



Universitat Oberta
de Catalunya

SEÑALES DE ENCEFALOGRAMA: Análisis y distinción de Canales Focales y No Focales.

Autor: ANTÍA TESTA PÉREZ

Tutor: JORDI BOSCH GARCIA

Profesor: JORDI SOLÉ CASALS

MASTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN
PROCESADO DE SEÑAL

09 JUNIO 2019



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial.

SinObraDerivada [3.0 España de Creative Commons.](https://creativecommons.org/licenses/by-nc/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Señales de encefalograma: Análisis y distinción de Canales Focales y No focales.
Nombre del autor:	Antía Testa Pérez
Nombre del colaborador/a docente :	Jordi Solé Casals
Nombre del PRA:	Jordi Solé Casals
Fecha de entrega (mm/aaaa):	06/2019
Titulación o programa:	Master Universitario en Ingeniería de Telecomunicación
Área del Trabajo Final:	Procesado de señal
Idioma del trabajo:	Español
Palabras clave	Epilepsia Señal focal/no focal

Resumen del Trabajo (máximo 250 palabras): Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo

Dentro de la biomedicina, la captación y procesado de señales fisiológicas, resulta fundamental para el diagnóstico o investigación de diferentes patologías que de alguna manera se originan en el cuerpo.

Existen múltiples tipos de señales a analizar, a lo largo de este proyecto nos centraremos en señales de Electroencefalograma y su tratamiento para la detección de focos epilépticos.

Partiendo de una base de datos de señales de electroencefalograma, captada en pacientes reales, buscaremos obtener una visión de las actividades del cerebro. Se procederá al procesado y tratado de las señales para su posterior clasificación en función de si se trata de una señal focal (foco epiléptico) o no. Para ello, nos apoyaremos en la herramienta Matlab, para llevar a cabo el procesado de las señales, la definición de parámetros de distinción y, por último su clasificación.

Abstract (in English, 250 words or less):

Within biomedicine, the capture and processing of physiological signals is fundamental for the diagnosis or investigation of different pathologies that somehow originate in the.

There are multiple types of signals to analyze, throughout this project we will focus on Electroencephalogram signals and their treatment for the detection of epileptic focus.

Starting from a database of electroencephalogram signals, captured in real patients, we will seek to obtain a vision of the activities of the brain. The processing and treatment of the signals will be proceeded for their subsequent classification according to whether it is a focal signal (epileptic focus) or not. To do this, we will rely on the Matlab tool, to carry out the signal processing, the definition of distinction parameters and, finally, its classification.

Índice

1. Introducción	7
1.1. Contexto y justificación del trabajo	7
1.2. Planificación	7
1.3. Estructura del resto del documento	8
2. ELECTROENCEFALOGRAFÍA	9
2.1. Electroencefalograma	9
2.2. Electroencefalograma intracraneal	11
2.3. Epilepsia	11
2.4. Base de datos de análisis	14
3. ALGORITMO DE ANÁLISIS	16
3.1. Punto de partida	16
3.2. Diseño del algoritmo	16
3.2.1. Obtención de señales.....	17
3.2.2. Filtrado paso banda.....	18
3.2.3. Cálculo de Parámetros	20
3.2.4. Fase de entrenamiento	28
3.2.5. Fase de test	32
3.3. Resultados	33
3.3.1. Análisis y mejora de resultados.....	33
3.3.2. Comparación con otros estudios.....	34
4. CONCLUSIONES	36
5. LÍNEAS FUTURAS	37
6. BIBLIOGRAFÍA	38
7. ANEXOS	39

Figuras y tablas

Índice de figuras

Figura 1: Captación y representación de la actividad eléctrica en el cerebro [3].....	9
Figura 2: La epilepsia, enfermedad del sistema nervioso.....	12
Figura 3: Señales focales y no focales	13
Figura 4: Señales de análisis	15
Figura 5: Esquema de definición del algoritmo	17
Figura 6: Representación de señales originales.....	18
Figura 7: Filtros y señal filtrada correspondiente	20
Figura 8: Tasa de cruces por cero en señal filtrada.....	21
Figura 9: Boxplot Tasa de cruces por cero	22
Figura 10: Boxplot Entropía de Shannon	23
Figura 11: Boxplot Entropía Aproximada	24
Figura 12: Boxplot Entropía de Muestra	25
Figura 13: Boxplot Entropía de Permutación	26
Figura 14: Boxplot Exponente de Hurst	27
Figura 15: Boxplot Dimensión Fractal	28
Figura 17: Métodos de clasificación.....	30

Índice de tablas

Tabla 1: Frecuencias de filtrado - Tipos de filtro.....	19
Tabla 2: Comparación de clasificadores	31
Tabla 3: Resultados finales.....	33

1. Introducción

1.1. Contexto y justificación del trabajo

La Ingeniería biomédica o Bioingeniería se trata de una disciplina que aúna ingeniería, biología y medicina. Surgió en la década de 1950 con el fin de combinar criterios y herramientas de análisis para la comprensión de distintas enfermedades y para el desarrollo de nuevas técnicas y productos en el ámbito de la Biomedicina.

La Bioingeniería es una de las principales áreas de aplicación del procesado de señales, que es uno de los pilares de la Ingeniería de Telecomunicación.

En este ámbito surge la Electroencefalografía (EEG), la cual permite obtener señales generadas por el cerebro a través de la grabación de la actividad eléctrica de éste durante un período de tiempo. Estas señales se deben a la actividad eléctrica generada por la actividad de miles de neuronas y se registra a partir de múltiples electrodos situados de manera más o menos invasiva en el paciente en cuestión.

El procesado de este tipo de señales, es fundamental en la detección de anomalías relacionadas con la actividad eléctrica del cerebro, permitiendo así definir diferentes enfermedades como puede ser la epilepsia, objetivo principal de este proyecto. De este modo, se buscará definir un algoritmo que permita diferenciar una señal focal (señal con síntomas de epilepsia) de una no focal (señal normal).

El procesado citado se desarrollará íntegramente con Matlab, considerado el lenguaje de más alto nivel para el cálculo matemático.

1.2. Planificación

PLAN DE TRABAJO TFM														
1. DEFINICIÓN PLAN DE TRABAJO			1. DEFINICIÓN PLAN DE TRABAJO											
Documentación y definición clara del alcance del proyecto	25/02/19	17/03/19	Documentación y definición clara del alcance del proyecto											
Planificación y realización del plan de trabajo	11/03/19	17/03/19	Planificación y realización del plan de trabajo											
2. DESARROLLO DEL PROYECTO			2. DESARROLLO DEL PROYECTO											
Familiarización con la Base de Datos	17/03/19	23/03/19	Familiarización con la Base de Datos											
Matlab: procesado de datos para su manipulación	23/03/19	07/04/19	Matlab: procesado de datos para su manipulación											
Matlab: familiarización con los clasificadores	07/04/19	21/04/19	Matlab: familiarización con los clasificadores											
Pruebas y comparación de clasificadores: obtención de resultados	21/04/19	26/05/19	Pruebas y comparación de clasificadores: obtención de resultados											
3. REALIZACIÓN DE LA MEMORIA			3. REALIZACIÓN DE LA MEMORIA											
4. REALIZACIÓN DE LA PRESENTACIÓN			4. REALIZACIÓN DE LA PRESENTACIÓN											
5. DEFENSA			5. DEFENSA											

1.3. Estructura del resto del documento

En primer lugar, se hará una breve introducción a la encefalografía, incluyendo las distintas técnicas de análisis de este tipo de señales. Seguidamente, nos centraremos en la epilepsia y en la base de datos en la que nos apoyaremos para el estudio global a realizar.

A continuación, pasaremos por explicar detalladamente el algoritmo de análisis, centrándonos en el procesado de las señales y en los parámetros fundamentales para el desarrollo del proyecto.

Finalmente detallaremos los resultados obtenidos, su comparación con estudios de índole similar y, por último, las conclusiones derivadas.

2. ELECTROENCEFALOGRAFÍA

En este capítulo se llevará a cabo una breve descripción de la Electroencefalografía, centrándonos en su importancia y en distintos modos de adquisición de señales de este tipo.

Además, nos centraremos en la Epilepsia, principal patología a detectar analizando este tipo de señales y, por último, se hará referencia a la base de datos que servirá de apoyo para el desarrollo global del proyecto.

2.1. Electroencefalograma

El cerebro está formado por miles de millones de células que se comunican entre ellas a través de impulsos eléctricos y están activas todo el tiempo, incluso mientras dormimos.

La mitad de estas células son neuronas, la mitad de las cuales ayudan y facilitan la actividad de estas. Estas neuronas están densamente interconectadas a través de sinapsis.

Cualquier actividad sináptica genera un impulso eléctrico denominado potencial postsináptico. La actividad de una única neurona, sin contacto directo con ella, es muy difícil de detectar. Sin embargo, cada vez que miles de neuronas actúan de forma sincronizada, generan un campo eléctrico que es lo suficientemente fuerte como para propagarse a través del tejido, los huesos y el cráneo, permitiendo de este modo su medida en la superficie de la cabeza [1]

De este modo, un electroencefalograma, consiste en el registro, a través del cráneo, de la actividad eléctrica (en voltios) producida por la descarga de neuronas en el cerebro. Esta actividad eléctrica se representa gráficamente mediante líneas ondulantes [2]

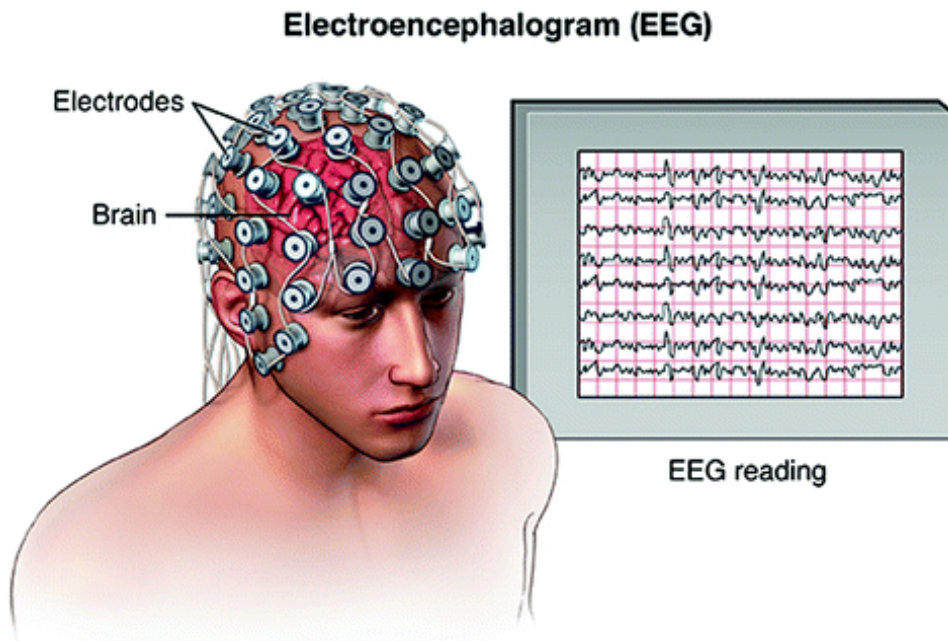


Figura 1: Captación y representación de la actividad eléctrica en el cerebro [3]

Analizando estas señales se consigue el diagnóstico de diferentes trastornos entre los que se encuentran:

- Epilepsia
- Tumor cerebral
- Daños cerebrales por lesiones en la cabeza
- Disfunciones cerebrales que pueden tener diversas causas (encefalopatía)
- Inflamación del cerebro (encefalitis)
- Accidente cerebrovascular
- Trastornos del sueño
- Confirmación de muerte cerebral en coma persistente

Se trata entonces, de un estudio que detecta la actividad eléctrica del cerebro mediante pequeños discos metálicos (electrodos) fijados sobre el cuero cabelludo. Los electrodos no transmiten ninguna sensación, únicamente registran las ondas cerebrales.

Para una aplicación más rápida, los electrodos se montan en mallas elásticas similares a un gorro de baño, lo que garantiza que los datos se pueden recopilar desde posiciones de cuero cabelludo idénticas en todos los pacientes.

Por último, como las fluctuaciones de voltaje medidas en los electrodos son muy pequeñas, los datos registrados se digitalizan y se envían a un amplificador. Los datos amplificados se pueden mostrar como una secuencia de valores de voltaje.

Las diferencias entre distintos sistemas EEG se deben generalmente a la cantidad de electrodos, la calidad de la digitalización, la calidad del amplificador y la cantidad de snapshots que el dispositivo puede tomar por segundo, es decir, la frecuencia de muestreo de las señales obtenidas.

Hay que tener en cuenta que, a pesar de la detallada información que este método proporciona a cerca de la actividad craneal, un electroencefalograma normal no suele resultar suficiente para poder dar un diagnóstico específico en relación a las patologías anteriormente citadas. Así, por ejemplo, en el diagnóstico de pacientes con epilepsia la posibilidad de encontrar actividad epileptiforme es del 29-55% y al realizar EEG sucesivos esta cifra aumenta al 69-77%. También se puede aumentar la sensibilidad realizándolo post privación de sueño, lo cual permitiría incrementar la actividad epileptiforme en un 13-31% adicional. Además, estos datos se ven alterados por múltiples factores como puede ser la edad del paciente.

Con todo esto, con el fin de poder afinar estos diagnósticos, se procede a usar métodos como son los electroencefalogramas profundos o intracraneales que serán detallados en el siguiente apartado.

2.2. Electroencefalograma intracraneal

Partiendo del método de obtención de señales explicado en el apartado anterior, existen técnicas más invasivas como son los Electroencefalogramas Intracraneales (iEEG).

Se basa igualmente en la disposición de electrodos pero directamente en contacto con el cerebro, conllevando la apertura del cráneo y la implantación de dichos dispositivos en su interior. De este modo, se realiza un procedimiento invasivo quirúrgico para implantar electrodos directamente sobre la superficie cerebral (grillas subdurales) o electrodos profundos cuyo propósito es poder llegar a zonas de sustancia gris no registrables en superficie como por ejemplo, la ínsula.

La implantación de electrodos intracraneales proporciona la fuente más fiable de neuroimagen, dotando de una resolución temporal y espacial excelentes, de modo que permite saber con gran exactitud lo que está ocurriendo en el área registrada.

El caso más frecuente de uso de este tipo de técnicas es la epilepsia en la que se usan los electrodos para encontrar el foco epiléptico antes de extirparlo así como para asegurarse de que no se tocan áreas vitales. De este modo, a diferencia del electroencefalograma convencional, permitiría obtener información relativa a:

- Diagnóstico relativo a padecer de epilepsia o no
- Ubicación de la actividad epiléptica
- Respuesta adecuado al tratamiento antiepiléptico

A lo largo de este proyecto nos basaremos en señales obtenidas apoyándose en este método, buscando diferenciar entre señales de iEEG con síntomas de epilepsia de señales que carezcan de esta patología.

2.3. Epilepsia

La Epilepsia es una enfermedad del sistema nervioso, debida a la aparición de actividad eléctrica anormal en la corteza cerebral, que provoca ataques repentinos caracterizados por convulsiones violentas y pérdida del conocimiento.



Figura 2: La epilepsia, enfermedad del sistema nervioso

Las convulsiones se deben a descargas eléctricas excesivas de grupos de células cerebrales que pueden producirse en diferentes partes del cerebro. Las convulsiones pueden ir desde episodios muy breves de ausencia o de contracciones musculares hasta convulsiones prolongadas y graves. Su frecuencia también puede variar desde menos de una al año hasta varias al día.

Las características de las convulsiones varían y dependen de en qué parte del cerebro comienza la alteración y cómo se propaga. Ocurren síntomas temporales, como pérdida del conocimiento o la conciencia, y alteraciones del movimiento, de los sentidos (incluyendo visión, audición y gusto), estado de ánimo u otras funciones cognitivas [4].

Según la Organización Mundial de la Salud (OMS), aproximadamente 50 millones de personas en todo el mundo tienen epilepsia, lo que la convierte en una de las enfermedades neurológicas más comunes en todo el mundo [5]. El 80% de las personas que padecen esta patología se encuentran en países subdesarrollados, o de ingresos bajos y medianos [4].

Respecto al origen de la epilepsia, esta puede darse por una causa identificable (epilepsia secundaria o sintomática) o bien sin que esta exista (epilepsia idiopática).

- Epilepsia idiopática: no existe una causa identificable para su aparición. Es el tipo más frecuente de epilepsia, afecta a 6 de cada 10 personas con la enfermedad.
- Epilepsia secundaria o sintomática: sus causas pueden ser muy variadas: daño cerebral por lesiones prenatales o perinatales, malformaciones congénitas o alteraciones genéticas con malformaciones cerebrales asociadas, traumatismos craneoencefálicos graves, accidentes cerebrovasculares que limitan la llegada del oxígeno al cerebro, infecciones cerebrales como las meningitis y encefalitis o la neurocisticercosis, algunos síndromes genéticos, tumores cerebrales...

En relación al tratamiento, por lo general, se pueden controlar los efectos asociados a estas enfermedades (convulsiones epilépticas) a través de medicamentos, pero algunos pacientes con epilepsia son resistentes a los medicamentos, por lo tanto, el tratamiento común es eliminar el foco de epilepsia con cirugía.

Previo a la cirugía, ha de localizarse el foco epiléptico de manera precisa mediante el uso de varias señales cerebrales o mediciones de neuroimagen.

Una vez localizado el foco epiléptico, existen dos tipos de tratamiento quirúrgico:

- Técnicas curativas: están indicadas en pacientes en las que el foco epiléptico está ubicado en zonas cerebrales bien localizadas y éstas pueden ser resecaadas con seguridad de no producir un déficit neurológico intolerable.
- Técnicas paliativas: se realizan en aquellos pacientes que tienen varias zonas corticales con propiedades epileptógenas o bien que tienen la zona epileptógena ubicada en una región que no puede ser resecaada. En este caso no se llega a eliminar el foco, pero se disminuye el efecto modificando sus características [2].

En este escenario, en el que resulta indispensable que encontrar con exactitud el foco epiléptico, surgen los análisis cuidadosos de los registros de electroencefalogramas intracraneales (iEEG) que fueron explicados en el apartado precedente. Los iEEG se han aplicado ampliamente en cirugías de epilepsia para la evaluación preoperatoria y el conocimiento cerebral; no solo ayudando a guiar la intervención quirúrgica de la epilepsia sino también proporcionando datos detallados relativos a las regiones cerebrales afectadas. Así, los iEEG se están convirtiendo en una visión clave para entender las actividades del cerebro.

En el registro de este tipo de señales definimos dos clases:

- Señales focales (F): señales registradas en un área epileptogénica
- Señales no focales (N): señales registradas en un área no epileptogénica

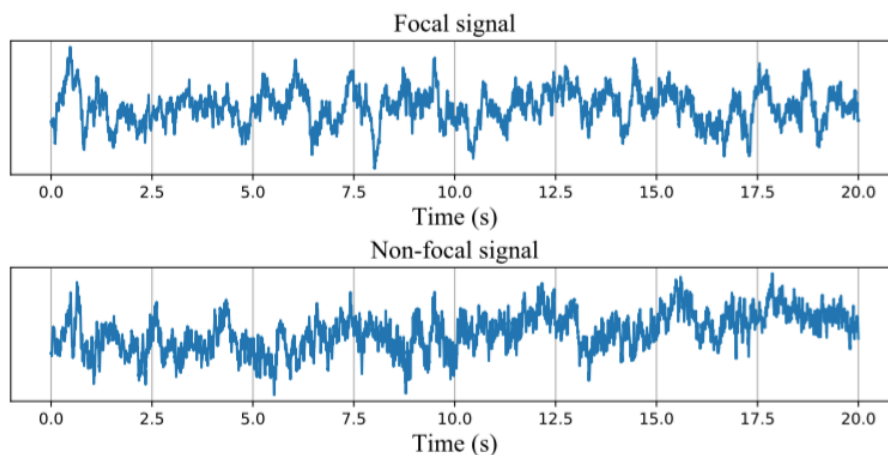


Figura 3: Señales focales y no focales

En la actualidad, los expertos clínicos diferencian estas señales a través del juicio visual basado en el análisis de las ondas, lo cual supone un proceso extremadamente lento y difícil. Además, los expertos clínicos calificados para dicho diagnóstico son bastante escasos. Por lo tanto, es una exigencia imperiosa aliviar el trabajo pesado de los expertos clínicos mediante la aplicación de métodos de aprendizaje automático eficaces. En esto se basará este proyecto, en la definición de un algoritmo que permita clasificar una señal dada, en base a un entrenamiento previo.

2.4. Base de datos de análisis

Para el análisis anteriormente citado, se usará un conjunto de datos que recopila las grabaciones de iEEG de cinco pacientes que sufren de epilepsia: "*Bern-Barcelona Dataset*"

Estas grabaciones se realizaron como parte de los diagnósticos de epilepsia en estos pacientes que padecían de esta patología y eran candidatos para cirugía de epilepsia.

En primer lugar, fueron se recurrió a estudios no invasivos, pero estos no permitieron la localización inequívoca de las áreas cerebrales en las que se originaron las convulsiones. Finalmente todos los pacientes fueron sometidos a registros de EEG intracraneal a largo plazo en el Departamento de Neurología de la Universidad de Berna.

Las zonas dañadas se encontraron en partes del cerebro que podrían ser tratadas quirúrgicamente sin el peligro de deficiencias neurológicas que serían inaceptables para los pacientes. Los cinco pacientes tuvieron un buen resultado quirúrgico. Tres pacientes lograron una completa libertad de crisis, y dos pacientes solo tenían auras pero ninguna otra crisis después de la cirugía. Además, todos los pacientes dieron su consentimiento de que sus datos de EEG a largo plazo podrían utilizarse para fines de investigación [7].

Este conjunto de análisis, correspondiente a los cinco pacientes a tratar, consta de 3.750 pares de señales focales y 3.750 pares de señales no focales. Cada señal tiene 10.240 muestras (20 segundos con una frecuencia de muestreo de 512 Hz), y todas las señales están filtradas con un filtro paso banda con frecuencias entre 0,5 y 150 Hz mediante el uso de un filtro Butterworth de cuarto orden.

Por lo tanto, comenzaremos nuestro análisis partiendo de 7.000 señales, la mitad de cada uno de los dos tipos, y con una morfología similar a la que se muestra en la figura 4:

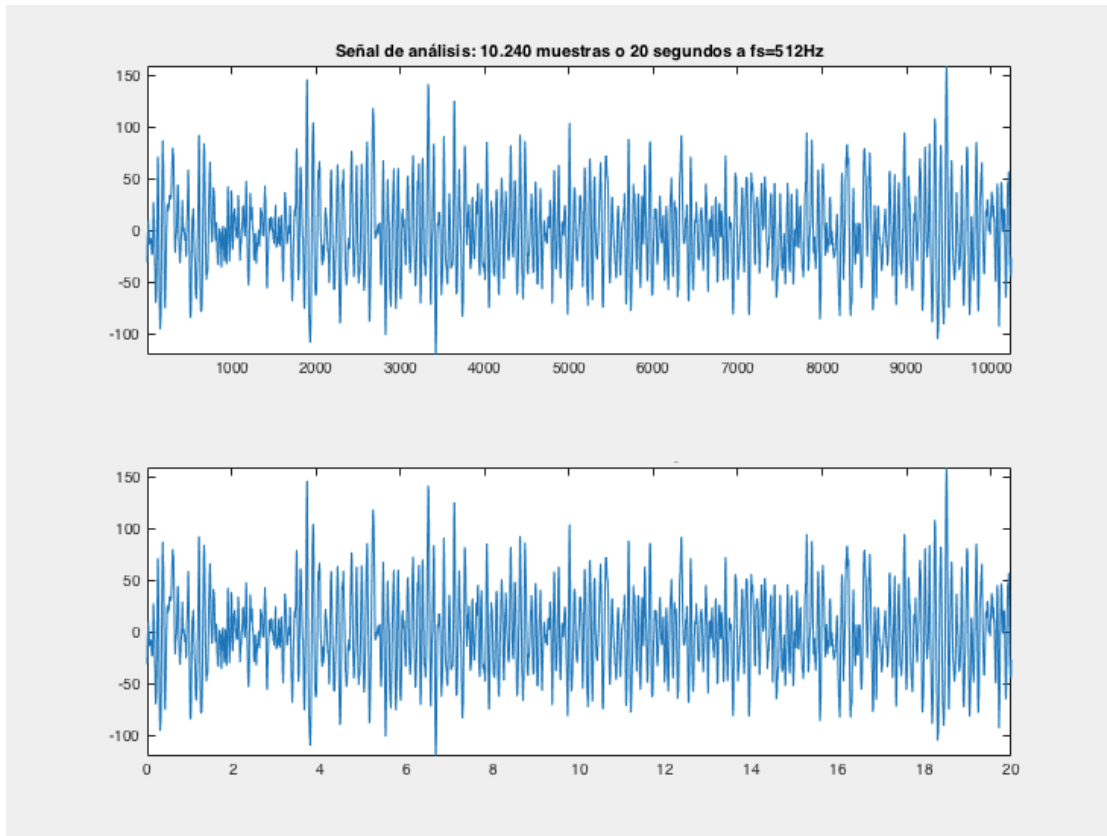


Figura 4: Señales de análisis

3. ALGORITMO DE ANÁLISIS

3.1. Punto de partida

En la actualidad existen múltiples métodos o algoritmos que permiten alcanzar resultados prometedores en la localización del foco epiléptico. Muchas de estas técnicas se basan en métodos de aprendizaje supervisados. Sin embargo, el inconveniente importante de estos métodos es que los expertos clínicos deben proporcionar una gran cantidad de datos ya etiquetados como focales o no focales y, para su obtención, cada 20 segundos (duración de cada una de nuestras señales de estudio), las señales iEEG deben ser evaluadas visualmente por múltiples expertos clínicos para así poder proporcionar la etiqueta final de este segmento, lo cual sigue siendo un proceso extremadamente lento y que requiere mucho tiempo.

Por lo tanto, en los escenarios reales, la adquisición de una gran cantidad de datos epilépticos fiables y etiquetados sigue siendo una tarea difícil. Esto indica que aunque los métodos de aprendizaje supervisados pueden obtener un excelente rendimiento en la localización del foco epiléptico, no son prácticos para el diagnóstico en el mundo real.

Con el fin de abordar este problema, este proyecto se basará en reducir significativamente la cantidad necesaria de datos etiquetados sin deteriorar el rendimiento del sistema global. Se llevará a cabo por tanto un entrenamiento previo basándonos en los datos que ya tenemos etiquetados y, posteriormente, su aplicación a señales permitiendo su distinción (focales o no focales). Para ello nos basaremos en el método estudiado en [5], "PU Learning (Positive Unlabeled Learning)"

3.2. Diseño del algoritmo

Partiremos de la base de datos citada anteriormente ("Bern-Barcelona Dataset") y procederemos al procesamiento de dichas señales con el fin de generar distintos parámetros que nos permitan diferenciarlas.

En primer lugar, se llevará a cabo el filtrado de la señal con distintos filtros que nos permitan, a partir de una señal, obtener 6 nuevas señales. De este modo multiplicamos la cantidad de datos de partida pudiendo aportar más datos etiquetados al análisis.

Posteriormente, se llevará a cabo el cálculo de una serie de parámetros cuyos valores nos permitan diferenciar señales focales de señales no focales. Estos datos, en la fase de entrenamiento nos permitirán, a partir de distintos métodos que se detallarán posteriormente, definir clasificadores que serán aplicados a señales no etiquetadas para definir su marca.

En definitiva, el esquema del procedimiento aplicado se muestra a continuación:

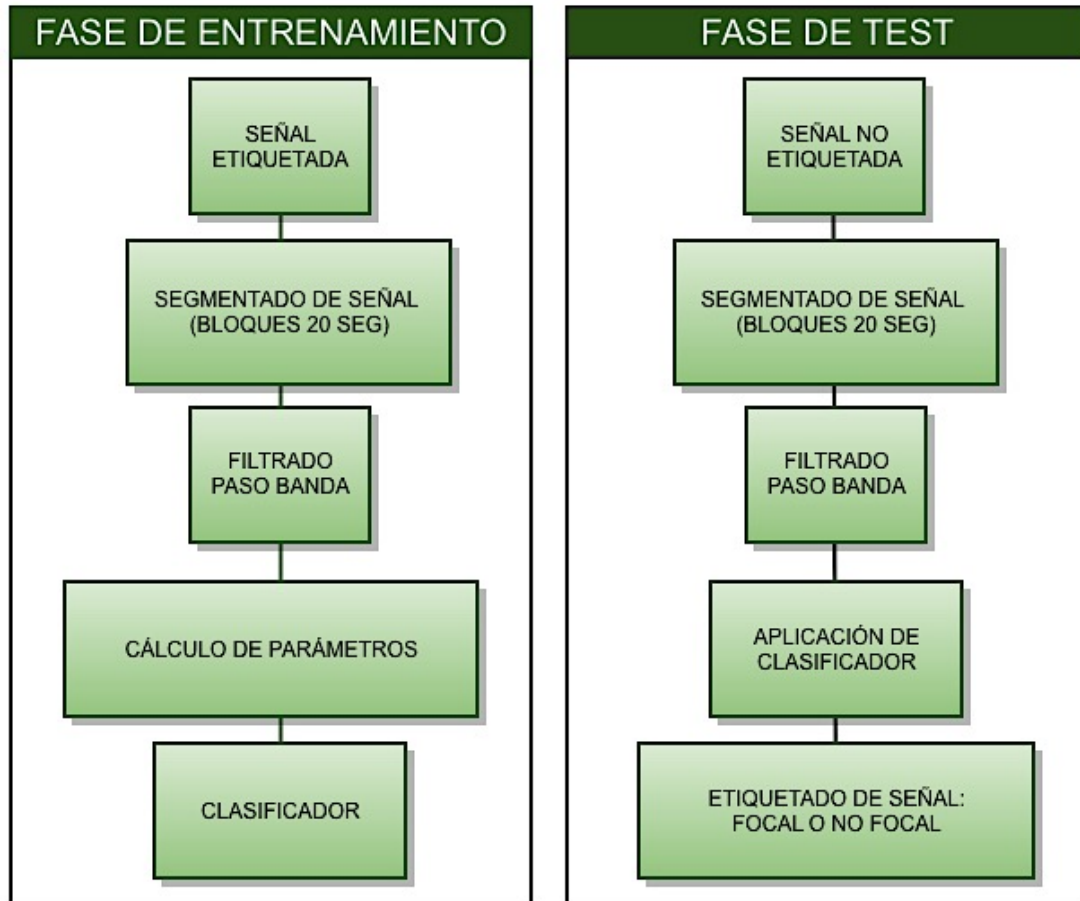


Figura 5: Esquema de definición del algoritmo

A continuación, se describirá detalladamente cada una de las fases llevadas a cabo en el diseño del algoritmo.

3.2.1. Obtención de señales

Partiendo de la base de datos anteriormente citada, se llevará a cabo su representación con el fin de conocer las señales con las que se acondicionará el estudio.

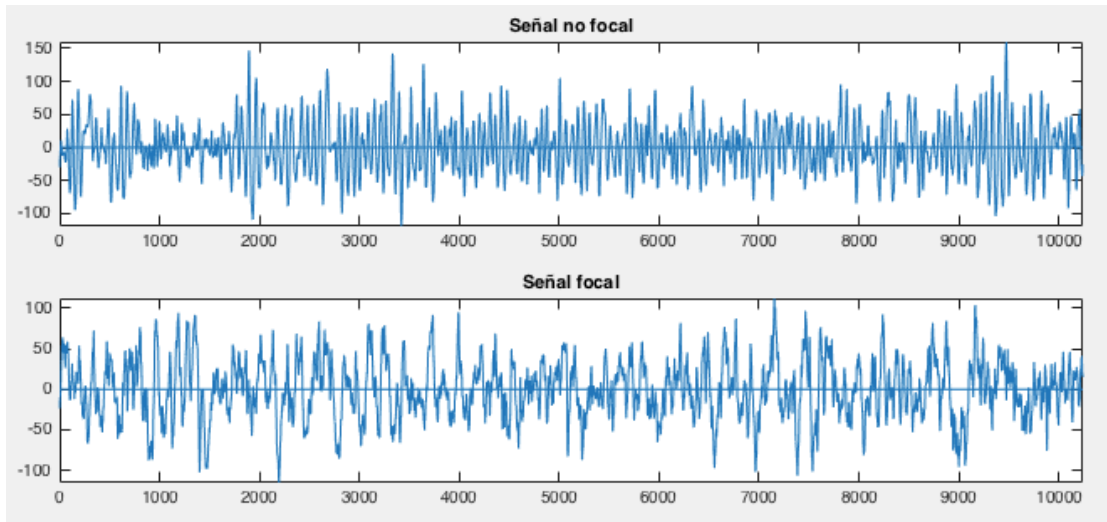


Figura 6: Representación de señales originales

Puede verse que cada señal ocupa 10.240 muestras, a una frecuencia de muestreo de 512 muestras/segundo. Por tanto, cada señal comprende 20 segundos de espectro.

Observamos además que, a simple vista, las señales focales y no focales tienen una morfología distinta, esto variará mucho dependiendo de la señal a tratar y veremos que, esta amplia variedad, será lo que le añada complejidad al análisis y a la distinción de este tipo de señales.

3.2.2. Filtrado paso banda

En primer lugar, se llevará a cabo un filtrado de las señales anteriores a distintas frecuencias. Concretamente se aplicarán seis filtros a distintas frecuencias pudiendo obtener de este modo seis nuevas señales a partir de una señal original.

El filtrado será paso banda, y el filtro diseñado será un filtro Butterworth. Los filtros se diferencian por las frecuencias que abarca la banda de paso de cada filtro y son los que se indican en la siguiente tabla:

TIPO DE FILTRO	FRECUENCIAS EN BANDA DE PASO
DELTA	0,5 – 4 Hz
THETA	4 – 8 Hz
ALPHA	8 – 13 Hz
BETA	13 – 30 Hz
GAMMA	30 – 80 Hz
RIPPLE	80 – 250 Hz

Tabla 1: Frecuencias de filtrado - Tipos de filtro

En la siguiente figura se muestra el resultado de aplicar cada uno de los filtros citados a la señal. Se refleja la forma de cada uno de los filtros aplicados y la forma de onda de la señal resultante:

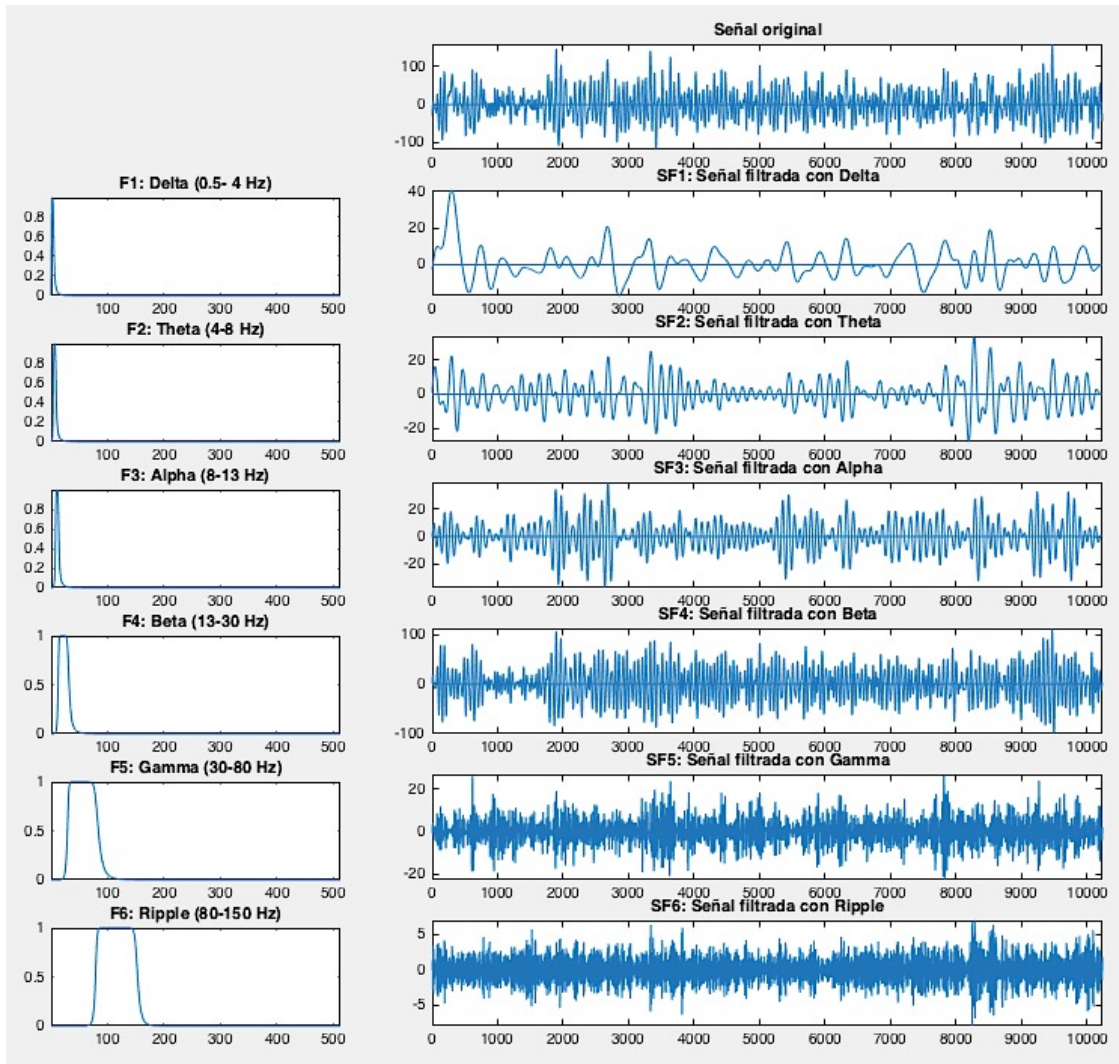


Figura 7: Filtros y señal filtrada correspondiente

Como resultado, a partir de ahora, tenemos 42.000 señales etiquetadas, en lugar de las 7.000 señales con las que hemos empezado el análisis.

3.2.3. Cálculo de Parámetros

Una vez tenemos las señales filtradas, se procederá al cálculo de siete parámetros por cada una de las señales de modo que esos siete parámetros caractericen a la señal y que permitan, a partir de su evaluación, clasificar la señal según sea focal o no focal. Estos parámetros son:

- Tasa de cruces por cero
- Entropía de Shannon
- Entropía Aproximada

- Entropía de Muestra
- Entropía de Permutación
- Exponente de Hurst
- Dimensión Fractal

A continuación, analizaremos cada uno de los parámetros, con el fin de ver sus características, cómo de útil es la información que aportan al estudio y a la definición final de la señal.

Tasa de cruces por cero

Consiste en evaluar el número de veces que la señal cruza la línea cero a lo largo de los 20 segundos de duración.

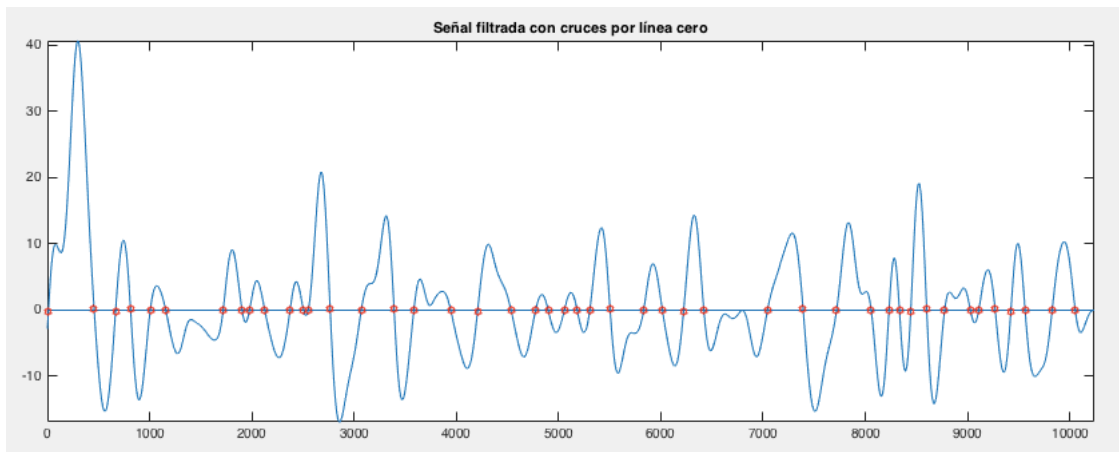


Figura 8: Tasa de cruces por cero en señal filtrada

En el ejemplo mostrado en la figura, se reproducen 46 cruces por cero a lo largo de la duración de la señal. Teniendo en cuenta que la duración de la señal comprende 10.240 muestras, la tasa de cruces por cero será del 0,45%.

Con el fin de ver si este parámetro resultará útil para la diferenciación de las señales focales y no focales, calculamos las tasas de cruces por cero para el total de señales de cada tipo. Así obtendremos lo siguiente:

- Media de Tasa de cruces por cero en señales focales: 3,2425%
- Media de Tasa de cruces por cero en señales no focales: 4,2483%

Procedemos a la representación de boxplot o diagrama de cajas, para poder ver no sólo la media sino el análisis de valores máximos, mínimos y la diferenciación de valores atípicos entre las muestras.

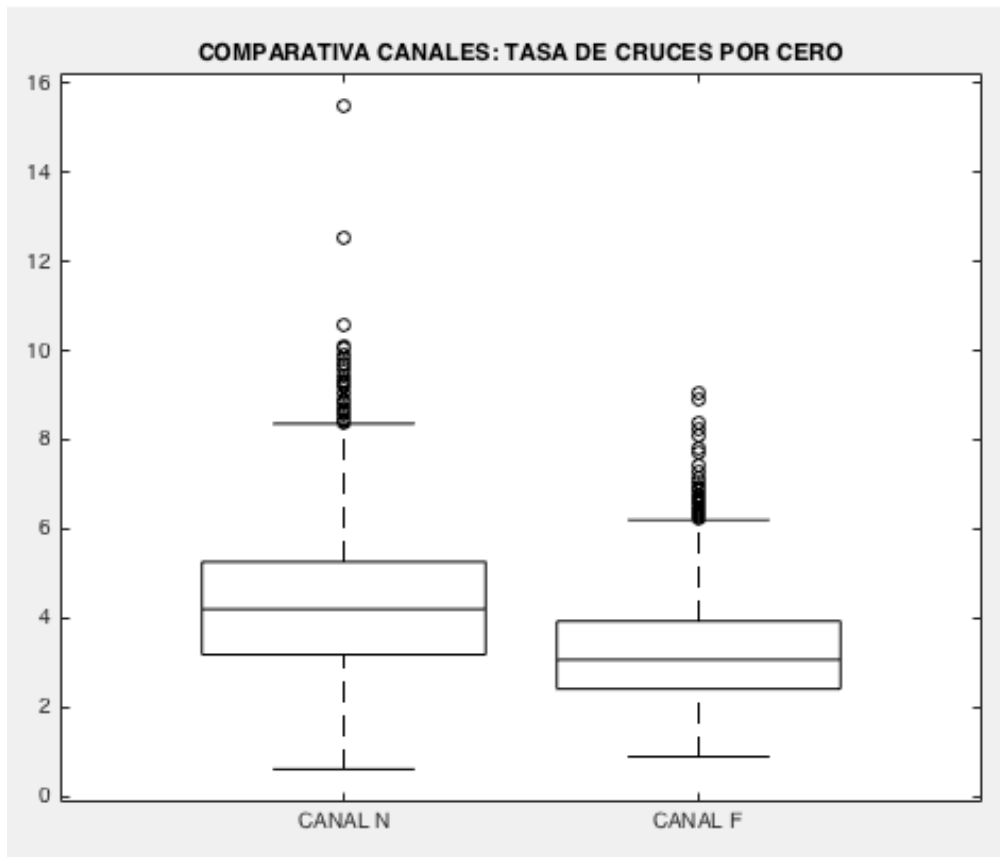


Figura 9: Boxplot Tasa de cruces por cero

En definitiva, en la mayoría de las señales de ensayo, la tasa de cruces por cero será mayor para señales no focales que focales. Además, mientras que en señales focales los valores atípicos son pocos y cercanos al máximo, en las señales no focales se ven varios valores atípicos que llegan a duplicar el valor máximo.

Entropías

La entropía hace referencia a la incertidumbre que existe en cualquier experimento o señal aleatoria. Es decir, evalúa la cantidad de información que contiene una señal y define la cantidad de “ruido” o “desorden” que contiene o libera el sistema. Se cumplen las siguientes premisas:

- Si todos los elementos de la señal son equiprobables (igual de probables) a la hora de aparecer, entonces la entropía será máxima.
- El cambio pequeño en una de las probabilidades de aparición de uno de los elementos de la señal afecta mínimamente a la entropía.

Existen múltiples formas y teorías de cálculo de entropía, en este proyecto se evaluarán cuatro diferentes:

- Entropía de Shannon (“Shannon Entropy”)

- Entropía Aproximada (“Approximate Entropy”)
- Entropía de Muestra (“Sample Entropy”)
- Entropía de Permutación (“Permutation Entropy”)

A continuación, procedemos a evaluar cómo de provechoso resultará cada uno de estos parámetros para nuestro análisis, viendo si permiten distinguir las señales determinativamente o si los valores de los parámetros para señales focales y no focales son prácticamente los mismos. En este segundo caso, el aporte del parámetro al análisis será mucho menor que en el primero de los planteamientos.

Entropía de Shannon

Puede verse que la media de valor de entropía de Shannon para señales focales y no focales es muy similar:

- Señales focales: $-3,3 \cdot 10^8$
- Señales no focales: $-2,7 \cdot 10^8$

Vemos, en la figura 10, que existen múltiples valores atípicos que hacen que este parámetro vaya a resultar poco determinativo. Concretamente en señales focales, 320 muestras son calificadas como atípicas, es decir, aproximadamente el 8,5% de las muestras.

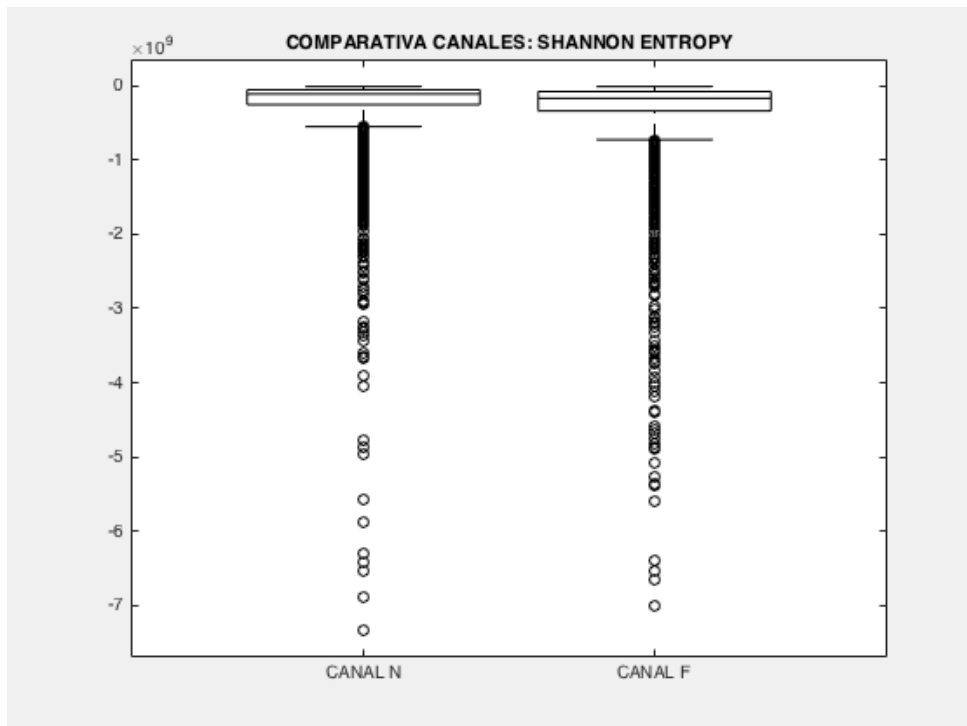


Figura 10: Boxplot Entropía de Shannon

Entropía Aproximada

Podemos ver que nuevamente la media se mantiene entre ambas señales, pero hay una gran diferencia en los cuartiles que presenta cada uno de los dos diagramas.

Para las señales focales, existe una gran acumulación de muestras cercanas a la media por encima (cuartil 3), mientras que en las señales no focales no existe esta distribución y la mayoría se trata de valores atípicos.

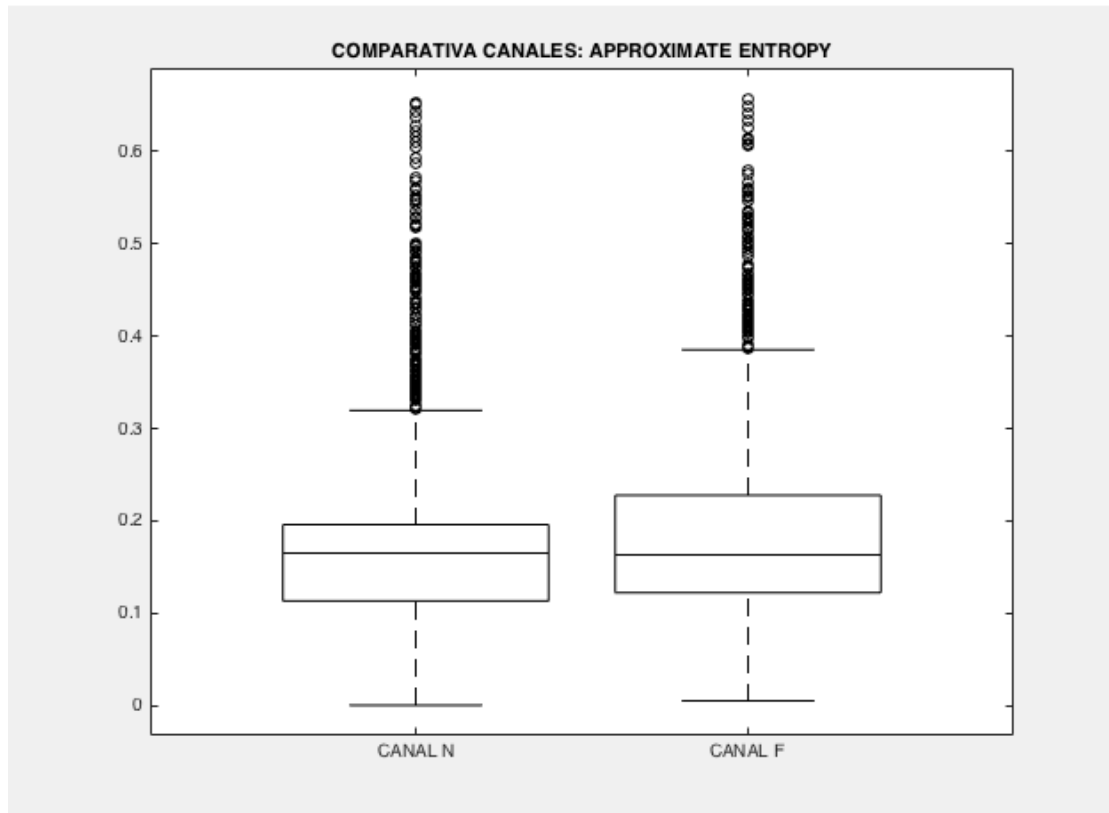


Figura 11: Boxplot Entropía Aproximada

Entropía de muestra

Observando la representación podemos ver que ambos diagramas son prácticamente iguales, de este modo podríamos decir a priori que no resultará un parámetro determinante en la decisión. Esto se analizará en la fase de pruebas.

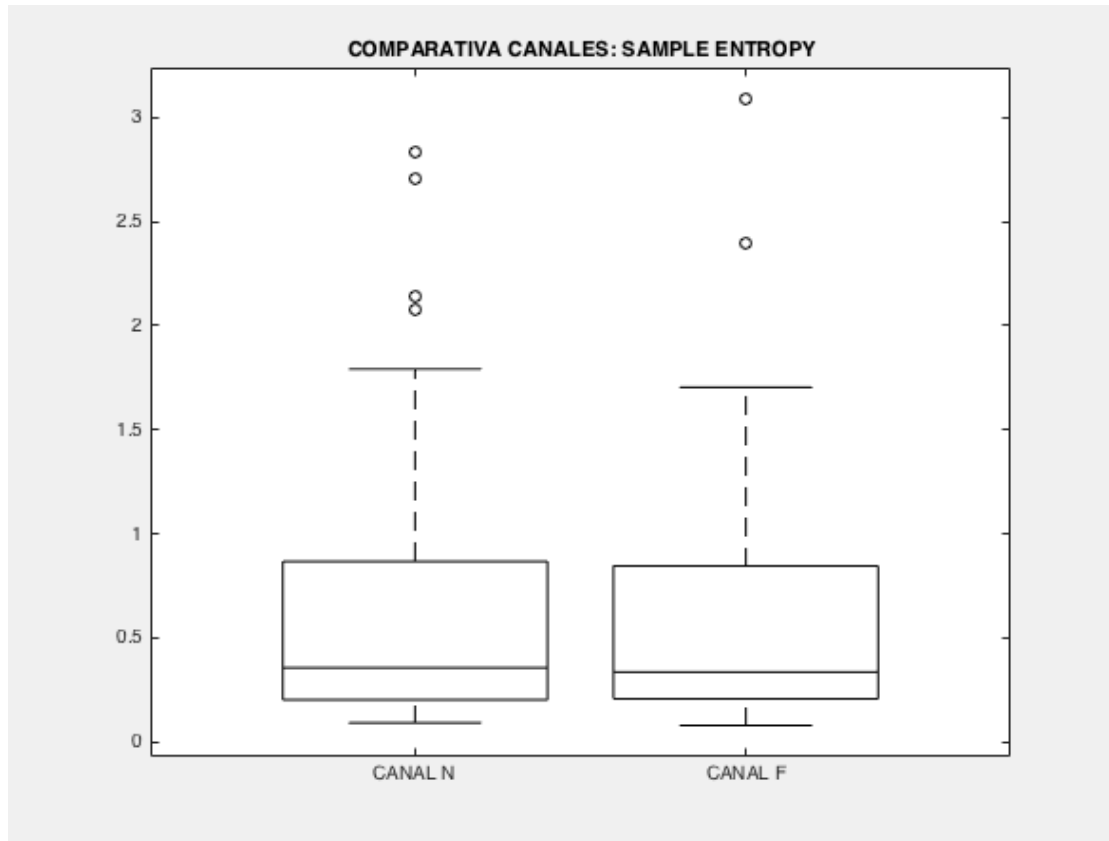


Figura 12: Boxplot Entropía de Muestra

Entropía de permutación

A partir del diagrama de cajas, podemos ver que la media de valores para señales focales y no focales es muy similar. Si nos fijamos en los valores atípicos, puede verse que en la primera se encuentran bajo el mínimo valor dentro de la distribución y para la segunda se encuentran en valores por encima del máximo. Este parámetro nos ayudará a diferenciar esos valores atípicos en señales focales o no focales.

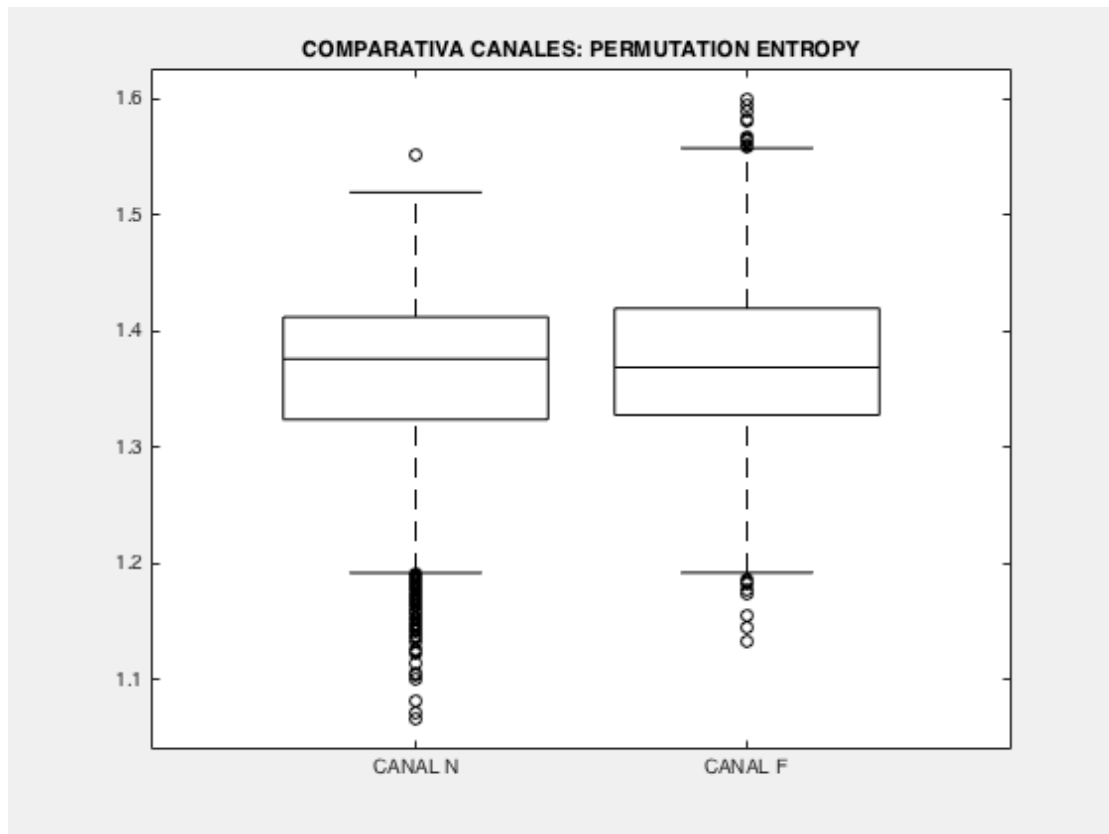


Figura 13: Boxplot Entropía de Permutación

Exponente de Hurst

Este parámetro permite ver, a partir del análisis de una señal, si ésta tiene un comportamiento tendencial, de reversión a la media o simplemente se mueve de forma aleatoria.

- Hurst = 0.5 \Rightarrow La serie es aleatoria
- Hurst < 0.5 \Rightarrow La serie tiene una dinámica de reversión a la media
- Hurst > 0.5 \Rightarrow La serie es tendencial.

Además, si la estimación de Hurst tiene un valor cercano a 1 estaremos frente a una serie con un comportamiento fuertemente tendencial. Por el contrario, si Hurst es cercano a cero tendremos mucha probabilidad de encontrarnos frente a serie que tiende a revertir a la media [8].

Si nuevamente observamos el diagrama de cajas asociado a este parámetro podemos ver que en la mayoría de los casos, para ambos tipos de señales, se trata de una serie tendencial, en la cual el parámetro de Hurst es superior a 0.5.

Puede verse también que, en señales focales, todos sus valores se encuentran muy cercanos a la media, mientras que en no focales son mas dispersos.

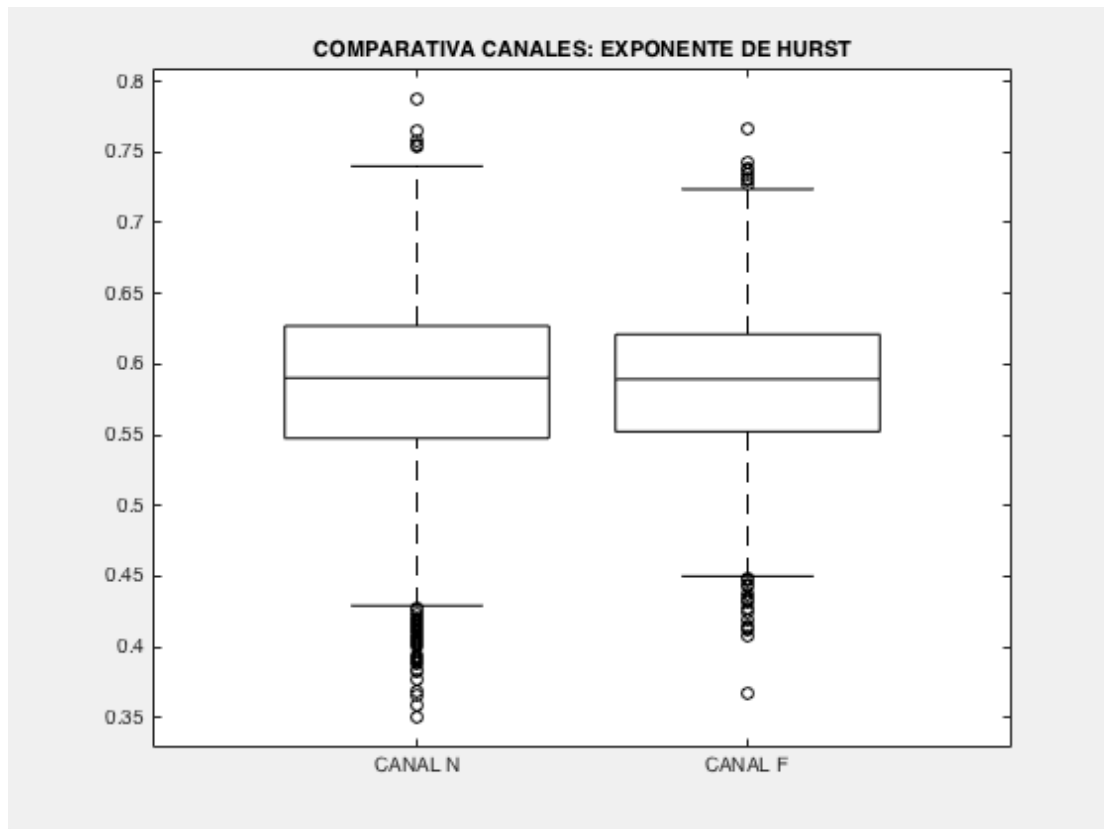


Figura 14: Boxplot Exponente de Hurst

Dimensión fractal

Este parámetro se refiere a una dimensión no entera o fraccionaria de un objeto geométrico y contiene información sobre la estructura geométrica de la señal en múltiples escalas [9].

La dimensión fractal de una forma de onda que representa una poderosa herramienta para la detección de transitorios. En particular, es utilizada en el análisis de señales biomédicas, concretamente en el análisis de electroencefalogramas y electrocardiogramas, utilizando esta característica para así identificar y distinguir estados específicos de función fisiológica.

Si analizamos el diagrama de cajas asociado, podemos ver que la media para ambas señales es muy similar (1,15 y 1,17 respectivamente). De todos modos, para señales no focales, contando con los valores atípicos, el 65% de los valores superan la media, en señales focales es sólo el 50%. Este parámetro será muy útil para la definición y distinción de las señales.

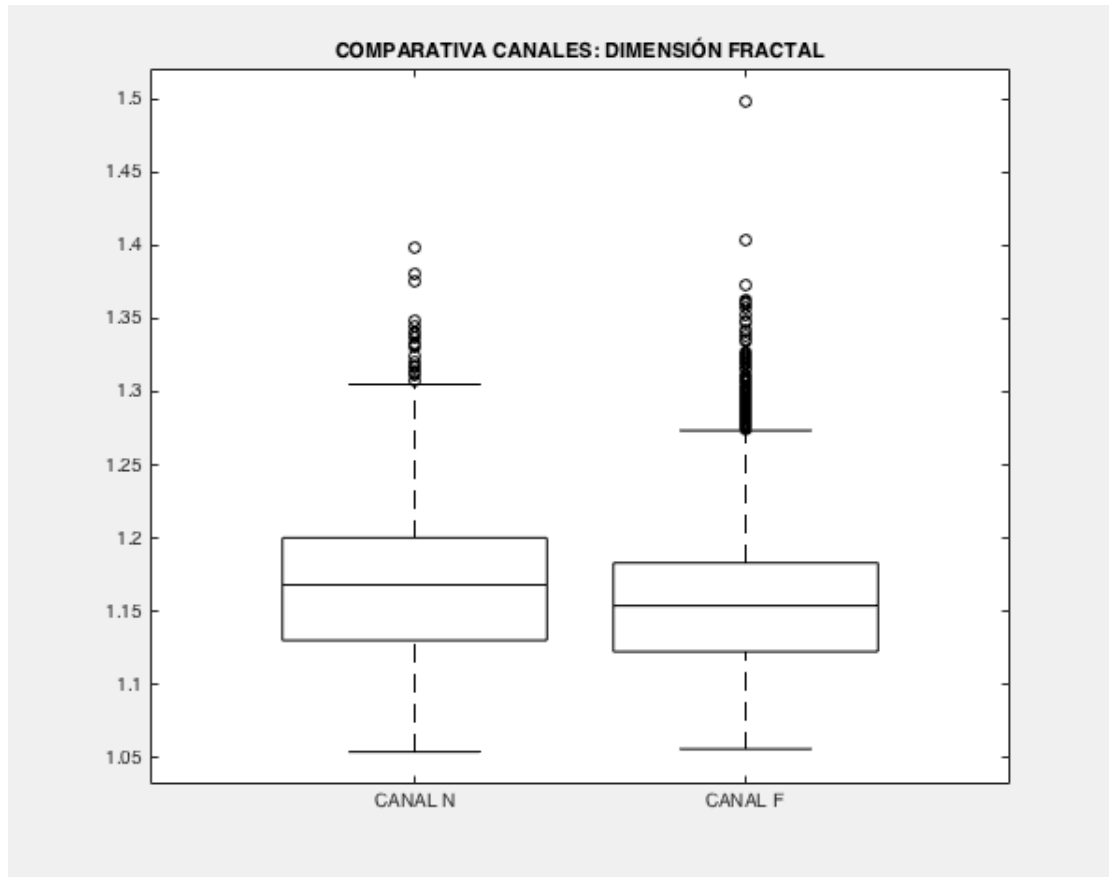


Figura 15: Boxplot Dimensión Fractal

3.2.4. Fase de entrenamiento

Con todo esto tendríamos, asociada a cada una de las señales filtradas, un conjunto de siete parámetros que la definen. Cabe tener en cuenta que estos datos, como ya dijimos, están ya etiquetados.

Por tanto nos encontramos con una matriz de datos, formada por los 42.000 registros y los 7 parámetros asociados (matriz de 42.000x7)

Dividiremos el conjunto de datos en dos bloques, definiendo un bloque de entrenamiento y un bloque para la fase de test. Las distribuciones que se llevarán a cabo se definen en la siguiente tabla:

FASE DE ENTRENAMIENTO	FASE DE TEST
90%	10%
80%	20%
70%	30%
60%	40%
50%	50%

Tabla 2: Distribución de los datos, fase de entrenamiento y fase de test

A partir del bloque de entrenamiento, evaluaremos distintos tipos de Clasificadores, haciendo uso de Matlab y de su aplicación "Classification Learner". Los Clasificadores que nos permitan obtener mejores resultados serán la base para clasificar los datos del bloque asociado a la fase de test. Dado que estos datos ya los tenemos clasificados, podremos ver qué tasa de aciertos se obtiene comparando los valores obtenidos con los ya registrados, generando así los resultados finales del estudio en cuestión.

Los diferentes métodos de clasificación son los que se muestran en la siguiente figura:

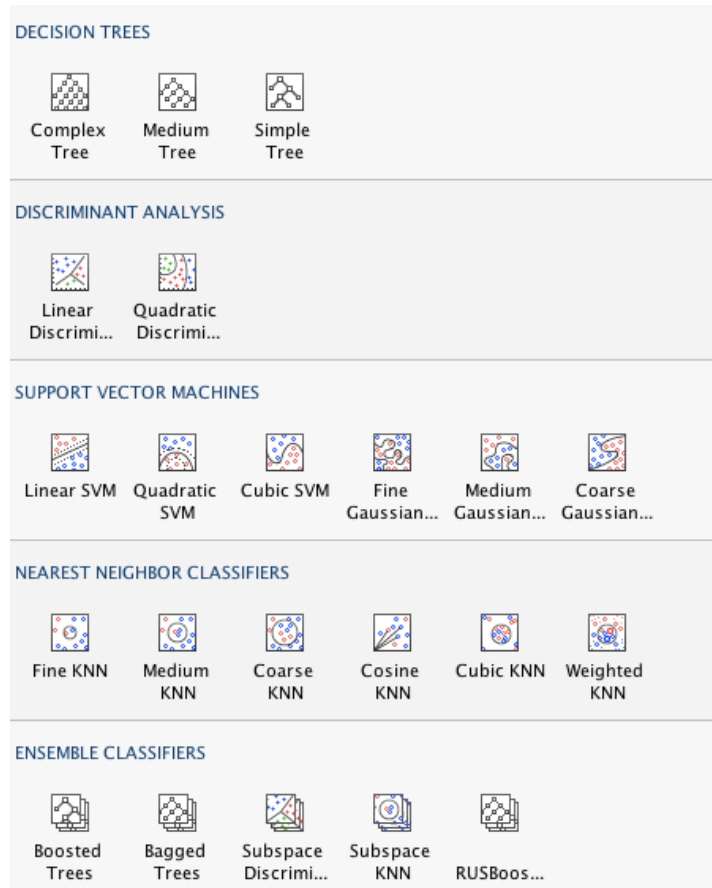


Figura 16: Métodos de clasificación

Árboles de decision (“Decision Trees”)

Es uno de los enfoques de modelado predictivo más utilizadas. Dado un conjunto de datos, se generan diagramas de construcciones lógicas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

Después de crear un árbol, se puede predecir fácilmente las respuestas para los nuevos datos [10].

Análisis discriminante (“Discriminant Analysis”)

El análisis discriminante asume que las diferentes clases generan datos basados en diferentes distribuciones gaussianas. Para entrenar a un clasificador, la función de ajuste estima los parámetros de una distribución gaussiana para cada clase [10].

Máquinas de vectores de soporte, SVM (“Support Vector Machines”)

SVM es un algoritmo de aprendizaje supervisado que se puede emplear para clasificación binaria o regresión. Se basa en la construcción de un hiperplano óptimo en forma de superficie de decisión, de modo que el margen de separación entre las dos clases en los datos existentes se

amplía al máximo. Los vectores de soporte hacen referencia a un pequeño subconjunto de las observaciones de entrenamiento que se utilizan como soporte para la ubicación óptima de la superficie de decisión. Consta de dos fases:

- Transformar los predictores (datos de entrada) en un espacio de características altamente dimensional (truco Kernel)
- Resolver un problema de optimización cuadrática que se ajuste a un hiperplano óptimo para clasificar las características transformadas en dos clases. El número de características transformadas está determinado por el número de vectores de soporte [10].

KNN (“Nearest Neighbor Classifiers”)

Consiste en la categorización de datos según su distancia/similitud a los puntos en un conjunto de datos de entrenamiento, permitiendo de este modo clasificarlos. Para el cálculo de esta distancia se usan múltiples métodos distintos (Chebychev, Hamming, correlación, ...).

Clasificador conjunto (“Ensemble Classifiers”)

Este clasificador ayuda a mejorar los resultados del aprendizaje automático al combinar varios modelos. Así, consiste en un conjunto de algoritmos que combinan varias técnicas de aprendizaje automático en un modelo predictivo para disminuir la varianza, el sesgo o mejorar las predicciones.

Finalmente, puede predecir una respuesta para un nuevo conjunto de datos al agregar las diferentes predicciones.

Las características relacionadas con la velocidad de predicción, uso de memoria e interpretabilidad de los distintos clasificadores, se muestran en la siguiente tabla.

	Velocidad de predicción	Uso de memoria	Interpretabilidad
Árboles de decision	Rápida	Bajo	Fácil
Análisis discriminante	Rápida	Bajo para lineal y grande para cuadrático	Fácil
SVM	Media para lineal y lenta el resto	Bajo para lineal y grande para el resto	Fácil para lineal y difícil para el resto
KNN	Lenta para cúbica y media para el resto	Media	Difícil
Ensemble Classifiers	Media/rápida dependiendo del tipo	Bajo/Alto dependiendo del tipo	Difícil

Tabla 3: Comparación de clasificadores

Procedemos a evaluar los datos de análisis con cada uno de los clasificadores.

Los datos de entrada serán, por cada registro, los siete parámetros calculados y un parámetro adicional binario que indicará si la señal asociada es focal o no focal. Esto nos permitirá entrenar el clasificador y, a partir de ahí, definir un algoritmo que aplicaremos al resto de datos sin clasificar.

En el momento de evaluar un clasificador en función de unos datos de entrada, se obtiene un parámetro que permite ver cual de ellos proporciona mejores resultados: Exactitud ("Accuracy")

Se evalúa para cada clasificador diferentes combinaciones de entrada, usando más (90%) o menos (50%) datos ya etiquetados. Los clasificadores que nos permiten obtener mejores resultados son los siguientes:

- SVM: Fine Gaussian
 - o Si se entrena el 90% de los datos, factor de exactitud: 67,30%
 - o Si se entrena el 50% de los datos, factor de exactitud: 66,40%
- Ensemble Classifiers: Bagged Trees
 - o Si se entrena el 90% de los datos, factor de exactitud: 67,20%
 - o Si se entrena el 50% de los datos, factor de exactitud: 66,00%
- Decision Trees: Complex Tree
 - o Si se entrena el 90% de los datos, factor de exactitud: 64,30%
 - o Si se entrena el 50% de los datos, factor de exactitud: 64,20%

A partir de ahora, en nuestro estudio nos centraremos en estos tres clasificadores, tratando de obtener óptimos resultados en cada una de las combinaciones a analizar.

3.2.5. Fase de test

Una vez tenemos el clasificador definido, damos por finalizada la fase de entrenamiento y pasamos a la fase de test. Partiendo de los clasificadores considerados como óptimos para nuestros datos, obtendremos, por cada registro de siete parámetros, el parámetro binario (octavo parámetro anteriormente citado) que determina el tipo de señal a tratar.

Dado que conocemos previamente la caracterización o etiqueta de cada registro podremos comparar si el resultado es correcto o no. De este modo, cotejaremos el resultado que obtiene nuestro clasificador con el resultado correcto pudiendo ver si se trata de un acierto o un falso positivo.

3.3. Resultados

En la siguiente tabla se muestra una comparación de resultados obtenidos en función del tipo de clasificador y de la repartición de datos entre fase de entrenamiento y fase de test.

		FINE GAUSSIAN SVM		BAGGED TREES		COMPLEX TREE	
		ACCUARACY	% ACIERTOS	ACCUARACY	% ACIERTOS	ACCUARACY	% ACIERTOS
FASE TRAINING	90	67,30%	66,85%	67,20%	65,99%	64,30%	62,70%
FASE TESTING	10						
FASE TRAINING	80	67,10%	67,20%	67,10%	63,47%	63,60%	63,47%
FASE TESTING	20						
FASE TRAINING	70	66,70%	67,76%	66,80%	67,42%	63,90%	64,54%
FASE TESTING	30						
FASE TRAINING	60	66,50%	67,43%	66,30%	67,11%	63,50%	64,21%
FASE TESTING	40						
FASE TRAINING	50	66,40%	67,33%	66,00%	66,77%	64,20%	64,01%
FASE TESTING	50						

Tabla 4: Resultados finales

Puede verse que los porcentajes de acierto mayores se asocian al uso del clasificador SVM, repartiendo los datos de modo que el 70% constituyan la parte de entrenamiento y el 30% formen la parte de testeo.

Podemos ver, como conclusión, que en el mejor de los casos obtendremos un 67,76% de aciertos, es decir, de cada 100 señales a analizar, 67 señales serán clasificadas correctamente como focales o no focales y las 33 restantes se estarán clasificando erróneamente.

3.3.1. Análisis y mejora de resultados

Una vez tenemos los resultados definidos en el apartado anterior recurriremos a técnicas que nos permitan mejorar estos resultados con el fin de obtener la clasificación óptima de los datos.

Dado que el tipo de clasificador ya ha sido escogido como el óptimo, las modificaciones que podemos hacer se centran en los datos que alimentan los clasificadores.

En el Capítulo 2.2 generamos, para cada señal, una serie de parámetros que permitían de manera más o menos determinativa definir el tipo de señal en cuestión. Los parámetros que aportaban menos información distintiva al estudio son:

- Entropía de Shannon
- Entropía de Muestra

Ahora bien, repetimos todo el análisis anteriormente explicado excluyendo del análisis estos dos parámetros. Es decir, para cada señal los parámetros que entrarán en juego serán únicamente:

- Tasa de cruces por cero
- Entropía Aproximada
- Entropía de Permutación
- Exponente de Hurst
- Dimensión Fractal

Analizando los resultados vemos que los valores de porcentaje de acierto se ven alterados a nivel de centésima y que la exclusión de estos parámetros no supone una diferencia sustancial.

La última técnica a la que podemos recurrir para mejorar estos resultados, sería aumentar el conjunto de datos de origen, de modo que el clasificador cuente con una mayor cantidad de datos etiquetados desde el origen, de esto se hablará en el Capítulo 4.

Finalmente, procedemos a comparar nuestros resultados con los obtenidos a partir de otros estudios paralelos.

3.3.2. Comparación con otros estudios

Existen varias tesis que, al igual que nuestro estudio, se basan en técnicas que permiten detectar un tipo de señal EEG desconocido usando simples números o parámetros.

En el artículo que tomamos como partida, [5], se lleva a cabo un procedimiento muy afín al que estamos mostrando en este proyecto. Parte de la misma base de datos y calcula parámetros muy similares, basados igualmente en entropías y cálculos numéricos. En el propio artículo se indica que se llega a obtener, con distintos porcentajes de datos de entrenamiento/test hasta un máximo de 76,6% de accuracy, dato que no dista en gran medida de los máximos obtenidos en nuestro análisis: 67,3%.

Otros estudios, reflejados en [11], manifiestan resultados obtenidos para diferentes combinaciones de parámetros de estudio y tipo de clasificador. Los resultados se reflejan a continuación:

- Entropía de permutación y clasificador SVM:
 - Max accuracy: 75%
- Entropía de muestra promedio (ASE) y varianza promedio de frecuencias instantáneas (AVIF). Clasificador SVM:
 - Max accuracy: 85%

- Entropía de Shannon, Aproximada, Rényi, muestra y fase. Aplicación de funciones de modo intrínseco (IMFs) a las señales. Clasificador SVM:
 - Max accuracy: 87%

Como se puede ver, existen múltiples combinaciones que permitirán obtener, en función de la base de datos origen, mejores o peores resultados en cuanto a la clasificación.

4. CONCLUSIONES

La electroencefalografía se ha utilizado desde hace más de 100 años y hoy en día sigue siendo una herramienta muy útil en el tratamiento, seguimiento y monitorización de las epilepsias y la función cerebral. Supone una práctica indispensable en la actividad clínica diaria del paciente neurológico o de alto riesgo de complicaciones neurológicas.

Como todo examen clínico, su optimización sigue resultando fundamental, surgiendo nuevos desafíos como el aúne de neuroimágenes con los registros electrocardiográficos y video. Otro punto de optimización es la caracterización más rápida y precisa posible, sin retardos y sin fallos, proporcionando un diagnóstico certero y precoz en todos los pacientes.

A lo largo de este proyecto se ha propuesto una forma de caracterización de este tipo de señales cuyos resultados distarían de una obtención exacta de los resultados que hablamos. Es muy importante tener en cuenta la poca distinción que existe morfológicamente entre estas señales, lo cual dificulta enormemente su caracterización.

Además, la necesidad de grandes bases de datos etiquetados, cuyo tratamiento genera largos tiempos de ejecución y crea la necesidad de máquinas con altas capacidades de computación, ha dificultado la ejecución final. Se ha tratado con múltiples técnicas de optimización de tiempo y código pero no se han conseguido reducir tiempos de ejecución que han alcanzado semanas.

5. LÍNEAS FUTURAS

Como principal línea futura estaría la aplicación de técnicas que permitan mejorar los resultados que actualmente tenemos. Entre estas técnicas estaría la aplicación de Data Augmentation, en [12] se hace referencia a este análisis. Este método consistiría en la creación de registros artificiales como sustitución de algunos reales, entrenar clasificadores con estos nuevos datos y compararlos con los resultados que se tienen actualmente. Según el artículo citado se lograría reemplazar hasta la mitad de los datos con datos artificiales.

Esta técnica sería interesante de tratar en este proyecto, pero la carga computacional requerida, dificulta notablemente su ejecución.

Otra línea futura de gran interés sería aplicar los métodos que se han empleado para la Epilepsia, en otro tipo de patologías relacionadas con señales EEG anormales o atípicas, como puede ser autismo o Alzheimer.

6. BIBLIOGRAFÍA

- [1] <https://imotions.com/blog/what-is-eeeg/>
- [2] Asociación de epilepsia:
<https://www.apicepilepsia.org/tratamiento-de-la-epilepsia/diagnostico-de-la-epilepsia/>
- [3] Siuly, S., Li, Y., & Zhang, Y. (2016). Electroencephalogram (EEG) and Its Background. In *EEG Signal Analysis and Classification* (pp. 3-21). Springer, Cham.
- [4] Organización mundial de la salud: <https://www.who.int/es/>
- [5] Zhao, X., Tanaka, T., Kong, W., Zhao, Q., Cao, J., Sugano, H., & Yoshiday, N. (2018, November). Epileptic Focus Localization Based on iEEG by Using Positive Unlabeled (PU) Learning. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 493-497). IEEE.
- [6] Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of neuroscience methods*, 123(1), 69-87.
- [7] R. G. Andrzejak, K. Schindler, and C. Rummel, "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients," *Physical Review E*, vol. 86, no. 4, p. 046206, 2012.
- [8] <https://estrategiastrading.com/exponente-de-hurst/>
- [9] Maragos, P., & Sun, F. K. (1993). Measuring the fractal dimension of signals: morphological covers and iterative optimization. *IEEE Transactions on signal Processing*, 41(1), 108.
- [10] <https://es.mathworks.com/help/stats/choose-a-classifier.html#bunt0ky>
- [11] Sharma, Rajeev & Pachori, Ram & Acharya, U Rajendra. (2015). An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures. *Entropy*. 17. 5218-5240. 10.3390/e17085218.
- [12] Dinarès-Ferran, J., Ortner, R., Guger, C., & Solé-Casals, J. (2018). A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI. *Frontiers in neuroscience*, 12, 308.

7. ANEXOS

Se incluye, por ultimo el conjunto de códigos empleados para la composición global del proyecto:

```
PARAMETROS=[];
for archivo=1:3750

%% CARGA DE DATOS NO FOCALES
if archivo < 10
    filename =sprintf('SENALES/Data_N_Ind_1_999/Data_N_Ind000%d.txt',archivo);
else
    if archivo < 100
        filename = sprintf('SENALES/Data_N_Ind_1_999/Data_N_Ind00%d.txt',archivo);
    else
        if archivo < 1000
            filename = sprintf('SENALES/Data_N_Ind_1_999/Data_N_Ind0%d.txt',archivo);
        else
            filename =
sprintf('SENALES/Data_N_Ind_1000_3750/Data_N_Ind%d.txt',archivo);
        end
    end
end
Data = load(filename);
Data = Data(:,1);
tipo=filename(14);
PARAM=Creador_Parametros(Data,tipo);
PARAMETROS=[PARAMETROS;PARAM];

%% CARGA DE DATOS FOCALES
if archivo < 10
    filename = sprintf('SENALES/Data_F_Ind_1_999/Data_F_Ind000%d.txt',archivo);
else
    if archivo < 100
        filename = sprintf('SENALES/Data_F_Ind_1_999/Data_F_Ind00%d.txt',archivo);
    else
        if archivo < 1000
            filename = sprintf('SENALES/Data_F_Ind_1_999/Data_F_Ind0%d.txt',archivo);
        else
            filename =
sprintf('SENALES/Data_F_Ind_1000_3750/Data_F_Ind%d.txt',archivo);
        end
    end
end
Data = load(filename);
Data = Data(:,1);
tipo=filename(14);
PARAM=Creador_Parametros(Data,tipo);
PARAMETROS=[PARAMETROS;PARAM];

end

fileID = fopen('PARAMETROS.txt','w');
[A,B]=size(PARAMETROS);
for i=1:A
    fprintf(fileID,'%d\t%d\t%d\t%d\t%d\t%d\t%d\t%d\n',PARAMETROS(i,1),PARAMETROS
(i,2),
    PARAMETROS(i,3),PARAMETROS(i,4),PARAMETROS(i,5),PARAMETROS(i,6),PARAMETROS(i
,7),
    PARAMETROS(i,8));
end
fclose(fileID);
39
```

```

function PARAMETROS=Creador_Parametros(Data,tipo)

%% FILTRADO PASO BANDA DE LA SEÑAL. 6 FILTROS DISTINTOS.
% A cada segmento de señal de 20segundos (10.240 muestras a una fs= 512Hz)
% se le aplican 6 filtros paso banda distintos, obteniendo 6 segmentos
% distintos (SF1 - SF6) a partir de uno `nico (S).

S=Data;
fs=512;    % Frecuencia de muestreo
fn=fs/2;  % Frecuencia de Nyquist
Rp=3;Rs=20; % Banda de paso y banda eliminada

subplot(7,3,[2 3]);plot(1:10240,S);axis tight;hold on;line([0 10240],[0 0]);
%Representación señal original
title('Señal original');

% Delta (0.5- 4 Hz)
wp=[0.5 4]/fs;ws=[0.2 8]/fs;
[n,wn]=buttord(wp,ws,Rp,Rs);[b,a]=butter(n,wn);
%Diseño filtro
subplot(7,3,4);[h,f]=freqz(b,a,[],fs);plot(abs(h));axis tight;
%Representación filtro
title('F1: Delta (0.5- 4 Hz)')
SF1 = filtfilt(b,a,S);
%Filtrado de la señal
subplot(7,3,[5 6]);plot(1:10240,SF1);axis tight;hold on;line([0 10240],[0 0]);
%Representación SF1
title('SF1: Señal filtrada con Delta');

% Theta (4-8 Hz)
wp=[4 8]/fs;ws=[1 12]/fs;
[n,wn]=buttord(wp,ws,Rp,Rs);[b,a]=butter(n,wn);
%Diseño filtro
subplot(7,3,7);[h,f]=freqz(b,a,[],fs);plot(abs(h));axis tight;
%Representación filtro
title('F2: Theta (4-8 Hz)')
SF2 = filtfilt(b,a,S);
%Filtrado de la señal
subplot(7,3,[8 9]);plot(1:10240,SF2);axis tight;hold on;line([0 10240],[0 0]);
%Representación SF2
title('SF2: Señal filtrada con Theta');

% Alpha (8-13 Hz)
wp=[8 13]/fs;ws=[6 17]/fs;
[n,wn]=buttord(wp,ws,Rp,Rs);[b,a]=butter(n,wn);
%Diseño filtro
subplot(7,3,10);[h,f]=freqz(b,a,[],fs);plot(abs(h));axis tight;
%Representación filtro
title('F3: Alpha (8-13 Hz)')
SF3 = filtfilt(b,a,S);
%Filtrado de la señal
subplot(7,3,[11 12]);plot(1:10240,SF3);axis tight;hold on;line([0 10240],[0 0]);
%Representación SF3
title('SF3: Señal filtrada con Alpha');

% Beta (13-30 Hz)
wp=[13 30]/fs;ws=[9 38]/fs;
[n,wn]=buttord(wp,ws,Rp,Rs);[b,a]=butter(n,wn);
%Diseño filtro
subplot(7,3,13);[h,f]=freqz(b,a,[],fs);plot(abs(h));axis tight;
%Representación filtro
title('F4: Beta (13-30 Hz)')
SF4 = filtfilt(b,a,S);
%Filtrado de la señal

```



```

subplot(7,3,[14 15]);plot(1:10240,SF4);axis tight;hold on;line([0 10240],[0 0]);
%Representación SF4
title('SF4: Señal filtrada con Beta');

% Gamma (30-80 Hz)
wp=[30 80]/fs;ws=[25 95]/fs;
[n,wn]=buttord(wp,ws,Rp,Rs);[b,a]=butter(n,wn);
%Diseño filtro
subplot(7,3,16);[h,f]=freqz(b,a,[],fs);plot(abs(h));axis tight;
%Representación filtro
title('F5: Gamma (30-80 Hz)')
SF5 = filtfilt(b,a,S);
%Filtrado de la señal
subplot(7,3,[17 18]);plot(1:10240,SF5);axis tight;hold on;line([0 10240],[0 0]);
%Representación SF5
title('SF5: Señal filtrada con Gamma');

% Ripple (80-150 Hz)
wp=[80 150]/fs;ws=[74 162]/fs;
[n,wn]=buttord(wp,ws,Rp,Rs);[b,a]=butter(n,wn);
%Diseño filtro
subplot(7,3,19);[h,f]=freqz(b,a,[],fs);plot(abs(h));axis tight;
%Representación filtro
title('F6: Ripple (80-150 Hz)')
SF6 = filtfilt(b,a,S);
%Filtrado de la señal
subplot(7,3,[20 21]);plot(1:10240,SF6);axis tight;hold on;line([0 10240],[0 0]);
%Representación SF6
title('SF6: Señal filtrada con Ripple');

%% OBTENCIÓN DE PARÁMETROS PARA CADA UNO DE LOS 6 SUBSEGMENTOS

SF = cellstr(['SF1';'SF2';'SF3';'SF4';'SF5';'SF6']);
PARAMETROS_SF=[];

emb_dim=5;r=0.2; %Embedding dimension and Tolerance parameter
order=3; %Orden de permutación de entropía
t=1; %Delay time of permutation entropy

for m=1:6
    SFX=eval(char(SF(m)));
    N=length(SFX);

    %% PARÁMETRO 1:TASA DE CRUCES POR CERO
    cruces=[];
    for i=1:10239
        if (SFX(i)<0 && SFX(i+1)>0) || (SFX(i)>0 && SFX(i+1)<0)
            cruces=[cruces i];
        end
    end
    RATIO_CRUCES_CERO=round(length(cruces)/length(Data)*100,2);
    PARAMETROS(m,1)=RATIO_CRUCES_CERO;
    %% PARÁMETRO 2:ENTROPÍA DE SHANNON (Shannon entropy)
    PARAMETROS(m,2)=wentropy(SFX,'shannon')/10^7;

    %% PARÁMETRO 3: ENTROPÍA APROXIMADA (Approximate entropy)
    tVecs = zeros(N - emb_dim+1, emb_dim + 1);
    for i = 1 : (N - emb_dim)
        tVecs(i, :) = SFX(i:(i + emb_dim));
    end
    phi_m_r = zeros(1,2);
    ss = 0;
    for mm = [emb_dim (emb_dim + 1)]
        ss = ss + 1;
        tVecsM = tVecs(1:(N - mm + 1), 1:mm);% Get the matrix that we need
    end
end

```

```

C = []; % Calculate distances between each pair of template vectors
for i = 1 : length(tVecsM)
    % Calculate the Chebyshev distance
    dsts = max(abs(tVecsM-repmat(tVecsM(i,:),size(tVecsM,1),1)),[],2);
    % It considers self-matching.
    % Compute the conditional probability
    C = [C sum(dsts < r)/(N - mm)];
end
phi_m_r(ss) = sum(log(C))/(N - mm); % compute sum of log probabilities
end
y = phi_m_r(1) - phi_m_r(2);
PARAMETROS(m,3)=y;

%% PARíMETRO 4: ENTROPÍA DE MUESTRA (Sample entropy)
tVecs = zeros(N - emb_dim+1, emb_dim + 1);
for i = 1 : (N - emb_dim)
    tVecs(i, :) = SFX(i:(i + emb_dim));
end
B_m_r = zeros(1,2);
ss = 0;
for mm = [emb_dim (emb_dim + 1)]
    ss = ss + 1;
    tVecsM = tVecs(1:(N - mm + 1), 1:mm); % Get the matrix that we need
    % Calculate distances between each pair of template vectors
    for i = 1 : length(tVecsM)
        % Calculate the Chebyshev distance
        dsts = max(abs(tVecsM-repmat(tVecsM(i,:),size(tVecsM,1),1)),[],2);
        dsts(i) = []; % Exclude self-matching
        % Compute the sum of conditional probabilities
        B_m_r(ss) = B_m_r(ss) + sum(dsts < r)/(N - mm - 1);
    end
end
% Compute log of summed probabilities
y = -log(B_m_r(2) / B_m_r(1));
PARAMETROS(m,4)=y;

%% PARíMETRO 5: ENTROPÍA DE PERMUTACIÓN (Permutation entropy)
permlist = perms(1:order);
c(1:length(permlist))=0;

for j=1:N-t*(order-1)
    [a,iv]=sort(SFX(j:t:j+t*(order-1)));
    for jj=1:length(permlist)
        if (abs(permlist(jj,:)-iv')==0
            c(jj) = c(jj) + 1 ;
        end
    end
end
hist = c;

c=c(find(c~=0));
p = c/sum(c);
pe = -sum(p .* log(p));
PARAMETROS(m,5)=pe;

%% PARíMETRO 6: EXPONENTE DE HURST
dat=SFX;
len=length(dat);
yvals=zeros(1,len);xvals=zeros(1,len);
data2=zeros(1,len);
index=0;
binsize=1;
while len>4

    y=std(dat');

```

```

index=index+1;
xvals(index)=binsize;
yvals(index)=binsize*y;

len=fix(len/2);
binsize=binsize*2;
for i=1:len % average adjacent points in pairs
    data2(i)=(dat(2*i)+dat((2*i)-1))*0.5;
end
dat=data2(1:len);

end % while
xvals=xvals(1:index);yvals=yvals(1:index);
logx=log(xvals);logy=log(yvals);
p2=polyfit(logx,logy,1);
PARAMETROS(m,6)=p2(1); % Hurst exponent is the slope of the linear fit of log-log
plot

%% PARAMETRO 7: DIMENSIÓN FRACTAL
L=0;
n = N - 1; %'n' is the number of steps in the waveform.
for i = 1:(N - 1)
    aux = sqrt(1 + ((SFX(i) - SFX(i+1))^2));
    L = L + aux;
    clear('aux');
end
% Computing 'd':
% 'd' is the planar extent or diameter of the waveform. It is estimated as
% the distance between the first point of the sequence and the point of the
% sequence that provides the farthest distance: d = max(dist(1,i)); i=2...N.
dist = NaN(1,N-1); %Predifinition variable for computatinoal efficiency.
for i = 2:N
    dist(i) = sqrt((1 - i)^2 + ((SFX(1) - SFX(i))^2));
end
d = max(dist);
PARAMETROS(m,7) = log10(n)/(log10(n) + log10(d/L));

%%RESPONSE
if tipo=='N'
    PARAMETROS(m,8)=0;
elseif tipo=='F'
    PARAMETROS(m,8)=1;
end
end
end

```