



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA ELECTRÓNICA

**MAESTRÍA EN INGENIERÍA ELECTRÓNICA,
OPCIÓN INSTRUMENTACIÓN ELECTRÓNICA**

TESIS

**PARA OBTENER EL GRADO DE MAESTRO EN INGENIERÍA
ELECTRÓNICA**

**IMPLEMENTACIÓN DE UN ALGORITMO DE CLASIFICACIÓN DE
EPILEPSIA A PARTIR DE SEÑALES ELECTROENCEFALOGRÁFICAS**

P R E S E N T A :

ING. SERGIO CADENA FLORES*

DIRECTORES:

DRA. MARÍA MONSERRAT MORÍN CASTILLO

DR. JOSÉ ÁNGEL ARIAS CRUZ

Agradecimientos

Agradecimiento a los miembros de mi jurado de tesis:

Dr. Josefina Castañeda Camacho, Dr. José Eligio Moisés Gutiérrez Arias, Dra. Selene Edith Maya Rueda. Por todo el tiempo empleado en mi proceso de evaluación y observaciones.

A mis asesores:

Dra. María Monserrat Morín Castillo y Dr. José Ángel Arias Cruz, les agradezco su confianza, el apoyo y la paciencia durante todo este tiempo de desarrollo y por permitirme trabajar con ustedes.

También extendiendo mi agradecimiento al Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCYT) por el apoyo económico que me brindó el tiempo en el que realice mis estudios de posgrado

Así como a la Vicerrectoría de Investigación y Estudios de Posgrado (VIEP) de la Benemérita Universidad Autónoma de Puebla.

Índice general

1. Introducción	8
Introducción	8
Historia de la electroencefalografía	8
Objetivos	10
Objetivo General	10
Objetivo Específicos	10
Estado del Arte	11
Justificación	14
Descripción y Diagrama a Bloques	15
2. Epilepsia	17
2.1. El Cerebro	17
2.2. ¿Que es el Electroencefalograma?	20
2.2.1. Sistema 10-20	20
2.2.2. Tipo de ondas cerebrales.	22
2.3. Clasificación de crisis epiléptica	24
2.4. Clasificación de la epilepsia	27
3. Marco Teórico	30
3.1. Ramas de la Inteligencia Artificial(IA)	30
3.2. Transformada de Wavelet	31
3.3. Características Matemáticas	34
3.4. Base de datos Berna Barcelona.	37
3.4.1. Extracción de características	39
3.5. Base de datos de la Universidad de Bonn	41
3.5.1. Características	41
3.5.2. Espacio multidimensional	42

4. Clasificación ML	44
4.1. ¿Que es el Machine Learning?	44
4.2. Tipos de Clasificadores	45
4.3. Matriz de Confusión	51
4.4. Validación Cruzada	53
5. Base de datos sintética	54
5.1. Generación de señales sintética epiléptica y basal	54
5.1.1. Parámetros de las señales sintéticas	55
5.1.2. Generación de señales sintéticas	56
5.2. Algoritmo para la generación de señales sintéticas	59
5.3. Desarrollo del algoritmo	62
5.3.1. Algoritmo de clasificación	62
5.3.2. Desarrollo del Algoritmo usando señales sintéticas	66
5.4. Resultados	68
5.5. Conclusiones	74
5.6. Trabajo a futuro	75
Bibliografía	75
Publicaciones	81
Apéndices	88
A.	88
A.1. Distintos enfoque de modelos matemático	88
A.1.1. Formas esféricas de la cabeza	88
A.1.2. Representaciones detalladas y precisas de la anatomía craneal	89
A.1.3. Modelo de medio conductor para la cerebro-cabeza	89
A.2. Representación esquemática del cerebro.	90
A.3. Problema directo de fuente localizada	91
A.4. Resolución del problema directo	96
A.5. Señales sintéticas de EEG en dos dimensiones	97
B.	99
B.1. Programa principal de clasificación de la Universidad de Bonn	99
B.2. Programa secundario de características en función del tiempo	101
B.3. Programa secundario de características en función de la frecuencia.	102

B.4. Programa principal para la creación de señales sintéticas	103
---	------------

Índice de figuras

1.1. Hans Berger (1873-1941) en 1927, con su primer registro de EEG.	9
1.2. Diagrama a bloques del procesamiento de señales EEG	16
2.1. Principales enfermedades neurológicas según la OMS	17
2.2. Hemisferios Cerebrales	18
2.3. Estructura de protección del cerebro	19
2.4. Forma y partes de una neurona	19
2.5. Ubicación de electrodos, sistema 10-20	21
2.6. Ritmo de ondas cerebrales así como la amplitud de cada onda	22
2.7. Interpretación de ritmo cerebral depende de la ubicación	24
2.8. Clasificación de las crisis convulsivas datos proporcionados por la ILAE	28
3.1. Ramas de la Inteligencia Artificial.	30
3.2. Descomposición de una señal muestra al aplicar la transformada wavelet, lado izquierdo diagrama de bloques y del lado derecho la aplicación del diagrama a la señal.	32
3.3. Señal azul electrodo (x) Señal Focal, señal naranja electrodo (y) señal focal	38
3.4. Señal azul electrodo (x) Señal No Focal, señal naranja electrodo (y) señal No focal	38
3.5. Los electrodos señalados como Fx (Focal x), Fy(focal y) y para No focal es (Nx y Ny)	39
3.6. Metodología del procesamiento de la señal EEG	40
3.7. Concentrado de datos de la universidad de Bonn	41
3.8. Características en un espacio multidimensional	42
4.1. Aplicando validación cruzada a conjunto de datos	53
5.1. Representación de una descarga focal de duración de 5 segundos.	57
5.2. Representación de un segmento de la señal sintética normal y epiléptica.	58
5.3. Descarga epiléptica focal en distintos momentos en función del tiempo	59

5.4. Diagrama de flujo de señales sintéticas	60
5.5. Barra de herramientas de Machine Learning incluida en MATLAB	64
5.6. Inicio para abrir información en Workspace	64
5.7. Workspace de matlab	65
5.8. Resultados finales del Workspace	66
5.9. Descomposición de una señal sintética de tipo normal	69
5.10. Descomposición de una señal sintética de tipo epiléptica	69
5.11. Representación del conjunto de 30 características	70
5.12. Matrices de confusión utilizados para la clasificación de los datos de la Universidad de Bonn	71
5.13. Resultados de la clasificación de la base de datos de la Universidad de Bonn	71
5.14. Matrices de confusión utilizados para la clasificación de la base de datos sintética .	72
5.15. Resultados de la clasificación de la base de datos sintética	72
5.16. Datos de la Universidad de Bonn con distintos Investigadores y sus Métricas	73
A.1. Representación de la cabeza, de forma eléctrica por capaz	90
A.2. Representación de los lóbulos del cerebro	98

Índice de tablas

Capítulo 1

Introducción

El propósito de este capítulo es presentar una reseña histórica de como el avance de la tecnología ha tenido una repercusión directa en el estudio de los registros electroencefalográficos (EEG), se mencionan aspectos como el desarrollo de los sistemas de registro de las señales EEG, sus orígenes, precursores, y su impacto en la medicina. Actualmente existen diferentes dispositivos que apoyan a un mejor y rápido diagnóstico en el padecimiento de trastornos neurológicos; esto ha llevado a utilizar nuevas herramientas entre como son la Inteligencia Artificial (IA) y de entre los que destacan el Machine Learning (ML) que ha permitido el desarrollo de clasificadores de patrones como un método alternativo que coadyuve al diagnóstico de tales enfermedades neurológicas, una de ellas es la epilepsia, por esta razón nuestro trabajo esta enfocado a la clasificación de la epilepsia.

Historia de la electroencefalografía

Es bien conocido que el electroencefalograma es el registro de las ondas cerebrales por medio de un electroencefalógrafo, esta investigación se inicio en el año de 1770 por el investigador italiano Luigi Galvani, quien presentó sus investigaciones sin tener ningún impacto hasta que su sobrino Luigi Giovanni Aldini, 50 años después confirmara los resultados. Esto despertó el interés en otros investigadores como lo fue al científico alemán Emil du Bois Raymond, quien describió el fenómeno eléctrico, y lo plasmó en su libro en dos volúmenes titulado “Investigaciones sobre la electricidad animal” en 1848 y 1849 respectivamente.

Por otra parte los primeros descubrimientos que describieron la existencia de una actividad eléctrica en el cerebro fue realizada por el fisiólogo Richard Caton, quien tuvo un gran interés en los estudios de Raymond. Caton también estuvo al tanto del trabajo de los doctores Edward Hitzig y

Gustav Theodor Fritsch demostrando que se presentaba actividad motora al aplicar choques eléctricos en algunos lugares específicos del cerebro de perros [1], [2].

Caton en 1875 fue el primero quien observó la actividad eléctrica cerebral, continua y espontánea, menciona haber encontrado “corrientes eléctricas en la sustancia gris” [2].

En su tesis doctoral (1891), el doctor Beck se enfocaba en la existencia corrientes eléctricas en el cerebro y en la médula espinal.

Para este tiempo aunque se utilizaban galvanómetros (los cuales fueron ocupados cuando Caton realizó sus experimentos), la respuesta en frecuencias del equipo utilizado por Beck y Cybulsky no tuvo la capacidad de respuesta esperada.

A principios del siglo pasado, los rusos Pavel Kaufman (1912) y Pradvich Neminski (1913) establecieron en sus investigaciones que los potenciales eléctricos cerebrales se pueden registrar sobre el cráneo. Para esto se realizaron experimentos con perros a los que se les realizaron craneotomías, colocando electrodos en la corteza cerebral, lograron visualizar la actividad electroencefalográfica epileptiforme que fue registrada por primera vez, como se muestra en [1], [2].

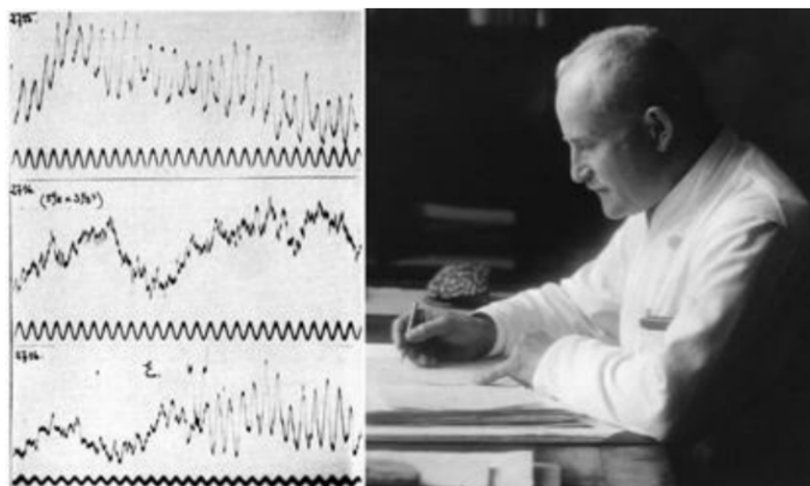


Figura 1.1: Hans Berger (1873-1941) en 1927, con su primer registro de EEG.

Cabe mencionar que, Hans Berger ver Figura 1.1 obtuvo su doctorado en la Universidad de Jena en el área psiquiátrica [3]. Berger, al inicio de sus investigaciones no contaba con el equipo necesario ni la experiencia como neurofisiólogo y sus conocimientos sobre electromecánica eran muy limitados. Para hacer el registro electroencefalográfico que provienen de humanos, utilizaba electrodos de aguja y un galvanómetro de cuerda con un espejo en el que se reflejaba luz, esto permitía la

exposición de papel fotográfico de bromuro de plata que hacia un recorrido de 3 cm por segundo (la misma velocidad que utilizamos hoy en día).

Berger bautizó a sus registros con el nombre de Electroencefalograma (que reemplazó al de electrocerebrograma el cual fue aportación de Nemiski), a pesar de saber la importancia de su descubrimiento, Berger publicó su descubrimiento 5 años después, debido a que él no estaba seguro de sus resultados, trabajó repitiendo muchas pruebas, en diferentes pacientes, incluso sobre sí mismo. En 1929 publicó su descubrimiento: Actividad eléctrica cerebral espontánea en humanos, relató en su publicación “En consecuencia, creo que he descubierto el Electroencefalograma del hombre” y lo reveló ahí por primera vez, descubriendo con esto los ritmos cerebrales y registrando y nombrando las de mayor voltaje y menor frecuencia fueron ondas alfa, las de menor voltaje y mayor frecuencia, ondas beta [4], [5].

En 1931 se refirió a la frecuencia con la que se observa la actividad electroencefalográfica anormal en pacientes con epilepsia y registró por primera vez actividad de punta-onda. Confirmó la presencia de las ondas alfa en región occipital y bautizó dicha actividad con el nombre de ritmo de Berger y se retiró del campo de la investigación 1938 [2], [5].

Objetivo General

- Implementar un algoritmo de clasificación de tipos de epilepsia a partir de sus características.

Objetivos Específicos

- Estudiar los diferentes tipos de epilepsia.
- Identificar las características de señales electroencefalográficas.
- Examinar varias técnicas de clasificación de Machine Learning para seleccionar la más apropiada.
- Analizar un modelo matemático que permita generar focos epilépticos sintéticos y aplicarle las técnicas de clasificación de Machine Learning .
- Desarrollar e implementar el algoritmo de clasificación.
- Comprobar los algoritmos mediante validación cruzada

Estado del Arte

El avance tan acelerado de las tecnologías en el área de la computación y la medicina en la última década ha traído como consecuencia que cada vez sea más sencillo detectar trastornos neurológicos, la detección temprana de la epilepsia da a los pacientes la oportunidad del control temprano y así poder suavizar los efectos de la enfermedad y dar un diagnóstico más exacto. En los siguientes trabajos de investigación se han abordado temas referentes a las técnicas de ML (Machine Learning), como ejemplo se mencionarán los trabajos en los cuales el procesamiento de las señales electroencefalográficas y el aprendizaje automático ha sido primordial para dar un diagnóstico más preciso, volviéndose un procedimiento crucial y desafiante.

En el año 2010 la investigación del doctor Abdulhamit, propone en su trabajo un marco de análisis y procesamiento de señales, para descomponer las frecuencias en sub-bandas utilizando la transformada discreta de wavalet (DWT) [6], para con ella clasificar estas frecuencias con el aprendizaje supervisado, donde utilizó la máquinas de vectores soporte (SVM) [7], y como aprendizaje no supervisado las técnicas para reducir un conjunto de datos con el análisis de componentes principales (PCA) [8], el análisis de componentes independientes (ICA) [9] y el análisis de discriminación lineal (LDA) [10], el doctor Abdulhamit explica que el diagnóstico de la epilepsia es una tarea difícil ya que se requiere de la observación del paciente. Como resultado de la clasificación de señales EEG se extrajeron las características no lineales mejorando el rendimiento del clasificador y el número de vectores [11].

En el año 2010 el doctor Kai fu Jianfeng, propone en su trabajo una nueva técnica para la clasificación de convulsiones obtenidas de las señales electroencefalográficas, (EEG) [12] utilizando la transformada Hilbert-Huang (HHT) [13] y la máquina de vectores soporte (SVM) [7] y los resultados presentados muestran que la mejor precisión de clasificación promedio de este algoritmo puede alcanzar el 99,125 % con el ritmo theta de las señales de EEG [14].

En el 2011 el doctor Weidong Zhou, propuso un nuevo enfoque de clasificación EEG basado en la máquina de aprendizaje extremo (ELM) [15] y características dinámicas no lineales. La teoría de la dinámica no lineal ha sido una poderosa herramienta para comprender las actividades eléctricas del cerebro, las características no lineales extraídas de las señales de EEG, como la entropía aproximada (ApEn) [16], el exponente de Hurst y el exponente de escala obtenido con el análisis de fluctuación sin tendencia (DFA) [17], se empleo para caracterizar los electroencefalogramas

(EEG) [12] interictales (relativo al intervalo de tiempo comprendido entre dos episodios de epilepsia), ictales (es el período de tiempo en el cual una persona esta teniendo una convulsión) y perictales (que indica los estados epilépticos preictal, ictal y posictal). Los experimentos del doctor Zhou demostraron que, en comparación con el algoritmo de retro propagación (BP) [18] y la máquina de vectores soporte (SVM) [7], el rendimiento de la máquina de vectores extremo (ELM) [15] es mejor en términos de tiempo de entrenamiento y precisión de clasificación del 95.5 % [19].

En el año 2017 el doctor Elly Matul Imah, demostró en su estudio al comparar algunos algoritmos con el algoritmo de aprendizaje automático que combinan la extracción de características y el algoritmo de clasificación para la identificación de ataques de epilepsia basados en datos de electroencefalogramas (EEG) [12]. Los algoritmos de clasificación comparados en este estudio son la cuantificación de vectores de aprendizaje de relevancia generalizada (GRLVQ) [20], la retro propagación, máquina de vectores soporte (SVM) [7] y bosque aleatorio, combinados con Wavelet y la extracción de características de análisis de componentes principales (PCA) [8], los resultados que obtuvo el doctor Elly Matul Imah se observan en el rendimiento y el reconocimiento de las cinco clases en el conjunto de datos de ataques epilépticos de EEG es GRLVQ, con una exactitud, precisión y recuperación de 0,9866 [21].

En el año 2020 el doctor Wail Mardini, propuso un marco para detectar ataques epilépticos a partir de señales EEG registradas en pacientes normales y epilépticos. El enfoque sugerido está diseñado para clasificar automáticamente la señal anormal de la normal. Este trabajo tiene como objetivo mejorar la precisión de la detección de ataques epilépticos y reducir los costos computacionales en donde el doctor Mardini utilizo los siguientes clasificadores Maquina de Vectores Soporte (SVM) [7], Vecinos mas cercanos (KNN) [22], Red neuronal artificial (ANN) [23] y Bayesiano ingenuo (NB) [24]. Los resultados experimentales muestran que los cuatro clasificadores producen resultados comparables para las características estadísticas derivadas de la wavelet [25].

En el año 2020 el doctor Ozlem Karabiber Cura, presento en su trabajo la clasificación del EEG epiléptico mediante la transformación sincronizada (SST) [26] con máquinas y técnicas de aprendizaje profundo, se calcularon las matrices SST de datos de EEG de convulsiones y pre-convulsiones de 16 pacientes con epilepsia. Se proponen dos enfoques basados en el aprendizaje automático y el aprendizaje profundo para clasificar las señales previas a la convulsión. En el enfoque basado en el aprendizaje automático, se calculan las diversas características, como los momentos conjuntos de orden superior, y estas características se clasifican mediante clasificadores Maquina de Vectores Soporte (SVM) [7], vecino mas cercano (kNN) [22] y Bayesiano ingenuo (NB) [24]. Los resulta-

dos de la simulación demuestran que ambos enfoques lograron índices de precisión de validación prometedores. Si bien se logra la máxima precisión de validación (90,2 %) para el enfoque basado en el aprendizaje automático, se logra la precisión de validación (90,3 %) para el enfoque basado en el aprendizaje profundo [27].

En el año 2021 el doctor Mohammad Masum, presentó un marco novedoso para clasificar de manera efectiva la epilepsia aprovechando el análisis estadístico de resumen de las características basadas en ventanas de las señales de EEG. Primero eliminó el ruido de las señales mediante el análisis de densidad del espectro de potencia, reemplazó los valores atípicos y luego las características de nivel de ventana extraídas de dominios estadísticos, temporales y espectrales. de esta manera aplicó diferentes clasificadores de ML, como Máquina de Vectores Soporte, Árbol de decisión, Bosque aleatorio y vecinos más cercanos, a las características extraídas. El marco propuesto que aplica el clasificador bosque aleatorio puede mejorar significativamente el rendimiento de la clasificación de señales EEG en comparación con otros métodos de clasificación de epilepsia en términos de exactitud, precisión y recuperación [28].

En el año 2021 el doctor Umme Salma M. el menciona en su trabajo que la mayoría de las personas diagnosticadas con epilepsia pueden manejarse con éxito, siempre que los episodios se descubran a tiempo, para ello, el aprendizaje automático juega un papel esencial en la detección. Algunos métodos usados por él fueron: Máquina de Vectores Soporte (SVM) [7], aumento de gradiente extremo (Xgboost) [29], árbol de decisión, análisis de discriminación lineal (LDA) [10], Bayesiano ingenuo, vecino mas cercano (k-NN) [22] y regresión logística. utilizó algoritmos de clasificación de aprendizaje automático utilizados para clasificar EEG que reflejaban epilepsia, los resultados muestran que (SVM) [7] es el método más efectivo para clasificar epilepsia, con una tasa de precisión del 96,84 % [30].

En el año 2021 la doctora Sricheta Pauri en su investigación menciona que recientemente, se han sugerido varios marcos de detección de convulsiones automatizados basados en técnicas de aprendizaje automático para reemplazar los enfoques tradicionales. Debido a que los datos de EEG multicanal son caóticos, elegir los canales óptimos y las características y categorizarlos siguen siendo desafíos abiertos. [31].

En el año 2021 el doctor Mosab A. A. Yousif quién en su trabajo menciona que el desarrollo de métodos de detección temprana para ataques epilépticos a partir de señales de electroencefalo-

grafía (EEG) [12], utilizando técnicas de procesamiento de señales y aprendizaje automático. En su trabajo menciona como utilizar algunas características en el dominio del tiempo junto con las mismas características extraídas para las subseñales de la señal original. Para obtener subseñales a partir de la señal de EEG, utilizó la descomposición de wavelet clásica en tiempo discreto, los resultados que se obtuvieron para la detección de los cinco estados de epilepsia se clasificaron utilizando los algoritmos bayesiana ingenua (NB) [24], árbol de decisión (DT) [32], bosque aleatorio (RF) [33], Multilayer Perceptrón (MLP) [34], k-vecinos cercanos (k-NN) y Máquina de Vectores Soporte (SVM) [7]. Los resultados muestran que bosque aleatorio, Multilayer Perceptrons, k-vecinos cercanos (k-NN) [22] con validación cruzada de 10 veces siendo los algoritmos más exitosos con un 92 % de precisión [35].

Justificación

El sistema de electroencefalografía (EEG), permite registrar ondas cerebrales sobre la superficie del cuero cabelludo, siendo esta una técnica no invasiva, esta es utilizada por los médicos y técnicos, para emitir un diagnóstico más preciso sobre alguna enfermedad neurológica. Las enfermedades cerebrales han aumentado en los últimos años. De acuerdo a la OMS, se puede destacar que habrá un incremento importante en cuanto a enfermedades neurológicas, lo que llevará a una alta demanda de atención médica y sistemas tecnológicos que puedan hacer una valoración más precisa. A continuación se enuncian algunas razones que permiten, dimensionar el porqué de este trabajo:

- La epilepsia es una enfermedad cerebral no transmisible crónica que afecta a personas de todas las edades.
- En todo el Mundo, unos 50 millones de personas padecen epilepsia, lo que la convierte en uno de los trastornos más comunes.
- Cerca del 80 % de los pacientes viven en países de ingresos bajos y medianos.
- Se estima que el 70 % de las personas con epilepsia podrían vivir sin eventos convulsivos, si se diagnostican y tratan adecuadamente
- El riesgo de la muerte prematura en personas con epilepsia es hasta tres veces mayor que en la población general
- Tres cuartas partes de las personas que viven en países de ingresos bajos no reciben el tratamiento que necesitan.

- En muchos lugares del mundo, los pacientes y sus familias pueden ser víctimas de la estigmatización y la discriminación.
- Los problemas de salud mental serán la principal causa de discapacidad en el mundo en el 2030.

La información estadística fue obtenida de los datos que ofrece la Organización Mundial de la Salud (OMS) [36]. Durante varias décadas el diagnóstico de esta enfermedad se ha robustecido con investigaciones de otras disciplinas, como lo es la Inteligencia Artificial (IA) que ofrece algunas opciones como los clasificadores para la identificación más temprana de la enfermedad, lo que podría ayudar en la prevención y su mejor cuidado posterior, por consiguiente se considera la clasificación, un medio por el cual se determina qué tipo de epilepsia padece la persona. Es conocido que cuando un paciente presenta crisis generalizadas hay una activación conjunta de todas las neuronas de la superficie cerebral, mientras que durante las crisis focales solo producen descargas anómalas las neuronas de una zona limitada de la superficie cerebral. No obstante, en algunas ocasiones la descarga puede iniciarse en una zona del cerebro y posteriormente extenderse a toda la superficie cerebral; esto es lo que se conoce como crisis secundariamente generalizadas. En términos generales, las crisis focales son más frecuentes que las generalizadas [37], [38].

Descripción y Diagrama a Bloques

El procesamiento de las señales EEG, conforman un análisis de las mismas, que permite identificar características de una buena o mala función cognitiva. Consiste en un estudio de las señales registradas, preprocesadas, para ser clasificadas.

Una metodología usada en la actualidad para la clasificación de las señales, es por medio de la extracción de características como se puede ver en la Figura 1.2, para este trabajo se utilizaran las herramientas que proporciona una de las ramas de la Inteligencia Artificial como es el Machine Learning (aprendizaje automático) el cual está dividido en aprendizaje supervisado y no supervisado. Por otra parte, el aprendizaje automático tiene varias técnicas de clasificación de las cuales podemos mencionar algunas como vecinos más cercanos, árbol de decisiones, bosque aleatorio, bayes ingenuos y máquina de vectores soporte.

En este trabajo se realizará la clasificación utilizando la máquina de vectores soporte, la cual es una de las técnicas más empleadas, consiste en implementar un hiper-plano en un espacio de N-dimensiones para poder clasificar todos los datos que se muestran en el plano en forma de puntos y así realizar la clasificación.

Para poder hacer el trabajo tanto de clasificación como la obtención de características se utilizó una base de datos pública, a saber, la base de datos de Berna Barcelona en la cual se utilizó clasificador de máquina de vectores soporte. Podemos observar, en la Figura 1.2 el diagrama de bloques para la obtención y procesamiento de las señales epilépticas.

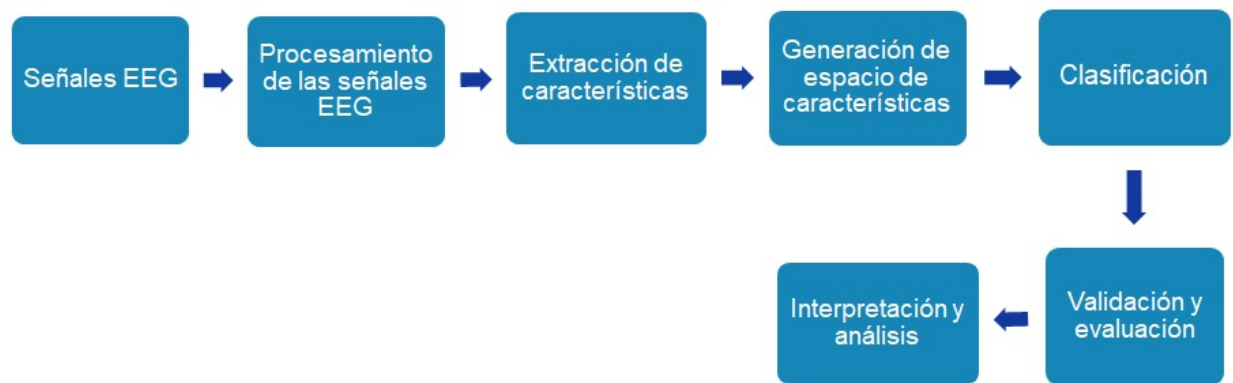


Figura 1.2: Diagrama a bloques del procesamiento de señales EEG

Capítulo 2

Epilepsia

La epilepsia es un trastorno neurológico crónico caracterizado por una constante persistencia a generar crisis epilépticas, consecuencias neurobiológicas, cognitivas, psicológicas y sociales, las estadísticas son claras como se puede observar en la Figura 2.1. Estas crisis resultan de una descarga anormal y excesiva de un conjunto de neuronas cerebrales. Desde el punto de vista electrofisiológico, las crisis son causadas por una alteración entre la excitación y la inhibición en redes neuronales, lo que lleva a episodios sincronizados de actividad neuronal. Estos episodios pueden originarse en una región específica del cerebro (crisis focales) o involucrar simultáneamente ambos hemisferios cerebrales (crisis generalizadas) [38].

Enfermedades más comunes del cerebro	
Enfermedad	Personas afectadas en el mundo
Epilepsia	50 Millones
Demencia	Cerca de 50 Millones
Pákinson	7 Millones
Enfermedades cerebrovasculares	Mueren más de 6 Millones anualmente

Figura 2.1: Principales enfermedades neurológicas según la OMS

2.1. El Cerebro

El cerebro es el órgano más importante del cuerpo humano, está protegido por el cráneo el cual protege a la corteza cerebral que es la parte del cerebro que recibe, integra y procesa información para el sistema nervioso central por esta razón mantiene el control total de todos los órganos del

cuerpo humano, mantiene las funciones motrices, las funciones cognitivas y una producción hormonal constante, el cerebro está compuesto por dos hemisferios como podemos ver en la Figura 2.2:

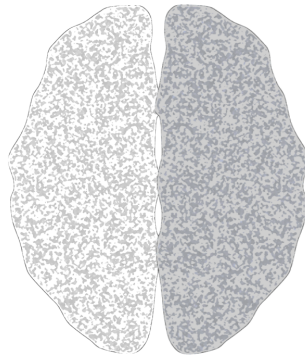


Figura 2.2: Hemisferios Cerebrales

- a) hemisferio derecho,
- b) hemisferio izquierdo.

Este cuenta con 4 lóbulos los cuales se encuentra ubicados en cada hemisferio del cerebro, los dos hemisferios cerebrales se encuentran divididos por el surco interhemisférico el cual divide los dos hemisferios cerebrales, sin embargo, además existen otras capas que protegen al cerebro, estas son las meninges y el líquido cefalorraquídeo, las meninges son las tres capas o membranas que protegen tanto a la médula espinal como al cerebro veamos la Figura 2.3:

- El cuero cabelludo es la parte de piel que cubre la cabeza y esta protege al cráneo y cerebro, es la primera capa que cubre el cerebro.
- El cráneo es la parte ósea superior del cuerpo humano y tiene una estructura compleja, su principal función es la de proteger el encéfalo que esta compuesto por cerebro, cerebelo y tronco encefálico y este a su vez esta dividido en dos el neurocráneo y bóveda craneal.
- Materia dura es la meninge más extensa y la que tiene la mayor resistencia y esta adherida firmemente al cráneo y esta es la que evita que se mueva demasiado, este contiene canales venosos que ayudan a drenar la sangre y el líquido cefalorraquídeo.
- El espacio aracnoideo esta debajo de la materia dura y se parece a una tela de araña y este esta lleno de líquido cefalorraquídeo.

- La piamadre es la capa que se adhiere a la superficie del cerebro, penetrando y siguiendo los contornos del cerebro.

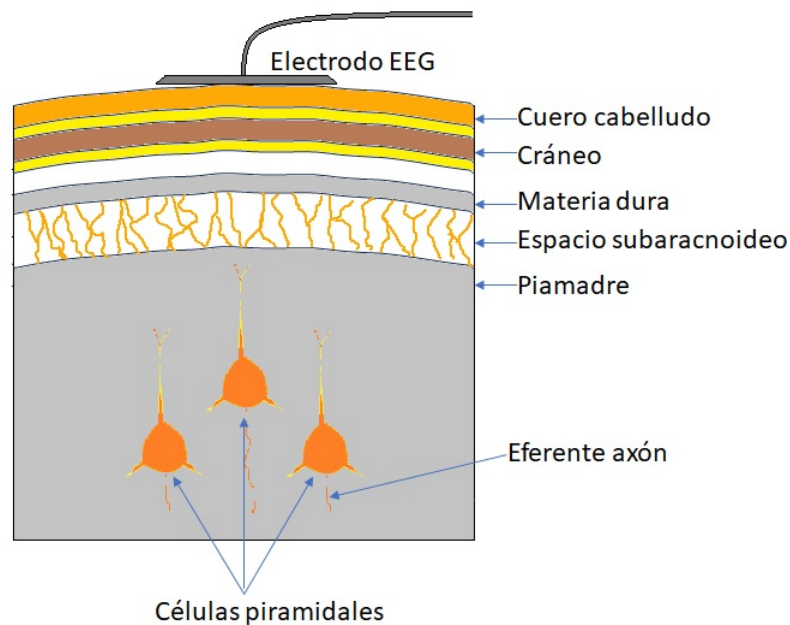


Figura 2.3: Estructura de protección del cerebro

Para medir los potenciales eléctricos que se producen por los conglomerados de neuronas, esto se realiza colocando los electrodos sobre el cuero cabelludo, lo cual es una forma no invasiva que, facilita la tarea de registro de actividad eléctrica de las neuronas. En la Figura 2.4 podemos ver las partes de una neurona y como esta realiza la comunicación (sinapsis) entre ellas.

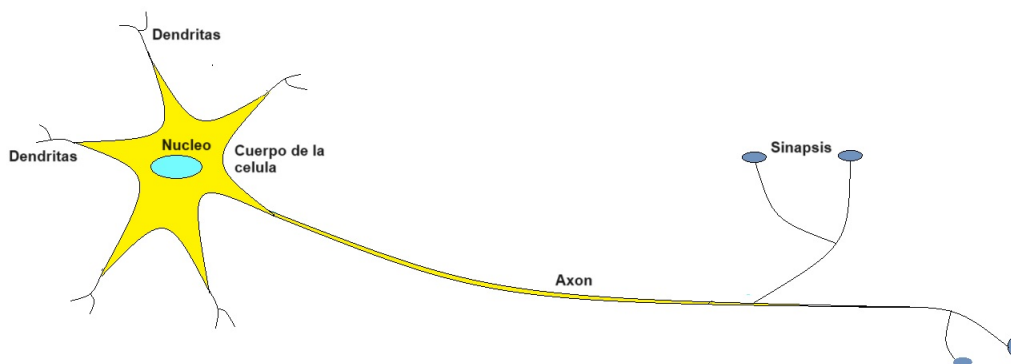


Figura 2.4: Forma y partes de una neurona

Las neuronas están conformadas por: Cuerpo celular, dendritas, axón, terminales axónicas y vesículas sinápticas, esta es la forma básica del sistema nervioso y su trabajo fundamenta es el de

recibir, procesar y transmitir información en forma de potenciales eléctricos, y a la comunicación entre neuronas se denomina sinapsis.

2.2. ¿Que es el Electroencefalograma?

El electroencefalograma es un registro médico que mide la actividad eléctrica del cerebro, mediante pequeños discos de oro o plata (electrodos), colocados sobre el cuero cabelludo y así obtener registros eléctricos de los pacientes en diferentes situaciones como: basal (es el estado del paciente, relajado y en estado normal) y con métodos de activación, como la hiper ventilación (esta es una respiración agitada)y la foto luminosidad (luminosidad intermitente). En ocasiones, y en función del paciente, será conveniente registrar la actividad del cerebro durante el sueño. Existen patrones normales y patrones anormales de funcionamiento del cerebro que permiten saber si el paciente puede tener una lesión o enfermedad cerebral.

Todos los canales de EEG tienen dos entradas. Cada canal de EEG representa la diferencia de potencial entre dos electrodos adyacentes. En los montajes referenciales, el electrodo activo está en la primera entrada y una referencia neutra está en la segunda entrada. Debido a esta relación, la primera entrada se ha denominado entrada “activa” y la segunda entrada “entrada de referencia”.

2.2.1. Sistema 10-20

El sistema internacional 10-20 es un método reconocido a nivel internacional y este nos describe la forma en la cual está dada la ubicación de los electrodos sobre el cuero cabelludo para que se efectuó un examen EEG. Este método fue desarrollado para mantener métodos estandarizados, y de esta manera garantizar que los resultados del estudio de un paciente de investigación o un paciente clínico puedan compilarse, reproducirse, analizarse y compararse. Una vez colocados los electrodos, conforme a la norma establecida como sistema 10-20 como se observa en la Figura 2.5. En la actualidad es el sistema más utilizado por los médicos como método no invasivo, ya que en la actualidad existen más sistemas para obtener este tipo de señales.

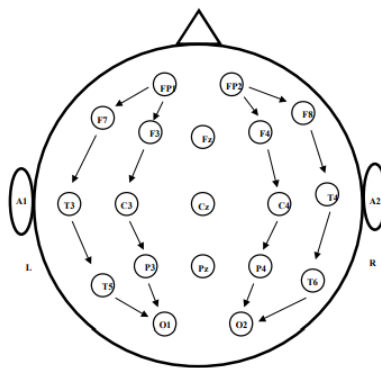


Figura 2.5: Ubicación de electrodos, sistema 10-20

El sistema 10-20 es un método reconocido internacionalmente utilizado para describir la colocación de electrodos en el cuero cabelludo en procedimientos como el electroencefalograma (EEG). Los números 10 y 20 hacen referencia a los porcentajes de distancia en los cuales los electrodos deben ser colocados en diferentes regiones del cráneo, incluyendo la parte frontal, occipital, y los lados derecho e izquierdo. La colocación de electrodos según el sistema 10-20 se basa en puntos de referencia craneales y utiliza una cuadrícula para determinar la ubicación precisa de cada electrodo. Los electrodos se colocan de manera que estén espaciados en un 10-20 entre sí, en relación con los puntos de referencia craneales específicos. Para colocar los electrodos según el sistema 10-20, se deben seguir los siguientes pasos:

1. Identificar los puntos de referencia craneales: Se utilizan puntos de referencia como el nasion (punto en el puente de la nariz), el inion (punto en la parte posterior de la cabeza) y los preauriculares (puntos delante de las orejas) para guiar la colocación de los electrodos.
2. Marcar los puntos de colocación de los electrodos: Utilizando la cuadrícula del sistema 10-20, se marcan los puntos donde se colocarán los electrodos en relación con los puntos de referencia craneales.
3. Colocar los electrodos: Se aplica una preparación adecuada en el cuero cabelludo para asegurar una buena conducción eléctrica. Luego, se colocan los electrodos en los puntos marcados según la cuadrícula 10-20.

Es importante destacar que la colocación precisa de los electrodos según el sistema 10-20 requiere experiencia y conocimiento técnico. En procedimientos clínicos o de investigación, un profesional

capacitado en EEG es el encargado de realizar esta tarea para obtener resultados confiables y precisos. En resumen, el sistema 10-20 es un método estandarizado utilizado para la colocación de electrodos en el cuero cabelludo en procedimientos como el EEG.

2.2.2. Tipo de ondas cerebrales.

Este tipo de ritmo cerebral es clasificado por su frecuencia y amplitud, se presentaran los 4 ritmos cerebrales como se observa en la Figura 2.6, si existen cambios significativos, se puede determinar si existe una afectación en el cerebro [39].

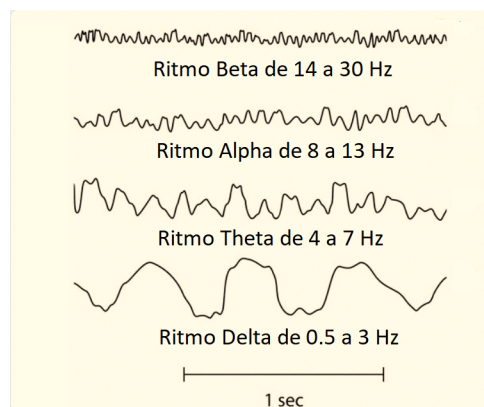


Figura 2.6: Ritmo de ondas cerebrales así como la amplitud de cada onda

■ Ondas Delta

Las ondas delta tienen una frecuencia de 0.5 y 4 Hz y tiene una amplitud de 20-200 μV , se asocian típicamente con el sueño profundo y reparador, así como con ciertos trastornos del sueño. Estas ondas se producen en la corteza cerebral, la capa exterior del cerebro, durante la fase de sueño llamada sueño de ondas lentas (SOL). Durante esta fase, la actividad cerebral disminuye y el cuerpo se relaja profundamente, permitiendo que los procesos de reparación y consolidación de la memoria se lleven a cabo. En el estado de vigilia (estado en el cual el paciente se encuentra despierto y alerta) estas ondas no son comunes, la presencia de ondas delta en el EEG también puede ser indicativa de ciertos trastornos como la epilepsia, este tipo de ondas se pueden presentar en una crisis y se pueden asociar con un mayor riesgo de desarrollar convulsiones.

■ Ondas Theta

Las ondas theta tienen una frecuencia de 4 a 7 Hz y tiene una amplitud de 20-100 μV , las ondas theta se desarrollan en el lóbulo temporal. Es usual en menores de 15 años y en el adulto somnoliento.

Se distribuye en todas las áreas con mayor amplitud en la región frontal central temporal bilateral. En mayores de 50 años, no representa anormalidad, la presencia de actividad theta intermitente y lateralizada a la región temporal media y anterior izquierda. La transición de vigilia a sueño se caracteriza por disminución de la amplitud del actividad alfa, en algunas ocasiones incremento de actividad rápida beta y más constantemente aparición de actividad lenta theta en todas las áreas. En menores de 15 años, es normal que alrededor del 30 por ciento de una época sea ocupado por actividad theta. De tal forma, que en este grupo de edad es más confiable identificar períodos de somnolencia basándose en la disminución de la actividad alfa más que en el incremento de la actividad theta.

Las ondas theta se originan en el lóbulo temporal del cerebro y se propagan hacia la superficie del cerebro, donde se registran en el cuero cabelludo por los electrodos del EEG. Las ondas theta pueden ser simétricas o asimétricas en ambos hemisferios cerebrales, y su amplitud y frecuencia pueden variar según la edad, el estado emocional, la actividad física, el sueño y la presencia de enfermedades neurológicas.

■ Ondas Alfa

En la persona despierta, recostada y con los ojos cerrados la actividad dominante es en la banda alfa de 8 - 13 Hz. Con una mayor amplitud en la región occipital donde en promedio y de acuerdo a la edad alcanza los 50 hasta 120 micro-voltios. Característicamente se atenúa o llega a bloquearse al abrir los ojos o con los estímulos auditivos. En resumen, para identificar esta actividad rítmica se requiere que el sujeto esté en vigilia, tenga los ojos cerrados y que la mayor amplitud sea occipital aunque no es infrecuente observarla en toda la región temporo-parieto-occipital de manera bilateral.

■ Ondas Beta

Esta actividad es de bajo voltaje, de 5 a 30 micro-voltios. Predomina en las áreas frontales y frontales superiores. Tiene una frecuencia mayor a los 14 a 30 Hz. Se opaca frecuentemente por artefacto de origen muscular sobrepuesto en estas mismas áreas. En niños, la transición de vigilia a sueño aumenta la incidencia de actividad beta. A cualquier edad, algunas medicaciones también incrementan significativamente la abundancia y la amplitud de estos ritmos rápidos. Por su frecuente formulación en personas que deben someterse a un estudio de EEG, los barbitúricos son de importante consideración.

■ Ondas Gama

Las ondas gama son un patrón de oscilación neuronal cuya frecuencia oscila entre los 30 y 100 Hz. Lo más típico es que se manifiesten en torno a 40 Hz. Estas ondas reflejan una alta actividad cerebral. Más alta que la que reflejan las ondas beta, que son aquellas presentes cuando estamos concentrados. Estas ondas indican que el cerebro está trabajando a una frecuencia que no es muy común, activando distintas regiones cerebrales. Se cree que estas ondas son las que se ponen en marcha ante funciones ejecutivas complejas o actividad mental superior. Estas actividades complejas como pueden ser la orientación, la atención, la consciencia o el razonamiento requieren la activación coordinada de las neuronas de varias regiones cerebrales para la integración de distintos tipos de información para lograr dar sentido a la realidad. Sin embargo, parece ser que estas ondas no solo reflejan este tipo de actividad, sino también otra que nada tiene que ver con estar concentrado o con el cerebro trabajando. Parecen estar relacionadas también con conductas explosivas, de altísima ansiedad o reacciones de terror. Para poder interpretar el tipo de ritmo cerebral que se está presentando, depende del lóbulo en donde se esté focalizando el ritmo cerebral es el hemisferio y la forma de onda que podemos observar, así mismo se describe todo en el siguiente diagrama de la Figura 2.7:

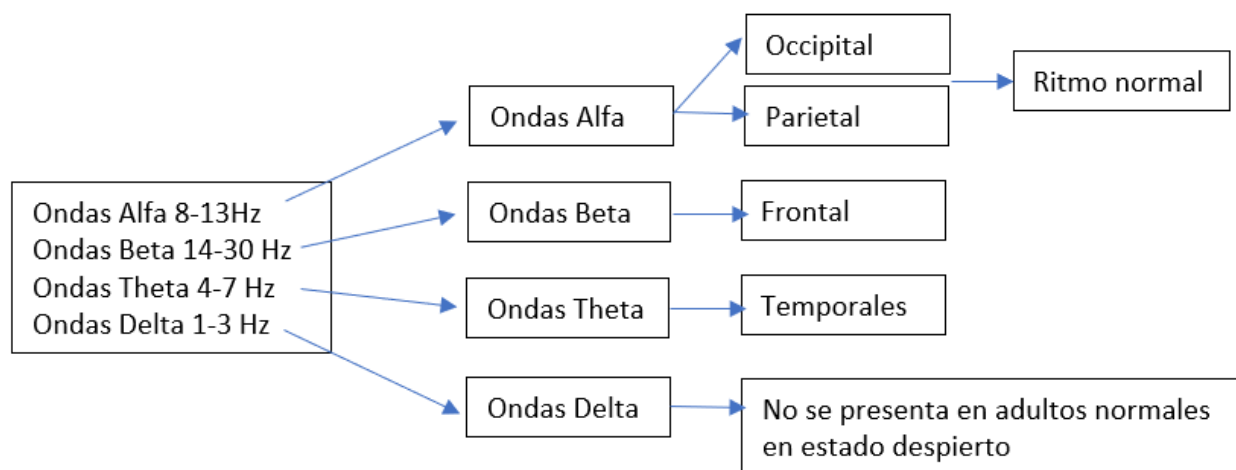


Figura 2.7: Interpretación de ritmo cerebral depende de la ubicación

2.3. Clasificación de crisis epiléptica

Las enfermedades neurológicas son trastornos que afectan al sistema nervioso central y al sistema nervioso periférico, existen diferentes tipos de enfermedades neurológicas entre las cuales esta presente la epilepsia. Las crisis epilépticas se pueden clasificar de diversas maneras según su origen,

manifestaciones y propagación en el cerebro. Según la clasificación de la Liga Internacional contra la Epilepsia (ILAE) [38], las crisis epilépticas se dividen principalmente en:

Crisis parciales

Las crisis parciales, también conocidas como crisis focales, se originan en una parte del cerebro y pueden o no propagarse a otras áreas. Estas crisis se clasifican de la siguiente manera:

- Crisis focales sin alteración de la conciencia (anteriormente conocidas como crisis parciales simples):
 - Estas crisis afectan una pequeña parte del cerebro, y la persona permanece consciente durante el evento.
 - Pueden manifestarse de diversas formas según la región del cerebro afectada: movimientos involuntarios, sensaciones anómalas, emociones intensas, entre otros.
 - Suelen durar menos de 2 minutos.
- Crisis focales con alteración de la conciencia (anteriormente conocidas como crisis parciales complejas):
 - Aunque se originan en una parte del cerebro, afectan la conciencia del individuo.
 - La persona puede aparecer aturdida o desconectada de su entorno y no responder a estímulos, aunque a menudo realiza acciones automáticas y repetitivas, como frotarse las manos o masticar.
 - Después de la crisis, es común que la persona no recuerde lo ocurrido durante el evento.
- Crisis focales que se generalizan:
 - Comienzan en una parte del cerebro pero rápidamente se propagan a todo el cerebro, resultando en una crisis generalizada, como una tónica, clónica o tónico-clónica.
 - La persona perderá la conciencia y puede presentar movimientos convulsivos.

Crisis generalizada

Las crisis generalizadas son aquellas que involucran a ambas partes del cerebro. Estas crisis afectan a todo el cerebro y, por lo general, resultan en síntomas que afectan a todo el cuerpo. Aquí está la clasificación de las crisis generalizadas:

- Crisis de ausencia (anteriormente conocidas como "pequeño mal"):

- **Ausencias típicas:** Son breves interrupciones de la conciencia, donde la persona puede tener una mirada fija y estar "ausente" durante unos segundos. Pueden ocurrir muchas veces al día.
- **Ausencias atípicas:** Son similares a las típicas pero pueden durar más tiempo y presentar síntomas adicionales como movimientos leves de las manos o los párpados
- **Crisis tónico-clónicas** (anteriormente conocidas como "gran mal"):
 - Combinan una fase tónica con rigidez muscular y una fase clónica con movimientos repetitivos de contracción-relajación.
- **Crisis tónicas:**
 - Se caracterizan por la rigidez súbita de los músculos, especialmente de la espalda, las extremidades y el cuello.
- **Crisis clónicas:**
 - Movimientos repetitivos de contracción-relajación que, a diferencia de las tónico-clónicas, no presentan una fase inicial rígida.
- **Crisis mioclónicas:**
 - Sacudidas o espasmos musculares rápidos y breves que pueden afectar a todo el cuerpo o solo a una parte.
- **Crisis atónicas** (o crisis de caída):
 - Se caracterizan por una pérdida súbita del tono muscular, lo que puede hacer que la persona colapse o caiga al suelo.
- **Crisis mioclónico-atónicas:**
 - Combinan las características de las crisis mioclónicas y atónicas.
- **Crisis mioclónico-tónicas:**
 - Combinan características de las crisis mioclónicas y tónicas.

Crisis desconocida

Las crisis de inicio desconocido se refieren a crisis epilépticas en las que, con base en la información disponible, no se puede determinar si se iniciaron como crisis focales o generalizadas. La clasificación de las crisis de inicio desconocido se subdivide en:

- Motoras:

- Tónico-clónicas: Estas son las crisis convulsivas clásicas que la mayoría de las personas asocia con la epilepsia. Involucran una fase tónica con rigidez muscular seguida de una fase clónica con movimientos repetitivos de contracción-relajación.
- Ausencia con componente motor (por ejemplo, parpadeo): Pérdida breve de la conciencia con manifestaciones motoras como parpadeo o movimientos ligeros de las manos.
- Atónicas: Pérdida súbita del tono muscular, lo que puede hacer que la persona colapse.
- Tónicas: Rigidez muscular.
- Clónicas: Movimientos repetitivos de contracción-relajación.
- Mioclónicas: Sacudidas o espasmos musculares rápidos y breves.
- Epilepsia mioclónica: Combina características de crisis mioclónicas y otros síntomas.

- No motoras:

- Ausencias típicas: Interrupciones breves de la conciencia, donde la persona parece ausente durante unos segundos.
- Ausencias atípicas: Similar a las ausencias típicas pero pueden durar más tiempo o presentar síntomas adicionales.
- De conducta (por ejemplo, ausencia automotora): La persona puede realizar movimientos o acciones automáticas sin tener conciencia de ellos [38].

2.4. Clasificación de la epilepsia

A lo largo de las décadas, la clasificación de la epilepsia ha pasado por numerosas revisiones y modificaciones. Estos cambios han sido impulsados por avances en la neurociencia, la genética y las tecnologías de imagen cerebral, así como por una mejor comprensión de las características clínicas y electroencefalográficas de las crisis y síndromes epilépticos. Con cada revisión, los expertos han buscado presentar una estructura más coherente y precisa que permita a los profesionales de la salud diagnosticar y tratar la epilepsia de manera más efectiva. Además de mejorar la atención al paciente, una clasificación más detallada y actualizada como se observa en la Figura 2.8 también facilita la investigación, permitiendo a los científicos investigar de forma más específica las distintas formas de epilepsia. En resumen, la continua evolución de la clasificación de la epilepsia refleja no solo el progreso en nuestro entendimiento de esta compleja enfermedad neurológica,

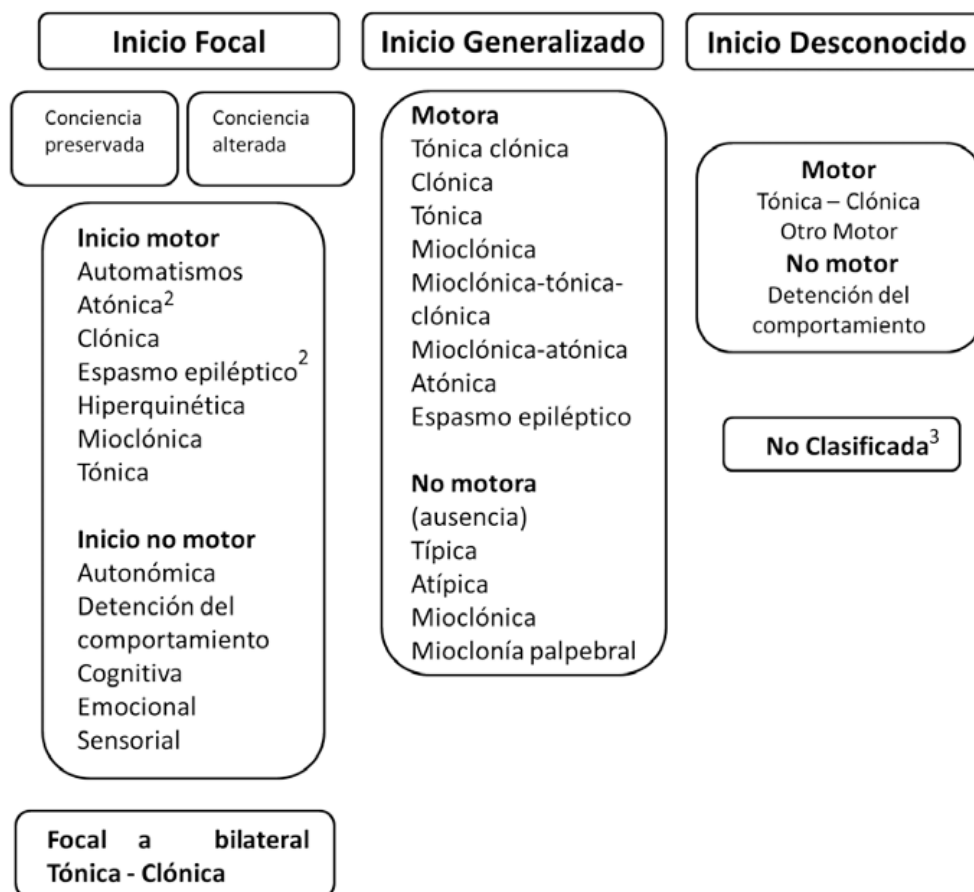


Figura 2.8: Clasificación de las crisis convulsivas datos proporcionados por la ILAE

sino también el deseo constante de la comunidad médica de proporcionar la mejor atención posible a los pacientes afectados. La epilepsia es una enfermedad del cerebro definida por cualquiera de las siguientes condiciones [38]:

- Al menos dos crisis no provocadas (o reflejas) que ocurren con más de 24 horas de diferencia.
- Una crisis no provocada (o refleja) y una probabilidad de nuevas crisis similares con riesgo general de recurrencia (al menos 60 %) luego de dos crisis no provocadas, que ocurran en los siguientes 10 años
- Diagnóstico de un síndrome epiléptico

La epilepsia se considera desaparecida o resuelta en las personas que tuvieron el síndrome epiléptico auto-limitado dependiente de la edad, pero que ahora han superado la edad aplicable, o para aquellos que han permanecido libres de ataques en los últimos 10 años, sin medicación para las crisis durante los últimos 5 años.

La clasificación de la epilepsia se lleva a cabo después de que se cumplen los criterios para diagnosticar la epilepsia (definición anterior). La clasificación se realiza utilizando un marco de clasificación multinivel, que implica la clasificación en tres niveles: El tipo de crisis, el tipo de epilepsia y el síndrome de epilepsia. Imágenes, EEG y otras investigaciones, cuando estén disponibles, contribuyen optimizar la clasificación en los tres niveles. Donde sea posible, se debe establecer un diagnóstico en los tres niveles. La etiología de la epilepsia debe considerarse desde el principio y en cada paso a lo largo de todo el camino del diagnóstico. Conocer la etiología puede informar una clasificación optimizada y tiene importantes implicaciones en el tratamiento para el paciente.

- Epilepsia generalizada.

Los pacientes con epilepsia generalizada presentan crisis generalizadas, y pueden presentar los resultados de encefalogramas interictales y/o ictales que acompañan a las crisis generalizadas (por ejemplo, punta-onda generalizada). Un antecedente familiar de crisis o epilepsia generalizada refuerza la susceptibilidad.

- Epilepsia focal.

Los pacientes con epilepsia focal tienen diferentes tipos de crisis focales, y pueden presentar hallazgos comunes en los encefalogramas interictales o ictales que acompañan a dichas crisis focales (como, por ejemplo, ondas agudas focales o reducción focal interictal). Las imágenes que muestran una anomalía cerebral estructural focal pueden resultar de apoyo, aunque los pacientes con etiologías genéticas e imágenes normales pueden también padecer epilepsia focal. Las epilepsias focales pueden ser unifocales, multifocales o hemisféricas.

- Epilepsia focal y generalizada.

Los pacientes pueden presentar crisis tanto generalizadas como focales, con hallazgos interictales y/o ictales en el encefalograma que pueden acompañar ambos tipos de crisis. Los pacientes con síndrome de Dravet y con síndrome de Lennox Gastaut pueden presentar epilepsia generalizada y focal.

- Epilepsia desconocida

El término 'desconocido' es utilizado para indicar que se entiende que el paciente tiene epilepsia, pero no es posible denotar si es focal, generalizada o una combinación de ambas. Esto puede ocurrir debido a falta de información para clasificar la epilepsia como, por ejemplo, si el EEG es normal o poco informativo [39].

Capítulo 3

Marco Teórico

3.1. Ramas de la Inteligencia Artificial(IA)

En este capítulo, se abordarán los fundamentos necesarios para comprender de manera más completa los aspectos relacionados con un sistema de clasificación de EEG (Electroencefalograma) aplicado a pacientes con epilepsia. Comenzaremos por explorar los conceptos de Machine Learning y matemáticos. Luego, posteriormente en la comprensión del EEG, explicando cómo se generan las señales, qué información proporcionan y cómo llevar a cabo la adquisición y análisis de estas señales de manera efectiva [40], [41].

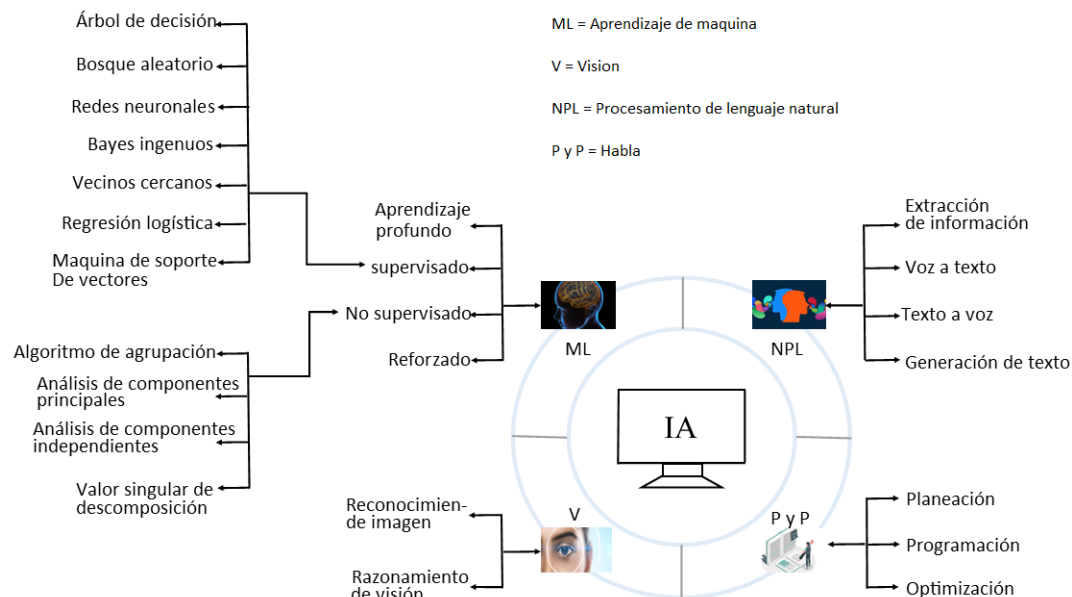


Figura 3.1: Ramas de la Inteligencia Artificial.

La Inteligencia Artificial (IA) es una combinación de algoritmos y máquinas que replican, amplifican y simulan las capacidades de los seres humanos, son herramientas que realizan tareas que tradicionalmente requerían inteligencia humana, las capacidades requeridas por la inteligencia artificial para realizar tareas abarcan el razonamiento lógico, la toma de decisiones, el lenguaje natural y la adaptación sobre el entorno. El reto primordial de la inteligencia artificial no es imitar la inteligencia humana, es la de crear sistemas a base de algoritmos que puedan realizar tareas específicas y que funcionen de manera óptima y que lleguen a superar el desempeño humano. La inteligencia artificial combina campos como las matemáticas, psicología, neurociencia, la informática y lingüística entre otras para crear modelos que puedan exhibir una forma de inteligencia. En la Figura 3.1 se muestra una clasificación completa de la inteligencia artificial de la cual solo se aplicará la parte de Machine Learning y la subsección de aprendizaje supervisado para emplear y aplicar la clasificación para nuestro sistema. En el próximo capítulo, exploraremos una poderosa herramienta que se ha convertido en la elección ideal para abordar el desafío que enfrentamos. Esta herramienta, conocida como 'wavelet', ha demostrado su eficacia en una amplia gama de aplicaciones y nos permitirá analizar y resolver nuestro problema de manera más precisa y eficiente.

3.2. Transformada de Wavelet

La transformada de wavelet es una herramienta matemática utilizada para analizar señales y funciones en diferentes escalas de tiempo y frecuencia. Proporciona una representación detallada de una señal al descomponerla en componentes de diferentes frecuencias y ubicaciones en el tiempo. La transformada de wavelet se basa en el uso de funciones llamadas Wavelets, que son funciones bien localizadas en el tiempo y la frecuencia. La aplicación de la transformada de wavelet implica los siguientes conceptos [42].

- Wavelet: Una wavelet es una función que tiene un tamaño y una forma bien definidos. Estas funciones están diseñadas para ser localizadas tanto en el dominio del tiempo como en el dominio de la frecuencia. Las wavelets se utilizan como bases para descomponer una señal en componentes de diferentes frecuencias y escalas.
- Descomposición y reconstrucción: La transformada de wavelet realiza la descomposición de una señal en diferentes escalas y frecuencias. Esto implica dividir la señal original en componentes de baja frecuencia (aproximación) y alta frecuencia (detalles). La descomposición se realiza mediante el cálculo de productos internos entre la señal y las wavelets en diferentes escalas y ubicaciones en el tiempo. La reconstrucción inversa combina estas componentes para reconstruir la señal original como en la Figura 3.2.

- Escalas y frecuencias: La transformada de wavelet permite analizar una señal en diferentes escalas y frecuencias. Las wavelets pueden expandirse y contraerse en el tiempo y la frecuencia, lo que proporciona una representación multi-resolución de la señal. Esto significa que la transformada de wavelet puede capturar tanto las características de alta frecuencia como las de baja frecuencia de una señal.
- Aplicaciones: La transformada de wavelet se utiliza en una variedad de campos, incluyendo procesamiento de señales, compresión de imágenes, análisis de datos, procesamiento de audio y video, entre otros. Es especialmente útil cuando se trabaja con señales que tienen características locales y cambiantes en el tiempo y la frecuencia.

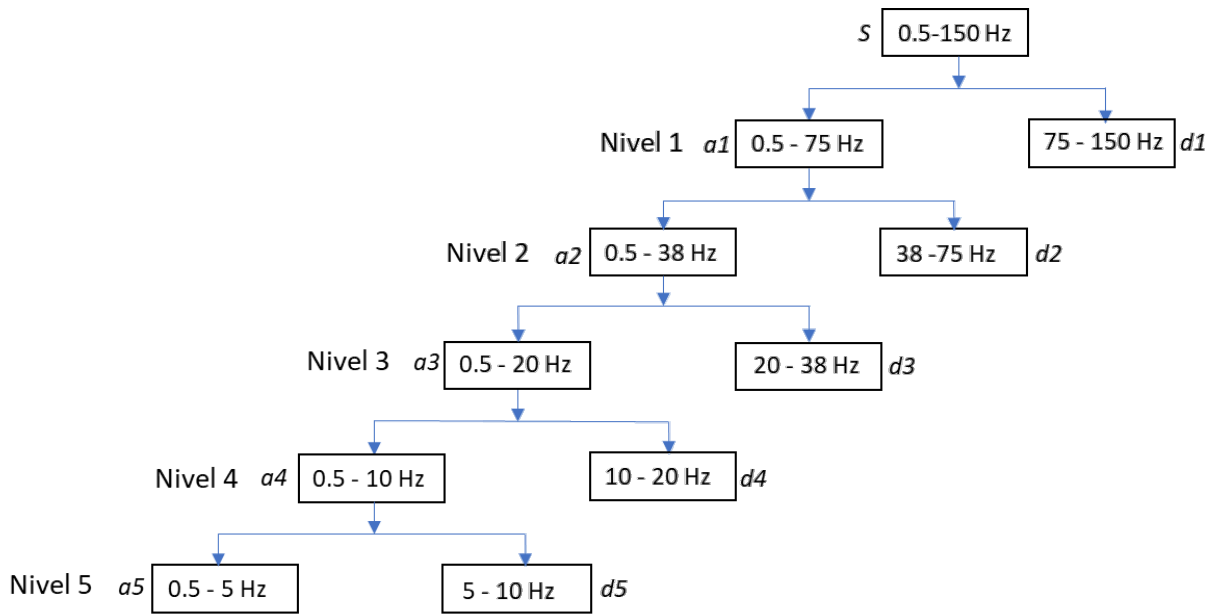


Figura 3.2: Descomposición de una señal muestra al aplicar la transformada wavelet, lado izquierdo diagrama de bloques y del lado derecho la aplicación del diagrama a la señal.

La transformada wavelet es una técnica ampliamente utilizada en el procesamiento de señales para la descomposición y análisis de señales EEG (Electroencefalografía), que registran la actividad eléctrica cerebral. Su aplicación en el dominio de las señales EEG permite una representación detallada de la señal en diferentes escalas de tiempo y frecuencia, lo que facilita la identificación de características específicas y patrones de actividad cerebral, se presenta la transformación Wavelet como una transformación lineal y continua de otra función dada, expresada como en la ecuación 3.1:

$$W(a, \tau) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} \psi^* \left(\frac{t - \tau}{a} \right) dt \quad (3.1)$$

Por lo tanto, la transformada de wavelet es una herramienta matemática utilizada para analizar señales en diferentes escalas de tiempo y frecuencia. Permite descomponer una señal en componentes de diferentes frecuencias y ubicaciones en el tiempo, lo que proporciona una representación detallada y multiresolución de la señal. Esta transformada tiene aplicaciones en diversos campos y es especialmente útil para analizar señales con características locales y cambiantes en el tiempo y la frecuencia.

Las wavelets de la familia "Daubechies"(comúnmente abreviada como "db") son ampliamente utilizadas en la clasificación de EEG epiléptico debido a varias características que las hacen efectivas para este propósito. Aquí hay algunas razones principales:

Localización en el dominio tiempo-frecuencia: Las wavelets de Daubechies están diseñadas para tener una propiedad de localización tanto en el dominio del tiempo como en el de la frecuencia. Esto significa que pueden capturar características transitorias y cambios rápidos en las señales de EEG, lo cual es importante para identificar eventos epilépticos que pueden tener una duración corta.

Banda de paso estrecha: Las wavelets Daubechies tienen una banda de paso estrecha en el dominio de la frecuencia, lo que les permite resaltar características específicas en diferentes bandas de frecuencia del EEG. Esto es particularmente útil para detectar patrones característicos de la actividad epiléptica en diferentes rangos de frecuencia.

Compactación de energía: Las wavelets Daubechies tienen una propiedad de compactación de energía, lo que significa que la mayor parte de la energía de la señal se concentra en un número reducido de coeficientes wavelet significativos. Esto facilita la extracción y representación de características relevantes del EEG, lo que a su vez ayuda en la clasificación de los patrones epilépticos.

Eficiencia computacional: Las wavelets Daubechies se pueden implementar eficientemente en términos de cálculos y memoria requerida, lo cual es importante para el procesamiento de grandes volúmenes de datos de EEG en tiempo real. Esto permite una clasificación rápida y en tiempo real de los eventos epilépticos, lo cual es crucial para el diagnóstico y la toma de decisiones clínicas.

Por lo tanto, las wavelets Daubechies, como la wavelet db, son ampliamente utilizadas en la clasificación de EEG epiléptico debido a su capacidad para capturar características transitorias, resaltar patrones en diferentes bandas de frecuencia, compactar la energía de la señal y su eficiencia computacional. Estas características las hacen efectivas en la detección y clasificación de eventos epilépticos, lo cual es fundamental para el diagnóstico y tratamiento de los pacientes.

3.3. Características Matemáticas

El EEG es una señal eléctrica que registra la actividad eléctrica del cerebro a lo largo del tiempo. Esta señal es inherentemente variable y ruidosa, lo que significa que contiene fluctuaciones en la amplitud y la frecuencia incluso cuando no ocurren eventos epilépticos. Las características estadísticas permiten resumir y analizar esta variabilidad para identificar patrones anormales. Los eventos epilépticos pueden tener diferentes características en comparación con la actividad cerebral normal. Las características estadísticas pueden ayudar a diferenciar entre eventos epilépticos y otros tipos de actividad cerebral, como artefactos de movimiento o actividad anormal, se van a mencionar las siguientes características [43]:

- **Media:** Representa el valor central o promedio de una señal EEG en un período específico y su representación es la siguiente:

$$\mu = \frac{1}{n} \sum_{i=1}^N x_i$$

- **Mediana:** Es una medida de tendencia central que indica el valor que ocupa la posición central en un conjunto de datos cuando estos están ordenados en forma ascendente o descendente y su representación es la siguiente:

$$Mediana = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

- **Moda:** Es una medida de tendencia central que indica el valor o los valores que ocurren con mayor frecuencia en un conjunto de datos, se basa en la frecuencia con la que aparecen los datos.
- **Cuartil:** En estadística, un cuartil es cada uno de los tres valores que dividen un conjunto de datos ordenados en cuatro partes iguales. Cada una de estas partes comprende aproximadamente un cuarto (o el 25 %) de los datos.

$$Cuartil = \frac{(n+1)xp}{4}$$

- **Desviación estándar:** Es una medida de la cantidad de variación o dispersión de un conjunto de valores. En estadística, es la raíz cuadrada de la varianza y proporciona una medida de la extensión en que los números del conjunto de datos se alejan de su media (promedio). Una desviación estándar baja indica que la mayoría de los datos tienden a estar cerca de la media,

mientras que una desviación estándar alta indica que los datos están más dispersos respecto a la media. Para población completa

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Para muestra de población

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n - 1}}$$

- **Varianza:** Es una medida estadística que describe la dispersión de los datos en un conjunto. Indica cuán lejos de la media (el promedio) se encuentran cada uno de los valores en el conjunto. En términos simples, la varianza mide qué tan extendidos están los datos. Una varianza grande sugiere que los números en el conjunto están más lejos de la media y entre sí. Por el contrario, una varianza pequeña indica que los números son más cercanos a la media y entre sí.

Varianza de población

$$\sigma^2 = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

Varianza de muestra

$$s^2 = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n - 1}}$$

- **Curtosis:** Es una medida estadística que describe la "agudeza" o "aplanamiento" de la distribución (forma de la curva) de los datos en comparación con la distribución normal (campana de Gauss). En términos técnicos, es una medida de las colas de la distribución. Indica la tendencia que tienen los valores de agruparse alrededor de las colas de la distribución.

La curtosis de una muestra:

$$G = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)+2}{(n-2)(n-3)}$$

La curtosis para una población:

$$\kappa = \frac{\sum (x_i - \mu)^4 / N}{(\sum (x_i - \mu)^2 / N)^2}$$

- **Oblicuidad ó asimetría,** en estadística es una medida que describe la distorsión o inclinación de la distribución de un conjunto de datos. Si la distribución es simétrica, como en el caso de

la distribución normal, la oblicuidad es cero. Una distribución con una oblicuidad positiva se inclina hacia la derecha, con una cola más larga hacia el lado derecho del pico de la distribución. Por el contrario, una oblicuidad negativa significa que la distribución se inclina hacia la izquierda, con una cola más larga en el lado izquierdo. Para el cálculo de una muestra

$$G = \frac{n}{(n-1)(n-2)}$$

$$\sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Para el cálculo de población

$$\gamma_1 = \frac{\sum (x_i - \mu)^3 / N}{(\sum (x_i - \mu)^2 / N)^{\frac{3}{2}}}$$

- Entropía: refleja la cantidad de información o sorpresa en un proceso estocástico. Alcanza su máximo cuando todos los posibles eventos tienen la misma probabilidad, implicando alta incertidumbre. Si un evento es mucho más probable que otros, la entropía será baja, indicando menor incertidumbre.

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

- Longitud de la curva: Si estás refiriéndote a la longitud de una curva en un contexto matemático, se trata de una medida que determina la extensión de una curva en el espacio.

$$s = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

- Raíz cuadrada media: También conocida como el valor cuadrático medio o RMS (del inglés, Root Mean Square), es una medida estadística de la magnitud de una cantidad variable que puede ser usada tanto para cantidades positivas como negativas. RMS es una forma de determinar el tamaño de una cantidad que varía en el tiempo. Es particularmente útil en contextos donde las variaciones tienen tanto direcciones positivas como negativas respecto de un punto de referencia.

$$RMS = \sqrt{\frac{1}{b-a} \int_a^b [f(t)]^2 dt}$$

- **Cruce por cero:** Es un término utilizado en diversas disciplinas, incluyendo la electrónica, el procesamiento de señales y la estadística, para describir un punto en el que una función o señal cambia de signo, es decir, pasa de ser positiva a negativa o viceversa. Este concepto es especialmente relevante en el análisis de señales de tiempo continuo o de tiempo discreto. No hay una "fórmula" única para determinar un cruce por cero, ya que es más un evento que una métrica calculable.
- **Potencia de banda:** Se refiere al cuadrado del valor absoluto de la transformada de Fourier de la señal, calculado en una banda de frecuencia específica. En el contexto de EEG, estas bandas de frecuencia suelen estar asociadas con ritmos cerebrales distintos, como las ondas delta, theta, alfa, beta y gama, cada una de las cuales corresponde a diferentes estados de actividad cerebral.

Matemáticamente, esto puede expresarse como:

$$P_{banda} = \int_{f_1}^{f_2} |X(f)|^2 df$$

potencia de banda para señales discretas:

$$P_{banda} = \sum_{n=f_1}^{f_2} |X(f)|^2$$

3.4. Base de datos Berna Barcelona.

En este capítulo se analiza la base de datos EEG de código abierto de Berna Barcelona que proporciona la información necesaria para la extracción de características del padecimiento que se está tratando en este trabajo y que contiene la información de pacientes con un registro focal (pacientes con epilepsia focal preoperatoria) y pacientes con registro no focal (pacientes sanos). En el ámbito del estudio de las oscilaciones cerebrales, se ha desarrollado un enfoque especializado utilizando lo que se denomina canales EEG focales. Estos canales tienen la capacidad de detectar variaciones iniciales en las trazas EEG, y su eficacia ha sido validada a través de revisiones visuales realizadas por expertos en el campo de la neurología. A pesar de que el método visual puede tener limitaciones subjetivas, su aplicación sigue siendo de gran relevancia en el campo de la neurociencia. Para garantizar la precisión de los resultados, se extrajeron 3750 combinaciones de señales, tomadas de segmentos específicos de grabaciones EEG, equivalentes a 20 segundos o 10240 muestras. Se establecieron criterios estrictos de calidad, omitiendo datos post-ictales o aquellos recopilados hasta tres horas después de un episodio convulsivo. Cada combinación fue sometida a un escrutinio

visual meticuloso, descartando aquellas que presentaban interferencias o artefactos notables. Sin embargo, ciertas interferencias menores, como las derivadas de la electricidad estándar a 50Hz, no fueron motivo de exclusión. Es importante destacar que la recopilación de datos no se vio influenciada por características clínicas específicas, permitiendo así un análisis más objetivo. Las trazas seleccionadas se categorizaron y archivaron para análisis subsiguientes, y se utilizó un enfoque similar para las trazas no focales [44] .

Una vez que se tiene claro los datos de partida de los datos ya mencionados podemos mostrar las gráficas de las señales como muestra de dos electrodos que se mencionan cada uno con un par de electrodo (x) y un par de electrodo (y) como se ve en la siguientes figuras Figura 3.3 y Figura 3.4 respectivamente:

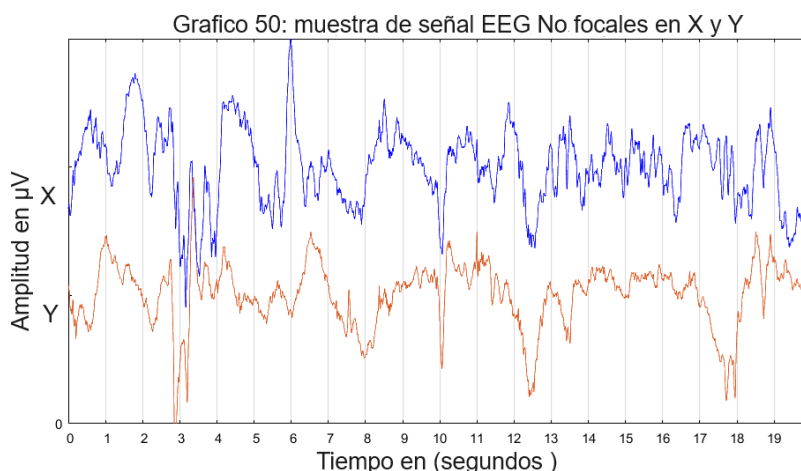


Figura 3.3: Señal azul electrodo (x) Señal Focal, señal naranja electrodo (y) señal focal

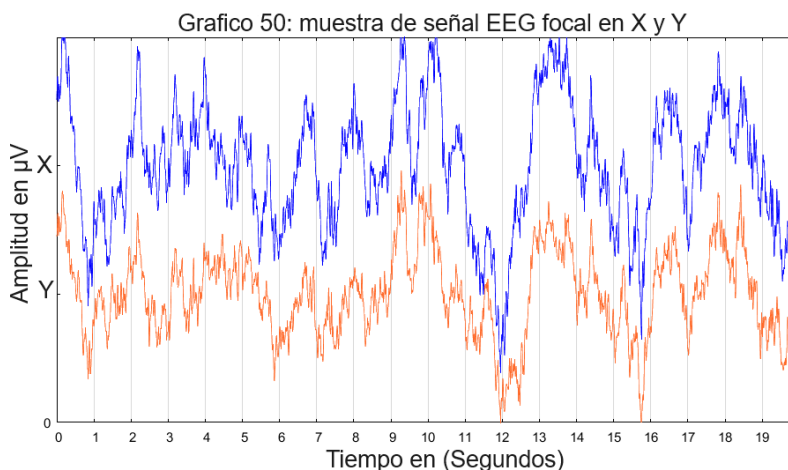


Figura 3.4: Señal azul electrodo (x) Señal No Focal, señal naranja electrodo (y) señal No focal

El registro de EEG focal y no focal para este estudio se tomó de la base de datos abierto ya antes mencionado. La motivación para que usemos esta base de datos se debe a su gran volumen y disponibilidad. Aparte de esto, muchas investigaciones han utilizado la misma base de datos para sus estudios [45], [46], [47], [48]. Este conjunto de datos incluye registros de EEG intracraneales de cinco pacientes con epilepsia farmacorresistente y se colocaron electrodos en las posiciones Fz y Pz. Las señales de EEG se muestrearon a 512 o 1024 Hz, para más o menos de 64 canales, respectivamente. Todas las señales de EEG se pasaron a través de un filtro de paso de banda Butterworth de cuarto orden (0,5 Hz 150 Hz).

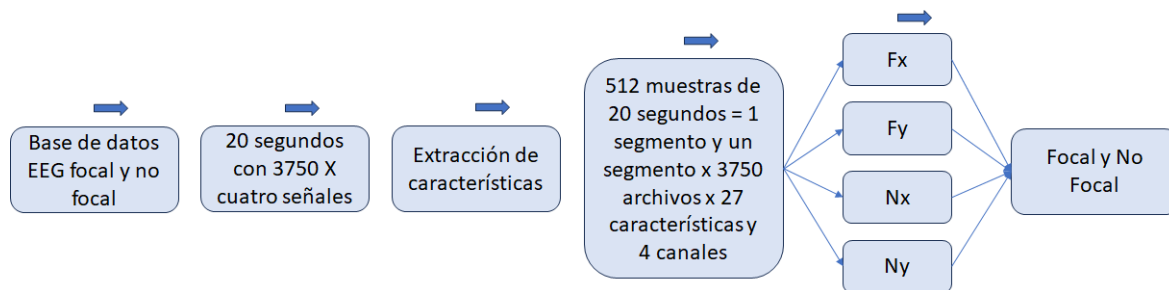


Figura 3.5: Los electrodos señalados como Fx (Focal x), Fy(focal y) y para No focal es (Nx y Ny)

La base de datos consta de 3750 pares de señales EEG focales y 3750 no focales de 20 segundos de duración cada una. Canal de EEG focal representado por la señal “x” y canal vecino representado por la señal “y”. Cada canal, tanto focal como no focal, consta de 10240 muestras como ya se había mencionado. La serie focal “x”, la serie focal “y”, la serie no focal “x” y la serie no focal “y” fueron representadas por Fx, Fy, Nx y Ny respectivamente. En la Figura 3.5 se muestran la señal EEG focal y no focal. De las cuatro series temporales de EEG disponibles, se formaron cuatro tareas de clasificación como se muestra en la figura anterior.

3.4.1. Extracción de características

Las señales EEG (electroencefalografía) registran la actividad eléctrica del cerebro y son ampliamente utilizadas en investigaciones y aplicaciones clínicas relacionadas con el cerebro y el sistema nervioso. La extracción de características estadísticas de estas señales puede proporcionar información valiosa sobre la actividad cerebral y ayudar en el análisis y la interpretación de los datos. Aquí se destacan algunos aspectos clave:

- **Análisis de patrones:** La extracción de características estadísticas permite identificar patrones o regularidades en las señales EEG. Al calcular medidas estadísticas como la media, la mediana, la desviación estándar, entre otras, se pueden obtener características que representan

la amplitud, la variabilidad y otras propiedades de las señales. Estas características estadísticas pueden revelar información sobre la actividad cerebral en diferentes estados, como la presencia de eventos o patrones anormales.

- **Reducción de dimensionalidad:** Las señales EEG a menudo contienen una gran cantidad de información, lo que puede dificultar el análisis y la interpretación. La extracción de características estadísticas permite reducir la dimensionalidad de los datos, es decir, representar las señales en un espacio de características más compacto y significativo. Al seleccionar características estadísticas relevantes, se puede reducir el tamaño de los datos y conservar solo la información más relevante para el análisis.
- **Detección de eventos y anomalías:** Las características estadísticas extraídas de las señales EEG pueden utilizarse para detectar eventos específicos o anomalías en la actividad cerebral. Por ejemplo, ciertas medidas estadísticas pueden ayudar a identificar picos de actividad, cambios abruptos en la amplitud o patrones anormales que pueden estar asociados con condiciones médicas o trastornos neurológicos.
- **Clasificación y diagnóstico:** La extracción de características estadísticas también es útil en la clasificación y el diagnóstico de diferentes estados o trastornos cerebrales. Al utilizar algoritmos de aprendizaje automático o técnicas de reconocimiento de patrones, las características estadísticas pueden ser utilizadas como entrada para entrenar modelos que puedan distinguir entre diferentes condiciones cerebrales, como la epilepsia entre otros.

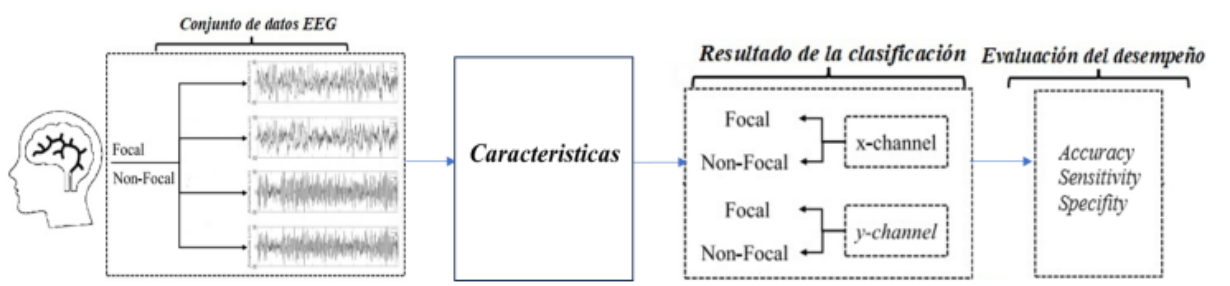


Figura 3.6: Metodología del procesamiento de la señal EEG

Por lo tanto, la extracción de características estadísticas en señales EEG es importante para analizar, interpretar y extraer información relevante de estas señales. Las características estadísticas proporcionan medidas cuantitativas que permiten identificar patrones, reducir la dimensionalidad de los datos, detectar eventos y anomalías, y ayudar en la clasificación y el diagnóstico de condiciones cerebrales. Esto contribuye a un mejor entendimiento de la actividad cerebral y a la aplicación de técnicas más efectivas en el ámbito de la medicina.

3.5. Base de datos de la Universidad de Bonn

Para este estudio, se obtuvo la base de datos de la Universidad de Bonn, que está disponible al público de forma gratuita. La estructura de la base de datos se divide en cinco conjuntos, etiquetados como A, B, C, D y E. Cada conjunto consta de 100 segmentos de EEG con una duración estandarizada de 23.6 segundos. Estos segmentos se seleccionaron cuidadosamente de registros de EEG, tras un análisis visual para eliminar cualquier interferencia o artefacto, como los causados por actividad muscular o movimientos oculares [49].

Los conjuntos A y B se derivaron de EEG superficial registrado en cinco voluntarios sanos, con procedimientos estándar de colocación de electrodos. El conjunto A corresponde a registros con los ojos abiertos, mientras que el conjunto B corresponde a registros con los ojos cerrados. Los conjuntos C-E se extrajeron de registros EEG prequirúrgicos. Los EEG seleccionados para los conjuntos C, D y E pertenecen a pacientes con control de convulsiones. Específicamente, el conjunto D proviene de una zona ictal que no se registró directamente, el conjunto C se registró en el hemisferio cerebral contra lateral y el conjunto E muestra actividad ictal. Por lo tanto, decidimos utilizar solo dos conjuntos en nuestro estudio, que fueron el conjunto C y E ver Figura 3.7.

Set	Simbolo	Pacientes	descripción	Segmento
A	Z	con ojos abiertos	sin crisis	100
B	O	con ojos cerrados	sin crisis	100
C	N	sin crisis	inter ictal	100
D	F	sin crisis	inter ictal	100
E	S	con crisis	ictal	100

Figura 3.7: Concentrado de datos de la universidad de Bonn

Todos los registros se realizaron utilizando un sistema de amplificación de 128 canales, con una referencia promedio común, y se eliminaron electrodos con actividad patológica o artefactos evidentes. Los datos se convirtieron de forma analógica a digital con una resolución de 12 bits y se almacenaron a una frecuencia de muestreo de 173.61 Hz, con un filtro de banda ajustado entre 0.53 y 40 Hz (Bonn repositorio 2023).

3.5.1. Características

Cada técnica de extracción de características nos proporciona información valiosa sobre los ritmos y potenciales neuronales que estamos investigando. A continuación, detallamos las características que hemos explorado en este estudio relacionados con los registros EEG.

Para abordar este análisis, hemos considerado 10 indicadores estadísticos que describen aspectos temporales de los datos, como la media, la mediana, la moda, la desviación estándar, el mínimo, el máximo, la asimetría, la curtosis, la movilidad y la complejidad. Además, hemos aplicado técnicas de Wavelet para calcular características como el máximo, el mínimo, la media y la desviación estándar en cada una de las sub-bandas de la señal. Esto nos permite capturar tanto la esencia como la variabilidad de los datos. A partir de esta segmentación, hemos derivado valores críticos y coeficientes, lo que nos ha llevado a un total de 30 características distintas.

En resumen, nuestra metodología incorpora un conjunto de 10 indicadores temporales y 20 indicadores orientados a la frecuencia para analizar exhaustivamente los registros EEG. Este enfoque se basa en la investigación realizada por A. Jaramillo García y su equipo en 2015, y nos permite obtener una comprensión más completa de los datos y los patrones que estamos estudiando.

3.5.2. Espacio multidimensional

El espacio multidimensional de características es un concepto que se utiliza para definir la dimensionalidad de un conjunto de datos. Esta dimensionalidad se refiere al número de características que se utilizan para describir un dato en particular. Por ejemplo, si se tiene un objeto que se describe con 20 características diferentes, se dice que se tiene un espacio de 20 dimensiones, como se representa en la Figura 3.8. Sin embargo, a medida que aumentamos la dimensionalidad, el espacio crece de manera exponencial, lo que significa que los datos se dispersan y su densidad disminuye. Este fenómeno se conoce como la "maldición de la dimensionalidad". La maldición de la dimensionalidad plantea un desafío porque a medida que agregamos más características, se requieren volúmenes cada vez mayores de datos para representar todas las posibles combinaciones de características. Esto puede hacer que sea más difícil identificar patrones en los datos y puede llevar a problemas de sobre-ajuste en los modelos [50].

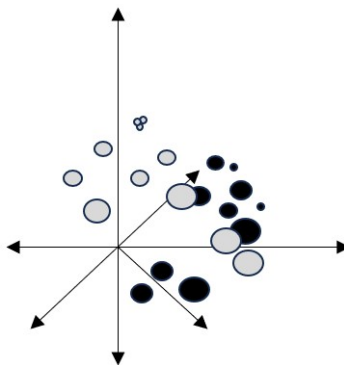


Figura 3.8: Características en un espacio multidimensional

En el contexto del aprendizaje automático, la selección adecuada del conjunto de características es de vital importancia. Un conjunto óptimo de características puede mejorar el rendimiento de un modelo, pero una elección inapropiada puede dificultar la eficacia del algoritmo. Por lo tanto, es esencial encontrar un equilibrio adecuado entre la cantidad de características y la calidad de las mismas. Nuestro estudio se enfoca en investigar la epilepsia, una enfermedad neurológica de gran relevancia. Para llevar a cabo nuestra investigación, hemos aprovechado la base de datos de la Universidad de Bonn, la cual contiene registros de señales EEG etiquetadas por expertos en la materia. En términos generales, el conjunto C comprende señales interictales, es decir, señales obtenidas de pacientes con epilepsia, pero capturadas en momentos libres de actividad epileptiforme, cada una con una duración de 23 segundos. Por otro lado, el conjunto E está compuesto por señales de pacientes que han experimentado episodios epilépticos.

En nuestro análisis, aplicamos la extracción de características y empleamos la transformada Wavelet, una herramienta altamente versátil en el procesamiento de señales. Una ventaja fundamental de la transformada Wavelet radica en su capacidad para adaptar los intervalos temporales según las necesidades específicas de las frecuencias que deseamos analizar. En situaciones que requieren una precisión mayor en frecuencias bajas, optamos por intervalos de tiempo más largos. En cambio, cuando es crucial captar información en frecuencias más altas, elegimos intervalos más cortos.

No obstante, previo a sumergirnos en los aspectos técnicos y procedimentales del análisis Wavelet, se torna imprescindible una elección meticulosa de una función Wavelet específica. Esta función, conocida como "Wavelet madre", actúa como un modelo o plantilla para todas las ventanas o segmentos que se aplicarán en el análisis. Existe una amplia variedad de funciones Wavelet, cada una con sus propias particularidades y aplicaciones. Entre las más notables y ampliamente reconocidas en la comunidad científica y técnica se incluyen la Haar, Daubechies, Biorthogonal, Coiflets, Symlets y Meyer. La elección de una familia en particular debe estar guiada por las características y requisitos particulares del problema en cuestión, ya que cada una posee propiedades que pueden resultar más adecuadas para ciertos contextos y aplicaciones específicas.

Capítulo 4

Clasificación ML

4.1. ¿Que es el Machine Learning?

El Machine Learning (Aprendizaje Automático) es una subdisciplina de la Inteligencia Artificial (IA) que se centra en desarrollar sistemas que aprenden y mejoran a partir de la experiencia, en lugar de depender de una programación explícita. Se basa en la idea de que las máquinas pueden identificar patrones, tomar decisiones y hacer predicciones basándose en datos, en lugar de instrucciones explícitas. El Machine Learning cuenta con las siguientes características [51] [52] [53]:

- **Datos:** Es fundamental para cualquier tarea de machine learning. A través del análisis de datos, los algoritmos pueden extraer patrones y hacer generalizaciones.
- **Modelos:** Una vez que se identifica un patrón en los datos, se crea un modelo que representa ese patrón. Este modelo se utiliza posteriormente para hacer predicciones o tomar decisiones sin ser explícitamente programado para hacerlo.
- **Entrenamiento:** Es el proceso de alimentar datos a un algoritmo de machine learning para aprender y adaptarse. A medida que se expone a más datos, el modelo ajusta sus parámetros internos para mejorar su rendimiento.
- **Predicción:** Una vez entrenado, el modelo puede hacer predicciones o decisiones basadas en nuevos datos.

También cuenta con una clasificación de los tipos de aprendizaje la cual mencionamos de una manera general, y mencionamos los enfoques más comunes [54], [55]:

- **Aprendizaje Supervisado:** Se utiliza un conjunto de datos etiquetado, es decir, cada entrada

en el conjunto de datos está asociada a una etiqueta o resultado conocido. El algoritmo aprende de estos datos y, una vez entrenado, puede hacer predicciones o clasificaciones en datos nuevos.

- **Aprendizaje No Supervisado:** Se trabaja con datos que no tienen etiquetas. El objetivo aquí es encontrar estructuras o patrones subyacentes en los datos, como la agrupación (clustering) o la reducción de dimensionalidad.
- **Aprendizaje Por Reforzamiento:** En este enfoque, un agente aprende cómo comportarse en un entorno realizando ciertas acciones y recibiendo recompensas o castigos en función de las consecuencias de esas acciones.
- **Aprendizaje Semi-supervisado:** Es una combinación de aprendizaje supervisado y no supervisado. Se utiliza cuando se tiene un gran volumen de datos y solo una pequeña parte de ellos está etiquetada.
- **Aprendizaje Activo:** En este tipo de aprendizaje, el modelo tiene la capacidad de solicitar etiquetas para ejemplos específicos en el conjunto de datos, por lo general cuando está menos seguro de la clasificación.
- **Aprendizaje de Transferencia:** Consiste en tomar un modelo previamente entrenado (en una tarea o conjunto de datos) y adaptarlo para una tarea diferente pero relacionada, aprovechando el aprendizaje previo.

Y se mencionara las aplicaciones del Machine Learning: El machine learning tiene numerosas aplicaciones, que van desde la recomendación de productos en sitios web, detección de fraudes en transacciones bancarias, reconocimiento de voz y de imágenes, vehículos autónomos, hasta la predicción de enfermedades en medicina.

En resumen, el Machine Learning es una herramienta poderosa y versátil que permite a las máquinas aprender y adaptarse a tareas complejas sin ser programadas explícitamente para cada una de ellas, abriendo la puerta a soluciones para problemas que antes eran difíciles o imposibles de resolver con la programación tradicional.

4.2. Tipos de Clasificadores

Los clasificadores son herramientas esenciales en el campo del Machine Learning que se utilizan para categorizar o etiquetar datos en clases específicas. A continuación, se presenta una breve descripción de los clasificadores más comunes [56]:

1. **Árbol de decisión.**

Los árboles de decisión no son modelos estadísticos basados en la estimación de parámetros de una ecuación dada. Así que, en lugar de estimar un modelo estadístico formal, actúan como algoritmos que clasifican a través de divisiones consecutivas. Estos son especialmente útiles cuando se maneja una gran cantidad de datos. Una de sus principales fortalezas es su naturaleza descriptiva, lo que facilita comprender e interpretar las decisiones que el modelo toma, exponiendo patrones complejos en los datos que no serían perceptibles con técnicas de regresión tradicionales.

Los árboles de decisión emergen del campo del aprendizaje automático y la inteligencia artificial. A partir de una base de datos dada, generan esquemas lógicos que sirven para solucionar problemas. A menudo se les llama segmentación jerárquica. Esta técnica es tanto explicativa como descomposición, y opera mediante un proceso iterativo y descendente que, desde una variable dependiente, genera grupos homogéneos a través de combinaciones de variables independientes, abarcando todos los casos en la muestra.

Los componentes fundamentales de un árbol de decisión son: nodos, ramificaciones y hojas. Mientras que los nodos simbolizan las variables iniciales, las ramificaciones se refieren a los distintos valores que estas variables pueden adoptar, y las hojas determinan los diferentes resultados posibles. Iniciando un árbol de decisión, nos encontramos con el nodo raíz, el cual representa la variable más significativa en el proceso de categorización. Si bien los algoritmos de aprendizaje de árboles de decisión generan modelos con variados grados de complejidad y coherencia, un conjunto de datos con inconsistencias puede llevar al modelo a adaptarse a esas fallas, resultando en lo que denominamos "sobre-ajuste". Para abordar este desafío, es esencial controlar la expansión del árbol adaptando los algoritmos, buscando crear modelos más universales. Este proceso es comúnmente referido como poda.

2. **Bosque aleatorio.**

El clasificador Random Forest es un método de aprendizaje supervisado que utiliza un conjunto de árboles de decisión para mejorar la precisión y controlar el sobreajuste, que es común en los árboles de decisión individuales. La forma en como funciona, se describe a continuación:

→ **Creación de un conjunto (forest) de árboles de decisión:**

- Se inicia seleccionando aleatoriamente "N" muestras del conjunto de datos de entrenamiento con reemplazo. Este método de muestreo se llama bootstrap sampling.
- Para cada muestra bootstrap, se construye un árbol de decisión. Al entrenar cada árbol, en cada división, se selecciona un subconjunto aleatorio de características

(variables) en lugar de usar todas las características disponibles. Esto añade diversidad a los árboles y es una clara diferencia entre Random Forest y el árbol de decisión clásico.

- El número de características que se pueden seleccionar es un hiper parámetro llamado Max-features y puede ser ajustado para optimizar el rendimiento del clasificador.

→ **Entrenamiento de cada árbol de decisión:**

- Cada árbol se construye de manera recursiva. En cada nodo del árbol, se elige la mejor división a partir del subconjunto de características basándose en alguna medida de pureza o ganancia de información, como la entropía o el índice de Gini.
- Este proceso continúa hasta que se cumplen ciertos criterios de parada, como alcanzar una profundidad máxima del árbol o cuando no se puede mejorar la pureza de los nodos.

→ **Predicción:**

- Para hacer una predicción, se hace pasar una nueva entrada por cada uno de los árboles del conjunto.
- Cada árbol da una predicción (en el caso de clasificación, sería la clase más frecuente en las hojas alcanzadas). Luego, la clase que reciba la mayoría de votos a través de todos los árboles es la que se asigna a la entrada.

→ **Ventajas del método Random Forest:**

- Reducción de sobreajuste: Al utilizar múltiples árboles y subconjuntos de características, el clasificador es menos propenso a sobreajustarse a los datos de entrenamiento.
- Mejora de la precisión: La combinación de múltiples árboles generalmente da lugar a una mejor precisión en comparación con los árboles de decisión individuales.
- Manejo de datos no balanceados: Random Forest puede manejar bien los conjuntos de datos con clases desequilibradas.
- Importancia de las características: Puede evaluar la importancia de cada característica en la clasificación.

→ **Hiperparámetros clave:**

- Número de árboles: cuántos árboles se deben construir antes de promediar las predicciones.
- Profundidad del árbol: cuán profundo debe ser cada árbol.
- Número de características a considerar para la mejor división: cuántas características se deben probar para la división en cada nodo.

3. Vecinos más cercanos (KNN).

El Clasificador de Vecinos Más Cercanos (k-NN, por sus siglas en inglés "k-Nearest Neighbors") es un método de aprendizaje supervisado, simple pero poderoso, que se utiliza tanto para la clasificación como para la regresión. A continuación, se presentan algunas características de este clasificador:

→ **Fundamento Básico:** El algoritmo k-NN se basa en la premisa de que las observaciones similares se encuentran cerca unas de otras en el espacio de características. Por tanto, la clase o el valor de una nueva observación se predice en función de las clases o valores de sus vecinos más cercanos en el conjunto de datos de entrenamiento.

→ **Funcionamiento:**

- Se elige un valor para 'k', que representa el número de vecinos más cercanos a considerar.
- Para una nueva observación, el algoritmo busca las 'k' observaciones más cercanas del conjunto de datos de entrenamiento.
- Para la clasificación, se asigna la clase que sea mayoría entre los 'k' vecinos más cercanos.
- Para la regresión, se toma un promedio (o mediana) de los valores de los 'k' vecinos más cercanos.

→ **Distancia:** La elección de la métrica de distancia es crucial. La distancia euclidiana es la más común, pero otras distancias como la de Manhattan, Minkowski o la de Mahalanobis también pueden ser utilizadas.

→ **Ventajas:**

- No requiere un entrenamiento explícito, ya que "memoriza" todo el conjunto de datos.
- Es intuitivo y fácil de entender.

- Funciona bien para conjuntos de datos pequeños y donde las fronteras de decisión son irregulares.

→ **Desventajas:**

- Es computacionalmente costoso para conjuntos de datos grandes, ya que necesita calcular la distancia a cada observación en el conjunto de datos para cada predicción.
- Es sensible a características irrelevantes y al ruido en los datos.
- La elección de 'k' y la métrica de distancia puede influir significativamente en los resultados.

El k-NN es un algoritmo fundamental en el aprendizaje automático y, a pesar de su simplicidad, puede ofrecer buenos resultados en muchos problemas prácticos, especialmente cuando los datos están bien estructurados y se elige una métrica de distancia adecuada.

4. Redes neuronales.

Las redes neuronales son un paradigma de aprendizaje automático inspirado en la forma en que funciona el cerebro humano. A continuación, se presenta un resumen simplificado sobre el clasificador de redes neuronales:

→ **Inspiración Biológica:** Las redes neuronales artificiales se inspiran en las neuronas del cerebro humano. Una neurona artificial recibe una serie de entradas, las procesa y produce una salida.

→ **Estructura Básica:**

- Neuronas: También conocidas como unidades o nodos, procesan la información.
- Capas: Las neuronas se organizan en capas: entrada, oculta(s) y salida.
- Conexiones: Las neuronas de una capa están conectadas a las de la siguiente. Cada conexión tiene un peso asociado.

→ **Funcionamiento:** Las entradas se introducen en la capa de entrada. La información se procesa a través de las capas ocultas. Se obtiene una salida, que es la predicción del modelo, en la capa de salida.

→ **Entrenamiento:**

- Se utiliza un conjunto de datos con entradas y salidas conocidas.

- La red aprende ajustando los pesos de las conexiones para minimizar el error entre sus predicciones y las salidas reales.
 - Este proceso suele realizarse mediante un algoritmo llamado backpropagation y optimizadores como el descenso del gradiente.
- **Función de Activación:** Las neuronas usan funciones de activación para introducir no linealidades en el modelo.
- **Ventajas:**
- Capacidad de aprender y modelar relaciones no lineales.
 - Adecuadas para datos de alta dimensionalidad, como imágenes.
 - Flexibles, ya que se pueden diseñar con múltiples capas y neuronas para adaptarse a diferentes tareas.
- **Desventajas:**
- Requieren grandes cantidades de datos para entrenarse adecuadamente.
 - Son "cajas negras", lo que significa que sus predicciones no siempre son fácilmente interpretables.
 - Sensibles a la inicialización de pesos y la arquitectura elegida.
- **Variantes:** Hay múltiples tipos de redes neuronales, como las redes neuronales convolucionales (CNNs) para el procesamiento de imágenes y las redes neuronales recurrentes (RNNs) para datos secuenciales.

5. Naive Bayes.

El clasificador Naive Bayes es un algoritmo de clasificación que se basa en el teorema de Bayes. Este teorema proporciona una forma de calcular la probabilidad de un evento en función de las probabilidades de ciertos eventos relacionados. Lo que hace ingenuo a Naive Bayes es su suposición de que las características de los datos son independientes entre sí. Aunque esta suposición puede no ser válida en muchos casos del mundo real, Naive Bayes puede ser sorprendentemente efectivo en la práctica. El algoritmo trabaja calculando la probabilidad de que un dato pertenezca a una clase específica dadas sus características. Luego compara estas probabilidades y asigna la clase con la probabilidad más alta. Es especialmente popular en aplicaciones como filtros de spam en correos electrónicos, clasificación de textos y análisis de sentimientos debido a su simplicidad y velocidad. Sin embargo, su precisión puede verse comprometida si las características del conjunto de datos están fuertemente correlacionadas.

Esta es la base matemática del algoritmo. Su fórmula es:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

donde:

- $P(A|B)$: es la probabilidad de que ocurra A dado de que B ha ocurrido.
- $P(B|A)$: es la probabilidad de que ocurra B dado de que A ha ocurrido.
- $P(A)$ y $P(B)$: son la probabilidad de que ocurra los eventos A y B respectivamente.

6. Máquina de Vector Soporte.

Las Máquina de Vectores Soporte (SVM, por sus siglas en inglés) son técnicas de aprendizaje supervisado diseñadas principalmente para tareas de clasificación y regresión. En la clasificación supervisada, los datos de entrada se conceptualizan como vectores con p dimensiones. Estos puntos se ubican dentro de un espacio más grande, cada uno asociado a una de dos categorías. El propósito de una SVM es crear un modelo que, basado en los puntos existentes, pueda determinar la categoría de un punto desconocido.

Su objetivo es distinguir estas categorías trazando un espacio (o margen) lo más amplio posible entre ellas. Así, cuando se introducen nuevos ejemplos, estos se categorizan en función de su ubicación respecto a ese margen.

Una característica esencial de las SVMs es la búsqueda del "hiperplano óptimo", que es básicamente una superficie que tiene la mayor separación con los puntos más cercanos a ella. Es por ello que a las SVMs a menudo se les denomina clasificadores de margen máximo". De este modo, los puntos que pertenecen a una categoría se situarán a un lado del hiperplano, mientras que los de la otra categoría estarán al lado opuesto.

4.3. Matriz de Confusión

Las matrices de confusión son herramientas fundamentales en el campo de aprendizaje automático y la ciencia de datos para evaluar el desempeño de los algoritmos de clasificación. A continuación, se ofrece un resumen sobre ellas [57]:

- **Definición:** Una matriz de confusión es una tabla que se utiliza para evaluar la calidad de las predicciones de un algoritmo de clasificación. Específicamente, compara las clasificaciones reales con las clasificaciones predichas por el modelo.

→ **Componentes básicos:**

- Verdaderos Positivos (VP): Casos en los que el modelo predijo positivo y la clase real es positiva.
- Verdaderos Negativos (VN): Casos en los que el modelo predijo negativo y la clase real es negativa.
- Falsos Positivos (FP): Casos en los que el modelo predijo positivo pero la clase real es negativa.
- Falsos Negativos (FN): Casos en los que el modelo predijo negativo pero la clase real es positiva.

→ **Métricas derivadas [58]:**

- Precisión: $VP + VN / (VP + VN + FP + FN)$. Mide la calidad de las predicciones positivas del modelo.
- Sensibilidad (Recall): $VP / (VP + FN)$. Mide la capacidad del modelo para detectar todos los positivos reales.
- Especificidad: $VN / (VN + FP)$. Mide la capacidad del modelo para detectar todos los negativos reales.

→ **Usos:**

- Evaluación del rendimiento: Ayuda a entender cómo se desempeña el modelo en términos de errores tipo I (FP) y errores tipo II (FN).
- Balance entre Precisión y Sensibilidad: En contextos donde uno de estos es más crítico que el otro (por ejemplo: detección de enfermedades), la matriz ayuda a ajustar el modelo de acuerdo con las necesidades.

→ **Multiclase:** Aunque se ha descrito en el contexto de la clasificación binaria, las matrices de confusión pueden extenderse a problemas de clasificación multiclase. En este caso, la matriz tiene dimensiones $M \times M$, siendo M el número de clases, y cada celda (i, j) indica cuántas veces la clase i fue clasificada como j .

→ **Visualización:** A menudo, las matrices de confusión se visualizan utilizando gráficos de calor (heatmaps) para facilitar su interpretación, especialmente en problemas multiclase.

4.4. Validación Cruzada

La validación cruzada, también conocida como "Cross-Validation", es un método empleado para verificar la eficacia de un análisis estadístico en situaciones donde los datos se dividen en conjuntos de entrenamiento y prueba. Esta técnica asegura que la eficiencia del modelo no depende de cómo se haya dividido el conjunto de datos como se ve en la Figura 4.1. Aunque se originó principalmente para evaluar modelos de regresión, su aplicación ahora abarca una variedad de tareas en Machine Learning. El enfoque más básico de validación cruzada implica entrenar el modelo con el conjunto de entrenamiento y luego probarlo con el conjunto de prueba, comparando las predicciones con los valores reales para determinar el error (conocido como método hold-out). Un enfoque más avanzado implica repetir este proceso múltiples veces con diferentes divisiones aleatorias, y luego promediar los resultados para obtener una medida de validación más robusta [59].

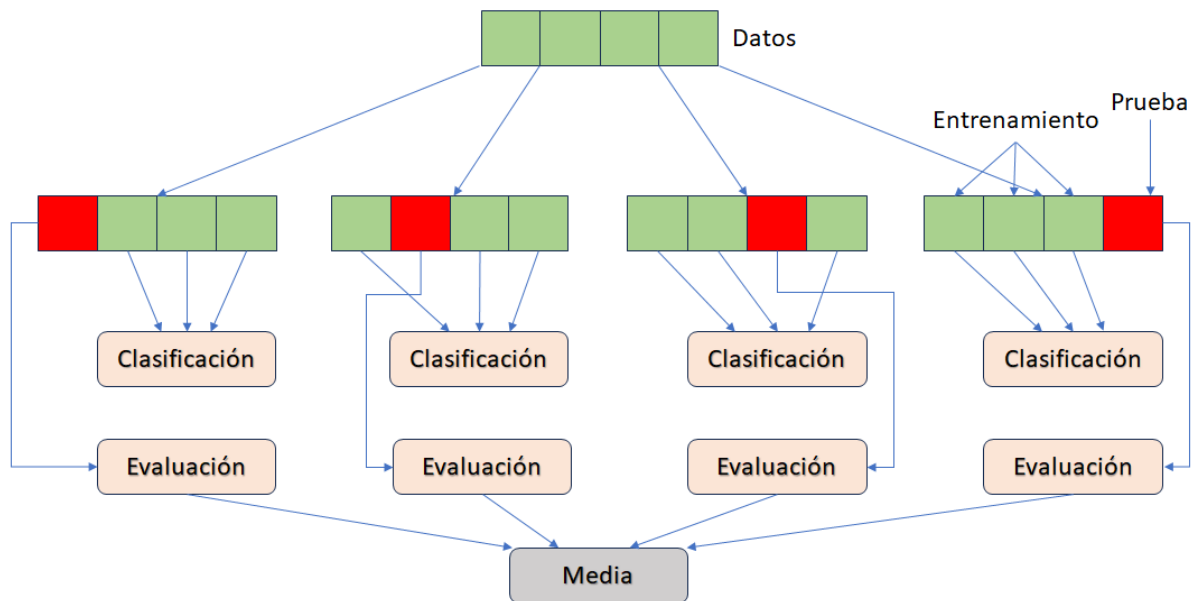


Figura 4.1: Aplicando validación cruzada a conjunto de datos

Capítulo 5

Base de datos sintética

5.1. Generación de señales sintética epiléptica y basal

Las señales sintéticas, también conocidas como señales artificiales, desempeñan un papel crucial en una amplia gama de campos y múltiples aplicaciones en la actualidad. Su importancia en la medicina radica en su capacidad para simular, recrear o generar información útil de manera controlada y predecible, a continuación se presenta las ventajas que ofrecen:

Ventajas de las señales sintéticas:

- Al no contar con bases de datos. Se generan señales que permiten conocer la característica de la patología, y por lo tanto tener información controlada o conocida y validar de manera teórica los algoritmos.
- Es posible generar bases de datos grandes para alimentar los algoritmos.
- Permiten comparar los resultados obtenidos con información real y realizar análisis de los mismos.

En este trabajo las señales sintéticas se generan basados en los ritmos cerebrales ya conocidos: Beta, Alfa, Theta Delta y Gamma de los cuales se realizó una descripción en la sección 1.2.2, estos ritmos cerebrales, dependiendo de sus parámetros en frecuencia y amplitud proporcionan en conjunto la información necesaria para poder determinar algunas patologías, tales como la epilepsia, la cual puede presentar patrones que salen de la normalidad. Para observar algunas alteraciones es necesario identificar o extraer información de los EEG, esto a través de la extracción de caracte-

rísticas y su clasificación. Cada señal ya sea sintética o normal es el resultado de estos ritmos y está asociado con diferentes estados mentales y niveles de actividad cerebral. Cada uno de estos ritmos proporciona información única sobre el funcionamiento de nuestro cerebro en diferentes circunstancias.

5.1.1. Parámetros de las señales sintéticas

En este trabajo se representa la señal epiléptica sintética como una suma de los cuatro ritmos cerebrales fundamentales ruido blanco adicional, así la señal generada solamente dependerá del tiempo y esto permite visualizarla como una señal unidimensional. La justificación de tal representación puede verse con más detalle en el Apéndice A. El hecho de que la señal pueda reducirse a un análisis unidimensional temporal, simplifica la problemática,

Como se explicó previamente, las señales neuronales de una persona sana se componen de los ritmos alfa, beta, delta y theta, en las frecuencias y amplitudes correspondientes que se manifiestan como oscilaciones cosenoidales con frecuencias específicas de 10 Hz, 20 Hz, 2 Hz y 6 Hz, respectivamente. Estas frecuencias son representativas de los distintos ritmos cerebrales como se observa en la ecuación 5.1, la cual representa una señal basal en condiciones normales y varían según la localización dentro de los lóbulos cerebrales. La señal basal se puede expresar como:

$$S_{\text{basal}} = S_{\text{alfa}} + S_{\text{beta}} + S_{\text{delta}} + S_{\text{theta}} \quad (5.1)$$

En el caso de que una señal represente actividad epiléptica, esta puede expresarse de acuerdo a la ecuación 5.2, en la que se incorpora la componente gamma con una frecuencia de 40 Hz, además de una descarga de sobreexcitación representada mediante dos perfiles gaussianos, uno positivo y otro negativo, de igual magnitud pero separados espacialmente por su ancho. Esto se representa como:

$$S_{\text{epile}} = S_{\text{alfa}} + S_{\text{beta}} + S_{\text{delta}} + S_{\text{theta}} + S_{\text{gamma}} + S_{\text{descar}} \quad (5.2)$$

Cada ritmo cerebral S_{ritmo} , se modela con una función senoidal como se ve en la ecuación 5.3 :

$$S_{\text{ritmo}} = C_{\text{ritmo}} \sin(2\pi f_{\text{ritmo}} t + \phi) \quad (5.3)$$

donde C_{ritmo} es la amplitud asignada al ritmo cerebral, f_{ritmo} la frecuencia específica, y ϕ una constante de fase con distribución uniforme entre 0 y 2π .

La descarga epiléptica S_{descar} se modela como:

$$S_{\text{descar}} = C_{\text{descar}} e^{-\frac{|t-t_e|^2}{2\gamma^2}} - C_{\text{descar}} e^{-\frac{|t-(t_e+2\gamma)|^2}{2\gamma^2}} \quad (5.4)$$

Aquí, t_e indica el momento de máxima descarga positiva y $t_e + 2\gamma$ el punto de descarga negativa mínima como se ve en la ecuación 5.4.

Finalmente, se añade ruido blanco a las señales S_{basal} y S_{epile} para simular las frecuencias presentes en cualquier sistema de adquisición, manteniendo una relación señal-ruido de 10 decibels. Este paso es crucial para reflejar las condiciones reales de medición y captura de señales neurofisiológicas.

5.1.2. Generación de señales sintéticas

Como se explicó en la sección anterior la señal sintética epiléptica se representa como suma de señales basales, esto permite analizarla como una distribución gaussiana, y observar la dispersión de los datos, lo que nos lleva a poder analizar los patrones de las descargas epilépticas focales sintéticas y por ende aplicar el algoritmo creado en MATLAB para así poder determinar el inicio del foco epiléptico.

En la Figura 5.1 se representa una señal sintética, la cual ha sido normalizada y filtrada. El análisis de esta señal se presenta en función del tiempo, la cual demuestra la presencia de una función gaussiana en dos ciclos, tanto negativo como positivo, generada a partir de la ecuación (5.4). Es crucial tener en cuenta que, al analizar descargas epilépticas en un electroencefalograma, los ciclos de estas pueden manifestarse tanto en valores negativos como positivos. La Figura 5.1 proporciona una visualización detallada de la descarga focal en azul y una señal normal en naranja, esta representa un evento epiléptico. En este contexto, la representación gráfica abarca un periodo de 10 segundos de la señal, destacando que el evento focal en cuestión tiene una duración de 5 segundos. Este análisis detallado contribuye a comprender mejor la dinámica temporal y las características particulares de la actividad cerebral sintetizada.

El término "descarga epiléptica" se sitúa en una perspectiva temporal y alude a un patrón de actividad anormal distintivo de la epilepsia. En condiciones normales, el estudio del electroencefalograma (EEG) revela una actividad eléctrica constante. La Figura 5.2 ejemplifica este estado de reposo o vigilia, presentando una señal sintética con características normales. En la misma representación gráfica, se observa otra trama que ilustra una descarga epiléptica anormal. Esta última exhibe notables variaciones en el tiempo y una amplitud en voltaje significativamente mayor. La diferencia

más evidente radica en las fluctuaciones de los picos, marcando una disparidad clara respecto a la actividad eléctrica típicamente constante en condiciones normales. Una comprensión detallada de estas variaciones temporales y amplitud anómala es esencial para identificar y entender la actividad epileptiforme en el EEG, destacando la importancia de la detección y caracterización de estas descargas en el diagnóstico de la epilepsia.

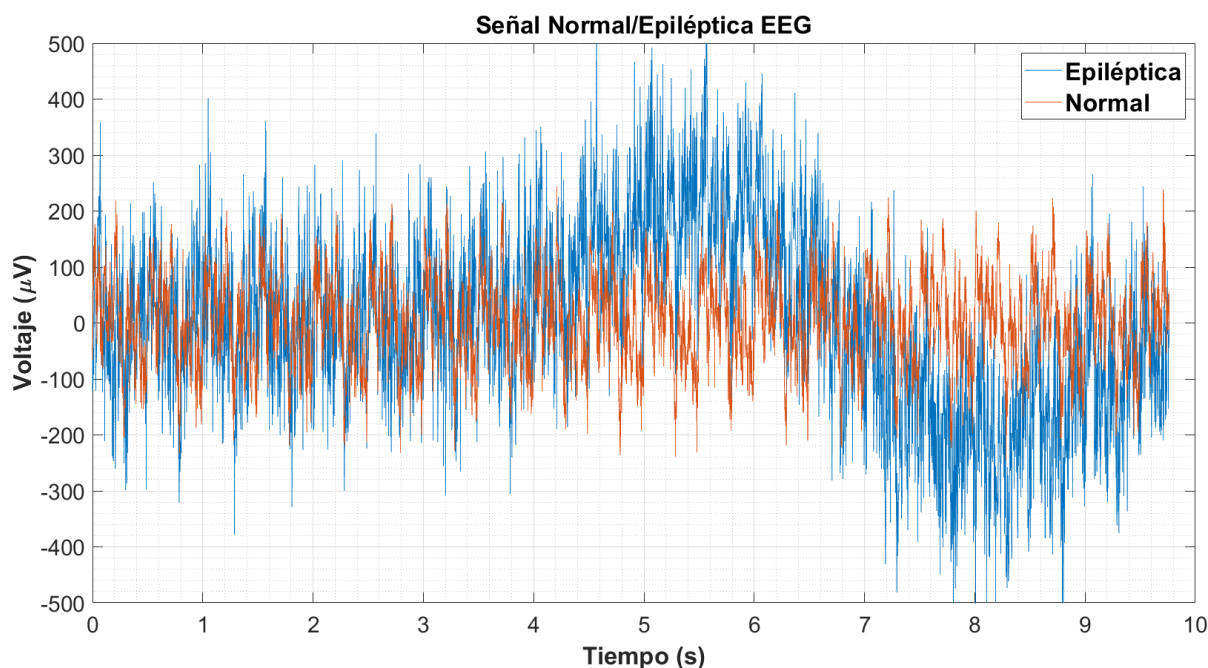


Figura 5.1: Representación de una descarga focal de duración de 5 segundos.

En el estudio de electroencefalogramas (EEG), la interpretación de las señales desempeña un papel crucial en la detección de patrones normales y anómalos en la actividad cerebral, lo cual podemos visualizar en la Figura 5.2 En la primera gráfica, se presenta una señal normal de EEG que refleja la actividad eléctrica cerebral en condiciones típicas de reposo o vigilia. Las ondas cerebrales exhiben una frecuencia regular y patrones de amplitud característicos. Se puede interpretar esta señal observando la distribución de las diferentes frecuencias de ondas y la amplitud correspondiente, destacando las variaciones esperadas según el estado mental del individuo.

En contraste, la segunda gráfica muestra una señal EEG focal con una forma gaussiana, indicando una actividad anómala en una región específica del cerebro. La interpretación se centra en identificar la localización focal de la actividad, describir la forma gaussiana en la señal y analizar la duración y amplitud de esta actividad anómala. En este caso, es esencial comparar estos patrones con la señal normal para destacar cualquier variación significativa.

Ambas interpretaciones deben considerar el contexto más amplio de la investigación o diagnóstico clínico, subrayando la importancia de estos análisis para comprender trastornos neurológicos, como la epilepsia. La conexión entre las características específicas de las señales y su relevancia

clínica contribuye a una interpretación más completa y precisa.

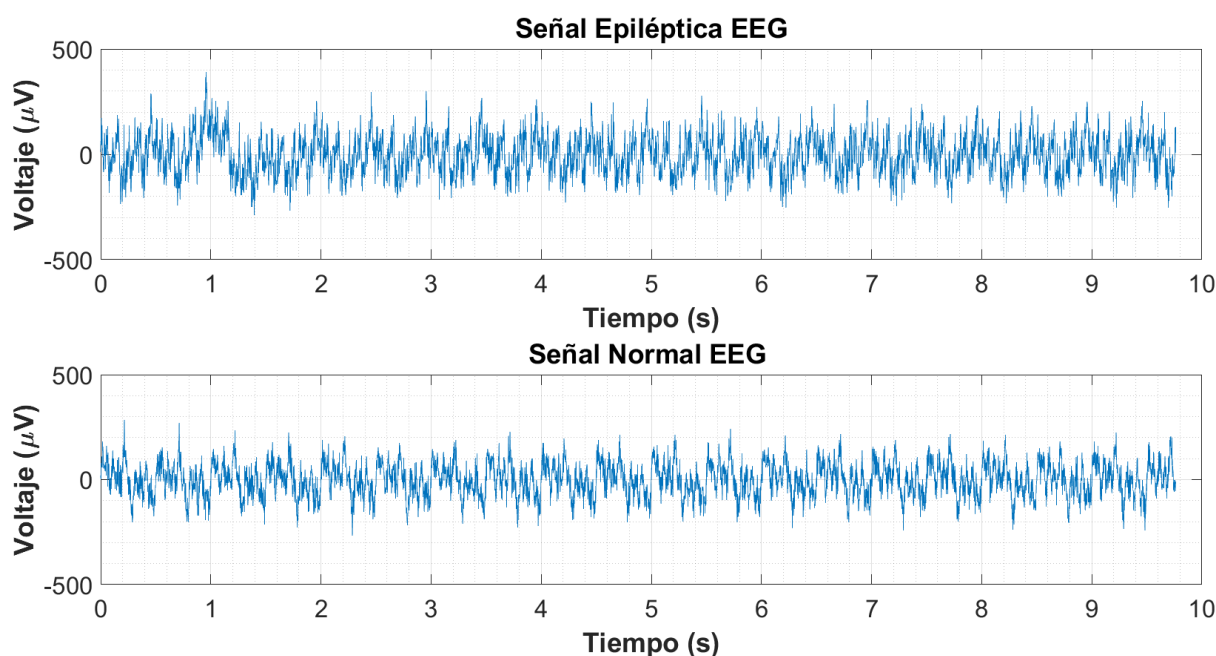


Figura 5.2: Representación de un segmento de la señal sintética normal y epiléptica.

En el análisis de eventos epilépticos, la función Gaussiana es una herramienta importante para representar y comprender la complejidad temporal de las descargas epileptiformes. En la Figura 5.3 se muestran cuatro momentos temporales distintos de un evento epiléptico con representaciones Gaussianas; esta función no sólo brinda una representación visual clara de la actividad eléctrica anormal, sino que también permite observar de manera precisa la evolución y variación de este evento a lo largo del tiempo. La función gaussiana se convierte así en un instrumento esencial para los cambios temporales de los eventos epilépticos, ofreciendo una valiosa ayuda en la identificación y caracterización de estos fenómenos en el estudio del electroencefalograma (EEG).

A la señal sintética generada se le aplicaron las 30 características, así se puede distinguir entre señales normales y epilépticas, la capacidad de discriminación se convierte en algo fundamental en la detección de patrones epilépticos durante el análisis de señales de electroencefalograma (EEG). La Figura 5.4, diseñada en función de 30 características específicas, representa las variaciones entre señales normales y epilépticas. Cada línea en la gráfica simboliza una de estas características.

Este conjunto de características surge de la exploración de 100 señales de EEG que exhiben actividad normal y las siguientes 100 señales que presentan actividad epiléptica, notoriamente marcada por una amplitud notable en la frecuencia. Cada característica captura la esencia única de las señales, y su variación se refleja en la gráfica. Notablemente, algunas de estas características muestran

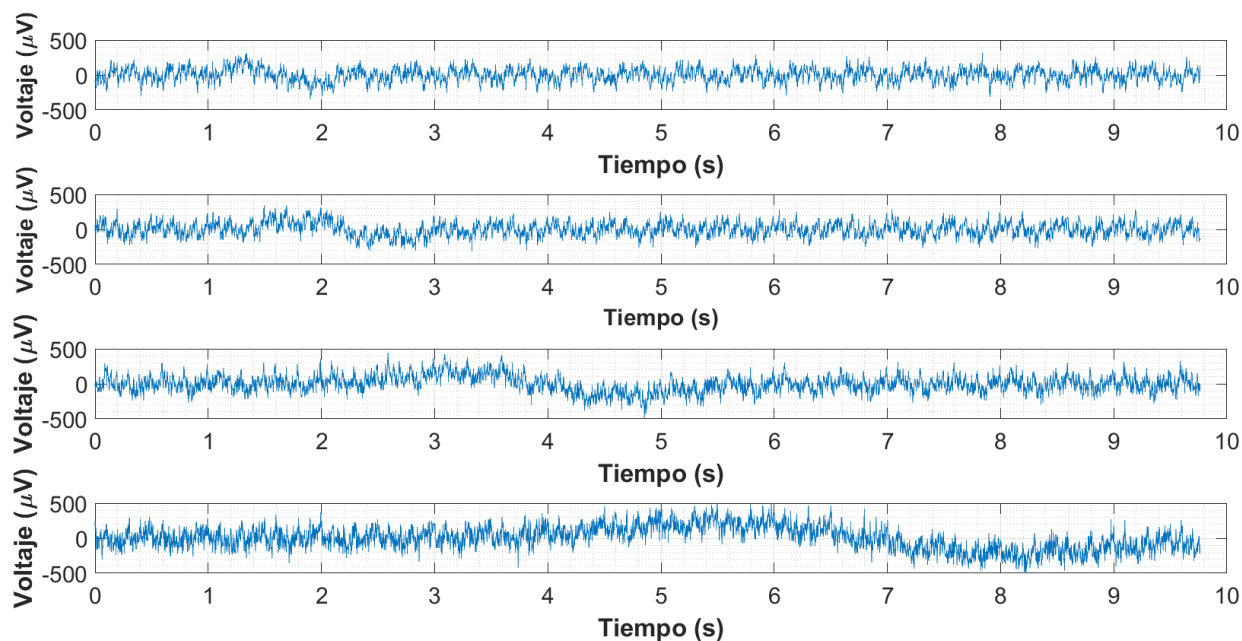


Figura 5.3: Descarga epiléptica focal en distintos momentos en función del tiempo

una actividad casi nula, manifestándose como líneas prácticamente rectas. Estos datos y características distintivas contribuye significativamente a la identificación precisa de patrones epilépticos en el EEG, reforzando la importancia de un análisis detallado y multi-facético en la interpretación de señales cerebrales.

5.2. Algoritmo para la generación de señales sintéticas

En la Figura 5.5, se muestra la estructura del algoritmo que permite generar la señal sintética :

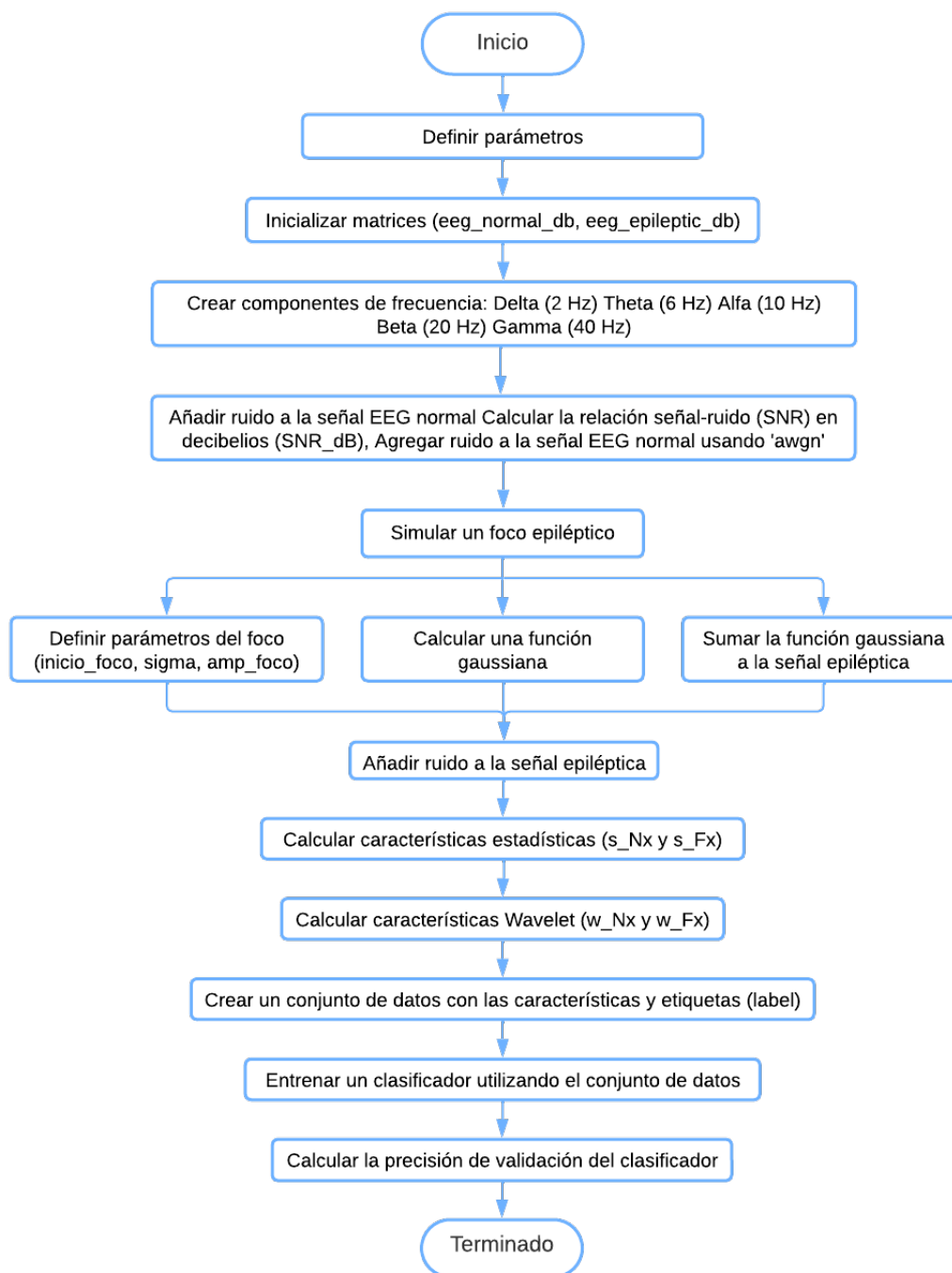


Figura 5.4: Diagrama de flujo de señales sintéticas

El desarrollo del algoritmos para el procesamiento de señales EEG sigue siendo un área de investigación vital debido a su impacto potencial en la medicina, la neurociencia, la tecnología y la mejora de la calidad de vida de los pacientes. Observando la continua evolución en este campo puede abrir nuevas posibilidades y contribuir significativamente al entendimiento del cerebro humano y sus

aplicaciones prácticas. En el presente trabajo se implementó un clasificador de Machine Learning. Este algoritmo simula señales EEG normales y epilépticas, donde se generan componentes de frecuencia, en esta parte se agrega ruido y es simulado un foco epiléptico. Donde, la señal es descompuesta en sub-bandas y se extraen sus características. Estas características se utilizan para entrenar un clasificador que distingue entre señales normales y epilépticas. La eficacia del clasificador se evalúa mediante un conjunto de datos de validación, proporcionando un enfoque completo para identificar patrones distintivos en las señales EEG.

El diseño del algoritmo destinado a la generación de señales EEG sintéticas es creado por las limitaciones de acceso a datos reales. Esta aproximación nos proporciona la capacidad de crear entornos controlados para evaluar con precisión la eficacia de algoritmos de clasificación y detección de epilepsia, como se destacó al principio de este capítulo. El programa desarrollado en MATLAB tiene la capacidad de generar una base de datos compuesta por 100 señales normales y 100 señales epilépticas. Los parámetros utilizados en la generación de estas señales incluyen una frecuencia de muestreo de 173Hz, el número de puntos de la señal de 4096 y un vector de tiempo, una vez propuestos los parámetros de inicio son generados los números de la base de datos a procesar, se crean matrices de almacenamiento. Estas matrices contienen los datos en 31 columnas de información, las cuales nos representan las 30 características utilizadas y 200 filas las cuales pertenecen a la base de datos, que representan las señales sintéticas.

El siguiente paso en nuestro proceso consiste en la descomposición de la señal mediante la transformada wavelet, la cual se explica detalladamente en la sección 3.2, en ella se destaca que la descomposición de estas señales se lleva a cabo en cinco sub-bandas con diferentes frecuencias, que incluyen cuatro de aproximación y una de detalle. Para este propósito, hemos implementado la wavelet madre Daubechies 4. Una vez descompuesta la señal, aplicamos características, como la media, la moda, la mediana y la desviación estándar, a las bandas de detalle. Posteriormente, con estas características aplicadas, construimos un espacio multidimensional de características, tal como se menciona en la sección 3.5.2. Tras la creación de este espacio, empleamos clasificadores, tales como la Máquina de Vectores Soporte, bosques aleatorios, vecinos más cercanos, redes neuronales, entre otros, según se detalla en la sección 4.2. En esta última, se exploran clasificadores ampliamente utilizados en el aprendizaje supervisado, formando parte de la rama de Machine Learning, como se expone en la sección 3.1 sobre las distintas ramas de la inteligencia artificial.

Con el propósito de obtener una clasificación precisa de las señales sintéticas, llevamos a cabo una validación cruzada. Esta técnica, mencionada en la sección 4.4 es fundamental en la evaluación de modelos de aprendizaje automático. Su principal objetivo radica en mejorar la precisión del sistema al exponerlo a datos no vistos. Al aplicar este enfoque, fortalecemos nuestro algoritmo y evitamos

una situación de sobre ajuste, asegurando así una evaluación más robusta de su rendimiento. Estas técnicas aplicadas al procesamiento de señales nos generan las llamadas matrices de confusión, las cuales se explican en la sección 4.3. Estas metodologías fueron empleadas tanto en las señales sintéticas como en la base de datos de la Universidad Bonn y así poder obtener los resultados de las matrices de confusión, como se muestra en las tablas siguientes.

5.3. Desarrollo del algoritmo

El tratamiento al cual son sometidas las señales EEG, en busca de propiedades particulares que solo se encuentran en función del cerebro han sido fundamentales para nuestra investigación, y esto consiste en extraer ciertas características que nos ofrecen información para su clasificación. Como primer paso de destaca la forma de medición de la actividad eléctrica del cerebro y esta se realiza en sus dos dominios tiempo y frecuencia, de esta forma podremos observar las variaciones en las señales EEG.

5.3.1. Algoritmo de clasificación

Se presenta el desarrollo del algoritmo de clasificación; esta es una base de datos real de la Universidad de Bonn [49], el cual se describe a continuación.

1.- Se cargan las muestras almacenadas de prueba desde el archivo `sets.mat` donde se contienen las matrices de los datos o grupos de la base de datos que son a b c d y e.

load sets.mat

2.- Se definen la frecuencia de muestreo (F_s) esta frecuencia fue seleccionada por la Universidad de Bonn así como los demás parámetros, el número de muestras por canal (*samples*), el vector tiempo en segundos (*time*) y el número de señales (*channels*) consultar Apéndice B.1. para obtener información detallada sobre el código.

$F_s = 173.61;$

samples = 4097;

time = *samples*/ F_s ;

channels = 100;

3.- Se inicializa el índice (k) y comienza un bucle sobre cada uno de los canales. Este índice se utiliza para llevar un seguimiento de la posición en la que se almacenarán las características estadísticas y de wavelet en las matrices

```

k = 1;
for j = 1 : channels

```

4.- En este paso, se generan las matrices específicas a b c d y e estas matrices representan los datos de las señales EEG en diferentes canales y se puede explicar de la siguiente manera, se extrae la columna j de cada matriz respectivamente.

```

data_a = a(:, j);
data_b = b(:, j);
data_c = c(:, j);
data_d = d(:, j);
data_e = e(:, j);

```

5.- En esta parte del código, se calculan las características estadísticas para cada segmento de señal en cada canal. en esta parte del código se crea una función independiente que se ejecuta al utilizar `stat_feat` aquí está contenido las características en función del tiempo, consultar Apéndice B.2 para obtener información detallada sobre el código.

```

s_a(k, :) = stat_feat(y_a);
s_b(k, :) = stat_feat(y_b);
s_c(k, :) = stat_feat(y_c);
s_d(k, :) = stat_feat(y_d);
s_e(k, :) = stat_feat(y_e);

```

6.- En esta parte del código, se calculan y almacenan características de la transformada de Wavelet para cada segmento de señal en cada canal. Esta es la segunda función que se crea para ejecutar en la función `wav_feat` y aquí están inmersas las características de frecuencia. En el siguiente paso se otorga solo se van a ocupado dos grupos a y e. En estos dos grupos de señales se encuentran el grupo c de señales que es un grupo de pacientes sanos y el grupo e que es de pacientes que presentan epilepsia focal y se asigna una etiqueta para poder cumplir con los requerimientos de sistema de aprendizaje supervisado para poder realizar una clasificación de estos dos grupos de señales, consultar Apéndice B.3 para obtener información detallada sobre el código.

```

w_a(k, :) = wav_feat(y_a);
w_b(k, :) = wav_feat(y_b);
w_c(k, :) = wav_feat(y_c);
w_d(k, :) = wav_feat(y_d);
w_e(k, :) = wav_feat(y_e);

```

7.- De esta forma podemos almacenar la información en un archivo conocido como `data` donde

se tiene las señales ya caracterizadas, para así poder usar esta información en las herramientas de Machine Learning.

```
statf = [s_a; s_e];
wavf = [w_a; w_e];
label = [zeros(2300, 1); ones(2300, 1)];
data = [statf wavf label];
```

8.- Se tiene acceso al programa de clasificación que proporciona MATLAB como se ve en la Figura 5.6, de estas herramientas se selecciona uno, esto es en el apartado de clasification Learner, una vez seleccionada esta herramienta se puede tener acceso al programa para anexar la información necesaria e introducir el archivo data para su procesamiento.

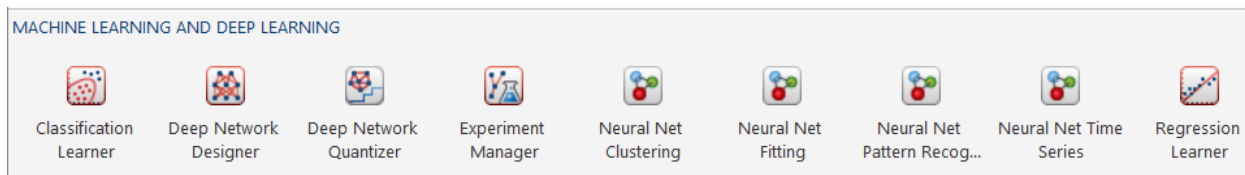


Figura 5.5: Barra de herramientas de Machine Learning incluida en MATLAB

9.- Ahora podemos introducir la información en la ventana principal del Clasification Learn, y se abre una nueva sesión como se observa en la Figura 5.7

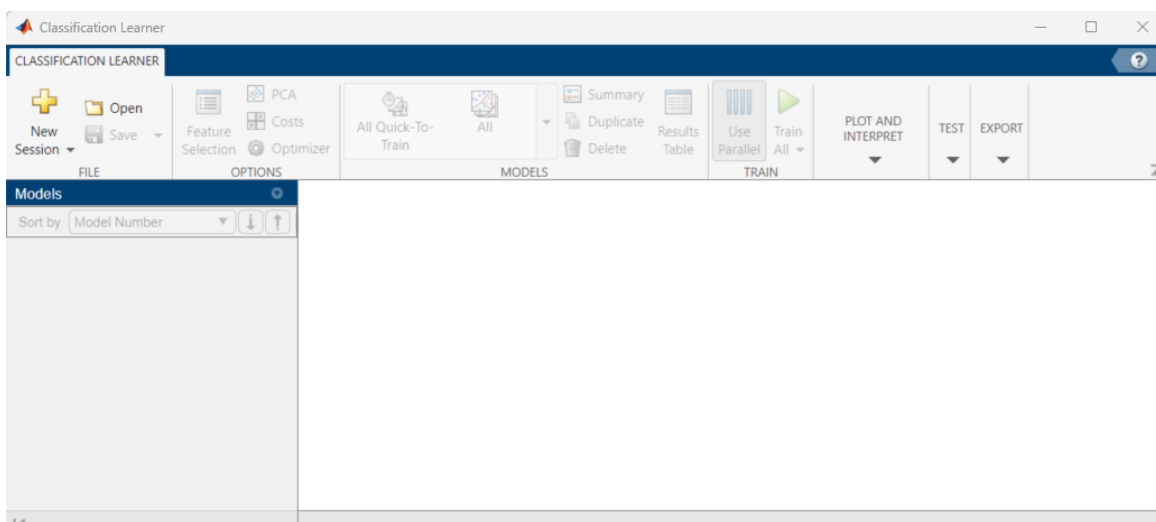


Figura 5.6: Inicio para abrir información en Workspace

Data set

Data Set Variable

data 4600x31 double

☒ Use columns as variables
☐ Use rows as variables

Response

☒ From data set variable
☐ From workspace

column_31 double 0 .. 1

Predictors

	Name	Type	Range
☒	column_1	double	-207.718 .. 256.103
☒	column_2	double	-304.5 .. 377.5
☒	column_3	double	-1868 .. 2047
☒	column_4	double	10.5702 .. 801.541

Add All Remove All

[How to prepare data](#) Refresh

Validation

Validation Scheme

Cross-Validation

Protects against overfitting. For data not set aside for testing, the app partitions the data into folds and estimates the accuracy on each fold.

Cross-validation folds 5

[Read about validation](#)

Test

☐ Set aside a test data set

Percent set aside 10

Use a test set to evaluate model performance after tuning and training models. To import a separate test set instead of partitioning the current data set, use the Test Data button after starting an app session.

[Read about test data](#)

Response variable is numeric. Distinct values will be interpreted as class labels.

Start Session Cancel

Figura 5.7: Workspace de matlab

10.- Una vez introducido el archivo data, recordemos que este archivo contiene toda la información de las características y como se puede observar vemos las etiquetas que se le dio a cada una de estas señales y así mismo podemos emplear la herramienta de validación cruzada. Así emplearemos esta interfaz de MATLAB que podemos seleccionar entre varios tipos de clasificadores y tendremos las opciones para entrenar el clasificador con la información que ya se proporcionó como se observa en la Figura 5.8, entrenaremos en clasificador y podremos obtener una gráfica de dispersión, la matriz de confusión como se observa en la Figura 5.9, podremos utilizar varios clasificadores para la comparación de métricas entre estos mismos.

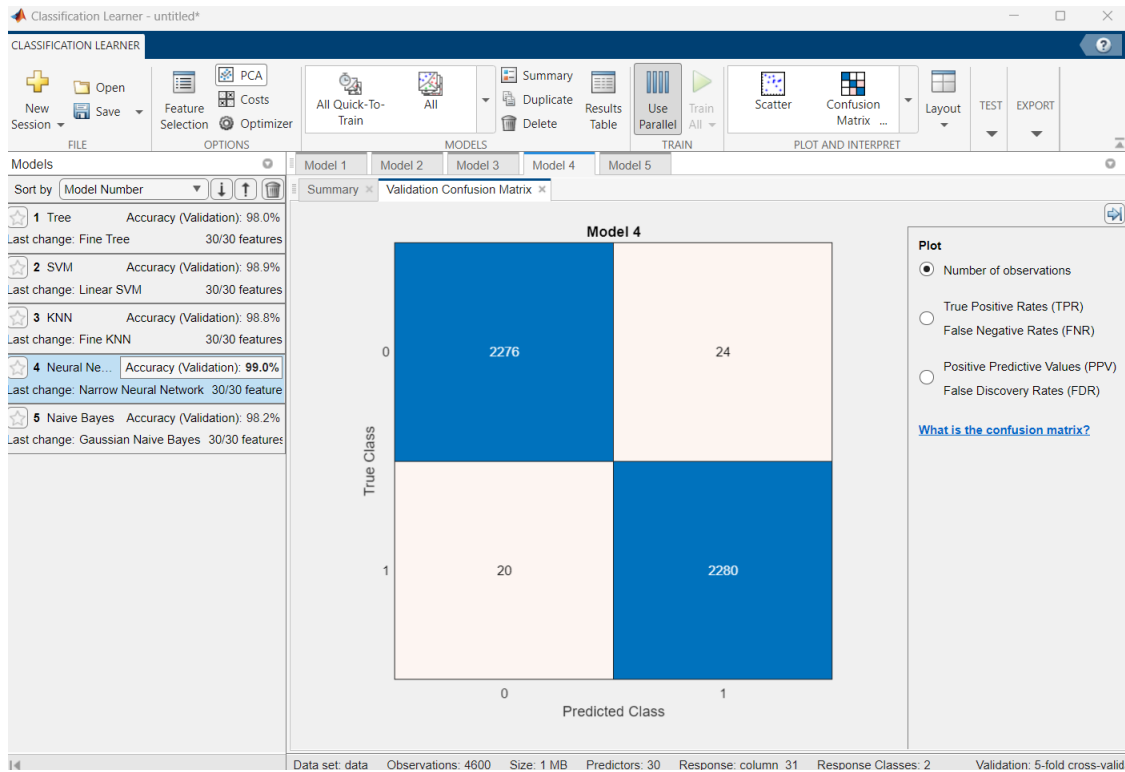


Figura 5.8: Resultados finales del Workspace

5.3.2. Desarrollo del Algoritmo usando señales sintéticas

Se emplean diversos componentes de frecuencia con el fin de simular tanto las señales EEG normales como las epilépticas. Estas señales se combinan y se les incorpora ruido para emularlas de una forma más realista. Adicionalmente, se introduce un evento epiléptico destinado a simular una señal epiléptica específica. Las señales resultantes se almacenan en matrices, contribuyendo así a la construcción de la base de datos de señales EEG.

1.- Se definen la frecuencia de muestreo (f_s), (N) nos proporciona el número de puntos de la señal, nuestro vector tiempo en segundos (t) y el número de señales a generar (num_signal), consultar Apéndice B.4 para obtener información detallada sobre el código.

$$N = 4097;$$

$$f_s = 173.61;$$

$$t = (0 : N - 1) / f_s;$$

$$num_signals = 100;$$

2.- Se crean 2 matrices para almacenar las señales generadas, normales y epilépticas:

$$eeg_normal_db = zeros(N, num_signals);$$

$$eeg_epileptic_db = zeros(N, num_signals);$$

- 3.- En este apartado nos basamos en la ecuación 5.3, la cual cuenta con una constante de fase, que esta uniformemente distribuida entre $0 \leq \phi \leq 2\pi$ radianes. Se definen los valores de C_{ritmo} para nuestro caso con un valor de 1 para cada ritmo cerebral y f_{ritmo} delta en 2 Hz con amplitud de $100 \mu\text{V}$, theta en 6 Hz con amplitud de $80 \mu\text{V}$, alfa en 10 Hz con amplitud de $60 \mu\text{V}$, beta en 20 Hz con amplitud de $20 \mu\text{V}$, gamma en 40 Hz con amplitud de $100 \mu\text{V}$.
- 4.- Para generar las señales con focos epilépticos, se definen los parámetros de acuerdo a la ecuación (5.4), C_{descar} de 100 a $400 \mu\text{V}$, t_e de 1 a 6 segundos, γ de 0.1 a 0.8 segundos.
- 5.- Para obtener señales epilépticas se suman los términos S_{gamma} y S_{descar} , a los demás ritmos. Se añade ruido blanco a todas las señales de 10 db con la intención de simular artefactos en las señales.
- 6.- Se guarda cada una de las señales en su respectiva matriz de almacenamiento mencionada como en el paso 2, para su posterior clasificación.

5.4. Resultados

En esta sección nos centramos en la clasificación de base de datos de la universidad de Bonn y la base de datos que nosotros generamos de forma sintética, que representan tanto patrones cerebrales normales como aquellos patrones denominados epilépticos. Estas señales, clasificadas mediante el diseño de nuestro algoritmo, permiten explorar el rendimiento de técnicas exploradas en secciones anteriores. El algoritmo propuesto que extrae características tanto en el dominio del tiempo como en el dominio de la frecuencia (con la herramienta DWT) definiendo un espacio de características en el que se aplican diferentes técnicas de clasificación automáticas, basadas en aprendizaje supervisado, para posteriormente comprobar su rendimiento mediante métricas como la precisión, exactitud, especificidad y sensibilidad.

Como fase inicial comenzamos con la extracción de características de cada señal disponible en nuestra base de datos (ya sea la de Bonn o la sintética), en esta etapa contamos con características en tiempo y frecuencia, en el caso del tiempo tomamos 10 características: Media, Mediana, Moda, Desviación estándar, Mínimo, Máximo, Oblicuidad, Curtosis, Movilidad y Complejidad. Para obtener las características en la frecuencia utilizamos la DWT con wavelet madre Db4 para descomponer las señales de EEG en diferentes niveles de detalle y aproximación, con la elección de 4 niveles de descomposición y tomando los coeficientes A4, D1, D2, D3 y D4, se extraen las características de media, mínimo, máximo y desviación estándar, con ello se Teniendo un total de 20 características obtenidas en el espacio de la frecuencia, definiendo así un espacio de características de 30 dimensiones. Es importante recalcar la ventaja de la transformada con la wavelet madre Db4 para descomponer las señales de EEG en diferentes niveles de detalle y aproximación, ya que permite una representación de las características frecuenciales de las señales EEG.

La elección de 4 niveles se basa en un compromiso entre la representación precisa de eventos temporales y la capacidad de resaltar patrones de frecuencia relevantes asociados a la epilepsia. La relación que existe en los anchos de bandas representativos de los ritmos cerebrales asociados a la condición de señales anormales y epilépticas. Esta configuración específica puede mejorar la capacidad de discernir entre patrones normales y epilépticos, contribuyendo así a un sistema de clasificación más preciso. A continuación, mostramos dos señales sintéticas generadas por el algoritmo, el cual se detalla en Apéndice B.4. Las señales son descompuestas por Db4 a cuatro niveles, esto se puede observar en el comportamiento de cada sub-banda, donde S es la señal original, A4, D1, D2, D3 y D4 definidas por el proceso de DWT. Como se puede observar en cada descomposición contamos con anchos de frecuencia distintos, que la señal original contiene. En la Figura 5.9 se presenta una señal sintética de tipo normal, y en la Figura 5.10 se presenta una señal sintética de tipo epiléptica.

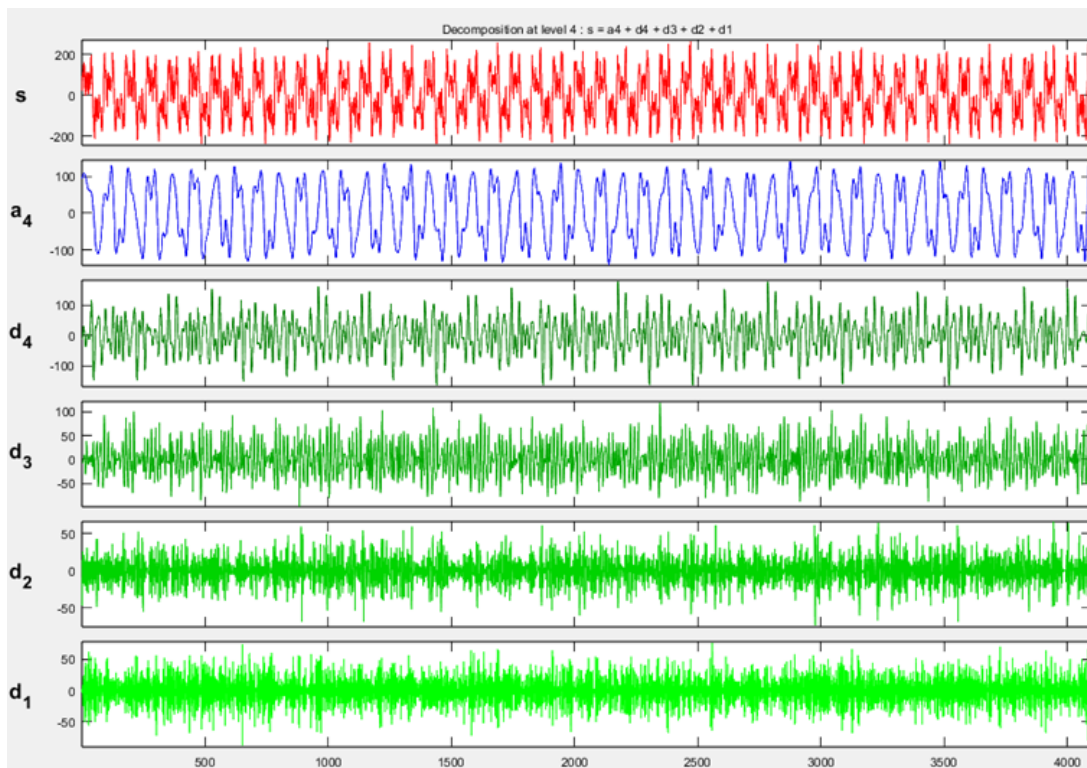


Figura 5.9: Descomposición de una señal sintética de tipo normal

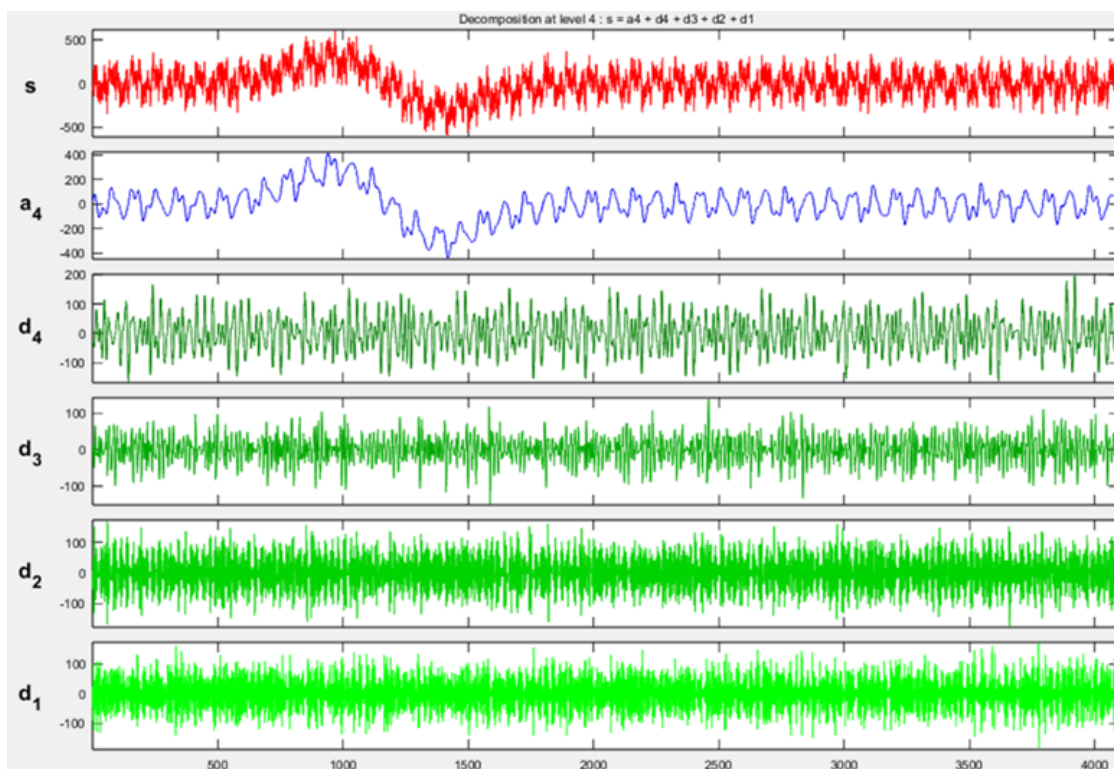


Figura 5.10: Descomposición de una señal sintética de tipo epiléptica

Una vez que se ha definido correctamente el espacio de características, se pueden observar sus valores obtenidos en la Figura 5.11, en esta figura se observa cómo difieren en valor las primeras 100 señales normales de las últimas 100, las cuales son epilépticas. Esta separación de valores nos permite realizar un primer análisis visual de los contrastes en el espacio de características que hemos definido.

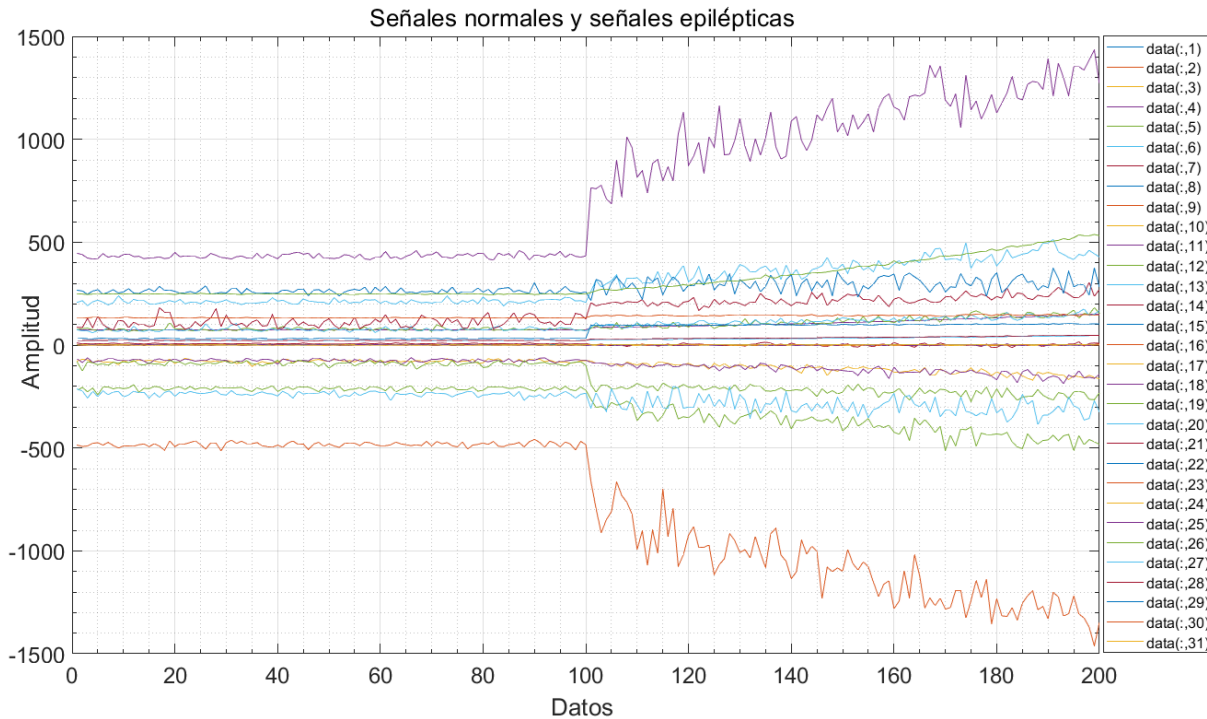


Figura 5.11: Representación del conjunto de 30 características

Debido a las observaciones realizadas en estos primeros análisis visuales en el espacio de características ver Figura 5.11, se procede a aplicar un conjunto de clasificadores. Estos clasificadores se eligen de acuerdo al tipo de datos que se manejan y se utilizan para procesos de clasificación, previamente definidos por características específicas. Se emplean seis clasificadores en total: Árbol de decisión (tree), Discriminante lineal, Máquina de Vectores de Soporte (SVM), Redes Neuronales Artificiales (ANN), k-Vecinos Más Cercanos (KNN) y Naive Bayes (NB).

Con el objetivo de evaluar el rendimiento de estos clasificadores en el espacio de características, se utilizan tanto la base de datos de la Universidad de Bonn como las señales sintéticas generadas por nuestra propuesta. Se busca encontrar un equilibrio entre las métricas de evaluación y el tiempo de procesamiento. Se selecciona finalmente un clasificador que cumpla con estos requisitos, los cuales son de vital importancia para su aplicación en un entorno práctico, con el fin de obtener las matrices de confusión, como se observa en la Figura 5.12. Estos clasificadores logran métricas

significativas, como se puede apreciar en la Figura 5.13.

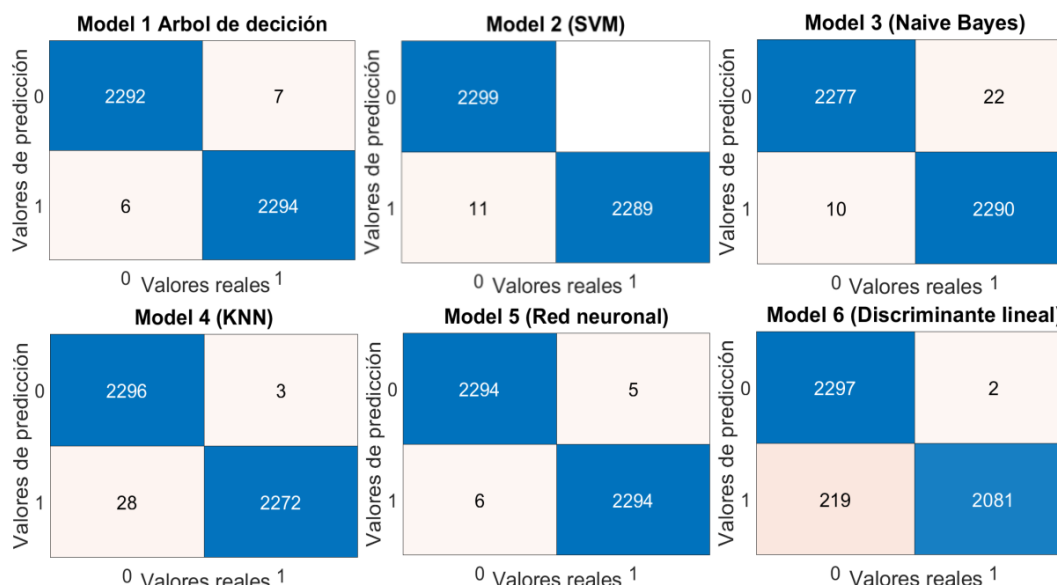


Figura 5.12: Matrices de confusión utilizados para la clasificación de los datos de la Universidad de Bonn

Conjunto de datos C-E Wavelet M. Daubechies, Bonn				
Clasificadores	Precisión	Exactitud	Especificidad	Sensibilidad
Árbol de decisión	99.60%	99.70%	99.60%	99.70%
SVM	100%	99.70%	100%	99.50%
Naive Bayes	99%	99.30%	99%	99.50%
KNN	99.80%	99.30%	99.80%	98.70%
Red neuronal	99.70%	99.70%	99.70%	99.70%
Discriminante lineal	99.90%	95.10%	99.90%	91.20%

Figura 5.13: Resultados de la clasificación de la base de datos de la Universidad de Bonn

De manera similar, se sigue la estrategia para las señales sintéticas con el fin de obtener las matrices de confusión, como se observa en la Figura 5.14. Estos clasificadores logran métricas significativas, como se puede apreciar en la Figura 5.15.

Es necesario discutir que, debido al control de cada uno de los parámetros de las señales sintéticas, no solo se logra comprender la presencia e importancia de cada ritmo cerebral, sino que también se permite ubicar una descarga en el tiempo junto con su amplitud y duración. Estos valores pueden definirse mediante rangos estadísticos encontrados en la patología o, específicamente, para una base de datos que cuente con una muestra significativa de pacientes con epilepsia.

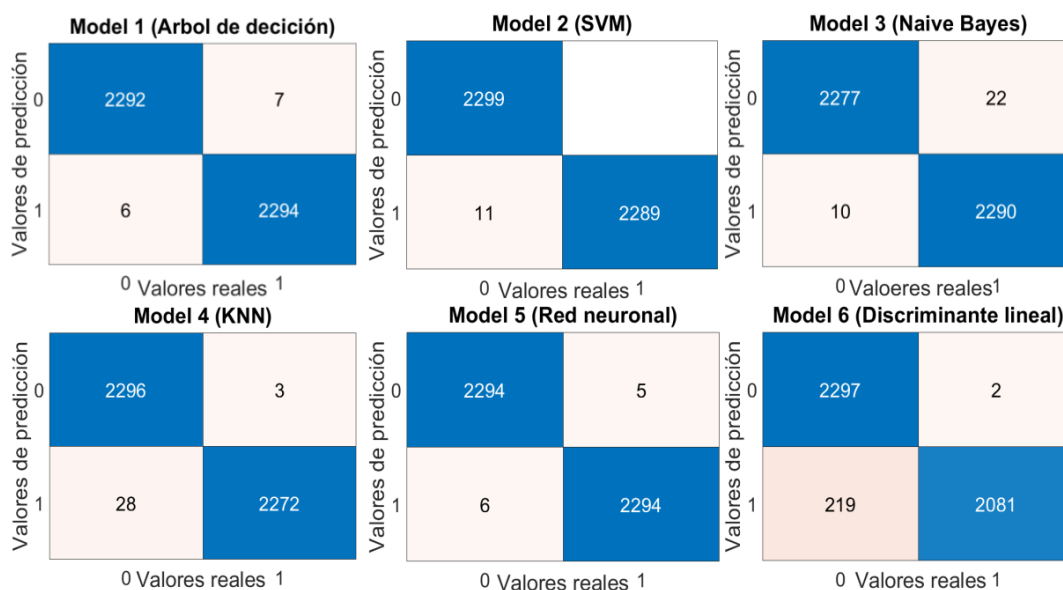


Figura 5.14: Matrices de confusión utilizados para la clasificación de la base de datos sintética

	Cojunto de datos C-E Wavelet M. Daubechies, sintéticas			
Clasificadores	Precisión	Exactitud	Especificidad	Sensibilidad
Árbol de decisión	99.60%	99.70%	99.60%	99.50%
SVM	100%	99.70%	100%	99.50%
Naive Bayes	99%	99.30%	99%	99.50%
KNN	99.80%	99.30%	99.80%	98.70%
Red neuronal	99.70%	99.70%	99.70%	99.70%
Discriminante lineal	99.90%	95.10%	99.90%	91.20%

Figura 5.15: Resultados de la clasificación de la base de datos sintética

En la aplicación de clasificadores de Machine Learning para analizar señales EEG reales (a través de bases de datos públicas) y señales sintéticas con el propósito de clasificar características, se exploraron diversos aspectos de las señales a tratar. Estos aspectos se derivan principalmente de la morfología que caracteriza una señal EEG como normal o epiléptica. Aunque un médico puede detectar visualmente estas diferencias, se requiere de una amplia experiencia y un uso significativo de tiempo, lo que genera un análisis costoso y prolongado.

Por lo tanto, mediante la recopilación de bases de datos etiquetadas por expertos en el tema, esta experiencia puede ser aprendida por un sistema de clasificación automática. Esto contribuye a mejorar la identificación de la epilepsia, guiando al profesional clínico para explorar con análisis más determinantes la certeza de este prediagnóstico sugerido por el clasificador. En la exploración de los clasificadores propuestos, se ha seleccionado el clasificador SVM debido a sus altas métricas

de rendimiento, su eficiente manejo de datos grandes y su buen desempeño en la base de datos real. La flexibilidad ofrecida por diferentes funciones de kernel le permite abordar el tratamiento de datos no lineales de manera efectiva.

A modo de comparación con otros trabajos consultados en nuestro estado del arte, llevamos a cabo un análisis comparativo del clasificador SVM. En la Figura 5.16, se presenta una tabla que muestra la métrica de Precisión, la cual es comparable con nuestra precisión del 100 % sobre la base de datos de la Universidad de Bonn.

Investigador	Kai Fu	Qi Yuan	Elly Matul	Wail Mardini	Ozlem K.	Umme Salma
Métrica	Precisión	Precisión	Precisión	Precisión	Precisión	Precisión
SVM	99.12	96.5	83.3	100	90.3	96.84

Figura 5.16: Datos de la Universidad de Bonn con distintos Investigadores y sus Métricas

5.5. Conclusiones

Se estudiaron los distintos tipos de epilepsia, con el objetivo de entender mejor cómo se manifiesta esta condición neurológica. Sin embargo, debido a la falta de bases de datos accesibles que incluyan EEG para varios tipos de epilepsia, se decidió simplificar el análisis a una clasificación fundamental: pacientes normales y aquellos con epilepsia focal (usando una base de datos específica). Se ha realizado un análisis detallado de las características en las señales EEG. Se destaca la importancia de la transformada wavelet para una descomposición eficiente en el dominio de la frecuencia al separar en sub-bandas de frecuencias la señal original. Esto facilita un análisis eficiente de los ritmos cerebrales, permitiendo extraer información importante. Este enfoque ha mejorado la capacidad para distinguir entre señales normales y señales de epilepsia focal.

Destacamos la importancia de elegir cuidadosamente la función madre de la wavelet, ya que esta elección puede tener un impacto significativo en los resultados obtenidos. Diferentes wavelets poseen propiedades distintas, y la selección debe basarse en la naturaleza específica de las señales EEG y el tipo de información que se busca extraer. En nuestro caso, nos enfocamos en la morfología que distingue las señales epilépticas de las normales. De igual manera es esencial determinar el número óptimo de niveles de descomposición; un número demasiado alto puede llevar a la sobreextracción de características, mientras que uno demasiado bajo podría perder información crucial. La elección puede depender de la frecuencia de muestreo y las características particulares del problema. En nuestro trabajo, nos basamos en los ritmos cerebrales y rangos de frecuencias máximos y mínimos que contienen información de la corteza cerebral, permitiéndonos comprender cómo cada sub-banda contribuye a la discriminación entre clases. Es importante tener en cuenta que la aplicación de la transformada wavelet puede ser exigente computacionalmente, especialmente con conjuntos de datos extensos o múltiples canales EEG, por lo que para trabajos futuros, es importante esta consideración.

En el trabajo se examinaron diferentes técnicas de Machine Learning el proceso de clasificación, centrándose especialmente en el clasificador Máquina de Vector Soporte (SVM), de las cuales fue seleccionada SVM, debido a que presentó mejores resultados respecto a otros clasificadores, bajo condiciones similares. Los resultados obtenidos en este trabajo son comparables a otros estudios [11, 14, 19, 21, 25, 27, 28, 30, 31 y 35]. Durante nuestras pruebas con varios clasificadores, el SVM destacó por su alta precisión, exactitud, especificidad y sensibilidad, particularmente, mostrando un buen rendimiento incluso en situaciones del mundo real; es versátil gracias a su capacidad para utilizar diferentes funciones de kernel, lo que permite manejar eficazmente patrones no lineales en los datos EEG si es preciso.

A la par, se desarrolló una simulación que involucra un control de parámetros de una señal definida, fundamentada en una aproximación con enfoque fisiológico en los ritmos cerebrales, tomando en cuenta la caracterización de dichos ritmos como señales senoidales con una frecuencia determinada la cual es la dominante en su respectiva banda. Se incluyó una descarga epiléptica caracterizada por un pulso gaussiano que permitió determinar la ubicación, duración y amplitud de una descarga epiléptica de tipo focal, mediante la definición de cada uno de estos parámetros y sus valores más probables extraídos de la literatura consultada (Apéndice A). Este enfoque ha demostrado ser eficaz al representar de manera aproximada y práctica los patrones de actividad epiléptica, simplificando los requisitos computacionales y mejorando la comprensión de las características de las señales cerebrales.

Se generó un algoritmo eficaz que facilita la clasificación y detección de señales EEG en 2 clases: normales y epilépticas, justificado por porcentajes mediante métricas de evaluación, comparándolo con los distintos trabajos. Cabe mencionar que las bases de datos públicas analizadas cuentan con tratamientos previos para su limpieza, y balanceo de los dos grupos a clasificar. Finalmente la combinación de técnicas de clasificación con un procesamiento previo en el dominio de la frecuencia, análisis de características y simulación ha contribuido significativamente a la robustez y precisión del estudio.

5.6. Trabajo a futuro

En el marco de la clasificación implica considerar varios aspectos, como la selección de características, donde se destacan técnicas como la prueba t de Student o la prueba chi-cuadrado para evaluar la importancia de dichas características. Asimismo, se contempla la aplicación del Análisis de Componentes Principales (PCA) con el propósito de transformar las características originales en un conjunto reducido y no correlacionado. Un aspecto es la identificación de datos atípicos, donde se emplean herramientas como gráficos de caja (box plots) y diagramas de dispersión, junto con técnicas estadísticas como la desviación estándar o el rango intercuartílico (IQR).

En las señal sintéticas se busca mejorar la versatilidad, se considera la inclusión de la distribución de múltiples canales de un arreglo EEG, incorporando artefactos de movimiento ocular, musculares, cardíacos, entre otros, que comúnmente están presentes durante la adquisición de datos. En la selección de kernels para SVM, se busca identificar aquel que mejor se adapte a la naturaleza específica de los datos y del problema abordado. Además, se plantea la necesidad de analizar un modelo matemático que refleje de manera fiel la complejidad inherente a la detección de señales cerebrales en el cuero cabelludo. Este análisis se orienta a retratar de manera precisa los desafíos asociados con la detección de patrones en datos cerebrales.

Bibliografía

- [1] L. Palacios and E. Palacios, “La epilepsia a través de los siglos”, 2001.
- [2] E. Goldesohn, “Historical perspectives and future directions. Willie E. Ed. The treatment of epilepsy, principles and practice”, 1993.
- [3] O. Zayachkivska, “Adolf Beck, co-founder of the EEG an essay in honour of his 150th birth-day”, p. 1, 03 2013.
- [4] P. Gloor, “Hans Berger on electroencephalography”, *American Journal of EEG Technology*, vol. 9, no. 1, pp. 1–8doi:10.1080/00029238.1969.11080728, 1969.
- [5] L. A.-Stevens, “Exploradores del cerebro”, 1974.
- [6] M. Shen, “An EEG based real-time epilepsy seizure detection approach using discrete Wavelet transform and machine learning methods”, *ELSEVIER*, vol. 77, no. 103820, 2022.
- [7] G. A. Betancourt, “Máquina de vectores soporte (SVM)”, *Scientia Et Technica*, vol. 1, no. 27, pp. 67–72, 2005.
- [8] K. M. Rabby, “Epileptic seizures classification in EEG using PCA based genetic algorithm through machine learning ”, *ACM Digital Library*, pp. 17–24, 2021.
- [9] M. W. S. H. B. W. Z. Lv, “A novel deep learning model based on the ICA and Riemannian manifold for EEG-based emotion recognition”, *ELSEVIER*, vol. 378, 2022.
- [10] F. R. C. Soldevilla, “Clasificación de señales EEG usando IDA y QDA aplicado a una interfaz cerebro computadora basada en p300”, *TECNIA*, vol. 28, no. 2, 2018.
- [11] A. Subasi and M. I. Gursoy, “EEG signal classification using PCA, ICA, IDA and support vector machines”, *Expert systems with applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [12] M. Savadkoohi, “A machine learning approach to epileptic seizure prediction using electroencephalogram (EEG) signal”, *ELSEVIER*, vol. 40, no. 3, pp. 1328–1341, 2020.

- [13] G. R. Sabrina López, “Evaluación del desempeño de la transformada ondícula y de Hilbert-Huang en la detección de los signos de somnolencia en el electroencefalograma”, *Ingeniería de Telecomunicaciones*.
- [14] K. Fu, J. Qu, Y. Chai, and Y. Dong, “Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM”, *Biomedical Signal Processing and Control*, vol. 13, pp. 15–22, 2014.
- [15] F. J. Martínez Albaladejo, “Evaluación experimental de aprendizaje máquina extremo aplicado a los sistemas de interfaz cerebro ordenador basados en imaginación de movimiento,” *Repositorio Digital*, 2018.
- [16] M. J. Luna de Luis, “Procesamiento digital de señales de EEG para clasificación de trastornos psicóticos mediante aprendizaje de máquinas”, *Repositorio académico de la Universidad de Chile*, 2021.
- [17] T. H. O. Bustillos, “Tendencias y perspectivas en la detección de depresión mediante el análisis de señales electroencefalográficas (EEG): Una revisión sistemática de la literatura”, *XIKUA Boletín Científico de la Escuela Superior de Tlahuelilpan*, vol. 11, no. 22, 2023.
- [18] J. L. Sarmiento-Ramos, “Aplicaciones de las redes neuronales y el Deep Learning a la Ingeniería biomédica”, vol. 19, no. 4, pp. 1–18, 2020.
- [19] Q. Yuan, W. Zhou, S. Li, and D. Cai, “Epileptic EEG classification based on extreme learning machine and nonlinear features”, *Epilepsy research*, vol. 96, no. 1-2, pp. 29–38, 2011.
- [20] M. Kastner, “Generalized functional relevance learning vector quantization”, pp. 27–29, 2011.
- [21] E. M. Imah and A. Widodo, “A comparative study of machine learning algorithms for epileptic seizure classification on EEG signals”, in *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 401–408, IEEE, 2017.
- [22] I. Ortiz Ramírez, “Machine learning para BCI orientada al procesamiento de las señales EEG en tiempo real”, 2021.
- [23] J. L. S. Ramos1, “Applications of neural networks and Deep Learning to biomedical engineering”, vol. 19, no. 4, pp. 1–18, 2019.
- [24] F. Soria Fernández, “Análisis de datos de electroencefalogramas para detección de epilepsia fotosensible”, 2021.

- [25] W. Mardini, M. M. B. Yassein, R. Al-Rawashdeh, S. Aljawarneh, Y. Khamayseh, and O. Meqdadi, “Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers”, *IEEE Access*, vol. 8, pp. 24046–24055, 2020.
- [26] BitBrain, “Cómo el Deep Learning está revolucionando el procesamiento de señal de EEG”, 2020.
- [27] O. K. Cura, M. A. Ozdemir, and A. Akan, “Epileptic EEG classification using synchrosquee-zing transform with machine and Deep Learning techniques”, in *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 1210–1214, 2021.
- [28] M. Masum, H. Shahriar, H. M. Haddad, and W. Song, “A statistical summary analysis of window based extracted features for EEG signal classification”, in *2021 IEEE International Conference on Digital Health (ICDH)*, pp. 293–298, 2021.
- [29] M. Illanes, “Influencia de los pulsos binaurales en los niveles de estrés inducido por realidad virtual, medidos por electroencefalograma y electrocardiograma”, vol. 21, no. 1, 2023.
- [30] U. S. M. and Najmusseher, “Classification algorithms used in the study of EEG based epileptic seizure detection,” pp. 1518–1521, 10 2021.
- [31] S. Parui and D. Basu, “An improved cognitive approach for automated epileptic seizure detection from multichannel EEG”, in *2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 68–71, 2021.
- [32] J. A. Morales Morante, “El proyecto xeeegal: sistema para la identificación de estados mentales a partir de señales bioeléctricas”, 2019.
- [33] C. Duque, “Metodología para la identificación biométrica a partir de señales EEG en múltiples estados emocionales”, 2023.
- [34] I. Chaos, “El perceptrón multicapa”, 1990.
- [35] M. A. Yousif and M. Ozturk, “Epileptic seizure classification for each of five classes using time domain and wavelet decomposition based features”, in *2021 Medical Technologies Congress (TIPTEKNO)*, pp. 1–4, IEEE, 2021.
- [36] OMS, “Epilepsia. world health organization (WHO)”, in *Organizacion Mundial de la Salud*, p. 1, OMS2023, 2023.
- [37] I. 2023, “Welcome to the international league against epilepsy”, in *ILAE 2023*, p. 1, 1909.
- [38] OMS, “Welcome to the international league against epilepsy. welcome to the international”, in <https://www.ilae.org/>, p. 1, 2023.

- [39] U. Fredes, “Detección de ataques epilépticos mediante análisis de señales de electroencefalograma”, 2023.
- [40] Medium, “AI mind map”, <https://medium.com/ml-ai-study-group/ai-mind-map-a70dafcf5a48>, 2017.
- [41] D. P. P. Cruz, *Inteligencia artificial con aplicaciones a la ingeniería*. 07 2010.
- [42] T. Wavelet, “Acervo para el mejoramiento del aprendizaje de los alumnos de ingeniería en IA”, https://virtual.cuautitlan.unam.mx/intar/?page_id=1108, 2018.
- [43] E. Matemáticas, “Fórmulas de estadística”, 2017.
- [44] B. Barcelona, “The Berna-Barcelona EEG database, nonlinear time series analysis”, https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2012-nonrandomness-nonlinear-dependence-and-nonstationarity-of-electroencephalographic-recordings-from-epilepsy-patients, 2013.
- [45] S. W. D. Chen and F. S. Bao, “Epileptic focus localization using EEG based on discrete Wavelet transform through full-level decomposition”, *25th International Workshop on Machine Learning for Signal Processing*, pp. 1–6, 2015.
- [46] R. B. P. Rajeev Sharma, “Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals”, *research*, vol. 17, pp. 669–691, 2015.
- [47] U. R. A. Rajeev Sharma, Ram Bilas Pachori, “An integrated index for the identification of focal electroencephalogram signals using discrete Wavelet transform and entropy measures”, *research*, vol. 17, pp. 5218–5240, 2015.
- [48] S. W. D. Chen and F. S. Bao, “Epileptic focus localization using discrete wavelet transform based on interictal intracranial EEG”, *research*, vol. 25, pp. 413–425, 2017.
- [49] U. de Bonn, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state”, https://www.upf.edu/documents/229517819/232450661/Andrzejak-P_hysicalReviewE2001.pdf/0e9a54b8-8993-b400-743e-4d64fa29fb63, 2001.
- [50] M. S. T. B. y Christian Lochmuller, “Propuesta de un espacio multidimensional para la gestión por procesos”, *research*, vol. 29, no. 1, pp. 222–230, 2013.

- [51] M. Learning, “El aprendizaje automático: objetivos y ejemplos”, [https : //www.unir.net/ingenieria/revista/machine – learning/](https://www.unir.net/ingenieria/revista/machine-learning/), 2023.
- [52] M. Learning, “¿qué es machine learning?”, [https : //www.ibm.com/es – es/topics/machine – learning](https://www.ibm.com/es-es/topics/machine-learning), 2023.
- [53] E. M. Rojas, “Machine learning: análisis de lenguajes de programación y herramientas para desarrollo”, *research*, vol. E28, pp. 586–599, 2020.
- [54] L. J. Sandoval, “Algoritmo de aprendizaje automático para análisis y predicción de datos”, *research*, vol. 11, 2018.
- [55] A. Géron, *Hands On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 09 2019.
- [56] G. Andrés Naranjo Oyola, *Implementación de aprendizaje automático en un modelo arquitectónico de IOT para la detección de convoluciones en pacientes con epilepsia*. 12 2021.
- [57] J. M. S. Muñoz, “Análisis de la calidad cartográfica mediante el estudio de la matriz de confusión”, *research*, vol. 6, 2016.
- [58] R. Borja-Robalinoz, “Estandarización de métricas de rendimiento para clasificadores Machine y Deep learning”, *research*, pp. 184–196, 2020.
- [59] D. Scientest, “Cross-validation : definición e importancia en machine learning”, [https : //datascientest.com/es/cross – validation – de finicion – e – importancia](https://datascientest.com/es/cross-validation-de-finicion-e-importancia), 2023.
- [60] M. A. Jatoi, “A survey of methods used for source localization using EEG signals”, pp. 142–52, 2013.
- [61] D. Cosandier-Rimélé, “The neuronal sources of EEG: modeling of simultaneous scalp and intracerebral recordings in epilepsy”, <https://pubmed.ncbi.nlm.nih.gov/18515148/>, 2008.
- [62] I. M. A. C. Bautista, “Diseño de sistemas embebidos en un FPGA para el análisis de señales electroencefalográficas empleadas en la detección de patologías”, 2019.
- [63] O.-O. J. Morín-Castillo MM, “Simplificación del problema inverso electroencefalográfico a una sola región homogénea con condición de Neumann nula”, 2012.
- [64] M. C. N. Bautista, “Problema inverso de identificación de fuentes en la corteza cerebral de tipo polar asociadas a focos epilépticos”, 2019.

Publicaciones

En base a los resultados obtenidos, se derivaron tres publicaciones que abordan aspectos específicos de la investigación. El primer artículo, titulado [Extracción de características a una base de datos de electroencefalogramas focales y no focales], se enfoca en el tema de extracción de características relevantes en las señales electroencefalográficas. El segundo artículo, [Clasificación de señales EEG mediante extracción de características en aprendizaje supervisado], se explora tanto las características como los clasificadores y se utiliza una familia de wavelet para la descomposición de las señales. Finalmente, el tercer artículo, [Caracterización de Señales de EEG y su Clasificación Utilizando las Familias de Wavelet y Machine Learning], examina la descomposición de señales electroencefalográficas y se realizó un comparativo entre clasificadores y familias de wavelets para ver la eficiencia del clasificador de SVM.

- Memoria de coloquio: *Extracción de características a una base de datos de electroencefalogramas focales y no focales*. En proceso de edición.
- Capítulo de libro: *Clasificación de señales EEG mediante extracción de características en aprendizaje supervisado*
Memorias de XIII Congreso Nacional de Ciencias de la Computación CONACIC 2023.
(ISBN 978-607-8857-77-7)
- Capítulo de libro: *Caracterización de Señales de EEG y su Clasificación Utilizando las Familias de Wavelet y Machine Learning*
Congreso Academia Journals Hidalgo 2023 (ISSN 1946-5351 online Vol.15, No.9) JOURNAL 2023

EXTRACCIÓN DE CARACTERÍSTICAS A UNA BASE DE DATOS DE ELECTROENCEFALOGRAMAS FOCALES Y NO FOCALES

Autores: Ing. Sergio Cadena Flores¹, Dra. María Monserrat Morín Castillo²,
Dr. José Ángel Arias Cruz³, Dr. José Jacobo Oliveros Oliveros⁴

RESUMEN

La detección temprana y precisa de patrones epilépticos en señales electroencefalográficas (EEG) desempeña un papel crítico en el manejo clínico de la epilepsia. En este artículo, presentamos un estado del arte sobre las investigaciones científicas más relevantes en la extracción de espacios de características para la detección de actividad epiléptica en EEG. Se utilizó una base de datos de registros EEG de la Universidad de Berna Barcelona, clasificando las señales en dos grupos principales: pacientes con epilepsia (focal) y pacientes sanos (no focal). El enfoque central de la investigación se basó en la extracción de características estadísticas para su futuro uso en la clasificación automática, con el propósito de mejorar la detección de actividad epiléptica. Los resultados obtenidos presentan prometedoras perspectivas para el desarrollo de herramientas de diagnóstico más precisas y efectivas en el manejo clínico de la epilepsia.

Palabras clave: EEG, focal, No focal, epilepsia

INTRODUCCIÓN

La detección temprana y precisa de patrones epilépticos en señales electroencefalográficas (EEG) es un desafío importante en el diagnóstico y manejo clínico de la epilepsia. Las técnicas de análisis de características en dominio de tiempo y frecuencia han demostrado ser una estrategia valiosa para identificar patrones específicos asociados con la actividad epiléptica. Esta revisión se enfoca en la exploración de investigaciones científicas que han aplicado diversas técnicas de análisis en el dominio del tiempo y la frecuencia para mejorar la detección de actividad epiléptica en EEG. La detección de patrones epilépticos en una señal EEG generalmente involucra una serie de pasos que permiten identificar y clasificar eventos epileptiformes. A continuación, se describen los pasos típicos en el proceso de detección de patrones epilépticos en una señal EEG:

Paso 1: Preprocesamiento de la señal EEG:

Este paso implica la preparación de la señal EEG para el análisis, lo que incluye la eliminación de artefactos y ruido. Se aplican técnicas de filtrado para eliminar interferencias externas, como ruido de línea y movimientos oculares. También se pueden utilizar técnicas de remontaje para mejorar la calidad de los datos y corregir posibles desplazamientos en los electrodos.

¹ Sergio Cadena Flores Estudiante de la Facultad de Electrónica de la Benemérita Universidad Autónoma de Puebla, México.
sergio.cadenafi@alumno.buap.mx (**autor de correspondencia**)

² La Dra. María Monserrat Morín Castillo es Profesora de la Facultad de Ciencias de la Electrónica de la Benemérita Universidad Autónoma de Puebla, México. maria.morin@correo.buap.mx

³ El Dr. José Ángel Arias Cruz es Profesor de la Facultad de Ciencias de la Electrónica de la Benemérita Universidad Autónoma de Puebla, México. angelarias01@gmail.com

⁴ El Dr. José Jacobo Oliveros Oliveros Profesor de las Facultad Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla, México. operadoradjunto@gmail.com

Clasificación de señales EEG mediante extracción de características en aprendizaje supervisado

Classification of EEG signals by feature extraction in supervised learning

Abstract

Epilepsy is a neurological disease that affects more than 50 million people worldwide, according to data from the WHO (World Health Organization) [1]. The electroencephalogram (EEG) is used in approximately 80% of cases for the diagnosis of epilepsy, being a tool used to detect abnormal electrical patterns in the brain. However, interpreting EEG signals is a complex and challenging task. With the aim of improving the usual medical procedure, in this article we present a method based on data mining and machine learning to detect epileptiform signals, using the database of electroencephalograms of the University of Bern Barcelona, which is open to the public. Our approach consists of the decomposition of EEG signals by means of the Wavelet Transform (DWT), which decomposes the original signals into frequency ranges in which a series of characteristics will be extracted. These were used as inputs in the Vector Support Machine (SVM) algorithm and compared with other supervised learning classification techniques, obtaining average results of 80% effectiveness in the classification of epileptic and non-epileptic signals. Showing remarkable performance in comparison. This method offers a promising improvement in the detection of epilepsy and can ultimately contribute significantly to more accurate diagnosis and personalized treatment for patients.

Resumen

La epilepsia es una enfermedad neurológica que afecta a más de 50 millones de personas en todo el mundo, según datos de la OMS (Organización Mundial de la Salud) [1]. El electroencefalograma (EEG) se utiliza en aproximadamente el 80% de los casos para el diagnóstico de la epilepsia, siendo una herramienta utilizada para detectar patrones eléctricos anormales en el cerebro. Sin embargo, interpretar las señales de EEG es una tarea compleja y desafiante. Con el objetivo de mejorar el procedimiento médico habitual, en este artículo presentamos un método basado en la minería de datos y el aprendizaje automático para detectar señales epileptiformes, utilizando la base de datos de electroencefalogramas de la Universidad de Berna Barcelona, que está abierta al público. Nuestro enfoque consiste en la descomposición de señales EEG por medio de la Transformada Wavelet (DWT), que descompone las señales originales en rangos de frecuencia en las que se extraerán una serie de característica. Las cuáles fueron utilizadas como entradas en el algoritmo Vector Maquina de Vectores Soporte (SVM) y comparado con otras técnicas de clasificación del aprendizaje supervisado, obteniendo resultados promedio de 80% de efectividad en la clasificación de señales epilépticas y no epilépticas. Mostrando un rendimiento destacable en su comparación. Este método ofrece una mejora prometedora en la detección de la epilepsia y, en

Fecha de recepción: julio 28, 2020 / Fecha de aceptación:



BUAP

La Benemérita Universidad Autónoma de Puebla

A través de la Facultad de Ciencias de la Computación
otorga la presente

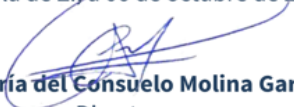
CONSTANCIA

**A: Sergio Cadena Flores, María Monserrat Morín
Castillo, José Ángel Arias Cruz, José Jacobo Oliveros
Oliveros**

Por haber impartido la ponencia simultanea "Clasificación de Señales EEG
Mediante Extracción de Características en Aprendizaje Supervisado", en el marco
del XIII Congreso Nacional de Ciencias de la Computación CONACIC 2023, de la
Facultad de Ciencias de la Computación del 02 al 06 de octubre de 2023.

"Pensar bien, para vivir mejor"

H. Puebla de Z., a 06 de octubre de 2023


M.I. María del Consuelo Molina García
Directora

Caracterización de Señales de EEG y su Clasificación Utilizando las Familias de Wavelet y Machine Learning

Ing. Sergio Cadena Flores¹, Dra. María Monserrat Morín Castillo²,
Dr. José Ángel Arias Cruz³, Dr. José Jacobo Oliveros Oliveros⁴

Resumen— Las enfermedades neurológicas, como la epilepsia, están en aumento. Esta última surge debido a la actividad eléctrica anormal del cerebro. Se presentó un algoritmo para detectar señales epileptiformes usando técnicas de minería de datos y Machine Learning, con datos públicos. Se emplearon seis familias de Wavelet para descomponer señales electroencefalográficas y extraer características. Luego, se utilizó aprendizaje supervisado para clasificarlas mediante cuatro técnicas: Máquina de soporte vectorial (SVM), vecinos más cercanos (KNN), redes neuronales (ANN) y bayesiano ingenuo (NB). El mejor resultado lo dio la combinación de redes neuronales con Wavelet Madre Coiflets, logrando un 99.2% de eficiencia. Este estudio demuestra que el aprendizaje supervisado es una opción eficiente y rápida frente a otras técnicas más robustas que requieren más tiempo de procesamiento.

Palabras clave—Machine Learning, Epilepsia, Minería de datos, Clasificadores, Wavelet.

Introducción

En este trabajo se hace uso de la base de datos de la Universidad de Bonn (University the Bonn 2023), para analizar la epilepsia siendo esta una patología neurológica que se manifiesta a través de descargas eléctricas anormales en el cerebro, afectando a aproximadamente 50 millones de individuos globalmente, según la Liga Internacional Contra la Epilepsia (ILAE) (ILAE 1909).

Esta condición se ubica como el tercer trastorno neurológico más prevalente, detrás del Alzheimer y los accidentes cerebrovasculares. Su heterogeneidad se refleja en la variedad de manifestaciones, que dependen directamente de la región cerebral comprometida y la gravedad del trastorno. A pesar de las significativas consecuencias fisiológicas y cognitivas que puede acarrear, un diagnóstico oportuno y un tratamiento adecuado permiten que la mayoría de los pacientes (más del 70%) mantengan calidad de vida y cotidianidad (Kuldeep y Malhotra 2022).

El diagnóstico convencional de la epilepsia ha dependido históricamente de la electroencefalografía (EEG), una técnica que registra la actividad eléctrica neuronal. Sin embargo, la interpretación de las señales de EEG ha requerido una revisión visual exhaustiva por neurólogos especializados, proceso que no solo es laborioso, sino que también es susceptible a errores humanos, ante esta problemática, surge una urgente demanda de técnicas más eficientes y precisas para la interpretación de dichos registros (Khubeb Siddiqui et al 2020).

Las tecnologías emergentes de aprendizaje automático, una subrama de la inteligencia artificial, están demostrando ser una solución prometedora para enfrentar esta problemática. Estos algoritmos, particularmente aquellos basados en aprendizaje supervisado, se entrenan utilizando registros EEG previamente etiquetados por expertos para identificar patrones eléctricos tanto normales como patológicos. De este modo, es posible una detección temprana y precisa de patrones epilépticos en los registros EEG, reduciendo notablemente el tiempo y la carga laboral de los especialistas, mientras se incrementa la exactitud en el diagnóstico (Milind Natu et al 2022).

Adicionalmente, en una perspectiva comparativa, aunque existen otros sistemas de monitoreo cerebral como tomografía computarizada por emisión (SPECT), tomografía por emisión de positrones (PET), resonancia magnética (RMf) y la magneto encefalografía (MEG), la electroencefalografía (EEG) sigue siendo una herramienta destacada, especialmente cuando se evalúan factores como resolución temporal y espacial, invasividad y costos asociados (Beatriz Fagundo et al 2007).

La convergencia de la neurología clínica con las técnicas avanzadas de aprendizaje automático no solo tiene el potencial de revolucionar el diagnóstico de la epilepsia, sino que también sienta las bases para futuros avances en la interpretación automatizada de datos biomédicos. Esta simbiosis entre medicina e inteligencia artificial augura una era de diagnósticos más precisos y tratamientos más efectivos, potenciando una mejor calidad de vida para los pacientes con epilepsia y otras afecciones neurológicas.

¹ Sergio Cadena Flores Estudiante de la Facultad de Electrónica de la Benemérita Universidad Autónoma de Puebla. México. sergio.cadenafi@alumno.buap.mx (autor de correspondencia)

² La Dra. María Monserrat Morín Castillo es Profesora de la Facultad de Ciencias de la Electrónica de la Benemérita Universidad Autónoma de Puebla, México. maria.morin@correo.buap.mx

³ El Dr. José Ángel Arias Cruz es Profesor de la Facultad de Ciencias de la Electrónica de la Benemérita Universidad Autónoma de Puebla, México. angelarias01@gmail.com

⁴ El Dr. José Jacobo Oliveros Oliveros Profesor de las Facultad Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla, México. operadoradjunto@gmail.com



CONGRESO INTERNACIONAL DE INVESTIGACIÓN ACADEMIA JOURNALS HIDALGO 2023

Arte, Tecnología, Equidad

CERTIFICADO

otorgado a

Lic. Sergio Cadena Flores
Dra. María Monserrat Morín Castillo
Dr. José Ángel Arias cruz
José Jacobo Oliveros Oliveros

por su artículo titulado

Caracterización de Señales de EEG y su Clasificación Utilizando las Familias de Wavelet y Machine Learning

(Artículo No. HLG231)

La ponencia correspondiente fue presentada en el Congreso Internacional de Investigación Academia Journals Hidalgo 2023, *Arte, Tecnología, Equidad*, desarrollado los días 18, 19, y 20 de octubre de 2023, co-organizado por el PYSEIP, *Productos y Servicios Estratégicos para Investigación y Posgrado*, y la Universidad Politécnica Metropolitana de Hidalgo. El artículo está incluido en las siguientes publicaciones:

- (1) en el portal de Internet AcademiaJournals.com, con ISSN 1946-5351 online, Vol. 15, No. 09, 2023 e indexación en la base de datos *Fuente Académica Plus de EBSCOHOST*, Massachusetts, Estados Unidos y
- (2) en libros ebook digitales compilados por área temática, con números ISBN online*. Se tiene acceso libre a todas las publicaciones del congreso en el portal de internet de Academia Journals.

Los organizadores del congreso reconocen la participación de los autores en el congreso, agradeciendo sus contribuciones.

DR. RAFAEL MORAS, P.E.

Director General
Academia Journals



Apéndices

Apéndice A

A.1. Distintos enfoque de modelos matemático

La cabeza humana, con su compleja estructura y funciones, ha sido objeto de numerosos estudios y representaciones. Estos modelos buscan comprender y representar tanto la geometría de la cabeza como las propiedades eléctricas inherentes a los diversos tejidos cerebrales. Estas representaciones son cruciales para diversas aplicaciones y estudios relacionados con el cerebro. Existen varios modelos que se han desarrollado a lo largo del tiempo, cada uno con sus particularidades y enfoques específicos [60], [61], [62].

A.1.1. Formas esféricas de la cabeza

La cabeza humana, esta conformada con una variedad de contornos y características. Una descripción de algunos de estos modelos:

- El modelo de esfera es una representación básica y directa que con propiedades de conductividad uniforme en su interior. En términos más técnicos, este modelo tiene una conductividad homogénea en todo su volumen interno y, en contraste, su superficie exterior no permite ningún tipo de conductividad, es decir, tiene una conductividad de valor cero, otros modelos utilizan múltiples capas conductoras que pueden ser simplificados a este modelo de esfera simple, como se menciona en la referencia [63].
- Los modelos esféricos de múltiples capas ofrecen una representación más detallada y avanzada. Estos modelos están diseñados teniendo en cuenta diferentes compartimentos o capas, cada uno con sus propias propiedades de conductividad. Estos modelos utilizan varias capas conductoras, que en modelos reducidos pueden ser como esferas concéntricas, es decir, esferas que comparten el mismo centro pero tienen diferentes radios, imitando así los distintos

niveles. Esta estructura permite una mayor precisión al considerar la variabilidad en las propiedades conductoras de diferentes regiones o secciones, pero con mucho más complejidad cuando se analizan los problemas matemáticos asociados.

A.1.2. Representaciones detalladas y precisas de la anatomía craneal

Los modelos realistas de la cabeza buscan replicar de la forma más fidedigna posible la anatomía y características específicas del cráneo y el cerebro. Para lograr este nivel de precisión, se emplea una significativa capacidad de cálculo y se hacen uso de avanzados algoritmos matemáticos. Estos modelos se basan frecuentemente en imágenes detalladas de la cabeza, muchas veces obtenidas a través de herramientas de imagenología como la resonancia magnética. Dentro de los enfoques más populares y utilizados para modelar la cabeza de manera realista encontramos:

- **Modelo de elementos finitos:** Este enfoque se caracteriza por su capacidad para describir meticulosamente todos los compartimentos del cráneo y cerebro, considerando variaciones en la geometría y conductividades de cada región.
- **Método de elementos de frontera:** Este método asume menos detalles que los basados en elementos finitos. Esto se traduce en una representación más general, pero también en una menor complejidad en su cálculo.
- **Métodos proyectivos:** Este enfoque se distingue por emplear descripciones matemáticas precisas de las superficies del cráneo y del cerebro. Además, al caracterizar el espacio en el que se encuentran los campos eléctricos o magnéticos que se están estudiando, permite una reducción significativa en la complejidad del problema matemático subyacente.
- **Método de diferencias finitas:** Aunque comparte similitudes con el método de elementos finitos, en este caso se describen detalles en todo el volumen del cráneo y cerebro sin la necesidad de segmentar el espacio en pequeños volúmenes o compartimentos.

A.1.3. Modelo de medio conductor para la cerebro-cabeza

La cabeza se representa usando un esquema basado en un medio conductor con el objetivo principal de trazar conexiones entre las fuentes bioeléctricas cerebrales y los potenciales que son detectados en la superficie del cuero cabelludo. Al hacerlo, se facilita un profundo análisis que tiene como meta principal entender y posteriormente identificar las distintas fuentes de actividad bioeléctrica en el cerebro. Este análisis es esencial para la decodificación y comprensión de la dinámica cerebral.

Para poder abordar este desafío de manera estructurada, se han establecido dos problemas clave relacionados con la electroencefalografía:

- Problema Directo Electroencefalográfico (PDE): En este escenario, se supone conocida información sobre la fuente específica de corriente neuronal. La tarea principal aquí es calcular, el potencial electrostático que se genera. Es importante destacar que dicha fuente de corriente esta concentrada en alguna de las dos partes del cerebro-. volumen o corteza cerebral.
- Problema Inverso Electroencefalográfico (PIE): Este problema es más complejo. A partir de datos capturados mediante EEG en la superficie del cuero cabelludo, se busca determinar y ubicar las fuentes bioeléctricas activas dentro del cerebro.

A.2. Representación esquemática del cerebro.

Cuando nos enfocamos solamente en las fuentes ubicadas en la superficie de las corteza cerebral esto tiene un efecto en como observamos el flujo de las corrientes eléctricas normales. En este contexto, una fuente de actividad cerebral aparece con un cambio abrupto o una discontinuidad en la dirección de estas corrientes eléctricas normales, si presentamos la densidad de estas corrientes como g entonces el problema de investigación se centra en entender este fenómeno específico.

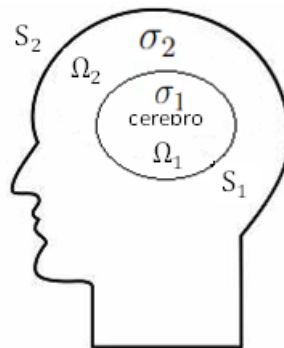


Figura A.1: Representación de la cabeza, de forma eléctrica por capas

Se definen de la siguiente manera las regiones de la cabeza:

- Ω_1 representa el cerebro
- Ω_2 representa las de más partes de la cabeza
- σ_1 y σ_2 son las conductividades de las regiones de Ω_1 y Ω_2 la cual es la conductividad del aire

- S_1 es la separación superficie entre Ω_1 y Ω_2
- S_2 es la parte donde se colocan los electrodos sobre el cuero cabelludo

A.3. Problema directo de fuente localizada

Cuando una fuente de corriente eléctrica se encuentra en el interior del cerebro, se plantea un problema de contorno en el que se analiza cómo se distribuye la corriente eléctrica en el cerebro. Esto es importante para comprender los efectos de la actividad eléctrica cerebral y tiene aplicaciones en campos como la estimulación cerebral y la neurociencia. El problema implica utilizar ecuaciones y condiciones de contorno que se mencionan abajo para estudiar la distribución de corriente eléctrica en el cerebro.

Es conocido que la corriente total en el cerebro que es J y esta se da como la suma de dos corrientes de diferente naturaleza, el cerebro, al ser un órgano eléctricamente activo, tiene corrientes que fluyen a través de él debido a la actividad neuronal. Estas corrientes son producto de las variaciones de potencial que ocurren a nivel celular. Al mencionar que J es la suma de dos corrientes de diferente naturaleza, se está haciendo referencia a que estas corrientes pueden tener diferentes orígenes o características, ya sea en términos de su magnitud, dirección, o la fuente de actividad neuronal que las produce. Identificar y entender estas corrientes separadas puede ofrecer una comprensión más detallada de la dinámica eléctrica cerebral y sus procesos asociados, como veremos en la ecuación A.1:

$$J = J^p + \sigma_i E \quad (\text{A.1})$$

donde

- J : Representa la densidad de corriente eléctrica, que es la cantidad de corriente eléctrica (carga por unidad de tiempo) que fluye a través de una unidad de área en un material conductor. Se mide en amperios por metro cuadrado
- J^p : es la corriente impresa, esta es la densidad de corriente debida a la conducción de portadores de carga, como electrones en un conductor metálico. Esta parte de la densidad de corriente está relacionada con la movilidad de los portadores de carga y la diferencia de potencial eléctrico a través del materia y es debida a la actividad biológica de las neuronas. Las neuronas, como células excitables, generan y transmiten señales eléctricas debido a los intercambios iónicos a través de sus membranas. Esta actividad eléctrica da lugar a

lo que llamamos corriente primaria o impuesta. En el contexto de medios dieléctricos, la polarización ocurre cuando un campo eléctrico externo induce un desplazamiento de cargas dentro del material, creando corrientes de polarización. De manera análoga, en el cerebro, la actividad biológica de las neuronas induce la corriente primaria, que es esencial para el funcionamiento y la comunicación neuronal. Es interesante observar cómo conceptos de la física y la biología pueden tener paralelismos en su comportamiento.

- σ : representa la conductividad eléctrica del material. La conductividad es una propiedad del material que indica cuán bien permite que la corriente eléctrica fluya a través de él. Cuanto mayor sea el valor de sigma, mejor conductor será el material. La conductividad se mide en siemens por metro.
- i : Representa la densidad de corriente debida a cualquier fuente externa de corriente eléctrica, como una fuente de alimentación o un generador de corriente. Esta parte de la densidad de corriente no está relacionada con la conductividad del material, sino que es una corriente adicional que se inyecta en el material.
- i : Es el campo eléctrico en el material. El campo eléctrico es una medida de la fuerza ejercida sobre una carga eléctrica en una ubicación dada. Se mide en voltios por metro (V/m). El campo eléctrico está relacionado con la diferencia de potencial eléctrico a través del material y la distancia entre los puntos de interés.
- La segunda parte $\sigma_i E$ se le denomina corriente óhmica, y es debida a la corriente eléctrica del medio.

Dado que las fluctuaciones temporales de la densidad de carga en las distintas áreas que constituyen la cabeza se manifiestan de la siguiente manera:

$$\frac{\partial \rho}{\partial t} = \frac{\sigma_i}{\epsilon_0} \rho_o e^{\frac{\sigma_i}{\epsilon_0}} \quad (\text{A.2})$$

La cabeza, siendo una estructura compleja, está compuesta de diversas regiones con propiedades eléctricas únicas. Estas propiedades, en particular la densidad de carga, pueden cambiar con el tiempo debido a diversos factores, como la actividad neuronal o cambios fisiológicos. Esta variación temporal es esencial para entender cómo las señales eléctricas se propagan y se interpretan en estudios relacionados con el cerebro, como la electroencefalografía.

Uno de los modelos utilizados que se ha venido describiendo considera: σ_1 la conductividad de la región Ω_1 y para la región del cráneo es aproximadamente cero.

Utilizando la ecuación de continuidad (una fórmula matemática)

$$\frac{\partial \rho}{\partial t} + \text{div} J = 0 \text{ en } \Omega \quad (\text{A.3})$$

- $\frac{\partial \rho}{\partial t}$ es la derivada parcial de la densidad de carga eléctrica ρ con respecto al tiempo.
- La variable temporal t representa cómo cambia la densidad de carga eléctrica en el tiempo.
- $\text{div}(j)$ La "divergencia" div de un vector de densidad de corriente eléctrica J en el espacio. La densidad de corriente eléctrica J describe la cantidad de corriente eléctrica (carga eléctrica por unidad de tiempo) que fluye a través de un punto en un material. La divergencia mide cómo se dispersa o se acumula la corriente eléctrica en un punto dado del espacio.
- El volumen en el espacio tridimensional se representa por ω .

Obteniéndose la siguiente ecuación:

$$\text{div} J = 0 \quad (\text{A.4})$$

Como se mencionó previamente, la densidad de corriente en el cerebro se define como $J = J^p + \sigma_1 E$ y para las otras áreas de la cabeza es $J = \sigma_2 E$. Al reemplazar la ecuación resulta en:

$$\begin{aligned} \text{div}(\sigma_i E) &= -\text{div} J^p \quad \text{en } \Omega_1 \\ \text{div} E &= 0 \quad \text{en } \Omega_2 \end{aligned} \quad (\text{A.5})$$

- Para la región Ω_1 , la ecuación $\text{div}(\sigma_i E) = -\text{div}(J^p)$ expresa la conservación de la carga eléctrica en esa región. La divergencia de $\sigma_i E$ representa cómo la densidad de corriente eléctrica cambia en función de la posición en la región Ω_1 . El término $-\text{div}(J^p)$ describe cómo las fuentes o sumideros de corriente eléctrica en el cerebro contribuyen a este cambio. En conclusión, esta ecuación describe cómo la densidad de corriente eléctrica se distribuye y cambia en el cerebro debido a la conductividad y las fuentes de corriente eléctrica.
- Para la región Ω_2 , la ecuación $\text{div} E = 0$ indica que la densidad de corriente eléctrica se conserva y no cambia en función de la posición en la región Ω_2 . Esto sugiere que en estas regiones, no hay fuentes o sumideros de corriente eléctrica, y la densidad de corriente eléctrica es constante o no se distribuye.

Sin embargo se pueden constatar en experimentos realizados que la ley de Gauss donde el rotacional se iguala a 0 por las bajas frecuencias utilizadas en las pruebas de los electroencefalogramas, donde

puede ser despreciado el campo magnético obteniendo así la relación para el campo eléctrico de la siguiente manera

$$\nabla \times E = 0 \quad (\text{A.6})$$

Así, a partir de las ecuaciones (5.6), se puede concluir que existe un potencial electrostático u de tal manera que obtenemos:

$$E = -\nabla u \quad (\text{A.7})$$

al reemplazar el valor en la ecuación , se obtienen nuevas ecuaciones que describen el comportamiento del potencial electrostático u en un contexto específico.

$$\text{div}(\sigma_i \nabla u) = -\text{div} J^P \quad \text{en} \quad \Omega_1$$

$$\Delta u = 0 \quad \text{en} \quad \Omega_2 \quad (\text{A.8})$$

Se basa en una suposición fundamental, la cual establece que la conductividad permanece constante en todas las regiones. Si consideramos esta premisa, podemos establecer las condiciones que el potencial electrostático u debe cumplir.

$$\begin{aligned} \Delta u &= 0 \quad \text{en} \quad \Omega_1 \\ \Delta u &= 0 \quad \text{en} \quad \Omega_2 \end{aligned} \quad (\text{A.9})$$

Con las condiciones de contorno:

$$\begin{aligned} u_1 &= u_2 \quad \text{en} \quad S_1 \\ \sigma_i \frac{\partial u_1}{\partial n_1} &= \sigma_i \frac{\partial u_2}{\partial n_2} + j \cdot n_1 \quad \text{en} \quad S_1 \\ \frac{\partial u_2}{\partial n_2} &= 0 \quad \text{en} \quad S_2 \end{aligned} \quad (\text{A.10})$$

La expresión $n_2 + j$ denota la contribución de una corriente localizada en la superficie de la corteza cerebral. Es importante tener en cuenta que los efectos causados por la densidad volumétrica de corriente, representados como J^P en el cerebro con una densidad J sobre la corteza cerebral, se ven reflejados tanto en la ecuación que describe el potencial en la región Ω_1 como en la con-

dición de contorno que se aplica sobre la superficie S_1 . Esto se discute en mayor detalle en la referencia[referencia].

Desde el problema (5.8) junto con las condiciones establecidas en (5.10), se derivan dos problemas fundamentales. Uno de ellos implica la consideración de una fuente concentrada en el interior de la cabeza, es decir, dentro del volumen cerebral.

$$\begin{aligned} \Delta u &= \frac{-div J^P}{\sigma_1} \quad en \quad \Omega_1 \\ \Delta u &= 0 \quad en \quad \Omega_2 \end{aligned} \quad (A.11)$$

Teniendo en considerando las restricciones de contorno:

$$\begin{aligned} u_1 &= u_2 \quad en \quad S_1 \\ \sigma_i \frac{\partial u_1}{\partial n_1} &= \sigma_i \frac{\partial u_2}{\partial n_2} \quad en \quad S_1 \\ \frac{\partial u_2}{\partial n_2} &= 0 \quad en \quad S_2 \end{aligned} \quad (A.12)$$

El segundo problema se enfoca en la concentración de la fuente en la corteza cerebral.

$$\begin{aligned} \Delta u &= 0 \quad en \quad \Omega_1 \\ \Delta u &= 0 \quad en \quad \Omega_2 \end{aligned} \quad (A.13)$$

Con las condiciones de contorno:

$$\begin{aligned} u_1 &= u_2 \quad en \quad S_1 \\ \sigma_i \frac{\partial u_1}{\partial n_1} &= \sigma_i \frac{\partial u_2}{\partial n_2} + j \cdot n_1 \quad en \quad S_1 \\ \frac{\partial u_2}{\partial n_2} &= 0 \quad en \quad S_2 \end{aligned} \quad (A.14)$$

Cada uno de estos enfoques permite abordar tanto el problema directo como el problema inversos que están relacionados.

Es fundamental destacar que los problemas inversos se enfocan en el desafío de descubrir una propiedad desconocida de un objeto o medio utilizando las respuestas observables proporcionadas por una señal de prueba. En consecuencia, los problemas inversos sientan las bases teóricas para llevar a cabo la detección y evaluación sin causar daño.

Por otro lado, los problemas directos se refieren a situaciones en las que se dispone de información sobre las causas que permiten describir los resultados generados por esas mismas causas. En otras

palabras, en los problemas inversos, se cuenta con información parcial sobre los resultados o efectos producidos en el medio, los cuales surgieron debido a ciertas causas que se intentan descubrir a través del análisis de estos resultados.

Es relevante notar que muchos problemas físicos, especialmente los relacionados con enfoques inversos, pueden representarse mediante operadores en el contexto de un modelo matemático.

El modelo que estamos utilizando para establecer cómo se relacionan las fuentes de actividad en el cerebro y las mediciones del EEG se llama cuasiestático. Esto significa que, en términos generales, asumimos que los cambios en la carga eléctrica en el cerebro a lo largo del tiempo son muy pequeños o insignificantes [22]. Debido a esta suposición, para capturar de manera precisa cómo cambian las señales del EEG con el tiempo, tratamos el problema como si estuviera en un estado estacionario en cada momento en el tiempo. En otras palabras, analizamos cómo se comporta el sistema en cada instante de tiempo, pero asumiendo que las variaciones temporales de la carga son casi nulas las presentamos de la siguiente manera:

$$\Delta u_1 = 0 \quad \text{en} \quad \Omega_1 \quad (\text{A.15})$$

$$\Delta u_1 = 0 \quad \text{en} \quad \Omega_2 \quad (\text{A.16})$$

$$u_1 = u_2 \quad \text{en} \quad S_1 \quad (\text{A.17})$$

$$\sigma_1 \frac{\partial u_1}{\partial n_1} = \sigma_2 \frac{\partial u_2}{\partial n_2} + g \quad \text{sobre} \quad S_1 \quad (\text{A.18})$$

$$\sigma_2 \frac{\partial u_2}{\partial n_2} = 0 \quad \text{sobre} \quad S_2 \quad (\text{A.19})$$

A.4. Resolución del problema directo

Pensemos en la situación en la que tenemos círculos que están perfectamente alineados en el mismo centro. Esta geometría sencilla nos proporciona información relevante, similar a lo que obtendríamos en el caso de esferas que también están concéntricas. En este contexto, hacemos la suposición que se indica a continuación y se encuentra detallada en [64]

$$g(\theta) = \sum_{n=1}^{\infty} g_k^1 \cos k\theta + g_k^2 \sin k\theta$$

la solución se busca en la forma:

$$u_1(r, \theta) = \sum_{k=1}^{\infty} a_k^1 r^k \cos k\theta + b_k^2 r^k \sin k\theta$$

$$u_1(r, \theta) = \sum_{k=1}^{\infty} (a_k^2 r^k + b_k^2 r^{-k}) \cos k\theta + \sum_{k=1}^{\infty} (c_k^2 r^k + d_k^2 r^{-k}) \sin k\theta$$

Y se emplean las condiciones (5.17)-(5.19) para calcular los coeficientes necesarios. Cuando aplicamos esta solución a la frontera S_2 , que corresponde a la evaluación de la solución en el punto $r = R_2$, obtenemos la solución al problema directo, que representamos como $V = (\theta, t)$. Esta solución está relacionada con el EEG registrado en la superficie del cuero cabelludo.

$$EEG(g) = V(\theta, t) = \sum_{k=1}^{\infty} \left(\frac{2g_k^1(t) R_1^{k+1} R_2^k}{k[(\sigma_1 - \sigma_2)R_1^{2k} + (\sigma_1 + \sigma_2)R_2^{2k}]} \right) \cos k\theta + \dots$$

$$\dots \sum_{k=1}^{\infty} \left(\frac{2g_k^1(t) R_1^{k+1} R_2^k}{k[(\sigma_1 - \sigma_2)R_1^{2k} + (\sigma_1 + \sigma_2)R_2^{2k}]} \right) \cos k\theta \quad (A.20)$$

donde R_1 representa el radio del cerebro, R_2 representa el radio de la cabeza, $g_k^1(t)$ y $g_k^2(t)$ son los coeficientes de Fourier de $g(\theta, t)$ para cada instante de tiempo, σ_1 y σ_2 son las conductividades de Ω_1 y Ω_2 , respectivamente.

A.5. Señales sintéticas de EEG en dos dimensiones

Para crear una simulación del EEG, se sugiere definir la función $g(\theta, t)$ de manera adecuada. Una de las opciones es la siguiente [65]:

$$g(\theta, t) = \left[P_i(t) \frac{e^{-\frac{|x-a|^2}{2\beta^2}}}{M_\beta} \right] \cdot n_i, i = 1, 2, \quad (A.21)$$

donde $P_i = \left(e^{-\frac{|t-t_e|^2}{2\gamma^2}} + c_1 \cos(\pi t f) + c_2 \sin(\pi t f) \right)$, $i = 1, 2$, c_1 y $c_2 \geq 0$ se puede explicar de la siguiente forma: La amplitud de la señal se refiere a la magnitud o intensidad de la señal.

- f es la "frecuencia" de la señal, que varía dependiendo de la ubicación de cada electrodo en el cuero cabelludo. Diferentes ubicaciones pueden tener diferentes frecuencias.
- t representa el tiempo en general y t_e se refiere al instante de tiempo específico en el que se visualiza un foco epiléptico en la señal.

- γ es el ancho de una función campana temporal, lo que significa que describe cuánto tiempo dura el evento.
- a es el "centro" de una función campana espacial y señala la posición del foco epiléptico en el espacio.
- β representa el ancho de una función campana espacial, además se va a denotar .

$P_i = \left(e^{-\frac{|t-t_e|^2}{2\gamma^2}} + c_1 \cos(\pi t f) + c_2 \sin(\pi t f) \right)$, $i = 1, 2$, cuando se tiene uno o más de un foco epiléptico, en caso de no tener foco epiléptico $P_i = c_1 \cos(\pi t f) + c_2 \sin(\pi t f)$ Al obtener los coeficientes de Fourier de la fuente (5.21), se obtiene la solución del problema directo:

$$\begin{aligned} EEG(g) = V(\theta, t) = & \sum_{k=1}^{\infty} \left(\frac{2g_k^1(t) R_1^{k+1} R_2^k}{k[(\sigma_1 - \sigma_2) R_1^{2k} + (\sigma_1 + \sigma_2) R_2^{2k}]} \right) \cos k\theta + \dots \\ & \dots \sum_{k=1}^{\infty} \left(\frac{2g_k^1(t) R_1^{k+1} R_2^k}{k[(\sigma_1 - \sigma_2) R_1^{2k} + (\sigma_1 + \sigma_2) R_2^{2k}]} \right) \cos k\theta \end{aligned} \quad (A.22)$$

El EEG sintético es generado por la función $g(t)$. En este contexto, R_1 se refiere al tamaño del cerebro, R_2 es el tamaño de la cabeza, σ_1 y σ_2 representan las conductividades promedio del cerebro y las capas externas de la cabeza. Los valores de θ_1 se determinan de la siguiente manera (ver Figura 5.2): dividimos la semicircunferencia en 5 intervalos equidistantes para obtener los puntos que representan la ubicación de cada lóbulo de la cabeza.

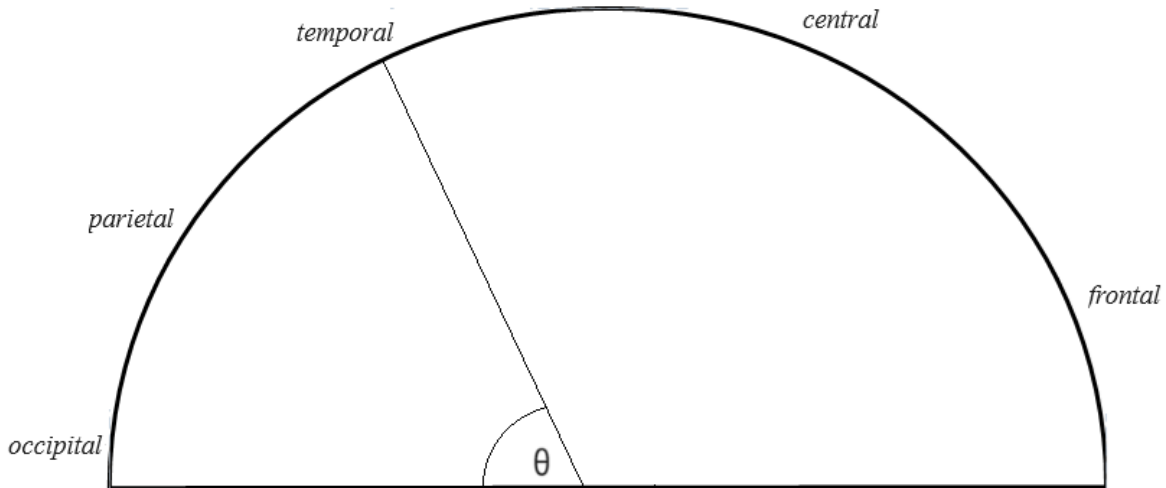


Figura A.2: Representación de los lóbulos del cerebro

Apéndice B

B.1. Programa principal de clasificación de la Universidad de Bonn

```
loadsets.mat;  
  
Fs = 173.61;  
  
samples = 4097;  
  
time = samples/Fs;  
  
channels = 100;  
  
k = 1;  
  
for j = 1 : channels  
  
  data_a = a(:, j);  
  data_b = b(:, j);  
  data_c = c(:, j);  
  data_d = d(:, j);  
  data_e = e(:, j);  
  
  x = 1 : ceil(Fs);  
  
  i = 0;
```

B.1. PROGRAMA PRINCIPAL DE CLASIFICACIÓN DE LA UNIVERSIDAD DE BONN100

Utilizado para acceder a datos de fila

```
for  $i = 1 : \text{floor}(\text{time})$ 
```

```
 $y\_a = \text{data\_a}(x);$ 
```

```
 $y\_b = \text{data\_b}(x);$ 
```

```
 $y\_c = \text{data\_c}(x);$ 
```

```
 $y\_d = \text{data\_d}(x);$ 
```

```
 $y\_e = \text{data\_e}(x);$ 
```

Son usadas características estadísticas- f1:mean, f2:median, f3:mode, f4:desviacion estandar, f5:minimum, f6:maximum, f7:skewness, f8:kurtosis, f9:mobility, f10:complexity, f11:hurst exponent

```
 $s\_a(k, :) = \text{stat\_feat}(y\_a);$ 
```

```
 $s\_b(k, :) = \text{stat\_feat}(y\_b);$ 
```

```
 $s\_c(k, :) = \text{stat\_feat}(y\_c);$ 
```

```
 $s\_d(k, :) = \text{stat\_feat}(y\_d);$ 
```

```
 $s\_e(k, :) = \text{stat\_feat}(y\_e);$ 
```

Características en 4 niveles db4 wavelet coeficientes(cd1,cd2,cd3,cd4,a4) son Utilizados f12-f16: max, f17-f21: min, f22-f26: mean, f27-f31:desviacion estandar

```
 $w\_a(k, :) = \text{wav\_feat}(y\_a);$ 
```

```
 $w\_b(k, :) = \text{wav\_feat}(y\_b);$ 
```

```
 $w\_c(k, :) = \text{wav\_feat}(y\_c);$ 
```

```
 $w\_d(k, :) = \text{wav\_feat}(y\_d);$ 
```

```
 $w\_e(k, :) = \text{wav\_feat}(y\_e);$ 
```

```
 $x = x + \text{ceil}(Fs);$ 
```

```
 $k = k + 1;$ 
```

End

End

```
 $\text{statf} = [s\_a; s\_e];$ 
```

```

wavf = [w_a; w_e];
label = [zeros(2300, 1); ones(2300, 1)];
data = [statf wavf label];
[trainedClassifier, validationAccuracy] = trainClassifier(data);
validationAccuracy

```

B.2. Programa secundario de características en función del tiempo

```

funcion statf = stat_feat(y)

```

Calcular las características estáticas para cada época o segmento

```

Mean = mean(y);

```

```

Median = median(y);

```

```

Mode = mode(y);

```

```

StdDev = std(y);

```

```

Min = min(y);

```

```

Max = max(y);

```

```

Skew = skewness(y);

```

```

Kurt = kurtosis(y);

```

```

[Mobility, Complexity] = (Hjorth_Parameters(y);

```

Complexity

```

statf = [Mean, Median, Mode, StdDev, Min, Max, Skew, Kurt, Mobility, Complexity];

```

```

,Complexity,Hurst

```

End

```

function[mobility, complexity] = Hjorth_parameters(y)

```

Computar los parámetros Hjorth mobility y complexity, calcula los parámetros de movilidad y complejidad de Hjorth.

```

n = length(y);
dy = diff([0; y]);
ddy = diff([0; dy]);
mx2 = mean(y.^ 2);
mdx2 = mean(dy.^ 2);
mddx2 = mean(ddy.^ 2);
mob = mdx2/mx2;
complexity = sqrt(mddx2/mdx2 - mob);
mobility = sqrt(mob);
End

```

B.3. Programa secundario de características en función de la frecuencia.

```

funcion wavef = wav_feat(y)
[c, l] = wavedec(y, 4, 'db4');
approx = appcoef(c, l, 'db4');
[cd1, cd2, cd3, cd4] = detcoef(c, l, [1234]);
[max_c, min_c, mean_c, std_c] = deal(zeros(1, 5));
[max_c(:)] = [max(approx(:)), max(cd1(:)), max(cd2(:)), max(cd3(:)), max(cd4(:))];
[min_c(:)] = [min(approx(:)), min(cd1(:)), min(cd2(:)), min(cd3(:)), min(cd4(:))];
[mean_c(:)] = [mean(approx(:)), mean(cd1(:)), mean(cd2(:)), mean(cd3(:)), mean(cd4(:))];
[std_c(:)] = [std(approx(:)), std(cd1(:)), std(cd2(:)), std(cd3(:)), std(cd4(:))];

```

```
wavef = [max_c,min_c,mean_c,std_c];
```

```
End
```

B.4. Programa principal para la creación de señales sintéticas

```
clear,clc,closeall
```

Parámetros.

$N = 4097$; Número de puntos en la señal.

$fs = 173.61$; Frecuencia de muestreo: cuántas veces se toma una muestra por segundo.

$t = (0 : N - 1)/fs$; Vector de tiempo, esencialmente una serie de puntos de tiempo. en los que se toma una muestra.

$num_signals = 100$; Número de señales que queremos generar

Inicio de foco

Matrices para almacenar las señales

$eeg_normal_db = zeros(N, num_signals)$; Almacén para señales EEG normales

$eeg_epileptic_db = zeros(N, num_signals)$; Almacén para señales EEG epilépticas

$F_delta = rand(num_signals, 1) * 2 * pi$;

$F_theta = rand(num_signals, 1) * 2 * pi$;

$F_alpha = rand(num_signals, 1) * 2 * pi$;

$F_beta = rand(num_signals, 1) * 2 * pi$;

$F_gamma = rand(num_signals, 1) * 2 * pi$;

Comenzamos a generar señales

$inicio_foco_c = linspace(1, 6, num_signals)$;

$sigma_c = linspace(.1, .8, num_signals)$;

$amp_foco_c = linspace(100, 400, num_signals)$;

```
for i = 1 : num_signals
```

Crear componentes de frecuencia basados en ritmos cerebrales:

Delta: oscilaciones lentas (como las olas del océano) que se encuentran comúnmente durante el sueño profundo.

$$\text{delta1} = 100 * \sin(2 * \pi * 2 * t); 2Hz$$

Theta: oscilaciones que a menudo se asocian con estados somnolientos o de ensueño.

$$\text{theta1} = 80 * \sin(2 * \pi * 6 * t); 6Hz$$

Alfa: ritmo comúnmente observado cuando una persona está relajada con los ojos cerrados.

$$\text{alpha1} = 60 * \sin(2 * \pi * 10 * t); 10Hz$$

Beta: oscilaciones más rápidas, asociadas con la actividad mental activa o la ansiedad.

$$\text{beta1} = 20 * \sin(2 * \pi * 20 * t); 20Hz$$

Gamma: oscilaciones muy rápidas asociadas con el procesamiento de la información.

$$\text{gamma1} = 100 * \sin(2 * \pi * 40 * t); 40Hz$$

Delta: oscilaciones lentas que se encuentran comúnmente durante el sueño profundo.

$$\text{delta} = 100 * \sin(2 * \pi * 2 * t + F_delta(i)); 2Hz$$

Theta: oscilaciones que a menudo se asocian con estados somnolientos o de ensueño.

$$\text{theta} = 80 * \sin(2 * \pi * 6 * t + F_theta(i)); 6Hz$$

Alfa: ritmo comúnmente observado cuando una persona está relajada con los ojos cerrados.

$$\text{alpha} = 60 * \sin(2 * \pi * 10 * t + F_alpha(i)); 10Hz$$

Beta: oscilaciones más rápidas, asociadas con la actividad mental activa o la ansiedad.

$$\text{beta} = 20 * \sin(2 * \pi * 20 * t + F_beta(i)); 20Hz$$

Gamma: oscilaciones muy rápidas asociadas con el procesamiento de la información.

$$\text{gamma} = 100 * \sin(2 * \pi * 40 * t + F_gamma(i)); 40Hz$$

Para el EEG normal, combinamos estos ritmos.

$$\text{eeg_normal} = \text{delta1} + \text{theta1} + \text{alpha1} + \text{beta1};$$

$SNR_dB = 10$; Relación señal ruido en decibelios

Añadir ruido: en la vida real, las señales no son perfectas y tienen interferencias o ruido".

$$\text{eeg_normal_db}(:, i) = \text{awgn}(\text{eeg_normal}, SNR_dB, 'measured');$$

Para EEG epiléptico: Basado en la investigación, las señales epilépticas a menudo muestran una disminución en el ritmo alfa y un aumento en las oscilaciones rápidas (como gamma). Simulamos esto ajustando los pesos de los ritmos. Simular un foco epiléptico usando una función gaussiana

```
inicio_foco = inicio_foco_c(i);
```

Tiempo de inicio del foco (s)

```
sigma = sigma_c(i);
```

Desviación estándar de la gaussiana

```
inicio_foco_2 = inicio_foco + 2 * sigma;
```

Tiempo de inicio del foco (s)

```
amp_foco = amp_foco_c(i);
```

Amplitud del foco

```
amp_foco_2 = 1 * amp_foco;
```

Suma

```
eeg_epileptic = delta + theta + alpha + beta + gamma;
```

```
foco = amp_foco * exp(-(abs(t - inicio_foco)).2/(2 * sigma).2) - ...
```

```
amp_foco_2 * exp(-(abs(t - inicio_foco_2)).2/(2 * sigma).2);
```

```
eeg_epileptic = delta + theta + alpha + beta + gamma + foco;
```

Agregar ruido al EEG

```
eeg_epileptic_db(:, i) = awgn(eeg_epileptic, SNR_db, 'measured');
```

End

Gráficas de las señales

```
figure, plot(t, eeg_epileptic_db(:, num_signals))
```

```
hold on, plot(t, eeg_normal_db(:, num_signals)), legend('Epileptic', 'Normal', 'FontSize', 10,  
'FontWeight', 'bold') y = eeg_epileptic_db(:, num_signals);
```

```
ylim([-500 500])
```

```
xlabel('Time(s)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje X

```
ylabel(Voltage $\mu$ V,'FontSize',12,'FontWeight','bold');
```

Etiqueta del eje Y con microvolts

```
title('EpilepticEEGsignal','FontSize',12,'FontWeight','bold');
```

Título del gráfico

```
figure
```

```
subplot(2,1,1);
```

```
plot(t,eeg_epileptic_db(:,20))
```

```
ylim([-500 500]) xlabel('Time(s)','FontSize',12,'FontWeight','bold');
```

Etiqueta del eje X

```
ylabel('Voltage $\mu$ V','FontSize',12,'FontWeight','bold');
```

Etiqueta del eje Y con microvolts

```
title('EpilepticEEGsignal','FontSize',12,'FontWeight','bold');
```

Título del gráfico

```
subplot(2,1,2);
```

```
plot(t,eeg_normal_db(:,20)),
```

```
ylim([-500 500])
```

```
xlabel('Time(s)','FontSize',12,'FontWeight','bold');
```

Etiqueta del eje X

```
ylabel('Voltage $\mu$ V','FontSize',12,'FontWeight','bold');
```

Etiqueta del eje Y con microvolts

```
title('NormalEEGsignal','FontSize',12,'FontWeight','bold');
```

Título del gráfico

```
figure;
```

Primera gráfica: Gráfico lineal

```
subplot(4,1,1);
```

```
plot(t,eeg_epileptic_db(:,10))
```

```
ylim([-500 500])
```

```
xlabel('Time(s)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje X

```
ylabel('Voltage( $\mu$ V)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje Y con microvolts

```
title('EpilepticEEGsignal', 'FontSize', 12, 'FontWeight', 'bold');
```

Título del gráfico

```
subplot(4, 1, 2);
```

```
plot(t, eeg_epileptic_db(:, 20))
```

```
ylim([-500 500])
```

```
xlabel('Time(s)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje X

```
ylabel('Voltage( $\mu$ V)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje Y con microvolts

```
title('EpilepticEEGsignal', 'FontSize', 12, 'FontWeight', 'bold');
```

Título del gráfico

```
subplot(4, 1, 3);
```

```
plot(t, eeg_epileptic_db(:, 50))
```

```
ylim([-500 500])
```

```
xlabel('Time(s)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje X

```
ylabel('Voltage( $\mu$ V)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje Y con microvolts

```
title('EpilepticEEGsignal', 'FontSize', 12, 'FontWeight', 'bold');
```

Título del gráfico

```
subplot(4, 1, 4);
```

```
plot(t, eeg_epileptic_db(:, 100))
```

```
ylim([-500, 500])
```

```
xlabel('Time(s)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje X

```
ylabel('Voltage( $\mu$ V)', 'FontSize', 12, 'FontWeight', 'bold');
```

Etiqueta del eje Y con microvolts

```
title('Epileptic EEG signal', 'FontSize', 12, 'FontWeight', 'bold');
```

Proceso de clasificación

Numero de señales N=100 todas

```
Ns = 100; fork = 1 : Ns
```

Temporal

```
N = eeg_normal_db(:, k);
```

```
F = eeg_epileptic_db(:, k);
```

```
s_Nx(k, :) = stat_feat(N);
```

```
s_Fx(k, :) = stat_feat(F);
```

DWT

```
w_Nx(k, :) = wav_feat(N);
```

```
w_Fx(k, :) = wav_feat(F);
```

End

```
statf = [s_Nx; s_Fx];
```

```
wavf = [w_Nx; w_Fx];
```

```
label = [zeros(Ns, 1); ones(Ns, 1)];
```

```
data = [statf wavf label];
```

```
[trainedClassifier, validationAccuracy] = trainClassifier(data)
```

```
validationAccuracy
```