

UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ingeniería



**Aplicación sistemática de algoritmos de aprendizaje  
automático para el estudio de la epilepsia y la detección de  
segmentos de interés en señales bioeléctricas**

Protocolo de trabajo de graduación presentado por Cristhofer Isaac  
Patzán Martínez, estudiante de Ingeniería Mecatrónica

Guatemala,

2023

## Resumen

El trabajo de graduación propuesto tiene como objetivo aplicar algoritmos de aprendizaje automático no supervisado a señales bioeléctricas, obteniendo los segmentos de actividad ictal de una grabación que puede durar horas, con el fin de facilitar a los doctores el análisis de señales bioeléctricas de un paciente que sufre episodios de epilepsia. Para cumplir con el objetivo se utilizarán datos EEG y EMG recopilados por UVG y HUMANA. Posteriormente se procesará la data utilizando la *Epileptic EEG Analysis Toolbox* desarrollada en fases anteriores dentro de la Universidad del Valle de Guatemala.

## Antecedentes

Las señales bioeléctricas y las señales de la epilepsia son temas de gran interés en la actualidad debido a su relevancia en la medicina y la neurociencia. La epilepsia es una enfermedad neurológica común que afecta a millones de personas en todo el mundo. La prevalencia de la enfermedad varía según el país y la región del mundo, siendo más común en niños y personas mayores, además, se estima que el 80 % de las personas con epilepsia vive en países de bajos y medianos ingresos, donde el acceso a los tratamientos puede ser limitado o inaccesible [1].

En Guatemala, se encuentra el Centro de Epilepsia y Neurocirugía Funcional HUMANA, una organización formada por profesionales en neurociencias que trabaja con pacientes que padecen problemas neurológicos de difícil control, incluyendo la epilepsia. El centro posee el único laboratorio de vídeo electroencefalograma en Guatemala y lleva a cabo anotaciones de señales EEG manualmente, lo que implica una gran cantidad de horas de trabajo realizadas por profesionales especializados del hospital. El tiempo de operación para resaltar los segmentos de interés y finalmente dar el diagnóstico depende de la duración del registro y puede durar desde cuatro horas hasta algunos días [2]. Según una entrevista realizada en 2020 al director de HUMANA, Dr. Juan Carlos Lara, las señales de donde se extraen los datos son obtenidas de personas sin y con padecimiento de epilepsias, lo que permite a los especialistas hacer las anotaciones competentes dentro de los registros [3].

En la Universidad del Valle de Guatemala (UVG) se ha estado desarrollando una herramienta de aprendizaje automático para la detección de crisis epilépticas en señales EEG. En el año 2020, María Jesús Angulo presentó un trabajo de graduación en el que se demostró que es posible detectar estas crisis mediante aprendizaje automático al caracterizar y dividir correctamente los segmentos de las señales [4].

Por su parte, en el año 2021, David Alejandro Vela presentó una segunda iteración de la herramienta en su trabajo de graduación. En este caso, se realizaron ajustes a la interfaz anterior, empleando algoritmos de aprendizaje automático supervisado para clasificar la señal EEG en una de cuatro clases [5]. El clasificador con mejor desempeño en tiempo continuo obtuvo un promedio de exactitud del 96.7 %, utilizando máquina de vectores de soporte con características en tiempo continuo y kernel gaussiano. Además, se agregó una nueva sección a la herramienta que se encarga de la generación de anotaciones dentro de un apartado con varias opciones de personalización para la visualización. Aunque se trabajó principalmente con señales EEG en dominio del tiempo en ambas fases, se reconoce la

Características	Tiempo Continuo			Wavelet		
Modelo	RNA	SVM		RNA	SVM	
Kernel	-	Gaussiano	Lineal	-	Gaussiano	Lineal
2 Clases	100.00%	99.80%	100.00%	97.70%	98.70%	97.90%
3 Clases	97.90%	98.90%	97.20%	98.20%	98.30%	97.20%
4 Clases	88.00%	91.30%	88.30%	81.20%	83.30%	77.10%
Promedio	95.30%	96.70%	95.20%	92.40%	93.40%	90.70%
Desv. Estándar	5.23%	3.81%	4.99%	7.90%	7.17%	9.64%

Figura 1: Resumen de los resultados de los clasificadores generados: red neuronal (RNA) y máquina de vectores de soporte (SVM) [5].

Extracción de características en dominio de la frecuencia				
Característica	Ubonn - Sano/Ictal		Kaggle - Interictal/Preictal	
	Porcentaje de rendimiento RNA	Tiempo	Porcentaje de rendimiento RNA	Tiempo
razón 1	91.00%	3.10 min	98.60%	4.52 min
razón 2	91.10%	3.08 min	84.20%	4.46 min
razón 3	77.00%	4.60 min	98.50%	6.76 min
razón 4	99.70%	2.88 min	96.10%	4.50 min
razón 5	90.90%	4.60 min	98.30%	6.16 min
std	99.60%	11.00 s	67.40%	9.73 s

Figura 2: Resumen del rendimiento de la RNA para dos clasificadores binarios utilizando características individuales en dominio de la frecuencia [3].

oportunidad de analizar las señales en otros dominios y de analizar otro tipo de señales bioeléctricas. Una de las limitaciones del trabajo fue la predominancia del uso de técnicas de aprendizaje supervisado, lo que abre espacio para profundizar en el análisis de los datos con aprendizaje no supervisado. Los resultados de esta segunda fase se observan en la Figura 1 [5]. Además en la Figura 3 se puede ver la ventana inicial de la app y en la Figura 4 se puede observar la forma en la que se clasifica un segmento de señal por color según su estado ictal.

En el año 2022 Camila Lemus [3] concluyó que las relaciones entre bandas de frecuencia son funcionales para la clasificación binaria, logrando un porcentaje de rendimiento superior al 99 % utilizando redes neuronales con dos o más características para las clases Ictal/Sano y superior a un 98.80 % para las clases Interictal/Preictal, dichos resultados se pueden observar en la Figura 2. Además, se encontró que es conveniente generar el vector de características combinando la razón 1 ( $\theta/\alpha$ ) y la razón 2 ( $\beta/\alpha$ ), ya que los clasificadores tienen un rendimiento igual o mayor a 98.90 % en un menor tiempo. Sin embargo, se observó que la extracción de características en dominio de la frecuencia no es muy eficiente, tardando en promedio aproximadamente 3 minutos por registro individual para procesar 409,700 muestras y 4.33 minutos para procesar 3 millones de muestras [3]. El algoritmo utilizado para el análisis y el reconocimiento de patrones en señales de ECG fue *K-means Clustering*. Los resultados mostraron que el algoritmo era capaz de identificar agrupaciones de señales de ECG asociadas a oscilaciones postictales de la frecuencia cardiaca. Esto sugiere que la agrupación de *K-means Clustering* podría utilizarse para desarrollar un nuevo método de diagnóstico y seguimiento de la epilepsia.



Figura 3: Ventana de la aplicación para iniciar sesión [5].

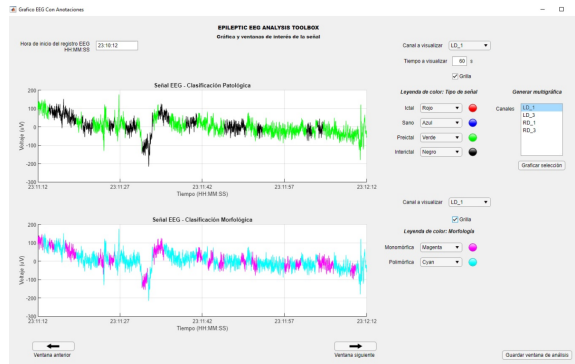


Figura 4: Ventana con la gráfica de una ventana del registro EEG del canal seleccionado [5].

## Justificación

En Guatemala, hay un acceso limitado a la atención médica, lo que se agrava en las zonas rurales donde la incidencia de la epilepsia es mayor. A pesar de la alta prevalencia de la epilepsia en el país, hay una escasa cantidad de investigaciones sobre esta afección, lo que hace necesario llenar este vacío en el conocimiento y obtener información importante sobre la epilepsia en esta población.

Para la continuación de esta línea de investigación se percibe la necesidad de ampliar la herramienta incorporando señales bioeléctricas de interés, como las de electromiograma

(EMG). Asimismo, se ha recomendado continuar experimentando el rendimiento de los clasificadores con diferentes técnicas de aprendizaje automático no supervisado para señales EEG, ECG y otras señales bioeléctricas relacionadas con la epilepsia. Para la investigación que se realizará en el presente proyecto, se aplicará la mejora del proceso de selección de características de las señales mediante asesoría médica y la validación constante con especialistas en el campo para mejorar la predicción de los clasificadores [3].

El acceso a un mayor número de señales proporciona una muestra más representativa de la actividad bioeléctrica cerebral y muscular de los pacientes epilépticos. Esto permite una comprensión más completa y precisa de los patrones y características de las señales relacionadas con la epilepsia, lo que conduce a una mejor comprensión de la enfermedad y sus diversas manifestaciones.

La aplicación de algoritmos de agrupación a un mayor número de señales de EEG y EMG puede ayudar a identificar subgrupos o patrones ocultos dentro de la población de pacientes con epilepsia. Estos subgrupos pueden tener relevancia clínica, como diferentes tipos de epilepsia, respuestas variadas al tratamiento o diferencias en la gravedad de los síntomas. Al identificar estos subgrupos, los profesionales sanitarios pueden adaptar el tratamiento y los cuidados de forma más precisa a cada paciente, personalizando así su enfoque médico.

En este trabajo se busca, además de, aplicar algoritmos de aprendizaje automático a una mayor cantidad de señales bioeléctricas también, implementar un algoritmo que pueda indicar en qué tiempo de la grabación se encuentran señales de interés para su estudio, con el fin de aprovechar de forma óptima el tiempo disponible. Actualmente, las grabaciones pueden durar hasta 24 horas y los médicos deben analizar toda la grabación para detectar la actividad bioeléctrica de interés. Este proceso resulta muy ineficiente.

## Objetivos

### Objetivo General

Aplicar los algoritmos de aprendizaje automático desarrollados en fases anteriores a una mayor cantidad de señales bioeléctricas, además de, mejorar el proceso de detección de segmentos de interés en las señales, para el estudio de la epilepsia.

### Objetivos Específicos

- Aplicar algoritmos de extracción de características y aprendizaje automático desarrollados en fases anteriores a las señales obtenidas.
- Mejorar el proceso de detección de segmentos de interés en las señales y la generación automática de anotaciones relevantes, según los parámetros de HUMANA.
- Realizar análisis estadísticos para evaluar el rendimiento de los algoritmos e identificar posibles mejoras a los mismos.
- Actualizar la herramienta de software para el estudio de la epilepsia desarrollada en fases anteriores, incorporando las mejoras a los algoritmos de clasificación y detección

de segmentos de interés de las señales bioeléctricas.

## Marco teórico

### Epilepsia

La epilepsia es uno de los primeros trastornos documentados en la historia de la neurología. Su aparición se remonta a más de 3.000 años atrás, siendo mencionada por primera vez en la antigua Babilonia. No obstante, fue en el año 400 a.C. cuando Hipócrates señaló que la epilepsia era un trastorno cerebral. La palabra “epilepsia” proviene del griego y significa “ataque”. A lo largo de la historia, el peculiar comportamiento provocado por ciertos tipos de crisis convulsivas ha dado lugar a numerosas supersticiones y prejuicios [6].

Una crisis epiléptica consiste en una alteración brusca y transitoria ocasionada por una actividad anormal de las neuronas, manifestándose a través de sensaciones, emociones y comportamientos extraños, espasmos musculares y pérdida de conciencia. La epilepsia implica una predisposición a experimentar crisis epilépticas recurrentes, siendo diagnosticada cuando una persona ha experimentado dos o más de estas crisis. Las crisis epilépticas se dividen en dos tipos principales: crisis generalizadas y crisis parciales o focales. En las crisis generalizadas, la descarga epiléptica afecta simultáneamente a toda la superficie del cerebro, mientras que en las crisis parciales o focales, la descarga epiléptica se origina en una parte específica del cerebro[7].

### Tipos de epilepsia

Existen varios tipos de epilepsia, que se clasifican en función del tipo de crisis que experimenta una persona. Algunos tipos comunes de epilepsia son

- Epilepsia focal: Las crisis focales, también llamadas crisis parciales, se originan en una parte específica del cerebro. Estas crisis pueden hacer que la persona experimente sensaciones o movimientos en un lado del cuerpo.
- Epilepsia generalizada: Las crisis generalizadas afectan a ambos lados del cerebro. Estas crisis pueden hacer que la persona pierda el conocimiento, caiga al suelo y experimente espasmos musculares o convulsiones.
- Epilepsia de ausencia: Las crisis de ausencia, también llamadas crisis de pequeño mal, hacen que la persona se quede con la mirada perdida y pierda temporalmente la conciencia de lo que le rodea.
- Epilepsia tónico-clónica: Las crisis tónico-clónicas, también llamadas crisis de gran mal, son el tipo de crisis más conocido. Implican pérdida de conocimiento, caída al suelo y movimientos espasmódicos repetitivos del cuerpo.
- Epilepsia mioclónica juvenil: Este tipo de epilepsia suele comenzar en la adolescencia. Se caracteriza por breves sacudidas o espasmos musculares, especialmente por la mañana después de despertarse.

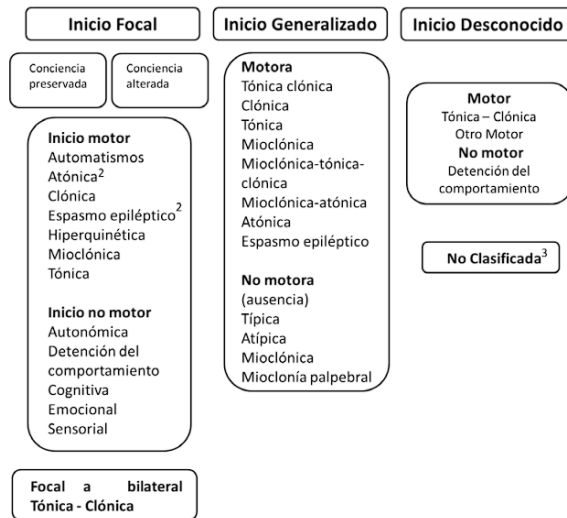


Figura 5: Clasificación Operacional Extendida de los tipos de crisis[8].

La Liga Internacional contra la Epilepsia (ILAE) ha introducido una categorización operativa actualizada de los tipos de crisis, mostradas en la Figura 5. La clasificación general incluye las crisis generalizadas, las crisis focales, también denominadas crisis de inicio parcial, y las crisis de origen desconocido. Estos tipos de crisis se clasifican a su vez en subcategorías basadas en síntomas motores y no motores. Además, en el caso de las crisis focales, existe otra distinción entre las que cursan con consciencia normal y las que presentan un nivel de consciencia alterado[8].

## Señales bioeléctricas

Las señales bioeléctricas son señales eléctricas producidas en los seres vivos que pueden medirse y controlarse continuamente. Estas señales suelen ser generadas por sistemas especializados de tejidos, órganos o células, como el sistema nervioso. Las señales bioeléctricas tienen características únicas que las hacen útiles en una amplia gama de aplicaciones, como la cardiología, la neurología y la ingeniería biomédica.

## Señales electroencefalográficas

La electroencefalografía (EEG) es una prueba de diagnóstico neurofisiológico que consiste en la medición de la actividad eléctrica cerebral en estado de reposo, vigilia o sueño, y durante diversas activaciones. En general, la electroencefalografía se utiliza para el análisis de la actividad eléctrica cerebral en humanos con el fin de obtener información para una inspección exhaustiva de la funcionalidad cerebral, ayudando a la detección y prevención de enfermedades y trastornos. Estas ondas cerebrales se conocen como señales electroencefalográficas (señales EEG), que proporcionan información indirecta relacionada con las funciones cerebrales, incluidas, entre otras, las tareas mentales, las acciones motoras y las expresiones faciales[9].

## Ritmos y formas de onda del EEG

A continuación se resumen brevemente las características de los ritmos y formas de onda más frecuentes, que también se puede observar en la Figura 7. Las señales EEG registradas tienen, en amplitudes que oscilan entre unos pocos micro-voltios y aproximadamente 100  $\mu\text{V}$  y un contenido frecuencial que oscila entre 0,5 y 30-40 Hz. Los ritmos electroencefálicos, también denominados ritmos de fondo, se clasifican convencionalmente en cinco bandas de frecuencia diferentes. La interpretación de estas bandas en términos de “normal” o “anormal” es relativa y depende de la edad y el estado mental del sujeto. Por ejemplo, el EEG de un recién nacido es drásticamente diferente del de un adulto y tiene, en general, un contenido de frecuencias considerablemente más alto.

- Ritmo delta,  $<4$  Hz. El ritmo delta suele aparecer durante el sueño profundo y tiene una gran amplitud. El adulto despierto y normal, pero es indicativo, por ejemplo, de daño cerebral o enfermedad cerebral (encefalopatía)[10].
- Ritmo theta, 4-7 Hz. El ritmo theta se produce durante la somnolencia y en ciertas fases del sueño[10].
- Ritmo alfa, 8-13 Hz. Este ritmo es más prominente en sujetos normales que están relajados y despiertos con los ojos cerrados; la actividad se suprime cuando los ojos están abiertos. Cuando los ojos están abiertos. La amplitud del ritmo alfa es mayor en las regiones occipitales[10].
- Ritmo beta, 14-30 Hz. Es un ritmo rápido de baja amplitud, asociado a la activación de la corteza cerebral. Asociado a un córtex activado y que puede observarse, por ejemplo, durante ciertas fases del sueño. El ritmo beta se observa principalmente en las áreas frontal y central del cuero cabelludo[10].
- Ritmo gamma,  $>30$  Hz. El ritmo gamma está relacionado con un estado de procesamiento activo de la información en el córtex. Utilizando un electrodo ubicado sobre el área sensoriomotora y conectado a una técnica de registro de alta sensibilidad, se puede observar el ritmo gamma[10].

La mayoría de estos ritmos pueden durar varios minutos, mientras que otros sólo unos segundos. Es importante darse cuenta de que un ritmo no está presente en todo momento, una señal irregular de aspecto “aritmico”.

## Picos y ondas agudas

Los picos y las ondas agudas, también conocidas como SSW, son patrones distintivos en las formas de onda de EEG que tienen un patrón temporal esporádico e impredecible que indica un comportamiento neural anormal. Estos patrones se observan a menudo en pacientes que padecen epilepsia y se denominan interictales, ya que ocurren entre ataques epilépticos.

La definición de SSW es un poco ambigua, pero generalmente se caracteriza por un fuerte ascenso inicial que lo diferencia de un pico, que dura entre 20 y 70 ms. Por otro lado, SSW



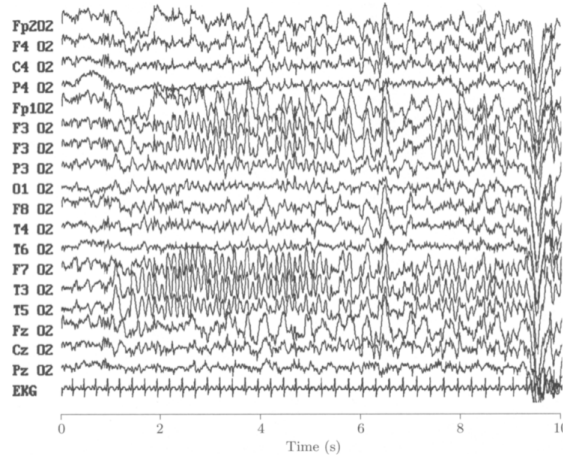


Figura 6: EEG multicanal que muestra el inicio de un ataque epiléptico después del primer segundo. El inicio se caracteriza por un aumento en la amplitud y un cambio en el contenido espectral. La convulsión es particularmente pronunciada en ciertos canales. Tenga en cuenta que el ECG se muestra en la parte inferior[10].

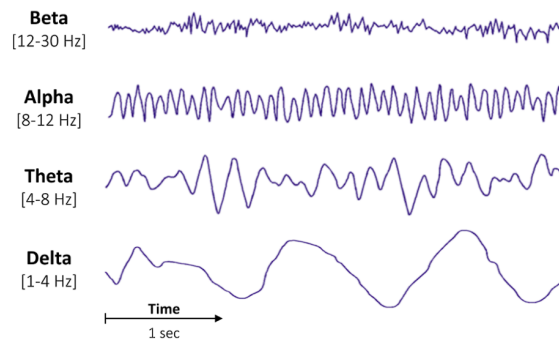


Figura 7: Ejemplo de ritmos cerebrales presentes en un EEG[11].

puede durar de 70 a 200 ms y exhibir formas de onda bifásicas y trifásicas. La morfología de SSW difiere según la ubicación del electrodo en el cuero cabelludo.

SSW puede ocurrir como eventos individuales o en series llamadas complejos de pico y onda, que consisten en un pico seguido de una onda lenta. La tasa de repetición de estos complejos puede oscilar entre menos de 3 a 6 Hz, y la tasa a menudo se asocia con diferentes interpretaciones clínicas.

Algunos artefactos EEG normales pueden parecerse a SSW, como la actividad cardíaca que puede hacerse pasar por un pico, particularmente las ondas del complejo QRS[10].

## Características de una señal EEG

Hay una serie de características que se pueden extraer de las señales de EEG para su procesamiento. Estas funciones se pueden dividir en dos categorías principales: funciones en el dominio del tiempo y funciones en el dominio de la frecuencia.

## EEG ictal

En el caso de una convulsión, el EEG se denomina EEG ictal y se caracteriza por un patrón inusual con un aumento repentino de la amplitud, como se muestra en la Figura 6. Además, el comienzo de una convulsión se acompaña de un cambio repentino en la amplitud. contenido de frecuencia que con frecuencia progresa a un patrón de picos y ondas. Debido a la variabilidad sustancial en el EEG ictal entre convulsiones, puede ser un desafío identificar el patrón de manera consistente, ya sea por medios manuales o automáticos[10].

## Procesamiento de señales EEG

Las diferentes formas de onda de la señal EEG llevan información clínicamente valiosa. Una forma de onda puede representar un evento aislado, o varias formas de onda pueden constituir un patrón de señal compuesto. En ambos casos, es fundamental desarrollar métodos para detectar y cuantificar objetivamente las características de la señal para facilitar la interpretación visual. La extracción de características de señales relevantes es particularmente crucial cuando el objetivo es diseñar un sistema de clasificación de EEG. La cancelación de ruido y artefactos es otro tema importante en el procesamiento de señales de EEG y un requisito previo para el análisis de señales posterior confiable[10].

## Señales electromiográficas

La electromiografía (EMG) es una herramienta de diagnóstico utilizada para analizar y registrar la actividad eléctrica producida por los músculos esqueléticos. La EMG se emplea en diversos campos de la medicina, como la neurología, la medicina deportiva, la rehabilitación y el diagnóstico. Los investigadores también utilizan la EMG para estudiar los patrones de activación muscular durante el movimiento, así como los patrones de fatiga y recuperación muscular. La EMG registra la actividad eléctrica llevada a cabo por las neuronas motoras durante las contracciones musculares y también mide la fuerza de las contracciones[12].

## Señales electrocardiográficas

Un electrocardiograma (ECG) es una prueba médica común que registra la actividad eléctrica del corazón. Se utiliza para evaluar la salud y el funcionamiento del corazón, y puede ayudar a diagnosticar problemas cardíacos o monitorizar afecciones existentes. Durante un ECG, se colocan electrodos en el pecho, las extremidades y a veces en otros lugares del cuerpo, para medir y registrar la actividad eléctrica del corazón [13].

La prueba es indolora y no invasiva, lo que significa que no se introduce nada en el cuerpo. Los resultados del ECG proporcionan información sobre la frecuencia cardíaca, el ritmo cardíaco, la regularidad de los latidos y la presencia de cualquier anomalía en la conducción eléctrica del corazón.

## Funciones en el dominio del tiempo

Las funciones en el dominio del tiempo se basan en el análisis de la amplitud de la señal y sus cambios a lo largo del tiempo. Algunas características comunes en el dominio del tiempo incluyen:

- Media: El valor promedio de la señal durante un intervalo de tiempo específico.
- Mediana: El valor medio de la señal durante un intervalo de tiempo específico.
- Varianza: La desviación cuadrada promedio de la señal de su valor medio.
- Desviación estándar: La raíz cuadrada de la varianza.
- Asimetría: Una medida de la asimetría de la distribución de la señal.
- Curtosis: Una medida del pico de la distribución de la señal.

## Funciones en el dominio de frecuencia

Las funciones de dominio de frecuencia se basan en el análisis del espectro de potencia de la señal. El espectro de potencia es un gráfico de la potencia de la señal en función de la frecuencia. Algunas características comunes en el dominio de la frecuencia incluyen:

- Potencia: La potencia total de la señal en una banda de frecuencia especificada.
- Energía: La energía total de la señal en una banda de frecuencia específica.
- Densidad espectral: La potencia de la señal por unidad de frecuencia en una banda de frecuencia especificada.
- Frecuencia centroide: La frecuencia en la que el espectro de potencia de la señal es más alto.
- Propagación: Una medida del ancho del espectro de potencia de la señal.

La elección de características para extraer de las señales de EEG depende de la aplicación. Por ejemplo, las funciones en el dominio del tiempo se usan a menudo para estudiar los cambios en la actividad cerebral a lo largo del tiempo, mientras que las funciones en el dominio de la frecuencia se usan a menudo para estudiar las diferentes bandas de frecuencia de la actividad cerebral.

## Propiedades de las señales deterministas y estocásticas

Las señales deterministas son aquellas que pueden describirse mediante un conjunto de ecuaciones que definen su comportamiento en el tiempo. Las señales estocásticas, por otro lado, son aquellas que se caracterizan por fluctuaciones aleatorias. Surge una pregunta fundamental con respecto a si el EEG debe verse como una señal determinista o estocástica.

Los intentos de responder a esta pregunta pueden proporcionar información sobre los mecanismos de generación de EEG, pero también tienen implicaciones para los métodos de análisis de señales apropiados considerados. Generalmente, las características exactas de la señal EEG en términos de amplitud, duración o morfología de las ondas individuales no se pueden predecir, por lo que es bastante natural percibir la señal EEG como la realización de un proceso estocástico. Esta perspectiva cobra mayor fuerza al observar que no es posible adquirir una señal EEG "pura" que refleje únicamente la actividad cerebral. De hecho, siempre hay un ruido aleatorio corruptor introducido, por ejemplo, por el ruido interno en el equipo de amplificación o en el proceso de digitalización, lo que, incluso si el EEG "puro" tuviera propiedades deterministas, en última instancia hace que sea razonable considerar el EEG como un proceso estocástico[10].

Hay una serie de propiedades que se pueden utilizar para distinguir entre señales deterministas y estocásticas en los datos de EEG. Una propiedad es el espectro de potencia de la señal. El espectro de potencia de una señal es un gráfico de la potencia de la señal en función de la frecuencia. Las señales deterministas suelen tener un espectro de potencia que se caracteriza por picos pronunciados en frecuencias específicas. Las señales estocásticas, por otro lado, suelen tener un espectro de potencia que se distribuye de manera más uniforme entre las frecuencias[10].

Otra propiedad que se puede utilizar para distinguir entre señales deterministas y estocásticas es la función de autocorrelación de la señal. La función de autocorrelación de una señal es un gráfico de la correlación entre la señal y ella misma en diferentes lapsos de tiempo. Las señales deterministas suelen tener una función de autocorrelación que decae rápidamente con el tiempo. Las señales estocásticas, por otro lado, suelen tener una función de autocorrelación que decae más lentamente con el tiempo[10].

Finalmente, las señales deterministas y estocásticas también se pueden distinguir por su previsibilidad. Las señales deterministas son predecibles en el sentido de que su comportamiento futuro puede predecirse a partir de su comportamiento pasado. Las señales estocásticas, por otro lado, no son predecibles en el sentido de que su comportamiento futuro no puede predecirse a partir de su comportamiento pasado[10].

## **Aprendizaje automático**

El aprendizaje automático (Machine Learning) es un subconjunto de la inteligencia artificial que permite a un sistema aprender y mejorar de forma autónoma mediante redes neuronales y aprendizaje profundo, sin necesidad de programación explícita, ingiriendo grandes cantidades de datos[14].

### **Aprendizaje supervisado**

El aprendizaje supervisado es un tipo de aprendizaje automático en el que el modelo se entrena en un conjunto de datos etiquetados. Esto significa que los datos se han etiquetado con la salida correcta. El modelo aprende a predecir la salida de nuevos datos en función de los datos etiquetados en los que se ha entrenado.

El aprendizaje supervisado se utiliza para una variedad de tareas, incluidas la clasificación, la regresión y la previsión. Las tareas de clasificación implican predecir la categoría de un punto de datos, como si un correo electrónico es spam o no. Las tareas de regresión implican predecir un valor numérico, como el precio de una casa. Las tareas de previsión implican predecir valores futuros, como el número de ventas en un mes determinado[14].

Hay muchos modelos diferentes de aprendizaje supervisado, cada uno con sus propias fortalezas y debilidades. Algunos de los modelos de aprendizaje supervisado más comunes incluyen:

- Regresión lineal
- Regresión logística
- Árboles de decisión
- Bosques aleatorios
- Máquinas de vectores de soporte

La elección de qué modelo de aprendizaje supervisado usar depende de la tarea específica en cuestión. Por ejemplo, la regresión lineal es una buena opción para tareas de regresión simples, mientras que los árboles de decisión son una buena opción para tareas más complejas.

El aprendizaje supervisado es una herramienta poderosa que se puede utilizar para resolver una variedad de problemas. Sin embargo, es importante tener en cuenta que los modelos de aprendizaje supervisado son tan buenos como los datos con los que se entrenan. Si los datos no son precisos o representativos del mundo real, el modelo no podrá hacer predicciones precisas.

Estas son algunas de las características del aprendizaje supervisado:

- Requiere datos etiquetados.
- Se utiliza para entrenar modelos para predecir valores de salida.
- Se puede utilizar para tareas de clasificación, regresión y previsión.
- Hay muchos modelos diferentes de aprendizaje supervisado disponibles.
- La elección del modelo depende de la tarea específica a realizar.

## **Aprendizaje no supervisado**

El aprendizaje no supervisado es un tipo de aprendizaje automático en el que el modelo se entrena en un conjunto de datos no etiquetados. Esto significa que los datos no tienen etiquetas asociadas. El modelo aprende a encontrar patrones en los datos y a agrupar puntos de datos similares.

El aprendizaje no supervisado se utiliza para una variedad de tareas, incluida la agrupación, la reducción de la dimensionalidad y la detección de anomalías. Las tareas de agrupamiento implican agrupar puntos de datos en función de sus similitudes. Las tareas de reducción de dimensionalidad implican reducir la cantidad de características en un conjunto de datos mientras se conserva la mayor cantidad de información posible. Las tareas de detección de anomalías implican identificar puntos de datos que son significativamente diferentes del resto de los datos[14].

Hay muchos modelos diferentes de aprendizaje no supervisado, cada uno con sus propias fortalezas y debilidades. Algunos de los modelos de aprendizaje no supervisado más comunes incluyen:

- Agrupamiento de K-medias
- Agrupación jerárquica
- Análisis de componentes principales (PCA)
- Descomposición en valores singulares (SVD)
- Modelos de mezcla gaussiana (GMM)

La elección de qué modelo de aprendizaje no supervisado usar depende de la tarea específica en cuestión. Por ejemplo, el agrupamiento de k-medias es una buena opción para tareas de agrupamiento simples, mientras que el agrupamiento jerárquico es una buena opción para tareas más complejas[14].

El aprendizaje no supervisado es una herramienta poderosa que se puede utilizar para resolver una variedad de problemas. Sin embargo, es importante tener en cuenta que los modelos de aprendizaje no supervisados son tan buenos como los datos con los que se entrenan. Si los datos no son precisos o representativos del mundo real, el modelo no podrá encontrar patrones precisos en los datos.

Estas son algunas de las características del aprendizaje no supervisado:

- No requiere datos etiquetados.
- Se utiliza para entrenar modelos para encontrar patrones en los datos.
- Se puede utilizar para tareas de agrupamiento, reducción de dimensionalidad y detección de anomalías.
- Hay muchos modelos diferentes de aprendizaje no supervisado disponibles.
- La elección del modelo depende de la tarea específica a realizar.

# Metodología

## Obtención señales EEG

Para desarrollar los algoritmos de aprendizaje automático, necesitamos señales que coincidan con las anotaciones de interés. La mayoría de estas señales se obtendrán como se ha hecho con anterioridad de la base de datos Seizure Prediction Challenges de la American Epilepsy Society en Kaggle. La base de datos contiene registros iEEG (electroencefalograma intracraneal) de un paciente, que se han clasificado en segmentos preictales e interictales. También se utilizarán señales proporcionadas por HUMANA y recolectadas en UVG. Estas señales no tendrán ninguna etiqueta previa, lo que da paso al uso del software desarrollado en UVG.

## Transformación de datos

La *Epileptic EEG Analysis Toolbox* se ha desarrollado en MATLAB, por lo que los datos deberán estar en formato “.mat”, además, los datos recolectados en UVG mediante BIOPAC se transformarán a tipo “Struct” para ser empleados en la herramienta de software.

## Procesamiento de señal

Con el fin de evitar tener datos erróneos, se procederá a filtrar las señales. Para se aplicarán algoritmos utilizados en las fases anteriores los cuales demostraron ser efectivos.

## Segmentación de datos, extracción y selección de características de la señal

Para disminuir el esfuerzo computacional se segmenta la señal en varias partes, ya que como se sabe se puede tener varias horas de grabación.

## Implementación de algoritmos de aprendizaje automático

Para la creación de modelos capaces de clasificar las señales se utilizarán las técnicas de “Aprendizaje supervisado” y “Aprendizaje no supervisado”.

- Aprendizaje supervisado: Redes neuronales y máquinas de vectores de soporte.
- Aprendizaje no supervisado: Agrupación jerárquica, agrupamiento de K-medias, modelos de mezcla gaussiana

## **Reconocimiento de segmento con actividad ictal**

Entrenados los modelos de aprendizaje automático se procesará y etiquetará la señal EEG. Esto con el fin de no tener una carga computacional alta. Posteriormente, se procederá a utilizar un algoritmo capaz de retornar el tiempo exacto donde se inicia la actividad preictal, ictal o postictal, según sea el interés del doctor. Para lo cual se apoyará de la función *find* nativa de MATLAB.

## **Validación de resultados con HUMANA**

El objetivo de la línea de investigación mencionada es ayudar a los médicos a optimizar el tiempo dedicado a analizar los datos de los pacientes que sufren epilepsia, mediante la herramienta de software. Para garantizar un correcto funcionamiento de la herramienta se espera una retroalimentación constante y verificar los resultados, con los especialistas.



## Cronograma de actividades

Actividad		Julio							Agosto				Septiembre			
		1-8	9-15	16-22	23-31	1-5	6-12	13-19	20-26	27-31	1-9	10-16	17-23	24-30		
A1	Familiarización señales bioeléctricas (EEG, EMG, ECG)															
A2	Analisis y comprensión de las señales bioeléctricas y sus características															
A3	Obtención y estandarización de señales bioeléctricas															
A4	Familiarización de la <i>Toolbox</i>															
B1	Pruebas de modelos de aprendizaje automatico															
B2	Generación de algoritmo para la detección de tiempo en el que se encuentra la actividad ictal de interes															
B3	Actualización del toolbox															
B4	Validación por profesionales y ajustes															
C1	Analisis y actualización de los metodos de entrenamiento para aprendizaje automatico															
C2	Investigación y comprensión de anotación de señales															
D1	Redacción de documento final de tesis															

# Índice preliminar

Prefacio	II
Lista de figuras	III
Lista de cuadros	IV
Resumen	V
Abstract	VI
1. Introducción	1
2. Antecedentes	2
3. Justificación	3
4. Objetivos	4
4.1. Objetivo General	4
4.2. Objetivos Específicos	4
5. Alcance	5
6. Marco Teórico	6
6.1. Epilepsia	6
6.2. Tipos de epilepsia	6 - 7
6.3. Señales bioeléctricas	7 - 10
6.3.1. Señales electroencefalográficas	7
6.3.1.1. Ritmos y formas de onda del EEG	8
6.3.1.2 Picos y ondas agudas	8 - 9
6.3.1.3. Características de una señal EEG	9
6.3.1.4. EEG ictal	10
6.3.1.5. Procesamiento de señales EEG	10
6.4. Señales electromiográficas	10
6.5. Señales electrocardiográficas	10
6.6. Funciones en el dominio del tiempo	11
6.7. Funciones en el dominio de frecuencia	11
6.8. Propiedades de las señales deterministas y estocásticas	12

6.9. Aprendizaje automático	12 - 14
6.9.1. Aprendizaje supervisado	12 - 13
6.9.2. Aprendizaje no supervisado	13 - 14
7. Diseño Experimental	15 - 17
7.1. Obtención de datos y estandarización	15
7.2. Segmentación de las señales y procesamiento	15
7.3. Extracción y selección de características	16
7.4. Entrenamiento de algoritmos de aprendizaje automático	16
7.5. Identificación de tiempo con data de interés	17
10. Resultados	19
11. Conclusiones	20
12. Recomendaciones	21
Bibliografía	22

## Referencias

- [1] O. organización mundial de la salud, *Epilepsia*, feb. de 2023. dirección: <https://www.who.int/es/news-room/fact-sheets/detail/epilepsy>.
- [2] C. de Epilepsia y Neurocirugía Funcional, *HUMANA, Epilepsia*, mar. de 2021. dirección: <https://humanagt.org/epilepsia/>.
- [3] C. Lemus, “Análisis y anotación de señales bioeléctricas de pacientes con epilepsia utilizando técnicas de aprendizaje automático supervisado y no supervisado,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2022.
- [4] M. J. Angulo, “Análisis y Reconocimiento de Patrones de Señales Biomédicas de Pacientes con Epilepsia,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2020.
- [5] D. A. Vela, “Automatización del Proceso de Anotación de Señales EEG de Pacientes con Epilepsia por Medio de Técnicas de Aprendizaje Automático,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2021.
- [6] E. Wyllie, A. Gupta y D. K. Lachhwani, *The treatment of epilepsy: principles & practice*. Lippincott Williams & Wilkins, 2006.
- [7] R. García-Ramos, A. G. Pastor, J. Masjuan, C. Sánchez y A. Gil, “FEEN: Informe sociosantario FEEN sobre la epilepsia en España,” *Neurología*, vol. 26, n.º 9, págs. 548-555, 2011.

- [8] R. S. Fisher, H. Cros, J. A. French, N. Higurashi, E. Hirsch y F. E. Jansen, *The ILAE classification of seizures and the epilepsies: Modification ...* 2017. dirección: <https://www.ilae.org/files/ilaeGuideline/Spanish-Traduccion-Fisher-Seizure-types.pdf>.
- [9] E. Elizondo, H. Manzo, D. Bermudez y E. Jiménez, *Manipulación de un brazo robótico mediante señales electroencefalográficas*, 2016. dirección: [https://www.ecorfan.org/bolivia/researchjournals/Tecnologia\\_e\\_innovacion/vol3num7/Revista\\_de\\_Tecnologia\\_e\\_Innovaci%C3%93n\\_V3\\_N7.pdf](https://www.ecorfan.org/bolivia/researchjournals/Tecnologia_e_innovacion/vol3num7/Revista_de_Tecnologia_e_Innovaci%C3%93n_V3_N7.pdf).
- [10] L. Sörnmo y P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic press, 2005, vol. 8.
- [11] P. R. Vallat, *Compute the average bandpower of an EEG signal*, 2018. dirección: <https://raphaelvallat.com/bandpower.html>.
- [12] L. Gila, A. Malanda, I. Rodríguez Carreño, J. Rodríguez Falces y J. Navallas, *Métodos de procesamiento y Análisis de Señales Electromiográficas*, 2009. dirección: [https://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1137-66272009000600003](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272009000600003).
- [13] M. MedlinePlus, *Electrocardiograma: Prueba de Laboratorio de Medlineplus*, 2023. dirección: [https://medlineplus.gov/spanish/pruebas-de-laboratorio/electrocardiograma/#:~:text=Un%20electrocardiograma%20\(ECG\)%20es%20un,circula%5C%20a%20trav%C3%A9s%20de%20%C3%A9l..](https://medlineplus.gov/spanish/pruebas-de-laboratorio/electrocardiograma/#:~:text=Un%20electrocardiograma%20(ECG)%20es%20un,circula%5C%20a%20trav%C3%A9s%20de%20%C3%A9l..)
- [14] G. G-Cloud, *¿Qué es el aprendizaje automático?* 2023. dirección: <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419#:~:text=Descargar%5C%20la%5C%20gu%C3%ADa-,Definici%C3%B3n%20de%5C%20aprendizaje%20autom%C3%A1tico,de%20grandes%20cantidades%20de%20datos..>