (b)35Hz-Total-J5-MLP, (c)15Hz-Total-J5-SVM, (d)35Hz-Total-J5-SVM, (e)15Hz-Total-AG-MLP y (f)15Hz-Total-AG-SVM.

Capítulo 7

Conclusiones y líneas futuras

El presente epígrafe pretende, en primer lugar, agrupar las diversas conclusiones extraídas del análisis de resultados realizado en el Capítulo ???. Finalmente, una vez comprendidos los resultados del TFM, se relatarán algunas líneas futuras que podrían contribuir a mejorar la aportación del presente trabajo.

7.1. Conclusiones

Este trabajo ha llevado a cabo una clasificación entre sujetos sanos y esquizofrénicos con unas altas tasas de acierto para la misma. A continuación se procede a la discusión de algunos de los resultados más significativos derivados de la investigación realizada en este TFM.

En primer lugar, se puede observar como la etapa de selección de características es crítica a la hora de implementar la clasificación de patrones. El hecho de utilizar las características más relevantes para la clasificación mejora los resultados de la misma como ya se había demostrado previamente en múltiples estudios [?, ?]. En primer lugar, se puede observar como el criterio J5, que no está ampliamente extendido como algoritmo de selección, y los AG mejoran en gran medida a criterios más clásicos basados en la teoría de la información como son el MIFS y el DISR. Este resultado encaja con otros artículos que ya utilizan otros métodos más avanzados para la selección de características. De esta forma, en [?] se utiliza un algoritmo basado en el método de Taguchi para mejorar la selección de características proporcionada por el algoritmo MIFS. Además, gran cantidad de investigaciones [?, ?, ?] implementan métodos no lineales como los basados en métodos evolutivos, a los que pertenecen los AG, para mejorar la calidad de la selección.

En segundo lugar, continuando con la comparación de algoritmos para la selección de características, es importante analizar en profundidad las diferencias entre los rendimientos de clasificación proporcionados por el criterio J5 y el AG ya que no son tan amplias como las de éstos con respecto a los algoritmos basados en la teoría de

la información (MIFS y DISR). Parece intuirse que el criterio J5 muestra selecciones más discriminantes para agrupaciones grandes de características, mientras que el AG se adapta mejor a la selección en agrupaciones más reducidas. De todas formas, esta diferencia puede ser debida a que el AG implementado no realiza una búsqueda lo suficientemente exhaustiva para las agrupaciones grandes. Por tanto, es inmediato pensar que podría existir una configuración más óptima del AG para estas agrupaciones ya que la teoría subyacente de estos algoritmos describe una gran cantidad de parámetros a variar para lograr adaptarse a los diferentes problemas permitiendo búsquedas más o menos exhaustivas. En cambio, el criterio J5 es un algoritmo mucho más cerrado para la selección de características ya que el único parámetro a variar es el número de características seleccionadas. Por lo tanto, parece evidente que, aunque la mayor tasa de acierto es proporcionada por el criterio J5, el AG se presenta como un método que puede obtener unas tasas de acierto iguales o superiores realizando en el mismo una optimización lo suficientemente profunda.

En tercer lugar, es importante analizar los resultados de forma espacial. Aunque la resolución espacial de la señal de EEG es limitada, las señales de los electrodos más discriminantes a la hora de diferenciar entre sujetos sanos y esquizofrénicos pueden dar una idea de la localización del potencial P3b. Este análisis espacial ya ha sido realizado en otras investigaciones. De esta forma, en [?] se detecta que la onda P3b se origina por la actividad parietal-temporal asociada a la atención y relacionada con el procesamiento cognitivo, verificando otros estudios posteriores. Los resultados proporcionados en este trabajo validan dicha hipótesis obteniendo unas mayores tasas de acierto en la clasificación de las características parietal-temporales. Además, cabe destacar que, dentro de los electrodos de la zona parietal-temporal analizada, la localización del electrodo P4 se alza como una zona de gran importancia para la detección de la onda P3b. El análisis estadístico para las características de la señal de EEG de dicho electrodo concluye con que 10 de sus 20 características extraídas pueden ser consideradas como discriminantes a la hora de diferenciar entre la población de sujetos sanos y la de esquizofrénicos. Además, el análisis de clasificación verifica los resultados estadísticos atribuyendo a la selección de las características del electrodo P4 las mayores tasas de acierto en la clasificación. Por lo tanto, parece evidente que la señal de EEG captada en la zona en la que se sitúa el electrodo P4 es la que mejor consigue distinguir entre sujetos sanos y esquizofrénicos en este estudio.

Otro resultado importante de extraído de esta investigación consiste en la verificación de la importancia de la etapa de pre-procesado del método de reconocimiento de patrones, y más en concreto, del filtrado realizado en el mismo. Se ha verificado como el filtrado más agresivo a 15Hz, como el propuesto en [?], proporciona tasas de acierto superiores

al filtrado más tradicional de 35Hz, como el implementado en [?]. El motivo de que el análisis a 15Hz proporciona resultados más óptimos en la clasificación puede deberse a que este estudio se ha pretendido centrar únicamente en las características de la onda P3b. De esta forma, teniendo en cuenta que el potencial P3b posee una frecuencia muy baja, parece que un filtrado a 15Hz consigue eliminar mayor cantidad de componentes de la señal que no pertenezcan a dicha onda conservando las características de la misma. Por lo tanto, el filtrado de 15Hz parece eliminar mayor cantidad de ruido lo que repercute en una detección más clara del potencial que pretende distinguir entre los controles y los esquizofrénicos y, por ende, en una mayor tasa de aciertos en la clasificación de los mismos.

Por otro lado, en cuanto a la importancia de las características extraídas de la señal pre-procesada de EEG, se ha verificado la importancia de la PSE como una característica que posee un gran valor discriminante a la hora de discernir entre sujetos sanos y esquizofrénicos. Esta característica ya había sido utilizada en [?] para la misma clasificación que la realizada en el presente estudio. Por lo tanto, los resultados de esta investigación parecen concordar en este sentido con el estado del arte actual. Además, siguiendo en la línea del análisis de características, es importante como verificar que las características seleccionadas para la clasificación más óptima incluyen tanto características temporales como frecuenciales. Por lo tanto, parece evidente que las características de ambos dominios se complementan con éxito para lograr unas tasas de clasificación elevadas.

Por último, es importante destacar que el principal inconveniente para realizar este análisis, ha sido el reducido número de sujetos a clasificar ya que las máquinas automáticas para la clasificación precisan de gran cantidad de ejemplos para su correcto entrenamiento. Pero, partiendo de esta limitación y reuniendo todas las deducciones obtenidas, se ha llegado a la conclusión de que el procesado más óptimo de los implementados para realizar esta tarea engloba un filtrado paso banda en el pre-procesado entre 0.1 y 15Hz, una agrupación de las características de todos los electrodos en conjunto, una selección de las mismas basada en el criterio J5 y el empleo de una red neuronal MLP como clasificador automático. Con este procedimiento óptimo, se ha logrado una tasa de aciertos de un 93,4 %, con una sensibilidad de un 87,7 %, y una especificidad del 96,3 %. Estos resultados han resultado ser superiores a otras investigaciones anteriores [?, ?, ?]. Esta efectividad a la hora de la clasificación hace pensar que el método utilizado puede ser implementado en aplicaciones de diagnóstico real como se pretendía en los objetivos de este TFM.

7.2. Líneas Futuras

A lo largo del desarrollo de este TFM se han ido cumpliendo los objetivos planteados en el Capítulo ???. No obstante, aún quedarían algunos aspectos que tratar en un futuro:

- **Aumento del banco de datos.** Para poder generalizar los resultados de la clasificación entre sujetos sanos y esquizofrénicos es condición necesaria poseer un número suficientemente alto de ejemplos. No obstante, se espera que en un futuro, el método de reconocimiento de patrones descrito se pueda testear mediante un mayor número de señales de EEG.
- **Optimización del Algoritmo Genético.** Teniendo en cuenta las conclusiones extraídas del presente estudio, parece evidente que una línea importante de investigación sería la realización de una optimización más en profundidad para la selección de características realizadas mediante el AG cuando el número de las mismas es elevado. Esta optimización podría consistir en testear los diferentes parámetros que puede incluir el AG o, incluso la modificación el orden de las diferentes etapas e su conjunto. Además, se podrían implementar otros métodos de selección de características basados también en la computación evolutiva como por ejemplo los algoritmos basados en colonia de hormigas o la búsqueda genética híbrida.
- **Combinación de EEG con otras técnicas de captación de potenciales cerebrales.** La señal de EEG posee algunas limitaciones, como su limitada resolución espacial. La combinación de las características extraídas de un EEG con otras extraídas de otras herramientas de captación de procesos cerebrales, como las resonancias magnéticas (MRI), podrían realizar una clasificación con unas tasas de acierto aún superiores.
- **Realización de otras clasificaciones.** La esquizofrenia es una enfermedad con multitud de variantes y con muchos factores a estudiar actualmente. Por todo ello, sería interesante poder realizar una clasificación automática entre diferentes variantes de la enfermedad. Además, también serían interesantes otros tipos de clasificaciones como por ejemplo la detección automática de la esquizofrenia en la adolescencia o la detección de individuos propensos a manifestar la enfermedad como familiares de pacientes. Una última clasificación que podría resultar de interés sería la relativa a la discriminación entre sujetos sometidos o no a alguna terapia con el fin de validar la eficiencia de la misma.

- **Implementación de una interfaz gráfica.** Todas las funciones del proceso realizado para la posterior clasificación pueden ser implementadas mediante una interfaz gráfica para que este proceso sea sencillo de ejecutar para usuarios con pocos conocimientos de programación como pueden ser los médicos o psiquiatras.

Bibliografía

- [1] L. Santos-Mayo. Diagnóstico automático de la esquizofrenia basado en el estudio de señales de eeg de actividad cortical en el cerebro. Master's thesis, Universidad de Valladolid, 2014.
- [2] D.L. Kelly. Treatment considerations in women with schizophrenia. *Journal of Women's Health*, 15(10):1132–1140, 2006.
- [3] American Psychiatric Association et al. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub, 2013.
- [4] J. Polich. On the relationship between eeg and p300: individual differences, aging, and ultradian rhythms. *International Journal of Psychophysiology*, 26(1-3):299–317, 1997.
- [5] D.H.R. Blackwood, A. Fordyce, M.T. Walker, D.M. St. Clair, D.J. Porteous, and W.J. Muir. Schizophrenia and affective disorders cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and p300 findings in a family. *The American Journal of Human Genetics*, 69(2):428–433, 2001.
- [6] R.W. McCarley, M.E. Shenton, B.F. O'Donnell, S.F. Faux, R. Kikinis, P.G. Nestor, and F.A. Jolesz. Auditory p300 abnormalities and left posterior superior temporal gyrus volume reduction in schizophrenia. *Archives of General Psychiatry*, 50(3):190–197, 1993.
- [7] B.I. Turetsky, E.A. Colbath, and R.E. Gur. P300 subcomponent abnormalities in schizophrenia: I. physiological evidence for gender and subtype specific differences in regional pathology. *Biological Psychiatry*, 43(2):84–96, 1998.
- [8] U.S. Department of health and human services. Technical report, National Institute of Mental Health, 2009.
- [9] A. Pfefferbaum, B.G. Wenegrat, J.M. Ford, W.T. Roth, and B. S. Kopell. Clinical application of the p3 component of event-related potentials. ii. dementia, depression

- and schizophrenia. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 59(2):104–124, 1984.
- [10] W.T. Roth, A. Pfefferbaum, T.B. Horvath, P.A. Berger, and B.S. Kopell. P3 reduction in auditory evoked potentials of schizophrenics. *Electroencephalography and Clinical Neurophysiology*, 49(5):497–505, 1980.
- [11] D.H. Mathalon, J.M. Ford, and A. Pfefferbaum. Trait and state aspects of p300 amplitude reduction in schizophrenia: a retrospective longitudinal study. *Biological psychiatry*, 47(5):434–449, 2000.
- [12] V.B. Souza, W.J. Muir, M.T. Walker, M.F. Glabus, H.M. Roxborough, C.W. Sharp, J.R. Dunan, and D.H. Blackwood. Auditory p300 event-related potentials and neuropsychological performance in schizophrenia and bipolar affective disorder. *Biological psychiatry*, 37(5):300–310, 1995.
- [13] N. Hazarika, J.Z. Chen, A.C. Tsoi, and A. Sergejew. Classification of eeg signals using the wavelet transform. In *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*, volume 1, pages 89–92. IEEE, 1997.
- [14] J.W. Kim, Y.S. Lee, D.H. Han, K.J. Min, J. Lee, and K. Lee. Diagnostic utility of quantitative eeg in un-medicated schizophrenia. *Neuroscience letters*, 589:126–131, 2015.
- [15] R. Boostani, K. Sadatnezhad, and M. Sabeti. An efficient classifier to diagnose of schizophrenia based on the eeg signals. *Expert Systems with Applications*, 36(3):6492–6499, 2009.
- [16] J.W. Kam, A.R. Bolbecker, B.F. O’Donnell, W.P. Hetrick, and C.A. Brenner. Resting state eeg power and coherence abnormalities in bipolar disorder and schizophrenia. *Journal of psychiatric research*, 47(12):1893–1901, 2013.
- [17] M. Ravan, G. Hasey, J.P. Reilly, D. MacCrimmon, and A. Khodayari-Rostamabad. A machine learning approach using auditory odd-ball responses to investigate the effect of clozapine therapy. *Clinical Neurophysiology*, 126(4):721–730, 2015.
- [18] N.N. Boutros, C. Arfken, S. Galderisi, J. Warrick, G. Pratt, and W. Iacono. The status of spectral eeg abnormality as a diagnostic test for schizophrenia. *Schizophrenia research*, 99(1):225–237, 2008.

- [19] E.J. Novella and R. Huertas. El síndrome de kraepelin-bleuler-schneider y la conciencia moderna: Una aproximación a la historia de la esquizofrenia kraepelin-bleuler-schneider syndrome from a modern perspective: An assessment of the history. *Clínica y Salud*, 21(3):205–219, 2010.
- [20] F.L. Ruch and P.G. Zimbardo. *Psicología y vida*. Biblioteca técnica de psicología. Trillas, 1979.
- [21] L. Palacios. Breve historia de la electroencefalografía. *Acta Neurológica Colombiana*, 2:104–7, 2002.
- [22] K.L. Moore, A.F. Dalley, and I.N. Benlloch. *Anatomía: con orientación clínica*. Médica Panamericana ; Lippincott Williams & Wilkins, 2002.
- [23] K. A Kooi, R. Paul Tucker, and R.E. Marshall. *Fundamentals of electroencephalography*. HarperCollins Publishers, 1978.
- [24] D.R. Ripoll. *Fundamentos de psicobiología*. Manuales (Universitat Oberta de Catalunya): Psicología. Editorial UOC, S.L., 2010.
- [25] H.H. Jasper. The ten twenty electrode system of the international federation. *Electroencephalography and clinical neurophysiology*, 10:371–375, 1958.
- [26] M.D. Rugg and M.G.H. Coles. *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford University Press, 1995.
- [27] W. Freeman and R.Q. Quiroga. *Imaging Brain Function With EEG: Advanced Temporal and Spatial Analysis of Electroencephalographic Signals*. Springer, 2012.
- [28] A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.
- [29] J.C. Hansen and S.A. Hillyard. Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, 49(3):277–290, 1980.
- [30] M.I. Núñez-Peña, M.J. Corral, C. Escera, et al. Potenciales evocados cerebrales en el contexto de la investigación psicológica: una actualización. *Anuario de Psicología*, 35(1):3–22, 2004.
- [31] J. Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical Neurophysiology*, 118(10):2128–2148, 2007.

- [32] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras. *Introduction to Pattern Recognition: A Matlab Approach: A Matlab Approach*. Academic Press, 2010.
- [33] J. Tohka. Sgn-2506: Introduction to pattern recognition. *Tampere University of Technology, Department of Signal Processing*, 2013.
- [34] J. Polich and A. Kok. Cognitive and biological determinants of p300: an integrative review. *Biological Psychology*, 41(2):103–146, 1995.
- [35] J. Polich. Neuropsychology of p300. *Oxford handbook of event-related potential components*, pages 159–188, 2012.
- [36] J. de la Torre Abaitua. Procesado de señales eeg para un interfaz cerebro-máquina (bci). Master’s thesis, Universidad Carlos III de Madrid, 2012.
- [37] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.
- [38] L. Bougrain, C. Saavedra, R. Ranta, et al. Finally, what is the best filter for p300 detection? In *TOBI Workshop III-Tools for Brain-Computer Interaction-2012*, 2012.
- [39] R.R. Vázquez Álvarez. *Áreas anatómicas cerebrales activadas en respuesta a estímulos con contenido moral*. PhD thesis, Universidad Autónoma Metropolitana, México, 2007.
- [40] W.K. Strik, A.J. Fallgater, A. Heidrich, H. Kulke, T.J. Mueller, and T. Dierks. Características de la onda auditiva p300 en la clasificación de leonhard.
- [41] D. Friedman and G.V. Simpson. Erp amplitude and scalp distribution to target and novel events: effects of temporal order in young, middle-aged and older adults. *Cognitive Brain Research*, 2(1):49–63, 1994.
- [42] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, 2000.
- [43] A. Mognon and M. Buiatti. Adjust tutorial an automatic eeg artifact detector based on the joint use of spatial and temporal features.
- [44] A. Salazar. *On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling: On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling*. Springer Theses. Springer Science+Business Media, 2012.

- [45] I. Kalatzis, N. Piliouras, E. Ventouras, C.C. Papageorgiou, A.D. 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*, 75(1):11–22, 2004.
- [46] 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*, 94(1):48–57, 2009.
- [47] M. Sabeti, S. Katebi, and R. Boostani. Entropy and complexity measures for eeg signal classification of schizophrenic and control participants. *Artificial Intelligence in Medicine*, 47(3):263–274, 2009.
- [48] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda, and M. Hirano. Quantification of eeg irregularity by use of the entropy of the power spectrum. *Electroencephalography and Clinical Neurophysiology*, 79(3):204–210, 1991.
- [49] G. Brown, A. Pocock, M.J. Zhao, and M. Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *The Journal of Machine Learning Research*, 13(1):27–66, 2012.
- [50] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *Neural Networks, IEEE Transactions On*, 5(4):537–550, 1994.
- [51] A.M. Fraser and H.L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A.*, 33(2):1134, 1986.
- [52] P.E. Meyer and G. Bontempi. On the use of variable complementarity for feature selection in cancer classification. In *Applications of Evolutionary Computing*, pages 91–102. Springer, 2006.
- [53] D.A. Bell and H. Wang. A formalism for relevance and its application in feature subset selection. *Machine Learning*, 41(2):175–195, 2000.
- [54] A. Tristán-Vega. Segmentación semiautomática de radiografías de la mano para la estimación de la edad ósea en la infancia. Master’s thesis, Universidad de Valladolid, 2004.
- [55] D.B. Fogel. *Evolutionary computation: toward a new philosophy of machine intelligence*, volume 1. John Wiley & Sons, 2006.

- [56] J.H. Holland. Adaptation in natural and artificial system: an introduction with application to biology, control and artificial intelligence. *Ann Arbor, University of Michigan Press*, 1975.
- [57] B. Oluleye, L. Armstrong, J. Leng, and D. Diepeveen. A genetic algorithm-based feature selection. *British Journal of Mathematics & Computer Science*, pages In-Press, 2014.
- [58] H. Fröhlich, O. Chapelle, and B. Schölkopf. Feature selection for support vector machines by means of genetic algorithm. In *Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on*, pages 142–148. IEEE, 2003.
- [59] T.T. Erguzel, S. Ozekes, O. Tan, and S. Gultekin. Feature selection and classification of electroencephalographic signals an artificial neural network and genetic algorithm based approach. *Clinical EEG and neuroscience*, page 1550059414523764, 2014.
- [60] I. Rejer and K. Lorenz. Genetic algorithm and forward method for feature selection in eeg feature space. *Journal of Theoretical and Applied Computer Science*, 7(2):72–82, 2013.
- [61] S.N. Sivanandam and S.N. Deepa. *Introduction to Genetic Algorithms*. Springer Berlin Heidelberg, 2007.
- [62] B. Oluleye, L. Armstrong, L. Jinsong, and D. Diepeveen. Zernike moments and genetic algorithm: Tutorial and application. *British Journal of Mathematics & Computer Science*, 4(15):2217–2236, 2014.
- [63] I MathWorks. Global optimization toolbox user’s guide (r2015a), 2015.
- [64] G.V. Glass and K.D. Hopkins. *Statistical methods in education and psychology*. Prentice-Hall Englewood Cliffs, NJ, 1970.
- [65] H.W. Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, 1967.
- [66] A.R. Ansari, R.A. Bradley, et al. Rank-sum tests for dispersions. *The Annals of Mathematical Statistics*, 31(4):1174–1189, 1960.
- [67] T.V. Perneger. What’s wrong with bonferroni adjustments. *BMJ: British Medical Journal*, 316(7139):1236, 1998.

- [68] D.A. Savitz and A.F. Olshan. Multiple comparisons and related issues in the interpretation of epidemiologic data. *American journal of epidemiology*, 142(9):904, 1995.
- [69] Inc The MathWorks. Statistics toolbox for use with matlab. Statistics Toolbox User's Guide.
- [70] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
- [71] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998.
- [72] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer, 2008.
- [73] C.C. Chang and C.J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [74] J.I. Arribas, V.D. Calhoun, T. Adali, et al. Automatic bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from fmri data. *Biomedical Engineering, IEEE Transactions on*, 57(12):2850–2860, 2010.
- [75] P.D. Wasserman. *Neural computing: Theory and Practice*. A van Nostrand Reinhold book. Van Nostrand Reinhold, 1989.
- [76] S.S. Haykin. *Neural Networks: A Comprehensive Foundation*. International edition. Prentice Hall, 1999.
- [77] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.
- [78] H. Demuth, M. Beale, and M. Hagan. Neural network toolbox 6. *User's guide*, 2008.
- [79] M.F. Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4):525–533, 1993.
- [80] S. Weerahandi. *Exact statistical methods for data analysis*. Springer Science & Business Media, 2013.
- [81] W.J. Krzanowski and D.J. Hand. *ROC Curves for Continuous Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2009.

-
- [82] C.E. Metz. Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier, 1978.
- [83] A.L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1):245–271, 1997.
- [84] R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.
- [85] N. Kwak and C.-H. Choi. Input feature selection for classification problems. *Neural Networks, IEEE Transactions on*, 13(1):143–159, 2002.
- [86] S. Bhattacharyya, A. Sengupta, T. Chakraborti, A. Konar, and D.N. Tibarewala. Automatic feature selection of motor imagery eeg signals using differential evolution and learning automata. *Medical & biological engineering & computing*, 52(2):131–139, 2014.
- [87] P. Rakshit, S. Bhattacharyya, A. Konar, A. Khasnobish, D.N. Tibarewala, and R. Janarthanan. Artificial bee colony based feature selection for motor imagery eeg data. In *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, pages 127–138. Springer, 2013.
- [88] L. Guo, D. Rivero, J. Dorado, C. R Munteanu, and A. Pazos. Automatic feature extraction using genetic programming: An application to epileptic eeg classification. *Expert Systems with Applications*, 38(8):10425–10436, 2011.

Apéndices

Apéndice A

Áreas bajo las curvas ROC para todos los procesados estudiados

En este apéndice se procede a mostrar las diferentes AUC producidas por los clasificadores del método de reconocimiento de patrones implementado en el presente trabajo para todos los tipos posibles de combinaciones de algoritmos en las diferentes etapas del mismo.

Agrupación	Procesamiento realizado			
	<i>15Hz-J5-MLP</i>	<i>35Hz-J5-MLP</i>	<i>15Hz-J5-SVM</i>	<i>35Hz-J5-SVM</i>
<i>Total</i>	0,9581	0,9293	0,9620	0,9582
<i>Central</i>	0,8296	0,7451	0,8096	0,7800
<i>Frontal</i>	0,9038	0,8670	0,9339	0,9562
<i>Occipital</i>	0,8113	0,6990	0,5910	0,7324
<i>Parietal-temporal</i>	0,9008	0,8219	0,9523	0,8015
<i>C3</i>	0,6214	0,5846	0,6944	0,5101
<i>C4</i>	0,6986	0,6125	0,6407	0,5811
<i>Cz</i>	0,6451	0,7098	0,6048	0,6723
<i>F3</i>	0,7354	0,7134	0,6759	0,6720
<i>F4</i>	0,7098	0,5695	0,7296	0,5695
<i>F7</i>	0,6544	0,7980	0,6146	0,7374
<i>F8</i>	0,6139	0,6387	0,7086	0,6098
<i>Fp1</i>	0,7292	0,5958	0,6874	0,6245
<i>Fp2</i>	0,6979	0,6207	0,6034	0,5062
<i>Fz</i>	0,6706	0,7393	0,6504	0,7392
<i>O1</i>	0,8145	0,6252	0,8281	0,5826
<i>O2</i>	0,7589	0,6182	0,6014	0,6771
<i>P3</i>	0,7021	0,6881	0,6528	0,6621
<i>P4</i>	0,7874	0,7617	0,7615	0,6532
<i>Pz</i>	0,6529	0,6200	0,5822	0,5404
<i>T5</i>	0,7823	0,6218	0,7802	0,5492
<i>T6</i>	0,7131	0,7352	0,6463	0,6415

Tabla A.1: AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (J5) y el tercero al tipo de clasificador utilizado (MLP o SVM).

Agrupación	Procesamiento realizado			
	15Hz-AG-MLP	35Hz-AG-MLP	15Hz-AG-SVM	35Hz-AG-SVM
<i>Total</i>	0,9300	0,8605	0,9589	0,8965
<i>Central</i>	0,7414	0,6685	0,7559	0,6444
<i>Frontal</i>	0,8462	0,8002	0,8639	0,8106
<i>Occipital</i>	0,8188	0,8130	0,9046	0,8280
<i>Parietal-temporal</i>	0,8837	0,8715	0,8663	0,8757
<i>C3</i>	0,6259	0,5556	0,6839	0,6179
<i>C4</i>	0,6549	0,6709	0,5886	0,6956
<i>Cz</i>	0,7022	0,6955	0,7229	0,7114
<i>F3</i>	0,7215	0,7643	0,6898	0,8022
<i>F4</i>	0,7073	0,5652	0,5518	0,5801
<i>F7</i>	0,6957	0,7555	0,6025	0,6907
<i>F8</i>	0,6016	0,6570	0,6996	0,6343
<i>Fp1</i>	0,7418	0,7197	0,6284	0,5723
<i>Fp2</i>	0,7150	0,6287	0,7103	0,5714
<i>Fz</i>	0,5616	0,6729	0,6285	0,6732
<i>O1</i>	0,8135	0,6335	0,8433	0,6041
<i>O2</i>	0,7420	0,6889	0,7450	0,6473
<i>P3</i>	0,7259	0,6508	0,6253	0,6886
<i>P4</i>	0,8394	0,8137	0,8386	0,7504
<i>Pz</i>	0,7022	0,6132	0,5834	0,5927
<i>T5</i>	0,7960	0,6793	0,8020	0,7048
<i>T6</i>	0,7204	0,7736	0,5925	0,6849

Tabla A.2: AUC obtenidas para los diferentes algoritmos de procesamiento de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (AG) y el tercero al tipo de clasificador utilizado (MLP o SVM).

Agrupación	Procesamiento realizado			
	<i>15Hz-MIFS-MLP</i>	<i>35Hz-MIFS-MLP</i>	<i>15Hz-MIFS-SVM</i>	<i>35Hz-MIFS-SVM</i>
<i>Total</i>	0,7603	0,7169	0,7056	0,6781
<i>Central</i>	0,5675	0,6289	0,6368	0,5431
<i>Frontal</i>	0,7177	0,7037	0,6508	0,5631
<i>Occipital</i>	0,6793	0,6872	0,6140	0,6263
<i>Parietal-temporal</i>	0,8409	0,7571	0,8438	0,6877
<i>C3</i>	0,5586	0,5750	0,7152	0,5804
<i>C4</i>	0,6571	0,6261	0,5447	0,5391
<i>Cz</i>	0,6088	0,6472	0,5762	0,6158
<i>F3</i>	0,6513	0,6541	0,5601	0,5550
<i>F4</i>	0,6494	0,5139	0,5610	0,6620
<i>F7</i>	0,6347	0,6918	0,5766	0,6077
<i>F8</i>	0,5645	0,6594	0,6437	0,6038
<i>Fp1</i>	0,7216	0,5429	0,6321	0,6995
<i>Fp2</i>	0,6984	0,5327	0,6821	0,5992
<i>Fz</i>	0,6491	0,6886	0,6377	0,7457
<i>O1</i>	0,6876	0,5935	0,6375	0,5288
<i>O2</i>	0,6608	0,6863	0,7093	0,6916
<i>P3</i>	0,6807	0,6819	0,5999	0,7453
<i>P4</i>	0,8178	0,7547	0,7797	0,6217
<i>Pz</i>	0,6178	0,6153	0,5875	0,5840
<i>T5</i>	0,7586	0,6278	0,7473	0,5523
<i>T6</i>	0,5716	0,5789	0,5571	0,5697

Tabla A.3: AUC obtenidas para los diferentes algoritmos de procesamiento de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (MIFS) y el tercero al tipo de clasificador utilizado (MLP o SVM).

Agrupación	Procesamiento realizado			
	15Hz-DISR-MLP	35Hz-DISR-MLP	15Hz-DISR-SVM	35Hz-DISR-SVM
<i>Total</i>	0,7495	0,6006	0,7942	0,5276
<i>Central</i>	0,5390	0,6086	0,6203	0,5516
<i>Frontal</i>	0,6519	0,6417	0,5363	0,5403
<i>Occipital</i>	0,7389	0,5206	0,7627	0,5825
<i>Parietal-temporal</i>	0,7645	0,6914	0,7801	0,6254
<i>C3</i>	0,5663	0,5933	0,6762	0,6110
<i>C4</i>	0,5984	0,5986	0,6051	0,6678
<i>Cz</i>	0,5562	0,5760	0,5368	0,5619
<i>F3</i>	0,6186	0,7112	0,6379	0,7215
<i>F4</i>	0,6458	0,5550	0,6119	0,5310
<i>F7</i>	0,6839	0,7780	0,6430	0,6793
<i>F8</i>	0,5146	0,6023	0,6529	0,5144
<i>Fp1</i>	0,6385	0,5905	0,6317	0,5642
<i>Fp2</i>	0,6292	0,5438	0,7331	0,5898
<i>Fz</i>	0,5360	0,5999	0,6368	0,6455
<i>O1</i>	0,7870	0,5571	0,7770	0,5936
<i>O2</i>	0,6821	0,5278	0,5570	0,6522
<i>P3</i>	0,5487	0,5808	0,6365	0,6851
<i>P4</i>	0,7371	0,6441	0,7008	0,6250
<i>Pz</i>	0,6547	0,6322	0,6062	0,5775
<i>T5</i>	0,6657	0,5790	0,6911	0,5905
<i>T6</i>	0,6610	0,6478	0,5453	0,7309

Tabla A.4: AUC obtenidas para los diferentes algoritmos de procesado de la señal de EEG en las distintas agrupaciones de electrodos. La primera columna describe la agrupación de características que se utiliza para la clasificación en cada caso. El resto de columnas hacen referencia a los distintos algoritmos utilizados en las diferentes etapas del método de reconocimiento de patrones: el primer número se refiere al filtrado utilizado (15Hz o 35Hz), el segundo al algoritmo de selección de características (DISR) y el tercero al tipo de clasificador utilizado (MLP o SVM).

Apéndice B

Resultados completos del análisis estadístico

En este apéndice se procede a la descripción de los resultados proporcionados por el análisis estadístico llevado a cabo en el presente estudio mediante el test U para las características extraídas de la señal de EEG filtrada a 15Hz y a 35Hz. Las siguientes tablas muestran los p-valores generados por el test-U para cada una de las 340 características extraídas de la señal de EEG en cada filtrado. Cabe destacar que se mostrará también el valor h devuelto por el test no paramétrico utilizado. Este valor h es 0 si el p-valor es mayor de 0.05 y, por lo tanto, no puede considerarse que existan diferencias significativas para ambas poblaciones utilizando la característica seleccionada para la realización del test U; y es 1 cuando sucede lo contrario.

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	<i>p-valor</i>	<i>h-valor</i>	<i>p-valor</i>	<i>h-valor</i>
<i>varianza_Fp1</i>	0,74477	0	0,38738	0
<i>varianza_Fp2</i>	0,55187	0	0,79626	0
<i>varianza_F3</i>	0,97313	0	0,25688	0
<i>varianza_F4</i>	0,15397	0	0,67789	0
<i>varianza_C3</i>	0,72784	0	0,24758	0
<i>varianza_C4</i>	0,42544	0	0,37517	0
<i>varianza_P3</i>	0,95524	0	0,56698	0
<i>varianza_P4</i>	0,90173	0	0,17437	0
<i>varianza_O1</i>	0,45198	0	0,81364	0
<i>varianza_O2</i>	0,62931	0	0,97313	0
<i>varianza_F7</i>	0,81364	0	0,30700	0
<i>varianza_F8</i>	0,53696	0	0,07428	0
<i>varianza_T5</i>	0,72784	0	0,43859	0
<i>varianza_T6</i>	0,91953	0	0,17437	0
<i>varianza_Fz</i>	0,37517	0	0,95524	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	<i>p-valor</i>	<i>h-valor</i>	<i>p-valor</i>	<i>h-valor</i>
<i>varianza_Pz</i>	0,22969	0	0,27620	0
<i>varianza_Cz</i>	0,72784	0	0,24758	0
<i>media_Fp1</i>	0,08186	0	0,08186	0
<i>media_Fp2</i>	0,15397	0	0,58228	0
<i>media_F3</i>	0,03027	1	0,03027	1
<i>media_F4</i>	0,16056	0	0,84865	0
<i>media_C3</i>	0,86628	0	0,50777	0
<i>media_C4</i>	0,88397	0	0,84865	0
<i>media_P3</i>	0,06399	0	0,10358	0
<i>media_P4</i>	0,00519	1	0,07428	0
<i>media_O1</i>	0,21275	0	0,22110	0
<i>media_O2</i>	0,09006	0	0,35148	0
<i>media_F7</i>	0,14759	0	0,08588	0
<i>media_F8</i>	0,84865	0	0,72784	0
<i>media_T5</i>	0,66153	0	0,33999	0
<i>media_T6</i>	0,55187	0	0,62931	0
<i>media_Fz</i>	0,26642	0	0,15397	0
<i>media_Pz</i>	0,05491	0	0,33999	0
<i>media_Cz</i>	0,37517	0	0,20462	0
<i>P300ValorPico_Fp1</i>	0,01154	1	0,25688	0
<i>P300ValorPico_Fp2</i>	0,19672	0	0,30170	0
<i>P300ValorPico_F3</i>	0,83110	0	0,90173	0
<i>P300ValorPico_F4</i>	0,81364	0	0,59777	0
<i>P300ValorPico_C3</i>	0,53696	0	0,95524	0
<i>P300ValorPico_C4</i>	0,17437	0	0,11346	0
<i>P300ValorPico_P3</i>	0,29649	0	0,37517	0
<i>P300ValorPico_P4</i>	0,14141	0	0,23852	0
<i>P300ValorPico_O1</i>	0,72784	0	0,67789	0
<i>P300ValorPico_O2</i>	0,90173	0	0,88397	0
<i>P300ValorPico_F7</i>	0,03027	1	0,49349	0
<i>P300ValorPico_F8</i>	0,81364	0	0,79626	0
<i>P300ValorPico_T5</i>	0,81364	0	0,74477	0
<i>P300ValorPico_T6</i>	0,58228	0	0,72784	0
<i>P300ValorPico_Fz</i>	0,90173	0	0,58228	0
<i>P300ValorPico_Pz</i>	0,02549	1	0,11346	0
<i>P300ValorPico_Cz</i>	0,35148	0	0,67789	0
<i>P300ValorPicoAbs_Fp1</i>	0,64534	0	0,41252	0
<i>P300ValorPicoAbs_Fp2</i>	1,00000	0	0,97313	0
<i>P300ValorPicoAbs_F3</i>	0,39983	0	0,15397	0
<i>P300ValorPicoAbs_F4</i>	0,28623	0	0,97313	0
<i>P300ValorPicoAbs_C3</i>	0,59777	0	0,77899	0
<i>P300ValorPicoAbs_C4</i>	0,61345	0	0,55187	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	p-valor	h-valor	p-valor	h-valor
<i>P300ValorPicoAbs_P3</i>	0,39983	0	0,31775	0
<i>P300ValorPicoAbs_P4</i>	0,16056	0	0,18160	0
<i>P300ValorPicoAbs_O1</i>	0,91953	0	0,67789	0
<i>P300ValorPicoAbs_O2</i>	0,79626	0	0,91953	0
<i>P300ValorPicoAbs_F7</i>	0,49349	0	0,47943	0
<i>P300ValorPicoAbs_F8</i>	0,66153	0	0,16735	0
<i>P300ValorPicoAbs_T5</i>	0,41252	0	0,38738	0
<i>P300ValorPicoAbs_T6</i>	0,46559	0	0,79626	0
<i>P300ValorPicoAbs_Fz</i>	0,50777	0	0,91953	0
<i>P300ValorPicoAbs_Pz</i>	0,03387	1	0,18905	0
<i>P300ValorPicoAbs_Cz</i>	0,86628	0	0,58228	0
<i>P300Media_Fp1</i>	0,14759	0	0,16735	0
<i>P300Media_Fp2</i>	0,39983	0	0,58228	0
<i>P300Media_F3</i>	0,55187	0	0,59777	0
<i>P300Media_F4</i>	0,45198	0	0,67789	0
<i>P300Media_C3</i>	0,74477	0	0,71105	0
<i>P300Media_C4</i>	0,18160	0	0,33999	0
<i>P300Media_P3</i>	0,31775	0	0,22110	0
<i>P300Media_P4</i>	0,03995	1	0,14141	0
<i>P300Media_O1</i>	0,59777	0	0,91953	0
<i>P300Media_O2</i>	0,93737	0	0,86628	0
<i>P300Media_F7</i>	0,02405	1	0,09006	0
<i>P300Media_F8</i>	0,81364	0	0,46559	0
<i>P300Media_T5</i>	0,64534	0	0,90173	0
<i>P300Media_T6</i>	0,29649	0	0,76182	0
<i>P300Media_Fz</i>	0,59777	0	0,83110	0
<i>P300Media_Pz</i>	0,04693	1	0,13543	0
<i>P300Media_Cz</i>	0,42544	0	0,90173	0
<i>energia_Fp1</i>	0,88397	0	0,55187	0
<i>energia_Fp2</i>	0,61345	0	0,93737	0
<i>energia_F3</i>	0,50777	0	0,18160	0
<i>energia_F4</i>	0,25688	0	0,77899	0
<i>energía_C3</i>	0,72784	0	0,59777	0
<i>energia_C4</i>	0,45198	0	0,32875	0
<i>energia_P3</i>	0,66153	0	0,29649	0
<i>energia_P4</i>	0,37517	0	0,09891	0
<i>energia_O1</i>	0,91953	0	0,79626	0
<i>energia_O2</i>	0,81364	0	0,83110	0
<i>energia_F7</i>	0,47943	0	0,13543	0
<i>energia_F8</i>	0,55187	0	0,04947	1
<i>energia_T5</i>	0,83110	0	0,36320	0
<i>energia_T6</i>	0,81364	0	0,18160	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	p-valor	h-valor	p-valor	h-valor
<i>energia_Fz</i>	0,35148	0	0,77899	0
<i>energia_Pz</i>	0,18160	0	0,29649	0
<i>energia_Cz</i>	0,81364	0	0,95524	0
<i>P300Latencia_Fp1</i>	0,65820	0	0,90120	0
<i>P300Latencia_Fp2</i>	0,91912	0	0,63488	0
<i>P300Latencia_F3</i>	0,29934	0	0,85677	0
<i>P300Latencia_F4</i>	0,63665	0	0,55128	0
<i>P300Latencia_C3</i>	0,68586	0	0,23286	0
<i>P300Latencia_C4</i>	0,94624	0	0,66948	0
<i>P300Latencia_P3</i>	0,23383	0	0,71927	0
<i>P300Latencia_P4</i>	0,10498	0	0,38104	0
<i>P300Latencia_O1</i>	0,08766	0	0,78752	0
<i>P300Latencia_O2</i>	0,65106	0	0,35699	0
<i>P300Latencia_F7</i>	0,98195	0	0,00919	1
<i>P300Latencia_F8</i>	0,99091	0	0,66752	0
<i>P300Latencia_T5</i>	0,24248	0	0,98208	0
<i>P300Latencia_T6</i>	0,94603	0	0,43741	0
<i>P300Latencia_Fz</i>	0,81332	0	0,45812	0
<i>P300Latencia_Pz</i>	0,66949	0	0,36290	0
<i>P300Latencia_Cz</i>	0,44297	0	0,05225	0
<i>P300LatenciaAbs_Fp1</i>	0,19651	0	0,43167	0
<i>P300LatenciaAbs_Fp2</i>	0,38707	0	0,36270	0
<i>P300LatenciaAbs_F3</i>	0,02051	1	0,02217	1
<i>P300LatenciaAbs_F4</i>	0,38549	0	0,46536	0
<i>P300LatenciaAbs_C3</i>	0,84733	0	0,13689	0
<i>P300LatenciaAbs_C4</i>	0,21628	0	0,77030	0
<i>P300LatenciaAbs_P3</i>	0,20438	0	0,83979	0
<i>P300LatenciaAbs_P4</i>	0,21510	0	0,18508	0
<i>P300LatenciaAbs_O1</i>	0,12943	0	0,28089	0
<i>P300LatenciaAbs_O2</i>	0,08293	0	0,68577	0
<i>P300LatenciaAbs_F7</i>	0,46546	0	0,78733	0
<i>P300LatenciaAbs_F8</i>	0,35692	0	0,73617	0
<i>P300LatenciaAbs_T5</i>	0,56653	0	0,59751	0
<i>P300LatenciaAbs_T6</i>	0,00974	1	0,00314	1
<i>P300LatenciaAbs_Fz</i>	0,93730	0	0,13209	0
<i>P300LatenciaAbs_Pz</i>	0,39328	0	0,20425	0
<i>P300LatenciaAbs_Cz</i>	0,97302	0	0,73592	0
<i>LAR_Fp1</i>	0,09006	0	0,28623	0
<i>LAR_Fp2</i>	0,18160	0	0,31775	0
<i>LAR_F3</i>	0,55187	0	0,97313	0
<i>LAR_F4</i>	0,50777	0	0,90173	0
<i>LAR_C3</i>	0,16056	0	0,35148	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	<i>p</i> -valor	<i>h</i> -valor	<i>p</i> -valor	<i>h</i> -valor
<i>LAR_C4</i>	0,53696	0	0,29649	0
<i>LAR_P3</i>	0,66153	0	0,22110	0
<i>LAR_P4</i>	0,07071	0	0,07800	0
<i>LAR_O1</i>	0,93737	0	0,97313	0
<i>LAR_O2</i>	0,76182	0	0,52226	0
<i>LAR_F7</i>	0,18160	0	0,25688	0
<i>LAR_F8</i>	0,35148	0	0,22969	0
<i>LAR_T5</i>	0,61345	0	0,56698	0
<i>LAR_T6</i>	0,81364	0	0,39983	0
<i>LAR_Fz</i>	0,99104	0	0,93737	0
<i>LAR_Pz</i>	0,02014	1	0,13543	0
<i>LAR_Cz</i>	0,13543	0	0,41252	0
<i>LARAbs_Fp1</i>	0,74477	0	0,30700	0
<i>LARAbs_Fp2</i>	0,81364	0	0,72784	0
<i>LARAbs_F3</i>	0,22969	0	0,01785	1
<i>LARAbs_F4</i>	0,32875	0	0,77899	0
<i>LARAbs_C3</i>	0,61345	0	0,36320	0
<i>LARAbs_C4</i>	0,76182	0	0,93737	0
<i>LARAbs_P3</i>	0,21275	0	0,21275	0
<i>LARAbs_P4</i>	0,06084	0	0,04450	1
<i>LARAbs_O1</i>	0,53696	0	0,77899	0
<i>LARAbs_O2</i>	0,58228	0	0,64534	0
<i>LARAbs_F7</i>	0,45198	0	0,26642	0
<i>LARAbs_F8</i>	0,81364	0	0,04693	1
<i>LARAbs_T5</i>	0,55187	0	0,23852	0
<i>LARAbs_T6</i>	0,19672	0	0,11867	0
<i>LARAbs_Fz</i>	0,49349	0	0,36320	0
<i>LARAbs_Pz</i>	0,01580	1	0,04217	1
<i>LARAbs_Cz</i>	0,84865	0	0,38738	0
<i>Apositiva_Fp1</i>	0,00526	1	0,26831	0
<i>Apositiva_Fp2</i>	0,16383	0	0,38434	0
<i>Apositiva_F3</i>	0,90149	0	0,64418	0
<i>Apositiva_F4</i>	0,93735	0	0,86620	0
<i>Apositiva_C3</i>	0,32870	0	0,99104	0
<i>Apositiva_C4</i>	0,37517	0	0,32321	0
<i>Apositiva_P3</i>	0,24758	0	0,22110	0
<i>Apositiva_P4</i>	0,04450	1	0,14141	0
<i>Apositiva_O1</i>	0,72784	0	0,85742	0
<i>Apositiva_O2</i>	0,84865	0	0,86628	0
<i>Apositiva_F7</i>	0,02821	1	0,50629	0
<i>Apositiva_F8</i>	0,97312	0	0,56635	0
<i>Apositiva_T5</i>	0,91953	0	0,62921	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	p-valor	h-valor	p-valor	h-valor
<i>Apositiva_T6</i>	0,42543	0	0,86628	0
<i>Apositiva_Fz</i>	1,00000	0	0,67739	0
<i>Apositiva_Pz</i>	0,04450	1	0,15722	0
<i>Apositiva_Cz</i>	0,22969	0	0,74477	0
<i>Anegativa_Fp1</i>	0,14141	0	0,23852	0
<i>Anegativa_Fp2</i>	0,55939	0	0,62931	0
<i>Anegativa_F3</i>	0,58228	0	0,46559	0
<i>Anegativa_F4</i>	0,49349	0	0,84865	0
<i>Anegativa_C3</i>	0,97313	0	0,79624	0
<i>Anegativa_C4</i>	0,22036	0	0,35120	0
<i>Anegativa_P3</i>	0,37119	0	0,16191	0
<i>Anegativa_P4</i>	0,13916	0	0,12740	0
<i>Anegativa_O1</i>	0,77886	0	0,91940	0
<i>Anegativa_O2</i>	0,58167	0	0,78710	0
<i>Anegativa_F7</i>	0,08186	0	0,11867	0
<i>Anegativa_F8</i>	0,76182	0	0,31775	0
<i>Anegativa_T5</i>	0,77886	0	0,90171	0
<i>Anegativa_T6</i>	0,51456	0	0,71089	0
<i>Anegativa_Fz</i>	0,50777	0	0,97313	0
<i>Anegativa_Pz</i>	0,13358	0	0,70810	0
<i>Anegativa_Cz</i>	0,55176	0	0,91944	0
<i>Atotal_Fp1</i>	0,14759	0	0,16735	0
<i>Atotal_Fp2</i>	0,39983	0	0,58228	0
<i>Atotal_F3</i>	0,55187	0	0,59777	0
<i>Atotal_F4</i>	0,45198	0	0,67789	0
<i>Atotal_C3</i>	0,74477	0	0,71105	0
<i>Atotal_C4</i>	0,18160	0	0,33999	0
<i>Atotal_P3</i>	0,31775	0	0,22110	0
<i>Atotal_P4</i>	0,03995	1	0,14141	0
<i>Atotal_O1</i>	0,59777	0	0,91953	0
<i>Atotal_O2</i>	0,93737	0	0,86628	0
<i>Atotal_F7</i>	0,02405	1	0,09006	0
<i>Atotal_F8</i>	0,81364	0	0,46559	0
<i>Atotal_T5</i>	0,64534	0	0,90173	0
<i>Atotal_T6</i>	0,29649	0	0,76182	0
<i>Atotal_Fz</i>	0,59777	0	0,83110	0
<i>Atotal_Pz</i>	0,04693	1	0,13543	0
<i>Atotal_Cz</i>	0,42544	0	0,90173	0
<i>ATAR_Fp1</i>	0,09891	0	0,09440	0
<i>ATAR_Fp2</i>	0,50777	0	0,77899	0
<i>ATAR_F3</i>	0,09891	0	0,90173	0
<i>ATAR_F4</i>	0,20462	0	0,58228	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	p-valor	h-valor	p-valor	h-valor
ATAR_C3	0,99104	0	0,14759	0
ATAR_C4	0,39983	0	0,43859	0
ATAR_P3	0,45198	0	0,31775	0
ATAR_P4	0,04217	1	0,09006	0
ATAR_O1	0,46559	0	0,55187	0
ATAR_O2	0,21275	0	0,62931	0
ATAR_F7	0,03202	1	0,22969	0
ATAR_F8	0,49349	0	0,20462	0
ATAR_T5	0,50777	0	0,42544	0
ATAR_T6	0,38738	0	0,69439	0
ATAR_Fz	0,25688	0	0,10843	0
ATAR_Pz	0,07428	0	0,25688	0
ATAR_Cz	0,36320	0	0,17437	0
TAAR_Fp1	0,39983	0	0,17437	0
TAAR_Fp2	0,93737	0	0,64534	0
TAAR_F3	0,36320	0	0,58228	0
TAAR_F4	0,08588	0	0,99104	0
TAAR_C3	0,74477	0	0,59777	0
TAAR_C4	0,33999	0	0,52226	0
TAAR_P3	0,36320	0	0,25688	0
TAAR_P4	0,03580	1	0,16056	0
TAAR_O1	0,86628	0	1,00000	0
TAAR_O2	0,53696	0	1,00000	0
TAAR_F7	0,55187	0	0,21275	0
TAAR_F8	0,64534	0	0,17437	0
TAAR_T5	0,31775	0	0,39983	0
TAAR_T6	0,88397	0	0,61345	0
TAAR_Fz	0,39983	0	0,28623	0
TAAR_Pz	0,04693	1	0,23852	0
TAAR_Cz	0,26642	0	0,58228	0
ZC_Fp1	0,00261	1	0,05617	0
ZC_Fp2	0,42002	0	0,04667	1
ZC_F3	0,08892	0	0,65841	0
ZC_F4	0,03419	1	0,83716	0
ZC_C3	0,58444	0	0,44551	0
ZC_C4	0,32298	0	0,64044	0
ZC_P3	0,74317	0	0,11252	0
ZC_P4	0,12734	0	0,04866	1
ZC_O1	0,98169	0	0,92747	0
ZC_O2	0,11979	0	0,98196	0
ZC_F7	0,44464	0	0,08224	0
ZC_F8	0,31527	0	0,37112	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	<i>p</i> -valor	<i>h</i> -valor	<i>p</i> -valor	<i>h</i> -valor
<i>ZC_T5</i>	0,34836	0	0,28663	0
<i>ZC_T6</i>	0,26438	0	0,15575	0
<i>ZC_Fz</i>	0,93424	0	0,93651	0
<i>ZC_Pz</i>	0,31900	0	0,77397	0
<i>ZC_Cz</i>	0,00992	1	0,13470	0
<i>PSE_Fp1</i>	0,05781	0	0,04450	1
<i>PSE_Fp2</i>	0,36320	0	0,53696	0
<i>PSE_F3</i>	0,01311	1	0,03202	1
<i>PSE_F4</i>	0,72784	0	0,43859	0
<i>PSE_C3</i>	0,90173	0	0,61345	0
<i>PSE_C4</i>	0,47943	0	0,14141	0
<i>PSE_P3</i>	0,20462	0	0,16056	0
<i>PSE_P4</i>	0,00020	1	0,00041	1
<i>PSE_O1</i>	0,07428	0	0,23852	0
<i>PSE_O2</i>	0,07428	0	0,46559	0
<i>PSE_F7</i>	0,03202	1	0,00595	1
<i>PSE_F8</i>	0,84865	0	0,00834	1
<i>PSE_T5</i>	0,00891	1	0,03782	1
<i>PSE_T6</i>	0,25688	0	0,22969	0
<i>PSE_Fz</i>	0,53696	0	0,49349	0
<i>PSE_Pz</i>	0,03995	1	0,04693	1
<i>PSE_Cz</i>	0,45198	0	0,58228	0
<i>fmediana_Fp1</i>	0,06646	0	0,15985	0
<i>fmediana_Fp2</i>	0,38355	0	0,27559	0
<i>fmediana_F3</i>	0,45610	0	0,05310	0
<i>fmediana_F4</i>	0,33857	0	0,53295	0
<i>fmediana_C3</i>	0,53296	0	0,83842	0
<i>fmediana_C4</i>	0,90943	0	0,48077	0
<i>fmediana_P3</i>	0,58906	0	0,30756	0
<i>fmediana_P4</i>	0,00835	1	0,01001	1
<i>fmediana_O1</i>	0,72529	0	0,54238	0
<i>fmediana_O2</i>	0,69984	0	0,72686	0
<i>fmediana_F7</i>	0,41080	0	0,27917	0
<i>fmediana_F8</i>	0,83737	0	0,21625	0
<i>fmediana_T5</i>	0,75139	0	0,03951	1
<i>fmediana_T6</i>	0,44989	0	0,27348	0
<i>fmediana_Fz</i>	0,20259	0	0,12088	0
<i>fmediana_Pz</i>	0,06840	0	0,19912	0
<i>fmediana_Cz</i>	0,46345	0	0,45401	0
<i>fmedia_Fp1</i>	0,05781	0	0,10843	0
<i>fmedia_Fp2</i>	0,31775	0	0,41252	0
<i>fmedia_F3</i>	0,04450	1	0,05781	0

Continúa en la siguiente página

Característica	Filtrado de 15Hz		Filtrado de 35Hz	
	<i>p</i> -valor	<i>h</i> -valor	<i>p</i> -valor	<i>h</i> -valor
<i>fmedia_F4</i>	0,67789	0	0,43859	0
<i>fmedia_C3</i>	0,62931	0	0,77899	0
<i>fmedia_C4</i>	0,62931	0	0,46559	0
<i>fmedia_P3</i>	0,13543	0	0,06728	0
<i>fmedia_P4</i>	0,00093	1	0,00364	1
<i>fmedia_O1</i>	0,11867	0	0,22110	0
<i>fmedia_O2</i>	0,45198	0	0,69439	0
<i>fmedia_F7</i>	0,05781	0	0,04217	1
<i>fmedia_F8</i>	0,43859	0	0,00780	1
<i>fmedia_T5</i>	0,01897	1	0,04217	1
<i>fmedia_T6</i>	0,32875	0	0,18160	0
<i>fmedia_Fz</i>	0,42544	0	0,66153	0
<i>fmedia_Pz</i>	0,07428	0	0,07800	0
<i>fmedia_Cz</i>	0,67789	0	0,90173	0
<i>fmoda_Fp1</i>	0,93942	0	0,89592	0
<i>fmoda_Fp2</i>	0,66244	0	0,87902	0
<i>fmoda_F3</i>	0,49731	0	0,42578	0
<i>fmoda_F4</i>	0,92356	0	0,96058	0
<i>fmoda_C3</i>	0,76059	0	0,88080	0
<i>fmoda_C4</i>	0,94177	0	0,68402	0
<i>fmoda_P3</i>	0,42394	0	0,01207	1
<i>fmoda_P4</i>	0,00428	1	0,19269	0
<i>fmoda_O1</i>	0,55387	0	0,35827	0
<i>fmoda_O2</i>	0,30098	0	0,06483	0
<i>fmoda_F7</i>	0,11622	0	0,05052	0
<i>fmoda_F8</i>	0,79660	0	0,04545	1
<i>fmoda_T5</i>	0,86190	0	0,88920	0
<i>fmoda_T6</i>	0,81331	0	1,00000	0
<i>fmoda_Fz</i>	0,60269	0	0,21262	0
<i>fmoda_Pz</i>	0,08165	0	0,22686	0
<i>fmoda_Cz</i>	0,34634	0	0,00591	1

Tabla B.1: Resultados del análisis estadístico obtenidos para los diferentes filtrados realizados en el pre-procesado. La primera columna describe la característica que se utiliza para realizar el test U: el primer término hace referencia a la característica y el segundo al electrodo del que se extrae la mismas. El resto de columnas hacen referencia a los distintos *p*-valores y *h*-valores para el filtrado de 15Hz y el de 35Hz.

Apéndice C

Parámetros de las señales de EEG incluidos en cada matriz de características

En este apéndice se procede a mostrar las características seleccionadas por los diferentes algoritmos de *feature selection* según el filtrado realizado en el pre-procesamiento para las diferentes agrupaciones de características. Cabe destacar que, para los criterios J5, MIFS y DISR, éstas características están ordenadas por relevancia, de esta forma la característica 1 es la primera que selecciona al algoritmo de selección. Por el contrario, para los AG, el no poseer estos conceptos de relevancia, se han expuesto en orden aleatorio.

Agrupación	Característica									
	1	2	3	4	5	6	7	8	9	10
Total	PSE (P4)	Atotal (F4)	fmediana (Cz)	PSE (T5)	P300Latencia (O1)	P300LatenciaAbs (P3)	P300Latencia (C3)	LARAbs (T3)	P300LatenciaAbs (C4)	LARAbs (F4)
Central	ZC (Fp1)	Anegativa (C4)	LAR (Cz)	media (Cz)	fmoda (C3)	ZC (C4)	PSE (C4)	PSE (C3)	LARAbs (C3)	ATAR (C3)
Frontal	ZC (Fp1)	varianza (Fp2)	P300LatenciaAbs (F3)	media (F4)	ATAR (F7)	ZC (F3)	P300Media (F7)	P300LatenciaAbs (F3)	fmoda (Fp2)	P300Latencia (Fp1)
Occipital	PSE (O1)	P300Latencia (O1)	P300LatenciaAbs (O1)	media (O2)	P300ValorPico (O2)	TAAR (O2)	energía (O2)	fmediana (O2)	P300LatenciaAbs (O2)	ATAR (O2)
Parietal-temporal	PSE (P4)	Apositiva (T6)	PSE (T5)	ATAR (T5)	LARAbs (P2)	PSE (P2)	LAR (T5)	LAR (P3)	fmediana (T6)	media (T5)
C3	LAR	fmoda	ATAR	LARAbs	fmediana	fmoda	P300LatenciaAbs	P300Latencia	media	PSE
C4	Anegativa	energía	media	fmoda	varianza	P300LatenciaAbs	PSE	ZC	P300ValorPicoAbs	P300ValorPico
Cz	ZC	LAR	media	varianza	energía	PSE	Apositiva	fmediana	fmoda	Anegativa
F3	P300LatenciaAbs	LAR	media	P300Latencia	ZC	P300ValorPico	P300ValorPicoAbs	ATAR	TAAR	fmediana
F4	varianza	media	P300Latencia	Anegativa	P300LatenciaAbs	LAR	ZC	fmoda	energía	P300ValorPicoAbs
F7	P300ValorPico	PSE	LAR	Apositiva	P300Latencia	fmediana	fmoda	energía	LARAbs	media
F8	fmoda	energía	P300LatenciaAbs	P300ValorPico	Apositiva	PSE	P300Latencia	varianza	ZC	media
Fp1	ZC	varianza	ATAR	TAAR	fmediana	P300ValorPicoAbs	fmoda	Atotal	P300LatenciaAbs	LAR
Fp2	varianza	media	ZC	PSE	fmoda	Apositiva	P300ValorPicoAbs	P300Media	ATAR	P300Latencia
Fz	fmediana	media	P300Latencia	P300Media	ZC	LAR	PSE	fmoda	energía	LARAbs
O1	PSE	P300Latencia	P300LatenciaAbs	TAAR	media	fmoda	ZC	fmediana	fmediana	LAR
O2	PSE	P300LatenciaAbs	media	P300ValorPico	TAAR	P300Latencia	LARAbs	energía	varianza	ZC
P3	media	varianza	fmoda	P300ValorPicoAbs	LAR	P300LatenciaAbs	energía	Atotal	ATAR	P300ValorPico
P4	PSE	varianza	LARAbs	fmoda	ZC	P300LatenciaAbs	media	energía	LAR	fmediana
Pz	LARAbs	varianza	P300ValorPico	P300LatenciaAbs	TAAR	ATAR	Anegativa	P300ValorPicoAbs	fmoda	fmediana
T5	PSE	Apositiva	P300Latencia	P300ValorPico	Atotal	P300ValorPicoAbs	fmediana	fmediana	ZC	ATAR
T6	P300LatenciaAbs	P300ValorPico	media	P300ValorPicoAbs	ZC	LAR	Apositiva	energía	LARAbs	varianza

Tabla C.1: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, algoritmo J5).

Agrupación	Característica									
	1	2	3	4	5	6	7	8	9	10
Total	PSE (P4)	fmoda (O2)	fmediana (T5)	P300LatenciaAbs (T6)	LAR (Fp2)	Mediana Frec. (F4)	ATAR (T5)	fmoda (T6)	PSE (F4)	P300Latencia (F4)
Central	fmoda (Cz)	Anegativa (C4)	PSE (C4)	media (Cz)	PSE (C3)	fmediana (C2)	P300Media (C3)	fmoda (C3)	fmoda (C3)	ZC (C3)
Frontal	P300LatenciaAbs (F7)	ZC (Fp1)	P300Latencia (Fp2)	PSE (F8)	energía (Fp2)	varianza (Fp1)	P300ValorPico (F3)	LARAbs (Fp2)	fmoda (F8)	PSE (Fp1)
Occipital	fmoda (O2)	fmedia (O1)	fmoda (O1)	media (O2)	P300ValorPico (O2)	LARAbs (O2)	ZC (O2)	PSE (O2)	LAR (O2)	LAR (O1)
Parietal-temporal	PSE (P4)	P300LatenciaAbs (T6)	fmoda (T5)	ZC (Pz)	Atotal (Pz)	P300LatenciaAbs (T6)	fmoda (T6)	PSE (T5)	ZC (P4)	fmediana (Pz)
C3	ATAR	P300ValorPicoAbs	P300Latencia	P300ValorPico	LAR	Apositiva	Atotal	LARAbs	PSE	media
C3	ATAR	P300ValorPicoAbs	P300Latencia	P300ValorPico	LAR	Apositiva	Atotal	LARAbs	PSE	media
C4	LAR	Anegativa	fmoda	fmediana	energía	P300ValorPicoAbs	P300ValorPico	ZC	media	LARAbs
Cz	fmoda	varianza	energía	ATAR	fmedia	fmediana	PSE	media	P300Latencia	LARAbs
F3	fmedia	P300Latencia	P300LatenciaAbs	media	P300ValorPico	LAR	PSE	P300ValorPicoAbs	Anegativa	ZC
F4	LAR	P300Latencia	PSE	fmediana	fmedia	ATAR	TAAR	ZC	P300LatenciaAbs	fmoda
F7	P300Latencia	PSE	P300LatenciaAbs	LAR	fmediana	fmedia	media	varianza	energía	LARAbs
F8	PSE	media	LAR	ZC	LARAbs	P300Latencia	energía	ATAR	P300ValorPicoAbs	P300LatenciaAbs
Fp1	ZC	P300LatenciaAbs	media	P300ValorPicoAbs	P300ValorPico	ATAR	LARAbs	varianza	fmoda	PSE
Fp2	ZC	varianza	LAR	fmoda	LARAbs	P300ValorPicoAbs	TAAR	PSE	P300Latencia	fmedia
Fz	fmoda	P300LatenciaAbs	varianza	media	energía	P300ValorPico	Apositiva	fmediana	fmedia	PSE
O1	fmedia	fmoda	P300LatenciaAbs	media	P300ValorPico	varianza	LARAbs	P300Media	ATAR	P300ValorPicoAbs
O2	fmoda	PSE	ZC	P300ValorPico	LARAbs	media	P300LatenciaAbs	ATAR	varianza	P300Latencia
P3	fmoda	fmedia	LAR	media	energía	Anegativa	PSE	fmediana	P300Media	ATAR
P4	PSE	LAR	LARAbs	energía	varianza	media	ZC	P300ValorPico	TAAR	fmoda
Pz	PSE	ZC	LAR	LARAbs	media	TAAR	P300Latencia	energía	Apositiva	ATAR
T5	PSE	LAR	fmediana	LARAbs	P300ValorPicoAbs	ATAR	ZC	TAAR	P300LatenciaAbs	P300Latencia
T6	P300LatenciaAbs	varianza	P300ValorPicoAbs	LARAbs	ZC	ATAR	TAAR	P300Latencia	fmediana	PSE

Tabla C.2: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, algoritmo J5).

Agrupación	Característica									
	1	2	3	4	5	6	7	8	9	10
Total	energía (F3)	PSE (Fp1)	PSE (Fp2)	PSE (F3)	PSE (F4)	PSE (C3)	PSE (C4)	PSE (P3)	PSE (P4)	PSE (O1)
Central	energía (C3)	PSE (C3)	PSE (C4)	PSE (Cz)	varianza (C4)	media (Cz)	varianza (C3)	media (C3)	P300Media (C4)	fmoda (C3)
Frontal	energía (F3)	PSE (Fp1)	PSE (Fp2)	PSE (F3)	PSE (F4)	PSE (F7)	PSE (F8)	PSE (Fz)	media (F4)	fmoda (F7)
Occipital	energía (O2)	PSE (O1)	PSE (O2)	media (O2)	media (O1)	varianza (O2)	fmoda (O1)	fmoda (O2)	P300Media (O2)	P300Media (O1)
Parietal-temporal	energía (P4)	PSE (P3)	PSE (P4)	PSE (Pz)	PSE (T5)	PSE (T6)	fmoda (P3)	media (P4)	fmoda (P4)	fmediana (Pz)
C3	energía	PSE	varianza	media	fmoda	P300ValorPico	ZC	P300Media	P300ValorPicoP300ValorPicoAbs	fmediana
C4	LAR	PSE	varianza	media	fmoda	P300Media	P300ValorPicoAbs	P300ValorPico	ZC	fmedia
Cz	energía	PSE	media	fmoda	varianza	ZC	P300ValorPicoAbs	fmediana	P300Media	fmedia
F3	energía	PSE	media	varianza	fmoda	P300Media	P300ValorPico	P300ValorPicoAbs	ZC	fmedia
F4	energía	PSE	media	varianza	fmoda	P300ValorPico	P300Media	fmedia	ZC	P300ValorPicoAbs
F7	energía	PSE	fmoda	media	varianza	ZC	fmedia	P300Media	P300ValorPico	P300ValorPicoAbs
F8	energía	PSE	media	fmoda	varianza	P300ValorPico	fmedia	P300Media	ZC	fmediana
Fp1	Atotal	PSE	fmoda	P300ValorPico	media	varianza	fmediana	P300Media	fmedia	ZC
Fp2	TAAR	PSE	fmoda	fmediana	varianza	P300ValorPico	fmedia	fmedia	ZC	P300Media
Fz	energía	PSE	media	fmoda	varianza	ZC	P300Media	fmedia	P300ValorPicoAbs	fmediana
O1	energía	PSE	media	fmoda	varianza	P300Media	P300ValorPicoAbs	ZC	fmedia	P300ValorPico
O2	energía	PSE	media	varianza	fmoda	P300Media	P300ValorPicoAbs	fmediana	ZC	P300ValorPico
P3	LAR	PSE	fmoda	varianza	media	ZC	fmediana	P300Media	P300ValorPicoAbs	P300ValorPicoAbs
P4	energía	PSE	media	fmoda	varianza	P300ValorPicoAbs	ZC	P300Media	fmediana	fmedia
Pz	Atotal	PSE	media	fmoda	varianza	fmediana	ZC	P300ValorPicoAbs	fmedia	P300Media
T5	energía	PSE	media	varianza	fmoda	P300Media	ZC	P300ValorPico	fmedia	fmediana
T6	energía	PSE	media	varianza	fmoda	P300Media	ZC	P300ValorPico	fmedia	P300ValorPicoAbs

Tabla C.3: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, algoritmo MIFS).

Agrupación	Característica									
	1	2	3	4	5	6	7	8	9	10
Total	Anegativa (F3)	PSE (Fp1)	PSE (Fp2)	PSE (F3)	PSE (F4)	PSE (C3)	PSE (C4)	PSE (P3)	PSE (P4)	PSE (O1)
Central	Atotal (C4)	PSE (C3)	PSE (C4)	PSE (Cz)	varianza (C3)	fmoda (C4)	varianza (C4)	fmoda (Cz)	media (C4)	media (C3)
Frontal	Anegativa (F3)	PSE (Fp1)	PSE (Fp2)	PSE (F3)	PSE (F4)	PSE (F7)	PSE (F8)	PSE (Fz)	varianza (F3)	fmoda (Fp1)
Occipital	energía (O1)	PSE (O1)	PSE (O2)	fmoda (O1)	varianza (O2)	media (O2)	fmoda (O2)	media (O1)	varianza (O1)	P300Media (O2)
Parietal-temporal	energía (P3)	PSE (P3)	PSE (P4)	PSE (Pz)	PSE (T5)	PSE (T6)	fmoda (P4)	varianza (P4)	fmoda (P3)	media (P3)
C3	LAR	PSE	varianza	media	fmoda	P300ValorPico	P300Media	P300ValorPicoAbs	fmedia	fmediana
C4	Atotal	PSE	varianza	fmoda	fmedia	P300ValorPicoAbs	P300ValorPico	fmediana	ZC	P300Media
Cz	energía	PSE	fmoda	media	varianza	P300ValorPico	ZC	fmediana	P300ValorPicoAbs	P300Media
F3	Anegativa	PSE	fmoda	fmedia	varianza	P300ValorPico	fmediana	P300Media	P300ValorPicoAbs	fmedia
F4	energía	PSE	media	varianza	fmoda	P300ValorPico	P300Media	ZC	P300ValorPicoAbs	fmediana
F7	ATAR	PSE	media	varianza	fmoda	P300ValorPico	P300Media	fmediana	P300ValorPicoAbs	ZC
F8	energía	PSE	fmoda	varianza	P300Media	P300ValorPico	media	fmediana	P300ValorPicoAbs	fmedia
Fp1	energía	PSE	fmoda	media	varianza	fmediana	P300ValorPico	P300Media	fmedia	P300ValorPicoAbs
Fp2	energía	PSE	fmoda	media	varianza	fmediana	P300ValorPico	P300Media	P300ValorPicoAbs	fmedia
Fz	energía	PSE	media	fmoda	varianza	fmedia	P300ValorPicoAbs	CZ	P300Media	fmediana
O1	energía	PSE	fmoda	varianza	media	P300Media	CZ	P300ValorPicoAbs	fmedia	fmediana
O2	Atotal	PSE	media	varianza	fmoda	P300Media	P300ValorPicoAbs	fmediana	P300ValorPico	fmedia
P3	energía	PSE	fmoda	media	varianza	P300Media	fmediana	P300ValorPicoAbs	ZC	P300ValorPico
P4	energía	PSE	fmoda	varianza	media	P300Media	ZC	P300ValorPico	ZC	P300ValorPicoAbs
Pz	energía	PSE	fmoda	varianza	fmedia	fmediana	ZC	P300ValorPicoAbs	fmedia	P300Media
T5	energía	PSE	fmoda	media	varianza	P300Media	fmedia	P300ValorPico	fmediana	P300ValorPicoAbs
T6	energía	PSE	varianza	fmoda	media	P300Media	P300ValorPicoAbs	P300ValorPico	fmedia	ZC

Tabla C.4: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, algoritmo MIFS).

Agrupación	Característica									
	1	2	3	4	5	6	7	8	9	10
Total	energía (F3)	varianza (Fp1)	fmediana (P4)	LARAbs (T5)	Apositiva (P3)	Atotal (F3)	P300LatenciaAbs (F8)	LAR (Pz)	LARAbs (F4)	TAR (Pz)
Central	energía (C3)	varianza (C3)	LAR (C4)	LAR (C3)	LARAbs (Cz)	Atotal (Cz)	LARAbs (C4)	energía (Cz)	energía (C4)	LAR (Cz)
Frontal	energía (F3)	varianza (Fp1)	P300LatenciaAbs (F8)	LARAbs (F4)	Atotal (F3)	ATAR (Fp2)	TAAR (F8)	LAR (F4)	P300Latencia (F3)	Apositiva (Fp1)
Occipital	energía (O2)	varianza (O1)	LARAbs (O2)	Atotal (O1)	TAAR (O2)	P300Latencia (O1)	ATAR (O1)	energía (O1)	Atotal (O2)	P300LatenciaAbs (O1)
Parietal-temporal	energía (P4)	varianza (P3)	LAR (T6)	ATAR (Pz)	fmediana (P4)	Atotal (Pz)	LAR (Pz)	Atotal (P4)	Apositiva (P3)	LARAbs (P4)
C3	energía	varianza	LAR	Atotal	LARAbs	Apositiva	TAAR	Anegativa	P300Latencia	ATAR
C4	LAR	P300Media	LARAbs	Apositiva	P300Latencia	Atotal	TAAR	P300LatenciaAbs	ATAR	Anegativa
Cz	energía	media	P300LatenciaAbs	LARAbs	Atotal	TAAR	LAR	ATAR	Anegativa	Apositiva
F3	energía	varianza	Atotal	TAAR	LARAbs	P300Latencia	LAR	ATAR	Anegativa	Apositiva
F4	energía	media	Anegativa	Atotal	LAR	TAAR	LARAbs	ATAR	P300LatenciaAbs	Apositiva
F7	energía	P300ValorPico	Anegativa	Atotal	LARAbs	LAR	ATAR	TAAR	P300LatenciaAbs	Apositiva
F8	energía	varianza	P300LatenciaAbs	Atotal	TAAR	LAR	Anegativa	ATAR	LARAbs	Apositiva
Fp1	Atotal	media	Apositiva	Anegativa	TAAR	ATAR	LARAbs	energía	LAR	P300LatenciaAbs
Fp2	TAAR	P300Media	LAR	P300LatenciaAbs	LARAbs	Atotal	energía	Anegativa	ATAR	Apositiva
Fz	energía	varianza	Atotal	ATAR	LARAbs	P300Latencia	Anegativa	LAR	Apositiva	TAAR
O1	energía	media	LAR	Atotal	P300LatenciaAbs	P300Latencia	TAAR	ATAR	LARAbs	Apositiva
O2	energía	varianza	LARAbs	TAAR	ATAR	Apositiva	P300LatenciaAbs	LAR	ATAR	fmediana
P3	LAR	Anegativa	Apositiva	Atotal	TAAR	energía	ATAR	LARAbs	P300Latencia	P300LatenciaAbs
P4	energía	varianza	Atotal	LARAbs	ATAR	Apositiva	TAAR	LAR	fmedia	Anegativa
Pz	Atotal	varianza	LAR	ATAR	P300Latencia	LARAbs	TAAR	energía	Apositiva	P300LatenciaAbs
T5	energía	media	Apositiva	TAAR	Atotal	LAR	P300LatenciaAbs	ATAR	Anegativa	LARAbs
T6	energía	varianza	P300LatenciaAbs	LAR	LARAbs	P300Latencia	Atotal	Apositiva	TAAR	Anegativa

Tabla C.5: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, algoritmo DISR).

Agrupación	Característica									
	1	2	3	4	5	6	7	8	9	10
Total	Anegativa (F3)	PSE (Fp1)	ATAR (F3)	TAAR (F8)	Atotal (F3)	Atotal (C4)	energía (O1)	energía (C4)	Atotal (O2)	ATAR (O2)
Central	Atotal (C4)	varianza (C4)	energía (C4)	LAR (C3)	Atotal (C3)	energía (C3)	energía (Cz)	ATAR (C4)	ATAR (C3)	LAR (C4)
Frontal	Anegativa (F3)	PSE (Fp1)	ATAR (F3)	TAAR (F8)	Atotal (F3)	energía (F4)	energía (F8)	energía (Fp1)	energía (Fp2)	LAR (F4)
Occipital	energía (O1)	media (O1)	ATAR (O2)	ATAR (O1)	Atotal (O2)	energía (O2)	Anegativa (O2)	P300Latencia (O2)	LARAbs (O1)	Atotal (O1)
Parietal-temporal	energía (P3)	varianza (P3)	LAR (Pz)	LARAbs (Pz)	LAR (O2)	Atotal (Pz)	energía (P4)	energía (T5)	LAR (T5)	LARAbs (P3)
C3	LAR	media	energía	Atotal	LARAbs	ATAR	P300Latencia	TAAR	P300LatenciaAbs	Anegativa
C4	Atotal	varianza	energía	ATAR	Apos	Aneg	TAAR	LAR	P300ValorPicoAbs	LARAbs
Cz	energía	fmedia	LAR	Atotal	TAAR	Aneg	LARAbs	Apos	ATAR	P300LatenciaAbs
F3	Anegativa	PSE	ATAR	Atotal	energía	LAR	TAAR	LARAbs	P300LatenciaAbs	P300Latencia
F4	energía	varianza	LAR	Atotal	LARAbs	P300LatenciaAbs	ATAR	Apos	Aneg	P300Latencia
F7	ATAR	media	Atotal	Aneg	LAR	LARAbs	energía	P300Latencia	TAAR	Apos
F8	energía	varianza	TAAR	LARAbs	Aneg	LAR	Atotal	P300LatenciaAbs	ATAR	Apos
Fp1	energía	varianza	Atotal	TAAR	Aneg	LARAbs	P300LatenciaAbs	ATAR	LAR	P300Latencia
Fp2	energía	varianza	ATAR	LAR	Atotal	Apos	LARAbs	Aneg	TAAR	P300LatenciaAbs
Fz	energía	P300ValorPicoAbs	Atotal	P300LatenciaAbs	ATAR	P300Latencia	Aneg	LAR	TAAR	LARAbs
O1	energía	media	ATAR	Atotal	LARAbs	TAAR	P300Latencia	Apos	LAR	P300LatenciaAbs
O2	Atotal	P300Media	ATAR	P300Latencia	ATAR	Apos	energía	LARAbs	LAR	Aneg
P3	energía	varianza	LAR	LARAbs	TAAR	Apos	P300LatenciaAbs	Atotal	Aneg	ATAR
P4	energía	varianza	Atotal	P300LatenciaAbs	ATAR	LAR	Apos	LARAbs	TAAR	P300Latencia
Pz	energía	varianza	LAR	LARAbs	P300Latencia	Apos	Atotal	TAAR	P300LatenciaAbs	ATAR
T5	energía	LAR	media	Atotal	LARAbs	fmedia	ATAR	TAAR	Apos	P300Latencia
T6	energía	P300Media	P300LatenciaAbs	P300Latencia	LARAbs	Atotal	Apos	TAAR	Aneg	ATAR

Tabla C.6: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, algoritmo DISR).

Agrupación	Características									
	1	2	3	4	5	6	7	8	9	10
Total	varianza (Fp2), P300ValorPico (Fp1), P300Latencia (P3), P300Latencia (O1), P300LatenciaAbs (F3), P300LatenciaAbs (O1), Anegativa (O1), TAAR (Fz) y ZC (O1).									
Central	P300ValorPico (C3), P300ValorPicoAbs (C3), P300fmedia (C3), energía (C3), P300Latencia (C3), LAR (Cz), Apositiva (Cz), ATAR (C4), TAAR (C4), TAAR (Cz) y ZC (C3).									
Frontal	varianza (Fp2), media (Fz), P300ValorPico (Fp1), P300Latencia (Fp1), P300Latencia (F3), P300LatenciaAbs (F3), P300LatenciaAbs (F4), LARAbs (F7), LARAbs (Fz), Apositiva (Fz) y ZC (F3).									
Occipital	media (O1), media (O2), P300ValorPico (O2), P300ValorPicoAbs (O1), P300Media (O2), energía (O1), P300Latencia (O1), P300Latencia (O2), P300LatenciaAbs (O1), LAR (O1), LARAbs (O2), Atotal (O1), Atotal (O2), ATAR (O1), ZC (O1), ZC (O2) y fmediana (O1).									
Parietal-temporal	varianza (P3), P300ValorPico (T6), P300ValorPicoAbs (T5), energía (P3), P300LatenciaAbs (T6), Apositiva (T6), ZC (P3), PSE (P4), fmediana (P4), fmediana (T5), fmedia (T5) y fmedia (P3).									
C3	P300ValorPico, P300LatenciaAbs, LARAbs, Atotal, ZC y fmedia.									
C4	P300ValorPico, P300ValorPicoAbs, P300LatenciaAbs, Apositiva, PSE, fmedia.									
Cz	LAR, Apositiva y ZC.									
F3	media, P300Latencia, LAR y ATAR.									
F4	media, P300Media, P300Latencia, P300LatenciaAbs, LAR, LARAbs, Apositiva, Anegativa, Atotal, ZC y fmedia.									
F7	varianza, P300Media, LAR, ATAR y TAAR.									
F8	varianza, P300Latencia, Apositiva, Apositiva, PSE y fmedia.									
Fp1	varianza, P300ValorPico, energía, ATAR, ZC, fmediana, fmedia y fmedia.									
Fp2	varianza, P300ValorPicoAbs y ATAR.									
Fz	energía, P300Latencia, P300LatenciaAbs, Atotal, ATAR, fmediana, fmedia y fmedia.									
O1	media, P300Latencia, P300LatenciaAbs, Apositiva y ZC.									
O2	varianza, media, ZC, fmedia y fmedia.									
P3	varianza, P300ValorPicoAbs, PSE, fmediana y fmedia.									
P4	LARAbs.									
Pz	P300ValorPico, P300Media, LARAbs, PSE y fmedia.									
T5	media.									
T6	media.									

Tabla C.7: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 15Hz, AG).

Agrupación	Características
<i>Total</i>	varianza (C4), P300ValorPico (C3), P300ValorPicoAbs (C4), LAR (F3), LAR (O2), ZC (Fp2), fmediana (O2), fmedia (P4), fmoda (C3), fmoda (F7) y fmoda (Cz).
<i>Central</i>	P300ValorPico (C3), P300media (Cz), energia (C3), energia (Cz), P300Latencia (Cz), LAR (C4), Apositiva (C3), Apositiva (Cz), Anegativa (Cz), Atotal (Cz), TAAR (C3), TAAR (C4), ZC (Cz), PSE (C4) y fmediana (Cz).
<i>Frontal</i>	media (F3), media (Fz), P300ValorPicoAbs (Fp2), P300Media (F8), energia (Fp1), P300Latencia (F7), Apositiva (Fp2), Anegativa (Fz), ZC (Fp2), fmediana (F4), fmedia (F8) y fmoda (F3).
<i>Occipital</i>	media (O1), P300ValorPico (O1), P300Latencia (O2), LARAbs (O2), fmediana (O2), fmoda (O1) y fmoda (O2).
<i>Parietal-temporal</i>	varianza (T6), P300LatenciaAbs (T6), LAR (Pz), LARAbs (T6), TAAR (P3), PSE (P4) y fmedia (P3).
<i>C3</i>	varianza, energia, fmediana, fmedia y fmoda.
<i>C4</i>	LAR, LARAbs, Apositiva y fmedia.
<i>Cz</i>	varianza, energia, LARAbs y ZC.
<i>F3</i>	P300ValorPico, P300Latencia y LARAbs.
<i>F4</i>	varianza, P300Media, LAR, ZC y fmedia.
<i>F7</i>	varianza, P300ValorPicoAbs, P300Latencia, P300LatenciaAbs, Apositiva, ATAR, ZC, fmedia y fmoda.
<i>F8</i>	P300ValorPicoAbs, energia, Apositiva y Atotal.
<i>Fp1</i>	Anegativa.
<i>Fp2</i>	varianza, energia, P300LatenciaAbs, ZC, fmediana y fmoda.
<i>Fz</i>	varianza, Anegativa, ZC, PSE y fmediana.
<i>O1</i>	varianza, media, P300Media, P300LatenciaAbs, Anegativa, ATAR, fmediana y fmoda.
<i>O2</i>	P300ValorPicoAbs, energia, ZC, PSE y fmoda.
<i>P3</i>	varianza y energia.
<i>P4</i>	media, energia, ZC y PSE.
<i>Pz</i>	P300ValorPicoAbs, energia, P300Latencia y Anegativa.
<i>T5</i>	P300LatenciaAbs, fmediana, fmedia y fmoda.
<i>T6</i>	P300ValorPico, P300LatenciaAbs y PSE.

Tabla C.8: Parámetros de la señal de EEG seleccionados en las diferentes agrupaciones (filtrado de 35Hz, AG).