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

Cristhofer Isaac Patzán Martínez



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**

Trabajo de graduación presentado por Cristhofer Isaac Patzán Martínez
para optar al grado académico de Licenciado en Ingeniería Mecatrónica

Guatemala,

2024

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



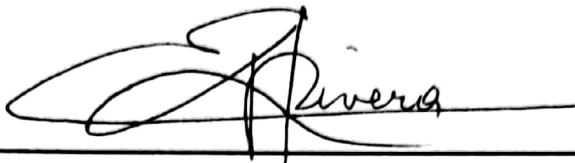
**Aplicación sistemática de algoritmos de aprendizaje
automático para el estudio de la epilepsia y la detección de
segmentos de interés en señales bioeléctricas**

Trabajo de graduación presentado por Cristhofer Isaac Patzán Martínez
para optar al grado académico de Licenciado en Ingeniería Mecatrónica

Guatemala,

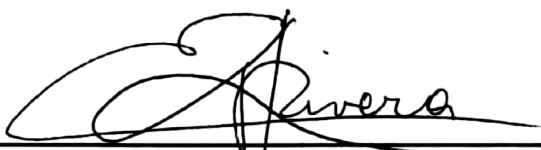
2024

Vo.Bo.:


(f) 

Dr. Luis Alberto Rivera Estrada


Tribunal Examinador:

(f) 

Dr. Luis Alberto Rivera Estrada

(f) 

Dr. Gabriel Barrientos

(f) 

M.Sc. Carlos Esquit

Fecha de aprobación: Guatemala, 13 de enero de 2024.

En el transcurso de esta ardua travesía académica que culmina con la presentación de esta tesis, es necesario detenerse un momento para expresar mi sincero agradecimiento a todas las personas que han sido pilares fundamentales en este camino lleno de desafíos y logros.

En primer lugar, quiero dedicar un profundo agradecimiento a la Universidad del Valle de Guatemala y SEGEPLAN, quienes me brindaron la oportunidad de realizar mis estudios de grado con una beca completa. Esta institución no solo me proporcionó una educación de alta calidad, sino que también me apoyó económicamente a lo largo de toda mi carrera. Quiero dedicar un profundo agradecimiento al Dr. Luis Rivera, quien desempeñó un papel esencial como mi asesor de tesis. Su apoyo incondicional, orientación experta y dedicación a este proyecto fueron invaluable. El Dr. Rivera no solo compartió su vasto conocimiento, sino que también brindó una fuente inagotable de inspiración y motivación. Su guía y compromiso fueron esenciales para llevar a cabo esta investigación de manera exitosa.

A mi familia, en especial a mi madre, tías y abuela, les debo un reconocimiento especial. Su apoyo constante, aliento inquebrantable y amor infinito fueron los cimientos sobre los cuales construí mi camino académico. Siempre estuvieron ahí para impulsarme en los momentos difíciles y celebrar conmigo en los triunfos. Su confianza en mí fue el motor que me llevó a perseverar y alcanzar este logro.

Esta tesis es el resultado de un esfuerzo colectivo y el apoyo de muchas personas que creyeron en mí a lo largo de los años. Cada página escrita y cada descubrimiento logrado son un testimonio de la dedicación y el trabajo arduo de todos aquellos que contribuyeron de alguna manera a este proyecto.

A todos ellos, mi más profundo agradecimiento.

Prefacio	III
Lista de figuras	XII
Lista de cuadros	XIII
Resumen	XIV
Abstract	XV
1. Introducción	1
2. Antecedentes	2
3. Justificación	5
4. Objetivos	7
4.1. Objetivo general	7
4.2. Objetivos específicos	7
5. Alcance	8
6. Marco teórico	9
6.1. Epilepsia	9
6.1.1. Tipos de epilepsia	9
6.2. Señales bioeléctricas	10
6.3. Señales electroencefalográficas	11
6.3.1. Ritmos y formas de onda del EEG	11
6.3.2. Picos y ondas agudas	12
6.3.3. Características de una señal EEG	13
6.3.4. EEG ictal	13
6.3.5. Procesamiento de señales EEG	13
6.4. Señales electromiográficas	13
6.5. Señales electrocardiográficas	14

6.6.	Características en el dominio del tiempo	14
6.7.	Características en el dominio de frecuencia	14
6.8.	Características en el dominio de tiempo-frecuencia	15
6.9.	Propiedades de las señales deterministas y estocásticas	16
6.10.	Aprendizaje automático	16
6.10.1.	Aprendizaje supervisado	17
6.10.2.	Aprendizaje no supervisado	18
6.10.3.	Aprendizaje por refuerzo	19
6.11.	Algoritmos jerárquicos	19
6.11.1.	Algoritmo Chameleon	20
6.12.	BIOPAC	20
6.13.	VAT	20
7.	Metodología	22
7.1.	Datos obtenidos con equipo <i>UVG</i>	22
7.1.1.	Señales <i>EMG</i>	22
7.1.2.	Señales <i>EEG</i>	24
7.1.3.	Estándar en recolección de datos BIOPAC	25
7.1.4.	Estructuración de los datos para ser usados en la herramienta de software para el estudio de la epilepsia	27
7.2.	Datos de HUMANA de pacientes con epilepsia	27
7.3.	Agrupamiento de datos	28
7.3.1.	Análisis de señales <i>EEG</i> de forma intersujeto	28
7.3.2.	Análisis de señales <i>EMG</i> de forma intrasujeto	28
7.4.	Aprendizaje automático	29
7.4.1.	Extracción de características	29
7.4.2.	Anotaciones automáticas	29
7.4.3.	Análisis estadístico	29
7.5.	Actualización de la herramienta de software para el estudio de la epilepsia	30
8.	Resultados de recolección de datos	34
8.1.	Señales bioeléctricas obtenidas con el equipo de <i>UVG</i>	34
8.1.1.	Resultados de la extracción de características	35
8.2.	Anotaciones automáticas	37
9.	Resultados estadísticos	40
9.1.	Clasificadores	40
9.2.	Algoritmos de agrupamiento	44
10.	Análisis de grupos	45
10.1.	Análisis de grupos en el dominio de tiempo	46
10.1.1.	Resultados K-means	46
10.1.2.	Resultados jerárquico	49
10.2.	Análisis de grupos en el dominio de frecuencia	54
10.2.1.	Resultados K-means	55
10.2.2.	Resultados jerárquico	63
10.3.	Análisis de grupos en el dominio de wavelets	70
10.3.1.	Resultados K-means	71

10.3.2. Resultados jerárquicos	79
10.4. Herramienta para la evaluación visual de la tendencia	86
11. Actualización de la herramienta de software para el estudio de la epilepsia	90
12. Exploración con algoritmo CHAMELEON	94
13. Conclusiones	97
14. Recomendaciones	98
15. Bibliografía	99
16. Anexos	102
16.1. Sitios web con herramienta de software y datos recolectados	102
16.2. Pruebas con algoritmo VAT	102
16.2.1. VAT características en el dominio de frecuencia	103
16.2.2. VAT características en el dominio de tiempo continuo	107
16.2.3. VAT características en el dominio de wavelets	112

Lista de figuras

1.	Resumen de los resultados de los clasificadores generados: red neuronal (RNA) y máquina de vectores de soporte (SVM) [6].	3
2.	Resumen del rendimiento de la RNA para dos clasificadores binarios utilizando características individuales en dominio de la frecuencia [4].	3
3.	Ventana de la aplicación para iniciar sesión [6].	4
4.	Ventana con la gráfica de una ventana del registro EEG del canal seleccionado [6].	4
5.	Clasificación operacional extendida de los tipos de crisis [12].	10
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 [15].	12
7.	Ejemplo de ritmos cerebrales presentes en un EEG [16].	12
8.	Pose de relajación para toma de señales EMG con equipo BIOPAC.	23
9.	Pose de actividad 1 en toma de grabación señal EMG.	23
10.	Pose de actividad 2 en toma de grabación señal EMG.	23
11.	Pose de actividad 3 en toma de grabación señal EMG.	23
12.	Actividad 4 movimiento controlado de cejas.	24
13.	Actividad 5 inclinación de la cabeza hacia abajo.	25
14.	Actividad 6 inclinación de la cabeza hacia arriba.	25
15.	Actividad 7 inclinación de la cabeza hacia el lateral derecho.	25
16.	Estándar para el almacenamiento de datos obtenidos con el equipo BIOPAC de la UVG.	25
17.	Señales bioeléctricas extraídas con el equipo BIOPAC almacenadas según la norma establecida.	26
18.	Contenido de <i>struct</i> para el uso de datos dentro de la herramienta de software para el estudio de la epilepsia.	27
19.	Ventana de anotaciones automáticas con nueva sección para exportar datos de interés.	29

20.	Error al utilizar botón de visualización de canales en la ventana de anotaciones automáticas.	30
21.	Alerta al emplear el algoritmo FeaturesMorf.	30
22.	Flujograma de la creación del vector de características versión 2022.	31
23.	Flujograma de la creación del vector de características versión 2023.	31
24.	Segmento de código para el calculo de características de una señal bioeléctrica.	32
25.	Flujograma de algoritmo ZC versión 2022.	33
26.	Flujograma de algoritmo ZC versión 2023.	33
27.	Structs generados a partir de los datos obtenidos con el equipo UVG.	35
28.	Características en el dominio del tiempo continuo.	36
29.	Características con wavelets.	36
30.	Características en el dominio de la frecuencia.	36
31.	Ventana de inicio para generar anotaciones automáticas.	37
32.	Anotaciones automáticas de 33 canales analizados.	38
33.	Anotaciones automáticas de 33 canales analizados con menor cantidad de EEG de personas sin actividad ictal.	38
34.	Matriz de confusión para las clases Ictal y Sano utilizando SVM con características wavelets.	41
35.	Matriz de confusión para las clases Ictal y Sano utilizando SVM con características en tiempo continuo.	41
36.	Matriz de confusión para las clases Ictal y Sano utilizando SVM con características frecuencia.	42
37.	Matriz de confusión para las clases Ictal y Sano utilizando RNA con características wavelets.	42
38.	Matriz de confusión para las clases Ictal y Sano utilizando RNA con características en tiempo continuo.	43
39.	Matriz de confusión para las clases Ictal y Sano utilizando RNA con características en frecuencia.	43
40.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 2 en el dominio del tiempo de actividad ictal y no ictal.	47
41.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 4 en el dominio del tiempo de actividad ictal y no ictal.	47
42.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 5 en el dominio del tiempo de actividad ictal y no ictal.	48
43.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 4 en el dominio del tiempo de actividad ictal y no ictal.	48
44.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 5 en el dominio del tiempo de actividad ictal y no ictal.	49
45.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 5 en el dominio del tiempo de actividad ictal y no ictal.	49
46.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 2 en el dominio del tiempo de actividad ictal y no ictal.	50
47.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 3 en el dominio del tiempo de actividad ictal y no ictal.	50

96.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 5 tipo wavelets de actividad ictal y no ictal.	76
97.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 6 tipo wavelets de actividad ictal y no ictal.	77
98.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 5 tipo wavelets de actividad ictal y no ictal.	77
99.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 6 tipo wavelets de actividad ictal y no ictal.	78
100.	Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 5 y 6 tipo wavelets de actividad ictal y no ictal.	78
101.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 2 tipo wavelets de actividad ictal y no ictal.	79
102.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 3 tipo wavelets de actividad ictal y no ictal.	79
103.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 4 tipo wavelets de actividad ictal y no ictal.	80
104.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 5 tipo wavelets de actividad ictal y no ictal.	80
105.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 6 tipo wavelets de actividad ictal y no ictal.	81
106.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 3 tipo wavelets de actividad ictal y no ictal.	81
107.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 4 tipo wavelets de actividad ictal y no ictal.	82
108.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 5 tipo wavelets de actividad ictal y no ictal.	82
109.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 6 tipo wavelets de actividad ictal y no ictal.	83
110.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 4 tipo wavelets de actividad ictal y no ictal.	83
111.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 5 tipo wavelets de actividad ictal y no ictal.	84
112.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 6 tipo wavelets de actividad ictal y no ictal.	84
113.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 5 tipo wavelets de actividad ictal y no ictal.	85
114.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 6 tipo wavelets de actividad ictal y no ictal.	85
115.	Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 5 y 6 tipo wavelets de actividad ictal y no ictal.	86
116.	VAT en la herramienta de software para el estudio de la epilepsia.	87
117.	Resultado tras aplicar VAT al vector de características de 5 dimensiones en el dominio tiempo.	88
118.	Resultado tras aplicar VAT al vector de características de 6 dimensiones en el dominio frecuencia.	88
119.	Resultado tras aplicar VAT al vector de características de 6 dimensiones en el dominio wavelets.	89

120.	Ventana de avisos.	91
121.	Notas en consola de MATLAB.	92
122.	Aviso de descarga finalizada.	92
123.	Archivos descargados.	93
124.	Funcionamiento de ventana multigráfico.	93
125.	Manual de instalación y modificación a dependencias para el uso del algoritmo CHAMELEON.	95
126.	Datos de prueba con secciones cónicas para algoritmo CHAMELEON.	96
127.	Datos de prueba con secciones no cónicas para algoritmo CHAMELEON.	96
128.	VAT usando 100 épocas de características tipo frecuencia.	103
129.	VAT usando 500 épocas de características tipo frecuencia.	104
130.	VAT usando 1000 épocas de características tipo frecuencia.	104
131.	VAT usando 1500 épocas de características tipo frecuencia.	105
132.	VAT usando 2000 épocas de características tipo frecuencia.	105
133.	VAT usando 2500 épocas de características tipo frecuencia.	106
134.	VAT usando 2600 épocas de características tipo frecuencia.	106
135.	VAT usando 2700 épocas de características tipo frecuencia.	107
136.	VAT usando 100 épocas de características tipo tiempo continuo.	108
137.	VAT usando 500 épocas de características tipo tiempo continuo.	108
138.	VAT usando 1000 épocas de características tipo tiempo continuo.	109
139.	VAT usando 1500 épocas de características tipo tiempo continuo.	109
140.	VAT usando 2000 épocas de características tipo tiempo continuo.	110
141.	VAT usando 2100 épocas de características tipo tiempo continuo.	110
142.	VAT usando 2200 épocas de características tipo tiempo continuo.	111
143.	VAT usando 2300 épocas de características tipo tiempo continuo.	111
144.	VAT usando 100 épocas de características tipo wavelets.	112
145.	VAT usando 500 épocas de características tipo wavelets.	113
146.	VAT usando 1000 épocas de características tipo wavelets.	113
147.	VAT usando 1500 épocas de características tipo wavelets.	114
148.	VAT usando 2000 épocas de características tipo wavelets.	114
149.	VAT usando 2100 épocas de características tipo wavelets.	115
150.	VAT usando 2200 épocas de características tipo wavelets.	115
151.	VAT usando 2300 épocas de características tipo wavelets.	116

Lista de cuadros

1.	Información de grabaciones dadas por HUMANA	28
2.	Datos en nube con equipo UVG	35
3.	Datos procesados en nube con equipo UVG	35
4.	Tiempos de extracción de características	36
5.	Tiempos de entrenamiento para clasificadores	37
6.	Resultado rand index para algoritmos de agrupamiento	44

En Guatemala no se cuenta con fácil acceso a nuevas tecnologías para el estudio a fondo de los casos de pacientes con epilepsia, que a la fecha se estiman que son alrededor de 325,000 pacientes. Una enfermedad que repercute tanto en la salud del paciente como en su desenvolvimiento social. El presente trabajo tiene como objetivo, aplicar los algoritmos de aprendizaje automático desarrollados en fases anteriores a una mayor cantidad de señales bioeléctricas, y mejorar el proceso de detección de segmentos de interés en las señales, para el estudio de la epilepsia.

Para lograrlo se inició con la obtención de señales bioeléctricas con el equipo BIOPAC de la Universidad del Valle de Guatemala (UVG), de personas que no sufren de ataques epilépticos y por parte de HUMANA, señales bioeléctricas de pacientes con ataques de epilepsia. Recolectar una mayor cantidad de datos que en años previos fue de gran necesidad, sirvió para una mejor clasificación al momento de entrenar el modelo de aprendizaje automático. Cabe destacar que a los datos obtenidos con el equipo de UVG, se les asignaba un nombre según la norma que se estableció en UVG. Posteriormente se extrajeron características en el dominio del tiempo, frecuencia y wavelets de las señales anteriormente mencionadas. Los experimentos para la validación de modelos, en el caso de las señales de electromiografía (EMG) fueron intrasujeto, mientras que, en el caso de las señales de electroencefalografía (EEG) fueron intersujeto.

Por último, se actualizó la herramienta de software para el estudio de la epilepsia que se ha desarrollado en los últimos años. Esto incluía el mejoramiento del proceso de detección de segmentos de interés en las señales bioeléctricas, la optimización de funciones para la extracción de características y el entrenamiento del modelo de aprendizaje automático y la generación automática de anotaciones relevantes. Cumplidas estas actividades se procedió con el análisis estadístico para evaluar el rendimiento de los algoritmos e identificar posibles mejoras a los mismos.

In Guatemala there is no easy access to new technologies for the in-depth study of the cases of patients with epilepsy, which to date are estimated to be around 325,000 patients. A disease that affects both the patient's health and social development. The present work aims to apply the machine learning algorithms developed in previous phases to a larger number of bioelectrical signals, and to improve the process of detecting segments of interest in the signals, for the study of epilepsy.

To achieve this, we began by obtaining bioelectrical signals with the BIOPAC equipment of the Universidad del Valle de Guatemala (UVG), from people who do not suffer from epileptic seizures, and by HUMANA, bioelectrical signals from patients with epileptic seizures. Collecting a larger amount of data, which in previous years was of great need, served for a better classification when training the machine learning model. It should be noted that the data obtained with the UVG equipment were assigned a name according to the standard that was established in UVG. Subsequently, time, frequency and wavelet domain features were extracted from the aforementioned signals. The experiments for model validation, in the case of electromyography (EMG) signals, were intrasubject, while in the case of electroencephalography (EEG) signals they were intersubject.

Finally, the software tool for the study of epilepsy that has been developed in recent years was updated. This included improving the process of detecting segments of interest in the bioelectrical signals, optimizing functions for feature extraction and training of the machine learning model and automatic generation of relevant annotations. Once these activities were completed, statistical analysis was performed to evaluate the performance of the algorithms and identify possible improvements to them.

En este trabajo de investigación, se aborda el tema de la epilepsia, un trastorno neurológico común caracterizado por convulsiones recurrentes e incapacitantes, conocidas como crisis epilépticas, médicamente descrito como actividad ictal. Se ha observado un creciente interés en la aplicación del aprendizaje automático en aplicaciones clínicas y experimentales relacionadas con el diagnóstico de la epilepsia y otros trastornos neurológicos y psiquiátricos. Los métodos de aprendizaje automático son considerados como potenciales herramientas capaces de proporcionar un rendimiento fiable y óptimo en el diagnóstico clínico, la predicción y la medicina personalizada, mediante la aplicación de algoritmos matemáticos y enfoques computacionales.

Entre las aplicaciones que se han empleado con los algoritmos de aprendizaje automático en el contexto de la epilepsia, se encuentra el reconocimiento de patrones de interés en señales bioeléctricas. Este enfoque ha demostrado ser efectivo en la reducción del tiempo de diagnóstico, en contraposición al método tradicional de análisis y anotaciones manuales realizadas por los médicos, que tiende a ser una tarea que consume mucho tiempo, especialmente cuando se tratan registros de varias horas de duración.

El objetivo de este proyecto de investigación es aplicar los algoritmos de aprendizaje automático desarrollados en fases anteriores a una mayor cantidad de señales bioeléctricas, y mejorar el proceso de detección de segmentos de interés en las señales, para el estudio de la epilepsia. Esto implica el análisis de diversas señales bioeléctricas y la evaluación de clasificadores. En trabajos previos, se emplearon características en el dominio del tiempo, frecuencia y se utilizaron transformadas Wavelet para el análisis. En este trabajo, se presenta una revisión de las bases teóricas utilizadas para abordar el problema, así como una descripción detallada de los experimentos realizados, la metodología empleada y los resultados obtenidos en el transcurso de esta investigación. Asimismo, se presentan las conclusiones y recomendaciones que surgen de este estudio, con miras a futuras investigaciones en esta área de estudio.

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 [1]. 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 [2].

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 [3]. 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 [4].

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 [5].

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 [6]. El clasificador con mejor desempeño en tiempo continuo obtuvo un promedio de exactitud del 96.7%, utilizando máquina de vectores de

Características	Tiempo Continuo			Wavelet		
	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) [6].

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 [4].

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 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 [6]. 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 [4] 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 (θ/α) y la razón 2 (β/α), 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 [4]. 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 [6].

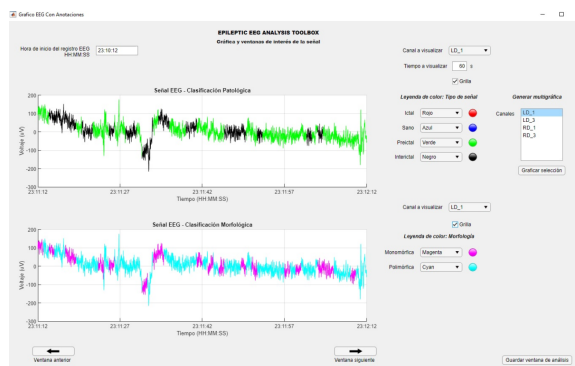


Figura 4: Ventana con la gráfica de una ventana del registro EEG del canal seleccionado [6].

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 [7].

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 realiza en el presente proyecto, se aplicó 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 [4].

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 [8].

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 [9]. 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.

4.1. Objetivo general

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

4.2. Objetivos específicos

- Obtener una buena cantidad de señales bioeléctricas capturadas con el equipo de la UVG, y de pacientes con epilepsia de HUMANA.
- Aplicar algoritmos de aprendizaje automático desarrollados en fases anteriores, para la extracción de características de las señales bioeléctricas del tipo EEG y EMG.
- 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.

Los resultados derivados de este estudio se aplicarán principalmente en el contexto de Guatemala, aunque las metodologías desarrolladas pueden ser de utilidad en otros entornos médicos similares. Para este trabajo de graduación se recopiló señales electroencefalográficas y electromiográficas capturadas con el equipo de la Universidad del Valle de Guatemala (UVG). Se obtuvieron señales bioeléctricas de pacientes con epilepsia atendidos en el Instituto HUMANA, ubicado en Guatemala.

Los algoritmos de aprendizaje automático se aplicaron específicamente para la extracción de características de las señales bioeléctricas, la detección de patrones de interés y la generación de anotaciones relevantes. Estos algoritmos se basaron en técnicas de aprendizaje supervisado y no supervisado. No se abordó la investigación de nuevos algoritmos de aprendizaje automático en este trabajo, sino la adaptación y optimización de técnicas existentes.

La actualización de la herramienta de software propuesta se enfocó en la mejora de los algoritmos empleados, así como aplicar buenas prácticas de programación, obteniendo una optimización en cuanto al costo computacional. Se incluyó la exportación de datos de los segmentos de interés a analizar. La implementación de tratamientos médicos y decisiones clínicas a partir de los resultados obtenidos con la herramienta, estuvo fuera del alcance de este trabajo.

6.1. 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 [10].

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 [11].

6.1.1. 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.

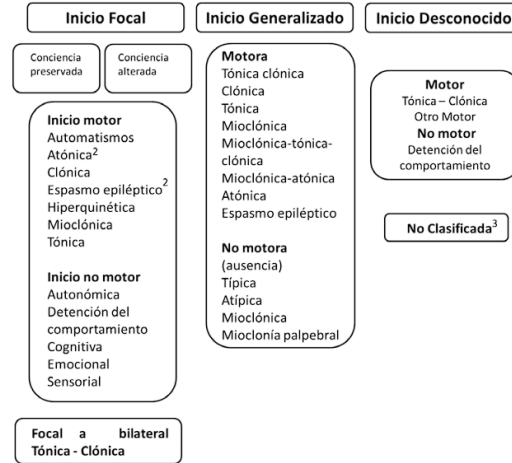


Figura 5: Clasificación operativa extendida de los tipos de crisis [12].

- 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.

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 [12].

6.2. 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 [13].

6.3. 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 [14].

6.3.1. 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 μV 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 durante profundo y tiene una gran amplitud. El adulto despierto y normal, pero es indicativo, por ejemplo, de daño cerebral o enfermedad cerebral (encefalopatía).
- Ritmo theta, 4-7 Hz. El ritmo theta se produce durante la somnolencia y en ciertas fases del sueño.
- 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.
- 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.
- 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.

La mayoría de estos ritmos pueden durar varios minutos, mientras que otros sólo unos

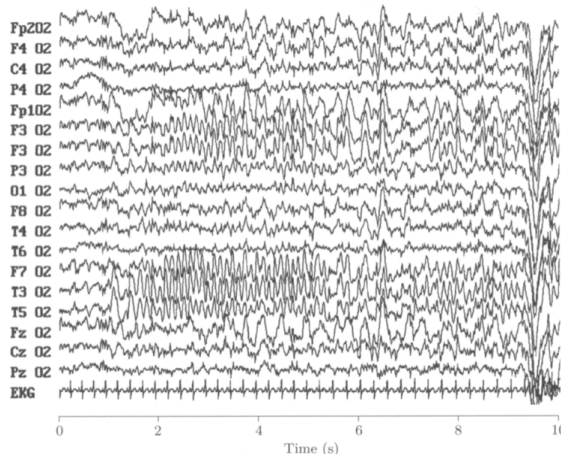


Figura 6: EEG multicanal que muestra el inicio de un ataque epiléptico después del primer segundo. El inicio se caracteriza por un aumento en la amplitud y un cambio en el contenido espectral. La convulsión es particularmente pronunciada en ciertos canales. Tenga en cuenta que el ECG se muestra en la parte inferior [15].

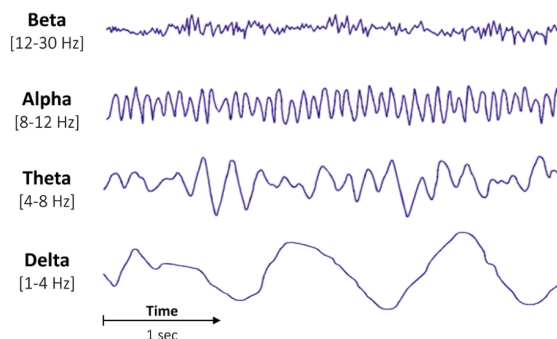


Figura 7: Ejemplo de ritmos cerebrales presentes en un EEG [16].

segundos. Es importante darse cuenta de que un ritmo no está presente en todo momento, una señal irregular de aspecto “arrítmico” [15].

6.3.2. 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 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 [15].

6.3.3. 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.

6.3.4. *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 [15].

6.3.5. 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 [15].

6.4. 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 [17].

6.5. 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 [18].

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 [19].

6.6. Características en el dominio del tiempo

Las características en el dominio del tiempo se basan en el análisis de la amplitud de la señal y sus cambios a lo largo del tiempo [20]. 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.

6.7. Características en el dominio de frecuencia

Las características en el dominio de la 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 [21]. 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 [22].

6.8. Características en el dominio de tiempo-frecuencia

Las wavelets, también conocidas como funciones de onda madre, son funciones matemáticas utilizadas en el procesamiento de señales y la análisis de datos. Se destacan por tener una duración finita y estar localizadas tanto en el dominio del tiempo como en el dominio de la frecuencia. La característica principal de las wavelets es que pueden representar tanto características de alta como de baja frecuencia en una señal, lo que las hace útiles para el análisis de señales complejas [23].

Algunas características comunes en el dominio de la frecuencia incluyen:

- Potencia: La potencia indica la cantidad de energía en una señal. En el contexto de las wavelets, se puede calcular la potencia en diferentes bandas de frecuencia, lo que proporciona información sobre cómo se distribuye la energía a lo largo de la señal.
- Media: La media es el promedio de los valores de una señal. Puede calcularse en diferentes niveles de descomposición wavelet para obtener la media en diferentes escalas.
- Desviación Estándar: La desviación estándar mide la dispersión de los valores con respecto a la media. En el análisis wavelet, puede indicar la variabilidad de la señal en diferentes escalas.
- Curtosis: La curtosis es una medida de la forma de la distribución de los valores de una señal. Indica la "cola" de la distribución y su nivel de apuntamiento. Valores altos de curtosis indican una distribución más concentrada alrededor de la media.
- Asimetría: La asimetría mide la falta de simetría en la distribución de los valores de una señal. Un valor positivo indica una cola más larga en el lado derecho de la distribución, mientras que un valor negativo indica una cola más larga en el lado izquierdo.
- Cruces por Cero (ZC): Las cruces por cero indican el número de veces que una señal cruza el eje horizontal. En el análisis wavelet, esto puede utilizarse para evaluar la cantidad de cambios de polaridad en diferentes niveles de descomposición.

La transformada wavelet, se utilizan para descomponer una señal en componentes de diferente frecuencia y amplitud en varios niveles de resolución. Esta descomposición multi-resolución es especialmente útil para analizar señales que contienen información en diferentes escalas de tiempo. Las wavelets han encontrado aplicaciones en una amplia gama de campos, como la compresión de imágenes, el procesamiento de señales biomédicas, el análisis de audio y la detección de patrones en datos [24].

6.9. 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 [15].

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 [15].

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 [15].

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 [15].

6.10. Aprendizaje automático

El aprendizaje automático (*Machine Learning*) es un tipo de inteligencia artificial (IA) que permite a las aplicaciones de software ser más precisas en la predicción de resultados sin estar explícitamente programadas para ello. Los algoritmos de aprendizaje automático utilizan datos históricos como entrada para predecir nuevos valores de salida [25]. Existen tres tipos principales de aprendizaje automático:

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje por refuerzo

6.10.1. 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 [25].

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 [26].

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.

6.10.2. 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 [25].

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 [25].

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 [27].

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.

6.10.3. Aprendizaje por refuerzo

El aprendizaje por refuerzo (*RL*) es un tipo de aprendizaje automático en el que un agente aprende a realizar acciones en un entorno para maximizar una recompensa. El agente aprende por prueba y error, y es recompensado por realizar acciones que conducen a los resultados deseados. Por ejemplo, un agente de aprendizaje por refuerzo podría ser entrenado para jugar un juego como el ajedrez jugando contra sí mismo una y otra vez. El agente sería recompensado por ganar juegos y penalizado por perder juegos. Con el tiempo, el agente aprendería a jugar el juego de manera más y más efectiva [25].

6.11. Algoritmos jerárquicos

Los algoritmos jerárquicos son un tipo de algoritmo de agrupamiento que construyen una jerarquía de grupos, desde grupos pequeños hasta grupos más grandes. Estos algoritmos se utilizan comúnmente en el análisis de datos, ya que pueden proporcionar una comprensión más profunda de los datos que los algoritmos de agrupamiento no jerárquicos [28].

En los algoritmos jerárquicos, los datos se inicializan como grupos individuales. Luego, los grupos se combinan de manera iterativa, de acuerdo con un criterio de agrupación. El criterio de agrupación puede ser cualquier función que mida la similitud entre dos grupos [28].

Los algoritmos jerárquicos se pueden clasificar en dos tipos principales: algoritmos aglomerativos y algoritmos de división. Los algoritmos aglomerativos comienzan con grupos individuales y luego combinan los grupos hasta que queda un solo grupo. Los algoritmos de división comienzan con un solo grupo y luego dividen los grupos hasta que quedan grupos individuales [29].

Los algoritmos jerárquicos tienen una serie de ventajas sobre los algoritmos de agrupamiento no jerárquicos. En primer lugar, los algoritmos jerárquicos pueden proporcionar una comprensión más profunda de los datos, ya que muestran cómo los grupos están relacionados entre sí. En segundo lugar, los algoritmos jerárquicos son más robustos a los valores atípicos, ya que los valores atípicos pueden ser asignados a grupos individuales [29].

Sin embargo, los algoritmos jerárquicos también tienen algunas desventajas. En primer lugar, los algoritmos jerárquicos pueden ser más costosos computacionalmente que los algoritmos de agrupamiento no jerárquicos. En segundo lugar, los algoritmos jerárquicos pueden ser más difíciles de interpretar que los algoritmos de agrupamiento no jerárquicos [28].

6.11.1. Algoritmo Chameleon

El algoritmo Chameleon es un método de agrupamiento jerárquico que fue diseñado para analizar y agrupar grandes conjuntos de datos de alta dimensionalidad, comúnmente utilizados en aplicaciones como minería de datos, reconocimiento de patrones, entre otros. Este algoritmo se centra en la identificación de estructuras jerárquicas complejas dentro de los datos [30].

Chameleon se destaca por su capacidad para lidiar con conjuntos de datos dispersos y de alta dimensionalidad, lo que lo hace útil en contextos donde los datos pueden tener múltiples atributos o características. Utiliza una medida de similitud específica que se adapta dinámicamente a la densidad y la distribución de los datos en el espacio, lo que le permite identificar clústeres de diferentes formas y tamaños [30].

Una de las particularidades del algoritmo Chameleon es su capacidad para ajustar dinámicamente la medida de similitud a diferentes escalas en el espacio de características, lo que le permite capturar estructuras complejas y variadas dentro de los datos [30].

6.12. BIOPAC

Los dispositivos BIOPAC son una línea de productos de calidad avanzada para investigadores y educadores en ciencias de la vida. Proporcionan una forma flexible e integrada de registrar y analizar datos fisiológicos como ECG, EDA (GSR), EEG, EGG, EMG, EOG, y más de 300 otras señales y medidas fisiológicas [31].

6.13. VAT

VAT (*Visual Assessment of (Cluster) Tendency*) es un método para evaluar visualmente la tendencia a la agrupación de un conjunto de objetos. La tendencia a la agrupación se refiere al grado en que un conjunto de objetos puede dividirse de forma natural en grupos o clusters. El VAT es una herramienta útil para determinar si un algoritmo de agrupación debe aplicarse a un conjunto de datos concreto [32].

El método VAT consta de dos pasos:

- Reordenación de los objetos: Los objetos se reordenan de forma que los objetos similares se coloquen más cerca unos de otros.
- Creación de una imagen de intensidad: La matriz reordenada de disimilitudes entre objetos se muestra como una imagen de intensidad. Las agrupaciones se indican mediante bloques oscuros de píxeles a lo largo de la diagonal.

El método VAT es fácil de implementar y puede aplicarse a cualquier conjunto de datos que pueda representarse como un conjunto de vectores de objetos o valores de disimilitud

por pares. El método también es eficaz para identificar conglomerados en conjuntos de datos con ruido y valores atípicos [32].

La metodología descrita a continuación contempla el proceso para la obtención de señales bioeléctricas con el equipo BIOPAC proporcionado en la UVG. Las señales bioeléctricas recopiladas en la UVG o brindadas por HUMANA, serán utilizadas para la implementación y evaluación de algoritmos de aprendizaje automático para el análisis y reconocimiento de segmentos de interés.

7.1. Datos obtenidos con equipo *UVG*

En la Universidad del Valle de Guatemala se cuenta con los modelos MP36 y MP41 de BIOPAC, siendo el primero una versión más completa y de escritorio, mientras que el segundo es una versión portátil. Utilizando el equipo MP36 y MP41, se recopilaron señales del tipo EEG y EMG. Para cada tipo de señal bioeléctrica se tuvo un mínimo de tres sets diferentes de acciones por individuo.

7.1.1. Señales *EMG*

Para las señales EMG, el sujeto de prueba debía de encontrarse en estado de relajación. La pose del brazo en dicho estado se puede observar en la Figura 8. Las grabaciones tuvieron una duración mínima de un minuto. Las personas mantuvieron la pose de relajación durante 2 segundos y posteriormente, durante otros 2 segundos realizaron la actividad solicitada, esto iteradas veces hasta cumplir con el mínimo de tiempo establecido. Las actividades fueron:

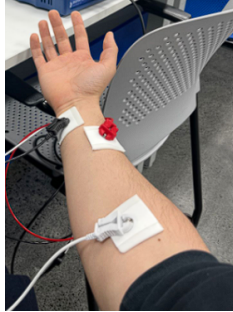


Figura 8: Pose de relajación para toma de señales EMG con equipo BIOPAC.



Figura 9: Pose de actividad 1 en toma de grabación señal EMG.

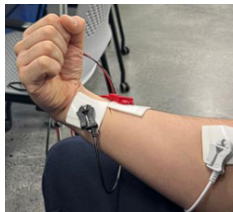


Figura 10: Pose de actividad 2 en toma de grabación señal EMG.



Figura 11: Pose de actividad 3 en toma de grabación señal EMG.

- Actividad 1: Elevar la muñeca hacia el exterior del cuerpo, como se observa en la Figura 9.
- Actividad 2: Contraer la muñeca hacia el interior del cuerpo, como se observa en la Figura 10.
- Actividad 3: Contraer paulatinamente cada uno de los dedos de la mano hacia el interior de la palma, como se observa en la Figura 11.

7.1.2. Señales *EEG*

Para las señales EEG, el sujeto de prueba debía de encontrarse en estado de relajación, lo cual implicaba ojos cerrados y haber realizado al menos 3 repeticiones de inhalación y exhalación profundas. Las grabaciones tuvieron una duración mínima de un minuto. Las personas mantuvieron la pose de relajación durante al menos 30 segundos al inicio, posteriormente se procedió a realizar una de las actividades específicas. Las actividades fueron:

- Actividad 1: Abrir los ojos durante 10 segundos y luego, volver a cerrar los ojos durante otros 10 segundos.
- Actividad 2: Abrir los ojos y realizar una prueba matemática nivel medio, la cual consistía en sumas, restas, multiplicaciones y divisiones de números enteros de hasta 7 dígitos.
- Actividad 3: En los trabajos de graduación presentados por Margareth Vela [33] y Oscar Fuentes [34], los cuales consistieron en el estudio cualitativo y cuantitativo del impacto de los pulsos binaurales en el estado de ánimo, concentración y calidad del sueño de las personas y aplicación de técnicas de aprendizaje automático y reconocimiento de patrones en las señales bioeléctricas. Se realizaron las actividades 1 y 2 nuevamente pero ahora, mediante audífonos se aplicaron pulsos binaurales.
- Actividad 4: Esta expresión involucra activar los músculos ubicados en la frente con el fin de levantar las cejas de forma rítmica y deliberada, tal como se evidencia en la Figura 12.
- Actividad 5: Inclinación de la cabeza hacia abajo. Este movimiento implica girar la cabeza hacia abajo, en dirección al pecho o hacia el suelo, como se ilustra en la Figura 13.
- Actividad 6: Inclinación de la cabeza hacia arriba. Este movimiento implica girar la cabeza hacia arriba, en dirección contraria al pecho, como se ilustra en la Figura 14.
- Actividad 7: Inclinación de la cabeza. Este movimiento consiste en girar la cabeza hacia la derecha, como se ilustra en la Figura 15.

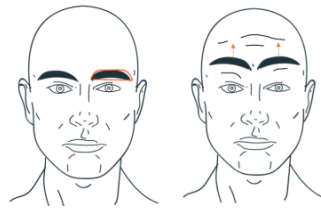


Figura 12: Actividad 4 movimiento controlado de cejas.



Figura 13: Actividad 5 inclinación de la cabeza hacia abajo.



Figura 14: Actividad 6 inclinación de la cabeza hacia arriba.



Figura 15: Actividad 7 inclinación de la cabeza hacia el lateral derecho.

7.1.3. Estándar en recolección de datos BIOPAC

Se definió un documento estándar de formato “.xlsx” para el almacenamiento de las señales EMG y EEG obtenidas con el equipo BIOPAC, como se observa en la Figura 16. Los datos principales que se necesitan son:

	A	B	C	D	E	F	G	H	I	J	K	L
1	t	Signal	Integrada	Alpha	Beta	Delta	Theta	Fs	T	N		t/N
2	0.00000	4.79736		-0.32511	-0.11804	4.68147	0.468235	200.0217	50.74450	10150		0.004999
3	0.00500	4.54102		0.075761	-0.05618	4.63558	0.095997					
4	0.01000	4.6875		0.387985	-0.50445	4.59246	-0.24966					
5	0.01500	4.20532		0.580913	-0.88796	4.55155	-0.5707					
6	0.02000	3.63159		0.65386	-0.71308	4.51163	-0.8697					
7	0.02500	4.23584		0.631016	-0.05307	4.47093	-1.14846					

Figura 16: Estándar para el almacenamiento de datos obtenidos con el equipo BIOPAC de la UVG.

- Señal (signal)

- Tiempo de grabación (T)
- Número de muestras (N)

La primer columna llamada “t” se auto genera al llenar los datos principales y consiste en el tiempo en el que se obtuvo la señal que se encuentra en la columna “*Signal*”. El parámetro que se encuentra en la columna “Fs” corresponde a la frecuencia de muestreo la cual se auto calcula. El parámetro que se encuentra en la columna “t/N” corresponde al tiempo dividido numero de muestras la cual se auto calcula, esto permite marcar la diferencia de tiempo que hay entre una muestra y la siguiente.

Las columnas siguientes son opcionales ya que dentro de la herramienta de software para el estudio de la epilepsia no se hace uso de ellas. Por lo que siempre habrá al menos una de estas columnas en blanco, ya que se llenan acorde al tipo de señal bioeléctrica a recolectar. Estas columnas son:

- Integrada: es el área situada bajo la curva de la señal EMG rectificada, lo que equivale a la integral matemática del valor absoluto de la señal EMG original [31].
- Alpha: señal EEG filtrada para observar frecuencias de 8 Hz a 12 Hz.
- Beta: señal EEG filtrada para observar frecuencias de 12 Hz a 30 Hz.
- Delta: señal EEG filtrada para observar frecuencias de 1 Hz a 4 Hz.
- Theta: señal EEG filtrada para observar frecuencias de 4Hz a 8 Hz.

Tras recolectar estos datos la norma de almacenamiento consiste en, nombrar el documento con el nombre y apellido de la persona a la que se le extrajo las señales bioeléctricas, seguido de un guion (bajo o normal), el tipo de actividad que se realizó y por último, el tipo de señal bioeléctrica que se extrajo. En la Figura 17 se observa un ejemplo de ello.

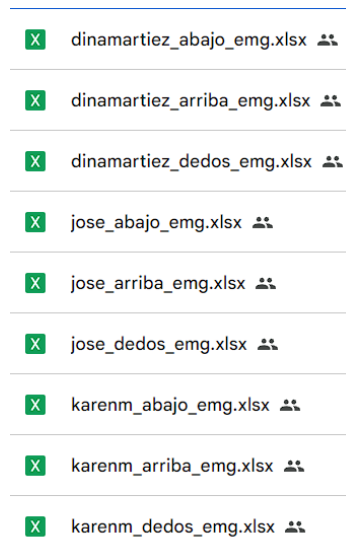


Figura 17: Señales bioeléctricas extraídas con el equipo BIOPAC almacenadas según la norma establecida.

7.1.4. Estructuración de los datos para ser usados en la herramienta de software para el estudio de la epilepsia

Para los datos obtenidos con el equipo BIOPAC de la UVG, es necesario convertirlos en un formato tipo *Struct*, de esta manera se podrá usar dicha información dentro de la herramienta de software para su análisis y estudio, como se muestra en la Figura 18.

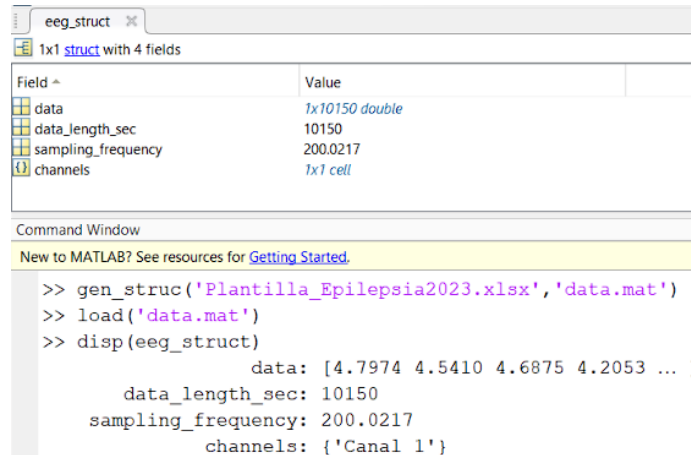


Figura 18: Contenido de *struct* para el uso de datos dentro de la herramienta de software para el estudio de la epilepsia.

Para ello se creó la función “gen_struct” la cual, posee como primer argumento el nombre del documento a extraer los datos y como segundo argumento el nombre y el formato con el que se almacenará el *struct*. Dicho formato contiene los siguientes espacios:

- Datos (mV)
- Cantidad de datos
- Frecuencia de muestreo (Hz)
- Cantidad de canales

7.2. Datos de HUMANA de pacientes con epilepsia

HUMANA ha compartido 4 grabaciones de señales EEG, la información de dichas grabaciones se puede ver en el Cuadro 1. Estos datos fueron de ayuda para el entrenamiento y verificación de exactitud de los modelos de aprendizaje automático.

Nombre	Canales	Duración (hh:mm:ss)	Frecuencia de muestreo
AL.edf	33	03:01:56	200 Hz
CLEA.edf	29	02:39:42	200 Hz
GIKA.edf	29	02:58:06	200 Hz
HCHC.edf	29	23:02:44	200 Hz
Ajczalar M.edf	30	15:57:13	300 Hz
GUADRON T.edf	30	03:02:01	300 Hz

Cuadro 1: Información de grabaciones de señales bioeléctricas de pacientes con epilepsia brindadas por HUMANA.

7.3. Agrupamiento de datos

Con el equipo UVG se recopilaron señales bioeléctricas del tipo EEG y EMG, las cuales se agruparon de forma Intersujeto e Intrasujeto respectivamente. En el caso de señales brindadas por HUMANA estas no necesitan agrupación, ya que por lo general son de una duración mayor a una hora de grabación contando como mínimo 29 canales, por lo que esto ya son datos suficientes para proceder a realizar su análisis.

El análisis de señales EEG de forma intersujeto y señales EMG de forma intrasujeto se realiza de esta manera debido a las diferencias fundamentales en la naturaleza de estas dos tipos de señales y los objetivos específicos de análisis en cada caso.

7.3.1. Análisis de señales *EEG* de forma intersujeto

Las señales EEG, que representan la actividad eléctrica del cerebro, tienden a mostrar una variabilidad significativa entre diferentes individuos. Las diferencias en la anatomía cerebral, la disposición de electrodos, la edad y otros factores pueden influir en las características de las señales EEG. Por lo tanto, se suele analizar de forma intersujeto para comprender cómo varían las respuestas entre diferentes personas.

El análisis intersujeto de las señales EEG es relevante en aplicaciones clínicas y de investigación que involucran poblaciones de pacientes o participantes diversos. Permite identificar patrones generales en grupos de personas y puede ayudar en la detección de trastornos neurológicos como la epilepsia en una población más amplia.

7.3.2. Análisis de señales *EMG* de forma intrasujeto

Las señales EMG, que registran la actividad eléctrica de los músculos, tienden a mostrar menos variabilidad intrasujeto, es decir, las características de la señal son relativamente consistentes en un individuo específico a lo largo del tiempo. Esto se debe a que la anatomía y la disposición de los músculos de una persona tienden a ser estables.

7.4. Aprendizaje automático

En la presente sección se presentan los algoritmos y métodos que se usaron para medir la efectividad de los algoritmos en el reconocimiento de patrones en señales bioeléctricas que sean de interés.

7.4.1. Extracción de características

Se extrajeron características en el dominio del tiempo, la frecuencia y Wavelets de cada uno de los conjuntos de datos obtenidos mediante el equipo UVG o proporcionados por HUMANA. Se aplicó la transformada a cada dominio a las señales, seguida de la extracción de características. A partir de este punto, estas características obtenidas mediante las transformadas serán referidas simplemente como “características frecuencia”, “características tiempo” o “características wavelets” por conveniencia. Este proceso de extracción de características fue analizado detalladamente con el objetivo de optimizar el tiempo de entrenamiento de los algoritmos de aprendizaje automático y, en la medida de lo posible, mejorar la clasificación de las señales bioeléctricas.

7.4.2. Anotaciones automáticas

Empleando los algoritmos de aprendizaje automático desarrollados en las fases anteriores, se clasificó la señal bioeléctrica entre, “Ictal”, “Sano”, “Interictal” o Preictal”, según fuera su caso. A su vez, se implemento una sección para exportar los segmentos de interés a un documento CSV, como se puede observar en la Figura 19.

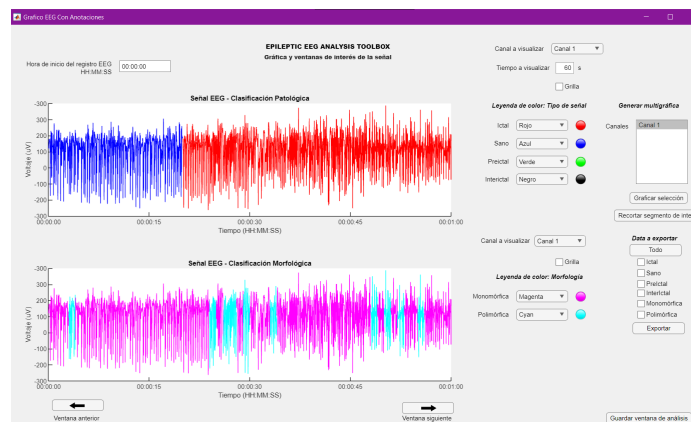


Figura 19: Ventana de anotaciones automáticas con nueva sección para exportar datos de interés.

7.4.3. Análisis estadístico

Mediante matrices de confusión se validó la efectividad de las redes neuronales y las SVM (algoritmos de aprendizaje supervisado). Para el caso de los algoritmos de agrupamiento de *K-means* y Jerárquico (algoritmos de aprendizaje no supervisado) se procedió a validar

```

EL usuario seleccionó C:\Users\Cris Pat\Desktop\U\Datos HUMANA\Señales HUMANA\AL.edf
Step 1 of 2: Reading requested records. (This may take a few minutes)...
Step 2 of 2: Parsing data...
Unrecognized method, property, or field 'cla' for class 'matlab.ui.control.UIAxes'.

Error in procesamiento/VisualizarButtonPushed (line 175)
    app.UIAxes_2.cla; %(CP)

Error in matlab.apps.AppBase>@(source,event)executeCallback(ams_app_callback_requires
    newCallback = @(source, event)executeCallback(ams, ...

Error while evaluating Button PrivateButtonPushedFcn.

```

Figura 20: Error al utilizar botón de visualización de canales en la ventana de anotaciones automáticas.

```

Warning: Integer operands are required for colon operator
when used as index.
> In FeaturesMorf (line 48)
In inicioanotacionesauto/ProcesarlasealEEGButtonPushed
(line 287)
In
matlab.apps.AppBase>@(source,event)executeCallback(ams_app_callback_requiresEventData,event)
(line 62)

```

Figura 21: Alerta al emplear el algoritmo FeaturesMorf.

su efectividad mediante el algoritmo “*rand index*”, el cual es una medida de validez para algoritmos de agrupamiento. Además se tabularon tiempos relevantes, desde la extracción de características hasta la generación de anotaciones automáticas.

7.5. Actualización de la herramienta de software para el estudio de la epilepsia

El primer paso para la actualización de la herramienta de software para el estudio de la epilepsia fue eliminar los mensajes de advertencia y arreglar errores que se encontraban dentro de la herramienta de software para el estudio de la epilepsia. En las Figuras 20 y 21 se puede observar algunos ejemplos, la mayoría de estos eran a causa de prácticas de programación no tan eficientes en MATLAB.

En todos los casos al momento de extraer las características de las señales bioeléctricas, se tenía que no se pre-creaba el vector donde se almacenarían dichas características, además de contar con variables y condicionales innecesarias, como se puede observar en la Figura 22. Esto resulta ser computacionalmente costoso, lo que da paso a pérdida de tiempo innecesaria. Por lo que se procedió a corregirlo, creando el vector de características con las dimensiones máximas que podría tener, siendo la cantidad de filas del tamaño del vector de datos y la cantidad de columnas, del tamaño de las posibles características de interés, como se puede observar en la Figura 23.

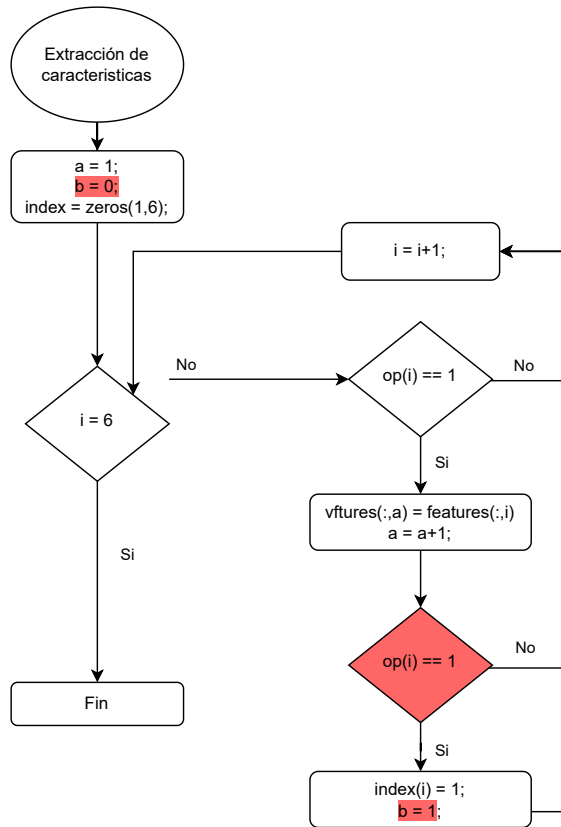


Figura 22: Flujograma de la creación del vector de características versión 2022.

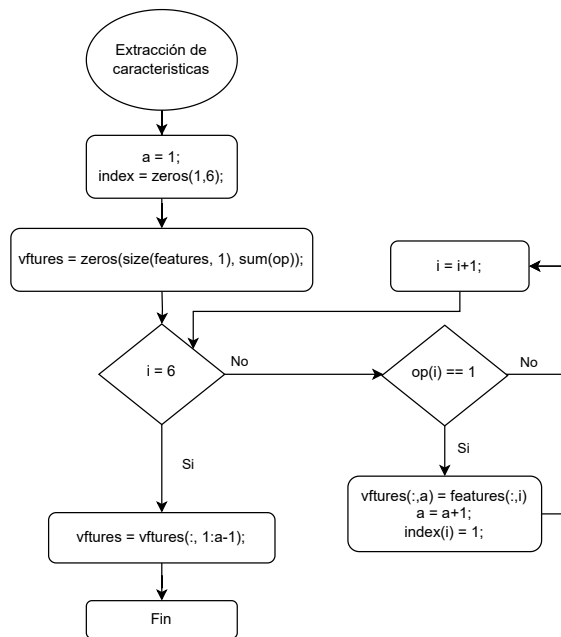


Figura 23: Flujograma de la creación del vector de características versión 2023.

Al momento de establecer rangos de datos a analizar por ventanas, se encontraba como

limite superior la expresión de la Ecuación(1), lo cual al tratarse de un vector los valores ingresados para los índices deben ser enteros, característica que no cumple el limite superior cuando la variable “muestras” es un número impar. Por lo que se procedió a redondear el valor al mínimo más próximo, dando la expresión de la Ecuación(2). Un claro ejemplo de este proceso se puede observar en la Figura 24.

$$Lim_sup = \frac{muestras}{2} + 1 \quad (1)$$

$$VectorVentana = VectorAnalizar(1 : floor(\frac{muestras}{2} + 1)); \quad (2)$$

```

%Cálculo de las características de morfología
WS = fft(ventana_interes);
WSn = abs(WS/muestras);
WS1_1=WSn(1:floor(muestras/2+1)); %corroborar
WS1_1(2:end-1) = 2*WS1_1(2:end-1);
WS1_1 = (WS1_1).^2; %Espectro de potencia

```

Figura 24: Segmento de código para el calculo de características de una señal bioeléctrica.

En el proceso de generación de característica, el algoritmo para determinar el “Zero Crossing” (ZC) no era el más adecuado, dicho algoritmo se puede ver en la Figura 25. El algoritmo creaba variables que no se utilizaban, sin embargo, si se realizaban operaciones matemáticas con ellos, lo que implica tiempo y calculo computacional innecesario. A su vez, el algoritmo verificaba el cambio de signo directamente, sin usar el argumento “umbral”, el cual es de gran relevancia ya que es este, quien permite el conteo adecuado de cambio de signo, sin tomar en cuenta aquellos cruces por cero que ocurren debido al ruido dentro de la señal. Por lo que se procedió a realizar las respectivas correcciones, como se muestra en la Figura 26. En la presente versión se toma en cuenta el umbral para un correcto conteo, además de, ser una versión simplificada lo que lo hace más simple de realizar los cálculos computacionales.

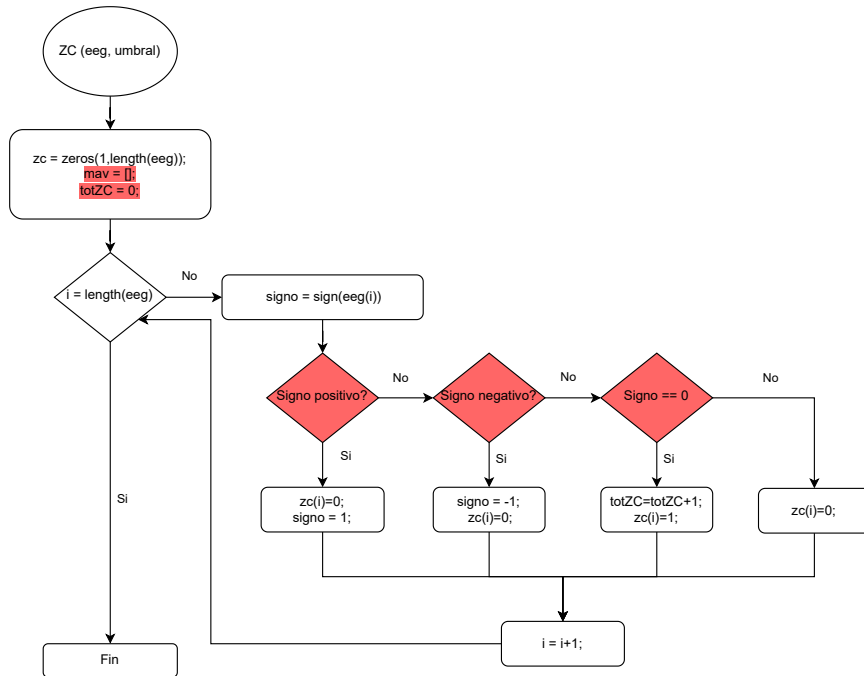


Figura 25: Flujograma de algoritmo ZC versión 2022.

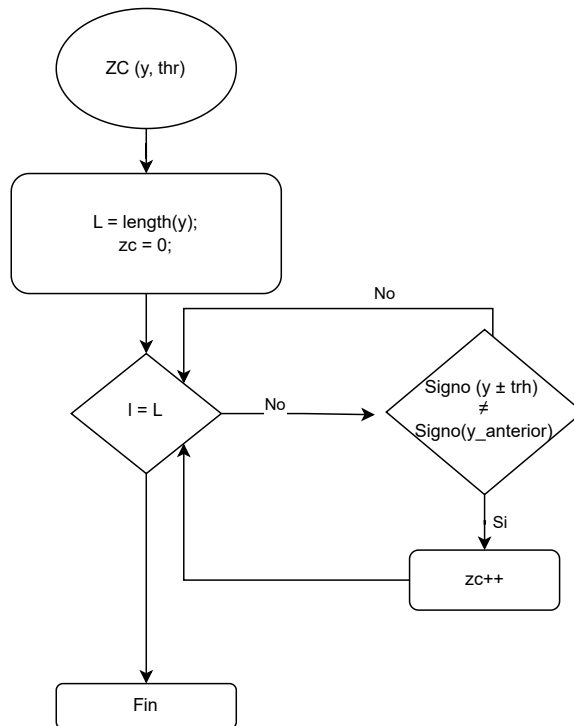


Figura 26: Flujograma de algoritmo ZC versión 2023.

Resultados de recolección de datos

En el presente capítulo se presentarán los datos de señales bioeléctricas obtenidas con el equipo de UVG y brindadas de HUMANA. Estas señales fueron la materia prima para la comprobación de funcionalidad de los algoritmos. En el capítulo de anexos, se incluyen los vínculos correspondientes a los datos recopilados.

8.1. Señales bioeléctricas obtenidas con el equipo de *UVG*

La recolección de señales bioeléctricas con el equipo de UVG ha dado como resultado un total de 187 grabaciones, como se puede ver en el Cuadro 2. Para el caso de las señales EEG, se cuenta con una duración promedio de 20 minutos por grabación. En el caso de las señales EMG, se cuenta con una duración promedio de 1 minuto por grabación. Por lo que ahora se cuenta con una base de datos de Señales EEG y EMG, recolectadas y procesadas en la Universidad del Valle de Guatemala.

En el Cuadro 2 de datos recolectados, la categoría “sin plantilla” se refiere a grabaciones de señales bioeléctricas que no se guardaron con el “Estándar en recolección de datos BIOPAC”, mientras que la categoría “con plantilla” hace referencia a grabaciones de señales bioeléctricas que sí utilizaban dicho estándar.

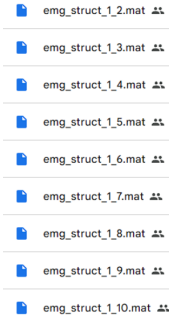


Figura 27: Structs generados a partir de los datos obtenidos con el equipo UVG.

Tipo	Cantidad	Formato
EEG sin plantilla	34	XLSX
EEG con plantilla	18	XLSX
EMG sin plantilla	18	XLSX y CSV
EMG con plantilla	117	XLSX

Cuadro 2: Datos recolectados con equipo UVG.

Una vez organizado los datos obtenidos con el equipo de UVG, se procedió a generar los *structs*, ya que este formato es el utilizado en la herramienta de software para el estudio de la epilepsia, siguiendo la norma que se describió en el capítulo anterior se almacenaron, se puede observar en la Figura 27. Este formato permite extraer características de las grabaciones realizadas.

8.1.1. Resultados de la extracción de características

Para los datos recolectados se procedió a extraer sus características. Cabe mencionar que los datos que se obtuvieron con el equipo UVG son de personas que no padecen de epilepsia, teniendo un total de 10 vectores de características y 24 *structs*, como se observa en el Cuadro 3.

Tipo	Cantidad	Formato
<i>Struct</i> de datos con señales bioeléctricas	24	MAT

Cuadro 3: Datos procesados de las grabaciones que se recolectaron con equipo UVG.

Para la generación de los vectores de características se cargaron solo 2 sets de datos, los de HUMANA tienen segmentos ictales (crisis) y no ictales (no crisis). Posteriormente se procedió a generar los vectores de características en tiempo continuo y wavelets, con todas las características seleccionadas, como se puede ver en las Figuras 28 y 29. Considerando que las grabaciones de HUMANA tienen una duración mayor a 2 horas, como se observa en el Cuadro 1, los tiempos en los que se extrajo las características fue favorable, como se

puede ver en el Cuadro 4.

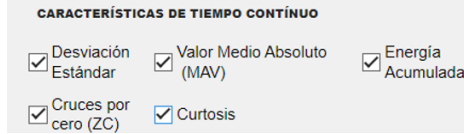


Figura 28: Características en el dominio del tiempo continuo.

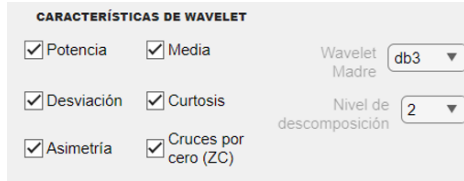


Figura 29: Características con wavelets.

En cuanto a las características en el dominio de la frecuencia, la primera corrida se intento recolectar todas las características a la vez, como se puede observar en la Figura 30; pero el proceso se interrumpió al transcurrir las 6 horas y no finalizar. Por lo que se procedió a extraer las características una a la vez, obteniendo periodos de duración muy grandes en comparación a las características en el dominio del tiempo y wavelets.

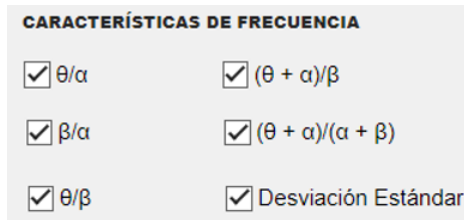


Figura 30: Características en el dominio de la frecuencia.

Tipo	Hora inicio (hh:mm:ss)	Hora fin (hh:mm:ss)	Duración (hh:mm:ss)
Tiempo continuo	14:54:00	15:33:07	00:39:07
Wavelet	15:35:00	15:35:32	00:00:32
Frecuencia	11:18:00	11:09:32	23:50:28
θ/α	17:10:25	18:26:52	01:16:27
β/α	18:30:00	19:34:27	01:04:27
θ/β	19:38:00	20:44:07	01:06:07

Cuadro 4: Tiempos de extracción de características con señales de HUMANA y UVG.

8.2. Anotaciones automáticas

El espacio de anotaciones automáticas necesita ser cargado con un set de datos provenientes de una señal bioeléctrica; un clasificador el cual es el encargado de segmentar y así poder diferenciar el tipo de estado en el que se encuentra el paciente (Sano, ictal, preictal, interictal) y posteriormente se debe seleccionar los canales a analizar, como se observa en la Figura 31.

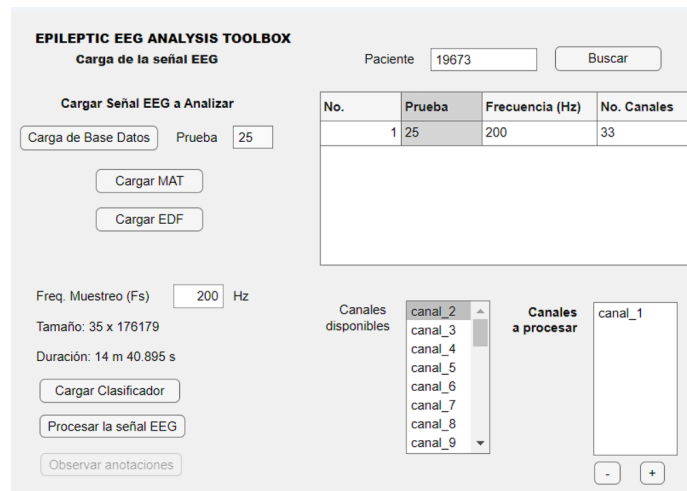


Figura 31: Ventana de inicio para generar anotaciones automáticas.

Para lo que se procedió a comprobar los tiempos de anotación según el clasificador, como se aprecia en el Cuadro 5. Como se puede observar nuevamente las wavelets son las características de menos duración, en este caso para la generación de anotaciones.

Clasificador	Duración (hh:mm:ss)	Grabación
RNA Tiempo	01:27:00	AL.edf
RNA Frecuencia	01:07:00	AL.edf
RNA Wavelet	00:12:00	AL.edf

Cuadro 5: Tiempos de entrenamiento para clasificadores con señales bioeléctricas de HUMANA y UVG.

En las Figuras 32 y 33, se observa la misma grabación brindada por HUMANA, la cual trata de una persona que tuvo iteradas ocasiones episodios epilépticos. La razón por la que visualmente difiere una de la otra es debido a la clasificación realizada, de lo cual se habla en los párrafos siguientes.

En la grabación de las Figuras 32 y 33, se percibe cambios abruptos, lo cual permite una diferencia visual de segmentos. Los segmento donde se encuentra una mayor densidad de oscilaciones y altas amplitudes (visualmente segmento más grueso), se debe a disparos de actividad bioeléctrica en el cerebro más movimiento físico debido a una convulsión. En cuanto

a los segmento donde se percibe menos oscilaciones a una baja amplitud, son segmentos donde la persona no se encuentra en actividad ictal.

Es en esta parte donde se puede notar la necesidad de una mayor cantidad de señales bioeléctricas por parte de HUMANA, ya que en la Figura 32, se muestra el resultado para la generación de anotaciones automáticas. Es casi nulo a simple vista, los segmentos que marcan actividad ictal (color rojo) versus los segmentos en estado sano (color azul). Esto en gran medida se debe a un fuerte sesgo de una mayor cantidad de datos EEG de personas sanas, que de personas con episodios epilépticos.

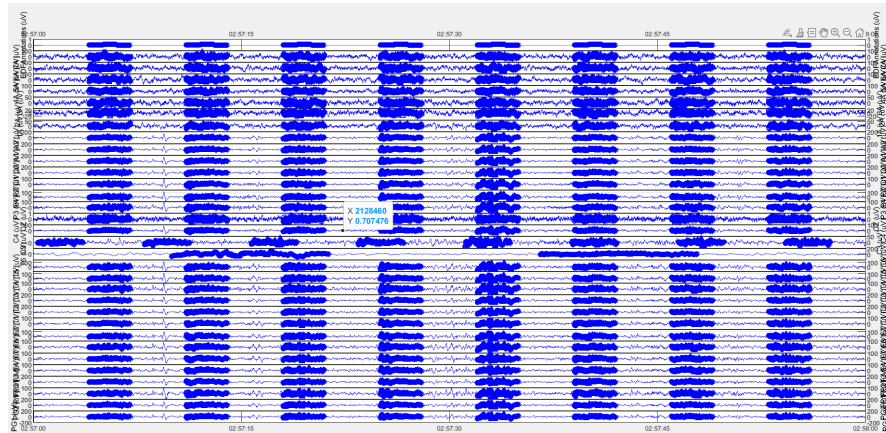


Figura 32: Anotaciones automáticas de 33 canales analizados.

Para solventar el inconveniente anterior sin contar con más datos por parte de HUMANA, se entreno al modelo con una menor cantidad de datos de personas sin actividad epiléptica. Desafortunadamente el resultado no fue el esperado. Como se puede observar en la Figura 33, ahora se contaba con el inconveniente que era mayor el sesgo por los datos de actividad epiléptica respecto a los segmentos en estado sano.



Figura 33: Anotaciones automáticas de 33 canales analizados con menor cantidad de EEG de personas sin actividad ictal.

Para solucionar este inconveniente, lo adecuado será entrenar el modelo con una cantidad

similar de datos de señales bioeléctricas, además de, contar con la misma cantidad de personas sanas y personas con episodios epilépticos de las que se extraen las señales bioeléctricas. Esto permitirá al modelo encontrar las características necesarias de cada estado de interés, sin importar el sujeto. Ya que actualmente el modelo es preciso solo con las mismas personas con las que se entreno el modelo. Esto quiere decir que la predicción del modelo se ve afectada si, las señales bioeléctricas provienen de personas distintas de las que se obtuvieron los datos con los que se entrenó el modelo.

Resultados estadísticos

En este capítulo, se presentan los resultados estadísticos obtenidos a partir de la aplicación de diversos métodos de análisis en el estudio de las señales bioeléctricas. Se analizan los resultados de los clasificadores empleados, se presentan las Matrices de confusión que reflejan la capacidad de predicción, y se evalúan los resultados de los algoritmos de *clustering* mediante el algoritmo *rand index*.

9.1. Clasificadores

Los clasificadores empleados para esta etapa mediante aprendizaje supervisado fueron las RNA y las SVM, dando resultados positivos.

Las SVM presentaron una mejor clasificación de los segmentos de las señales bioeléctricas, como se puede observar en las Figuras 34, 35 y 36. Estando por encima de un 85 % de exactitud. Presentando que las características de mejor comportamiento para clasificación son las “wavelets” seguidas de las características de “tiempo continuo” y por ultimo las características de “frecuencia”.

Las RNA por su parte, también produjeron resultados positivos con ligera diferencia respecto a las SVM, como se puede observar en las Figuras 37, 38 y 39. Las cuales también se encuentran por encima del 85 % de exactitud. Se percibe nuevamente que las mejores características para la clasificación de señales bioeléctricas son “wavelets”, en cuanto a las características de “tiempo continuo” se nota una pérdida del 0.4 % de exactitud respecto a la clasificación por SVM y una pérdida del 0.2 % de exactitud empleando las características de “frecuencia” respecto a la clasificación por SVM.

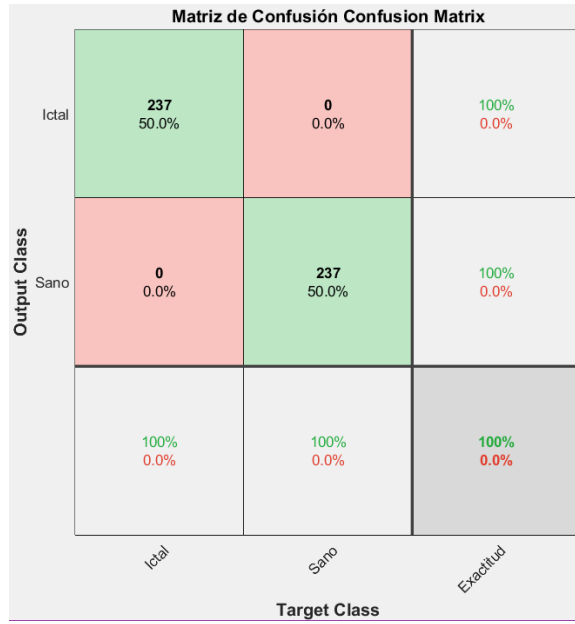


Figura 34: Matriz de confusión para las clases Ictal y Sano utilizando SVM con características wavelets.

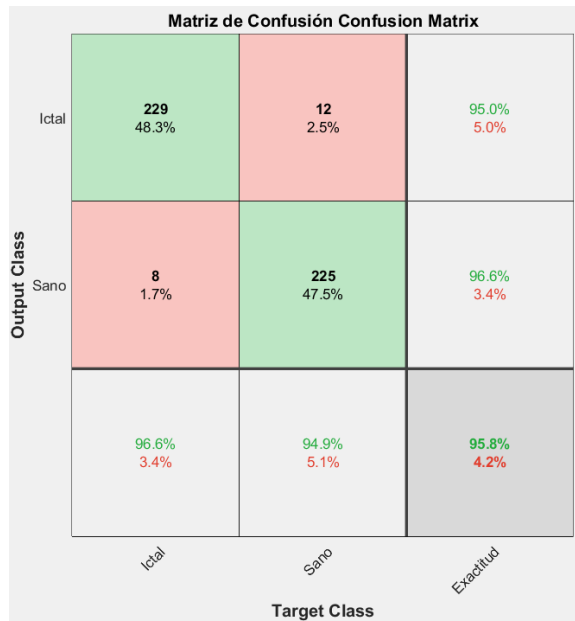


Figura 35: Matriz de confusión para las clases Ictal y Sano utilizando SVM con características en tiempo continuo.

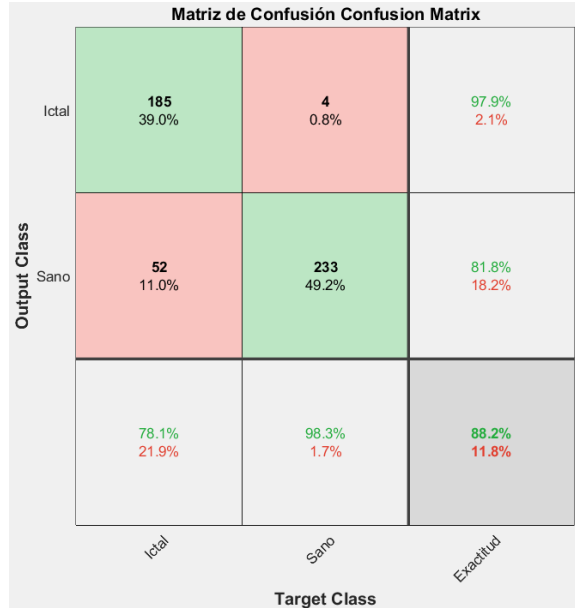


Figura 36: Matriz de confusión para las clases Ictal y Sano utilizando SVM con características frecuencia.

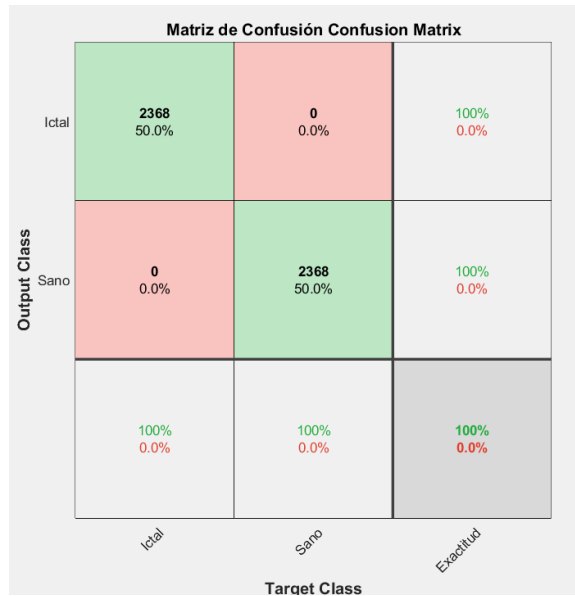


Figura 37: Matriz de confusión para las clases Ictal y Sano utilizando RNA con características wavelets.

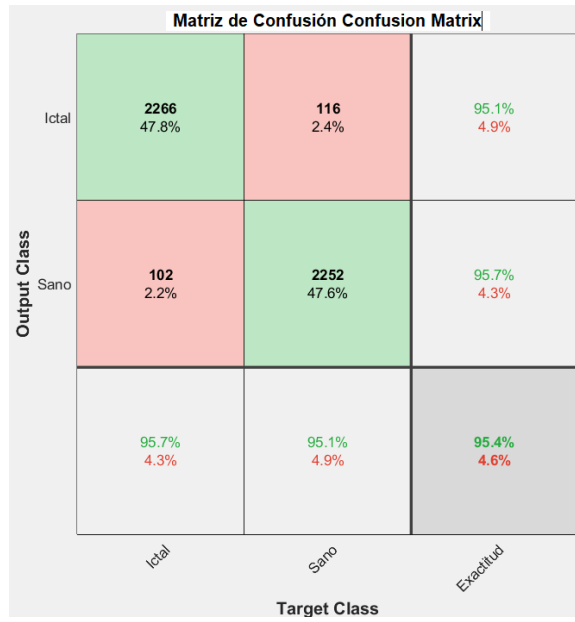


Figura 38: Matriz de confusión para las clases Ictal y Sano utilizando RNA con características en tiempo continuo.

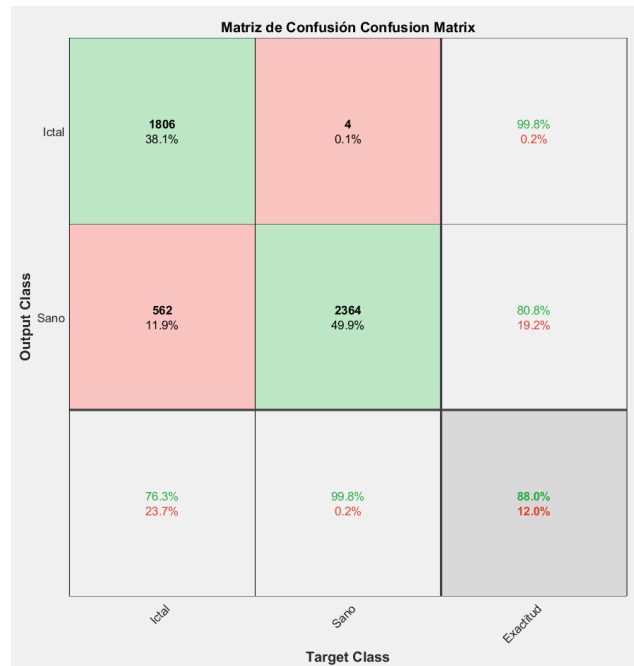


Figura 39: Matriz de confusión para las clases Ictal y Sano utilizando RNA con características en frecuencia.

9.2. Algoritmos de agrupamiento

Los algoritmos de agrupamiento empleados para esta etapa mediante aprendizaje no supervisado fueron *K-means* y jerárquico, obteniendo resultados positivos.

Como se puede observar en el Cuadro 6, las características “wavelets” y “tiempo continuo” independientemente del algoritmo de agrupación se obtienen los mismos resultados, sin embargo, para las características de “frecuencia” si se nota una gran mejora de clasificación con el algoritmo de agrupamiento jerárquico dando un 30.46 % de superioridad versus el algoritmo de agrupamiento *K-means*.

Por lo que nuevamente las características “wavelets” son las que presentan mejores resultado para la clasificación de segmentos de interés en señales bioeléctricas. Para las características en “tiempo continuo” se puede mejorar el porcentaje de clasificación obteniendo una mayor cantidad de características. En cuanto a las características de “frecuencia” si es notorio una mejora para el algoritmo de agrupamiento jerárquico. Las características en el dominio de la frecuencia a menudo tienen una estructura más jerárquica y pueden exhibir patrones de similitud a diferentes escalas. El agrupamiento jerárquico se adapta bien a la estructura jerárquica de los datos y puede identificar grupos en diferentes niveles de detalle, lo que puede ser beneficioso en datos con patrones complejos.

Algoritmo de agrupamiento	Característica	Rand index
K-means	Wavelets	100.00 %
K-means	Tiempo continuo	70.68 %
K-means	Frecuencia	50.01 %
Jerárquico	Wavelets	100.00 %
Jerárquico	Tiempo continuo	70.68 %
Jerárquico	Frecuencia	91.55 %

Cuadro 6: Porcentaje de validación para los algoritmos de agrupamiento mediante *rand index*.

El análisis de grupos (*clustering*) desempeña un papel fundamental en la investigación, ya que permite explorar la estructura subyacente de las señales bioeléctricas recopiladas. En este capítulo, se presentan los resultados del uso de algoritmos de agrupamiento, en particular el método de K-means y jerárquico, aplicados a las señales bioeléctricas en diferentes dominios: tiempo, frecuencia y wavelets. Cabe mencionar que los datos utilizados son el resultado de una época de 0.4 segundos, por lo que la densidad de muestras cambiara según la frecuencia con la que se haya grabado la señal bioeléctrica para las clases ictal, sano, preictal e interictal.

En el enfoque de análisis, se aplicó el método de *clustering* utilizando todas las características extraídas de las señales bioeléctricas. La decisión de emplear el conjunto completo de características se basa en la premisa de que cada característica contribuye de manera única a la representación de la información en los datos. Utilizando todas las características, se busca capturar la complejidad y la variabilidad completa presente en las señales.

Posteriormente, para la visualización de los resultados, se graficaron pares específicos de características seleccionadas. Dada la naturaleza multidimensional de los datos, la representación gráfica de las características se limita a graficar pares específicos, ya que la visualización directa de espacios de características con más de tres dimensiones se vuelve inherentemente compleja. La elección de graficar pares específicos permite explorar las relaciones bidimensionales entre características seleccionadas, ofreciendo una perspectiva visual que facilita la interpretación de patrones y tendencias. A pesar de la limitación en la representación tridimensional, la estrategia de graficar pares específicos proporciona una visión detallada de las interacciones entre características, contribuyendo a la comprensión global de la estructura de los datos y respaldando el análisis de los resultados obtenidos mediante los algoritmos de *clustering* aplicados al conjunto completo de características.

En la última sección de este capítulo se hablará acerca de la incorporación de la herramienta VAT, la cual ha resultado fundamental para abordar el desafío de visualizar las

clases en un espacio multidimensional sin perder información relevante debido a la alta dimensionalidad de los datos. VAT proporciona una representación visual eficaz al reducir la complejidad de los datos, permitiendo la observación de patrones y relaciones entre clases. Esta herramienta ha facilitado la identificación de estructuras y la evaluación de la tendencia de agrupamiento en el conjunto de datos, mejorando así la comprensión y la interpretación de los resultados de los algoritmos de clustering aplicados en contextos de alta dimensionalidad.

10.1. Análisis de grupos en el dominio de tiempo

En esta sección, se presenta el análisis de grupos aplicado a las características extraídas en el dominio del tiempo. Las características que se extrajeron en el dominio del tiempo fueron:

1. Desviación estándar
2. Valor medio absoluto (MAV)
3. Cruces por cero (ZC)
4. Curtosis
5. Energía acumulada

10.1.1. Resultados K-means

Los resultados del proceso de agrupamiento en el dominio de tiempo revelan una interesante distribución de los datos. En particular, como se observa en las Figuras 40 y 41 la distribución de los datos son muy continuos, lo cual indica que las características 1 respecto de 2 y 1 respecto de 4 podrían no ser muy efectivas para la predicción del tipo de señal bioeléctrica. Mientras que en las Figuras 42, 43, 44 y 45. Los datos tienden a concentrarse de manera significativa en uno de los grupos, mientras que el otro grupo muestra una menor concentración de datos.

La gran concentración de datos en un grupo y la menor concentración en el otro grupo indican que las señales bioeléctricas pueden categorizarse de manera efectiva en función de estas características en el dominio del tiempo. Esto sugiere la existencia de patrones distintivos que permiten la separación de las señales en dos grupos significativos. Aunque cabe destacar que es aquí donde se percibe la necesidad de una mayor cantidad de datos de pacientes con epilepsia, lo que permitiría tener dos grupos igual de densos en datos.

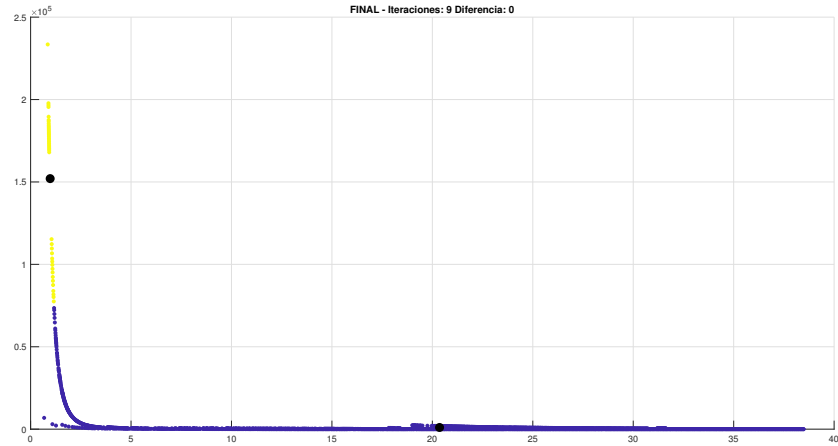


Figura 40: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 2 en el dominio del tiempo de actividad ictal y no ictal.

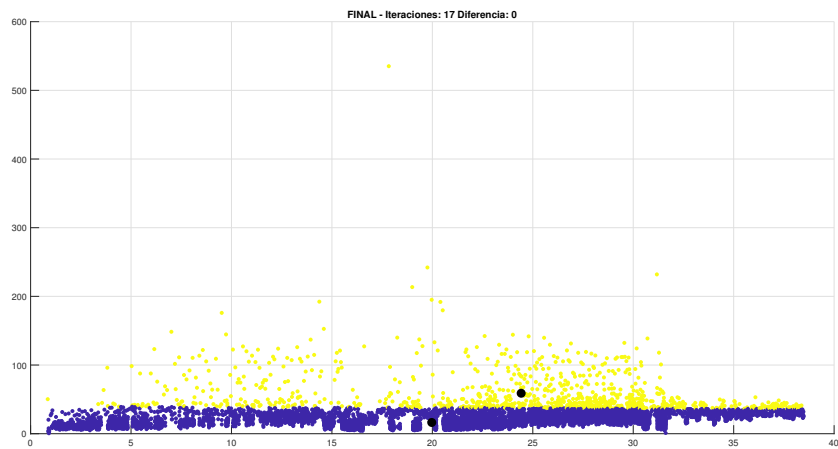


Figura 41: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 4 en el dominio del tiempo de actividad ictal y no ictal.

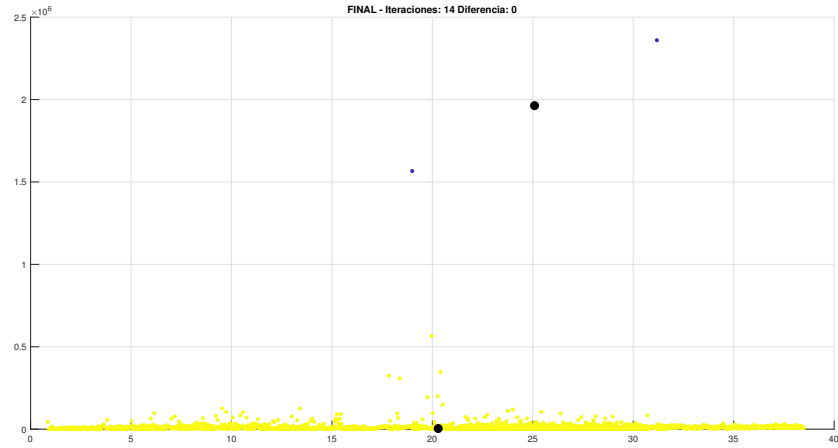


Figura 42: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 5 en el dominio del tiempo de actividad ictal y no ictal.

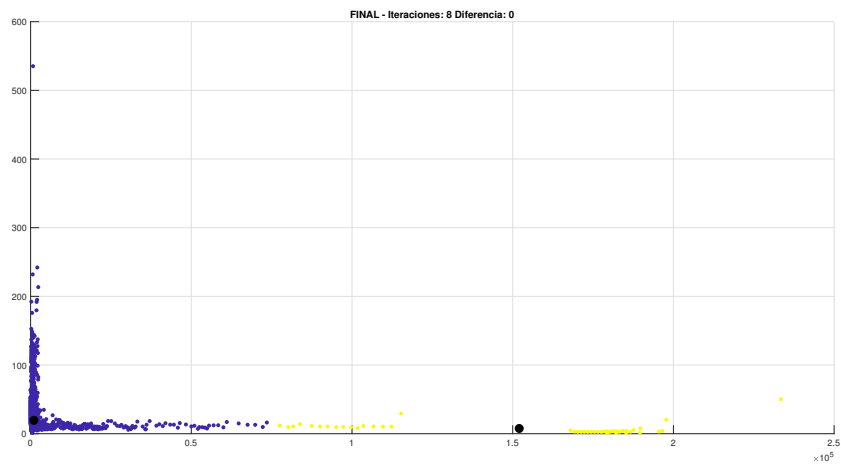


Figura 43: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 4 en el dominio del tiempo de actividad ictal y no ictal.

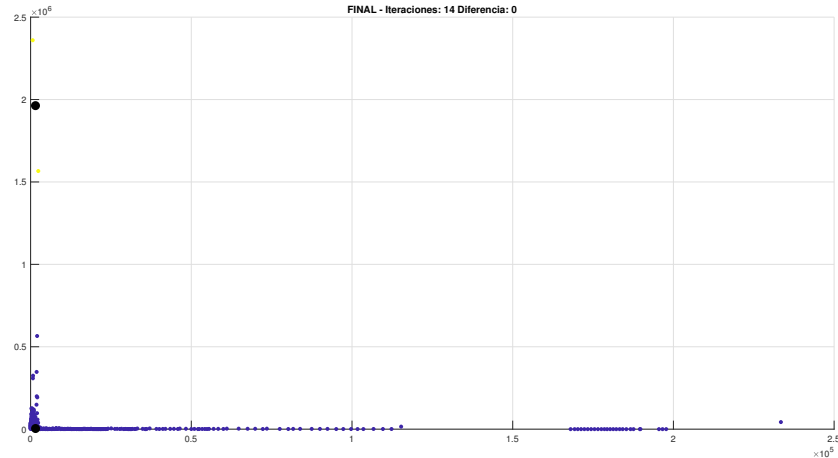


Figura 44: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 5 en el dominio del tiempo de actividad ictal y no ictal.

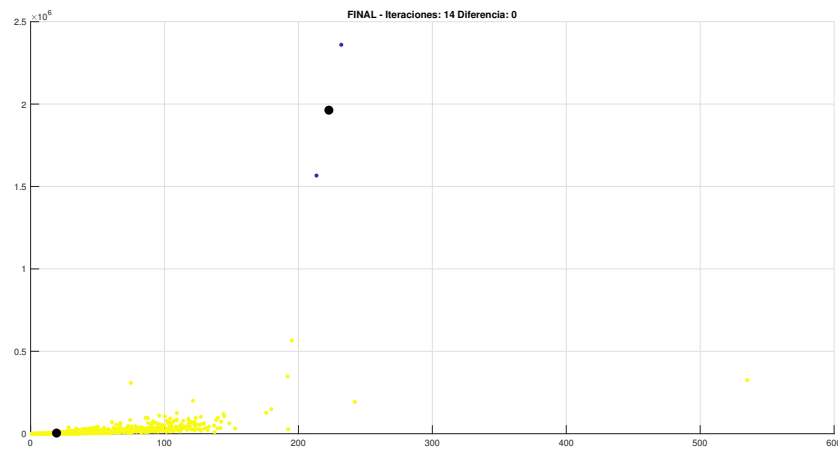


Figura 45: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 5 en el dominio del tiempo de actividad ictal y no ictal.

10.1.2. Resultados jerárquico

Se observó que al aplicar el algoritmo de agrupación jerárquica a las características extraídas en el dominio del tiempo, los datos resultantes mostraron una disposición continua.

En las Figuras 46 - 50, sugiriendo un comportamiento lineal en la distribución de las señales bioeléctricas. En cuanto a la Figura 51, se percibe que los datos son muy similares entre ellos. Por lo que diferenciar clases es complejo utilizando las características 1 y 4. Por ultimo, como se observa en las Figuras 52 - 55, la combinación de la característica 5 con cualquier otra es altamente discriminatoria entre clases, además de, tener la peculiaridad que los datos se superponen, para la clase ictal.

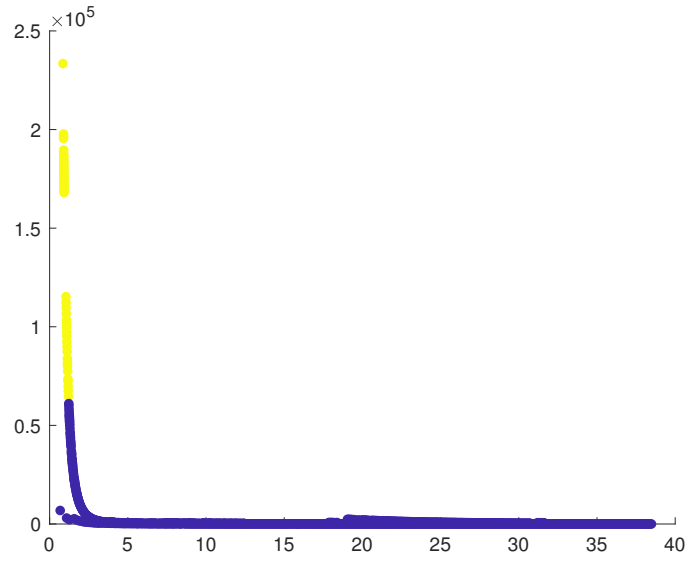


Figura 46: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 2 en el dominio del tiempo de actividad ictal y no ictal.

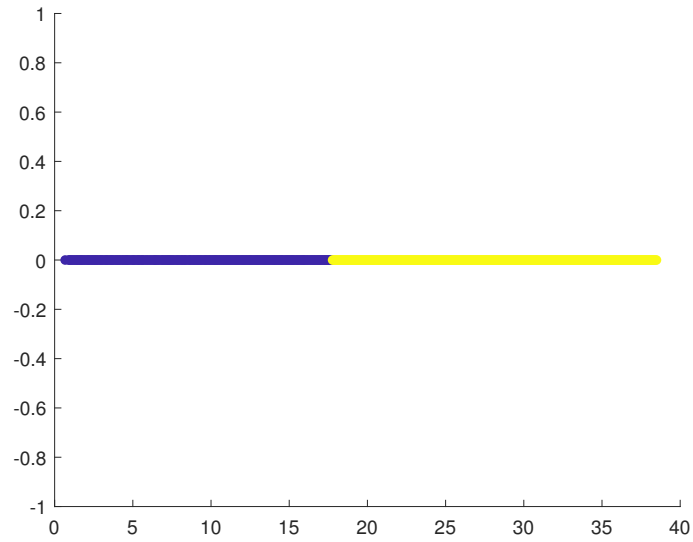


Figura 47: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 3 en el dominio del tiempo de actividad ictal y no ictal.

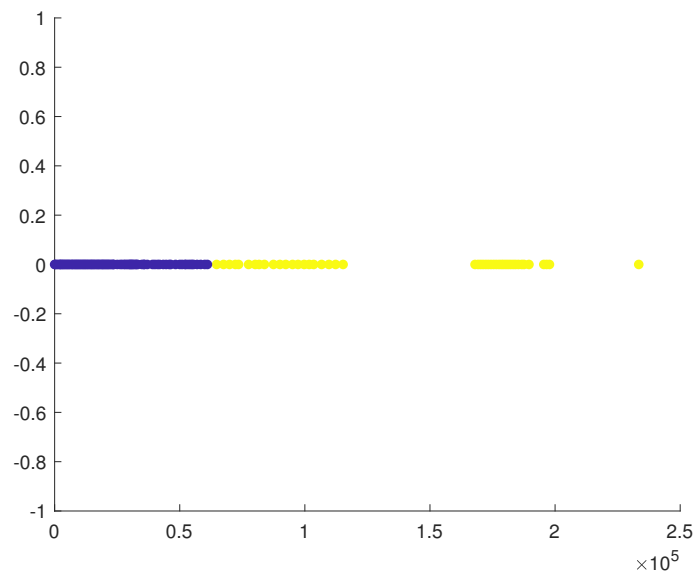


Figura 48: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 3 en el dominio del tiempo de actividad ictal y no ictal.

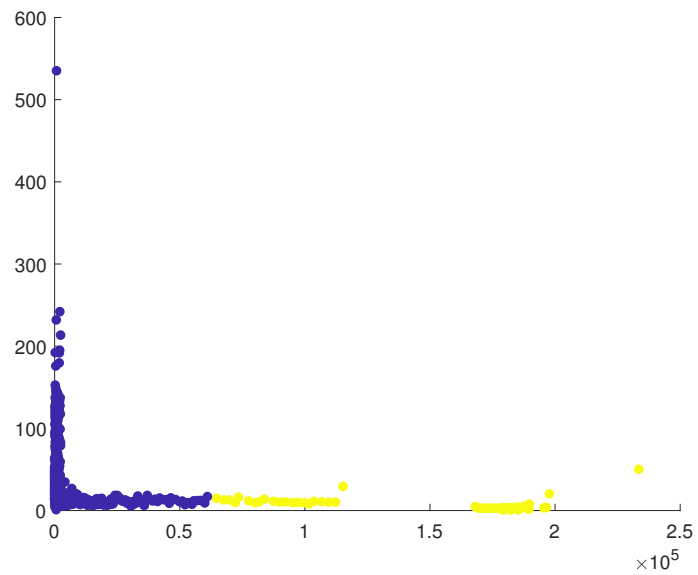


Figura 49: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 4 en el dominio del tiempo de actividad ictal y no ictal.

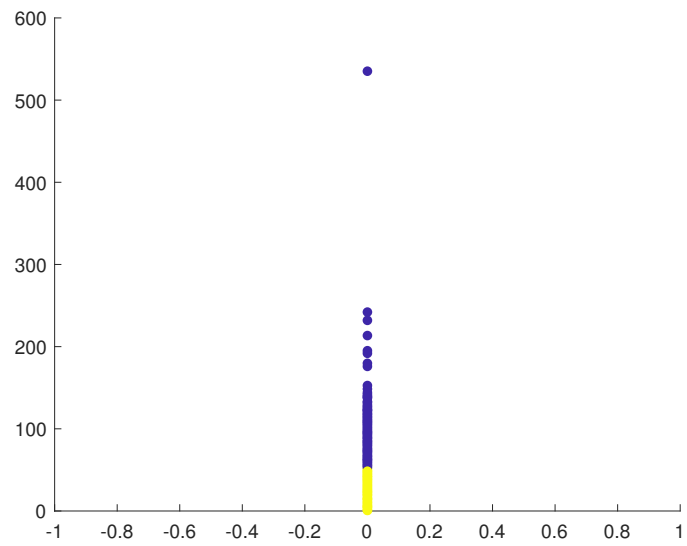


Figura 50: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 4 en el dominio del tiempo de actividad ictal y no ictal.

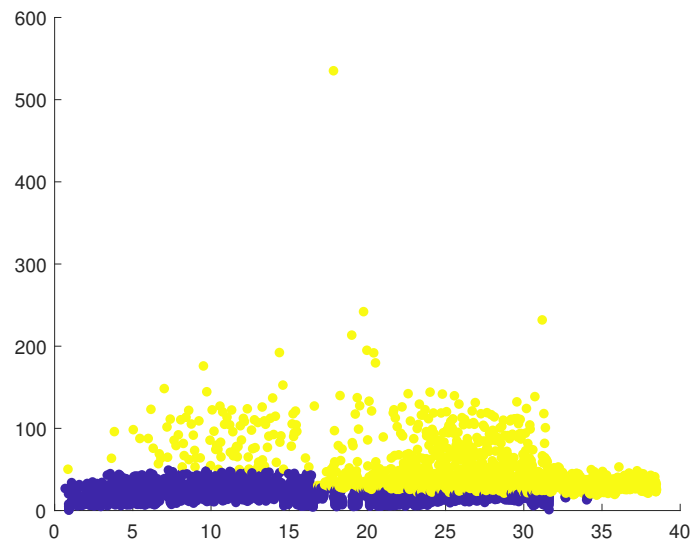


Figura 51: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 4 en el dominio del tiempo de actividad ictal y no ictal.

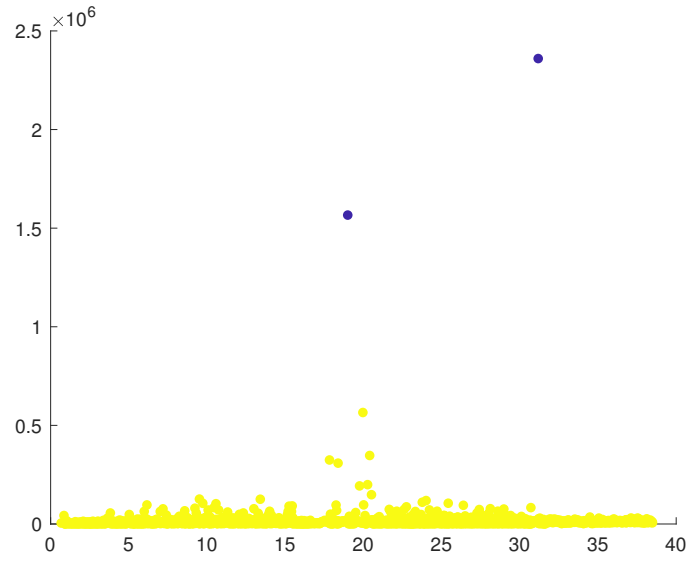


Figura 52: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 5 en el dominio del tiempo de actividad ictal y no ictal.

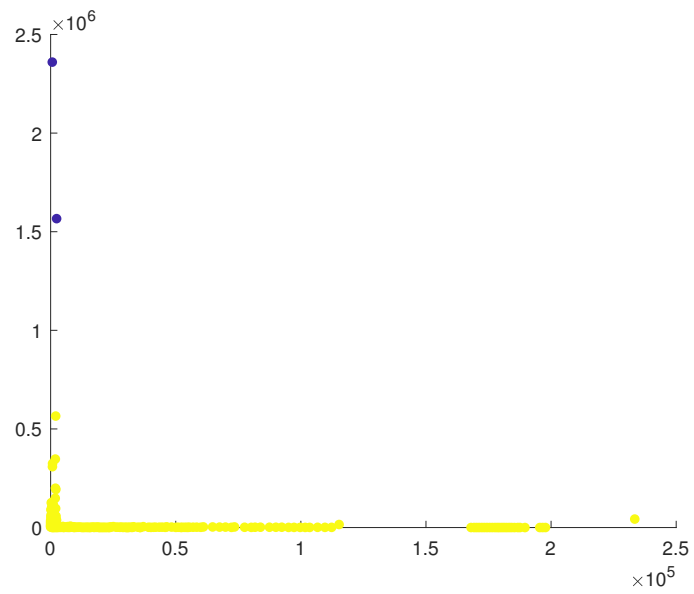


Figura 53: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 5 en el dominio del tiempo de actividad ictal y no ictal.

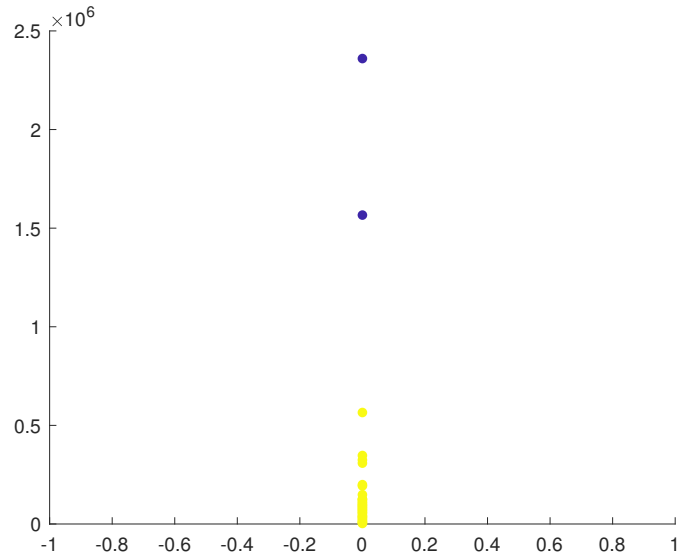


Figura 54: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 5 en el dominio del tiempo de actividad ictal y no ictal.

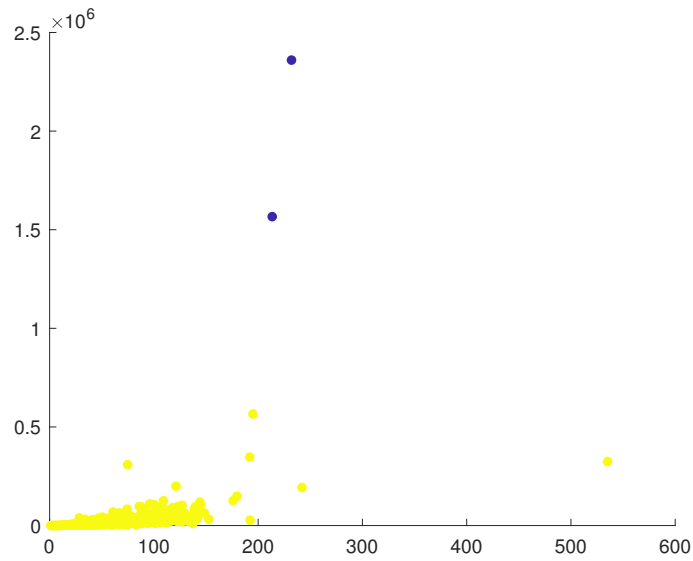


Figura 55: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 5 en el dominio del tiempo de actividad ictal y no ictal.

10.2. Análisis de grupos en el dominio de frecuencia

En el análisis de agrupamiento de señales EEG en el dominio de la frecuencia se trabajaron con las siguientes características:

1. $\frac{\theta}{\alpha}$
2. $\frac{\beta}{\alpha}$
3. $\frac{\theta}{\beta}$
4. $\frac{(\theta+\alpha)}{\beta}$
5. $\frac{(\theta+\alpha)}{(\alpha+\beta)}$
6. Desviación estándar

10.2.1. Resultados K-means

Los resultados fueron satisfactorios, como se puede observar en las Figuras 56 - 70. Los datos forman un contorno similar a una elipse estirada por uno o dos bordes. Este patrón de agrupación resultó en una correcta separación de los datos en dos grupos, que consisten en señales EEG de individuos con actividad epiléptica y personas sanas.

La capacidad de K-means para realizar esta separación respalda la validez de los resultados y sugiere la utilidad de las características de frecuencia en la discriminación entre individuos con actividad epiléptica y personas sanas, además de, poder notar que para este dominio la densidad de datos es muy similar para cada grupo. Por lo que el tener una mayor cantidad de datos y de manera diversificada (distintos pacientes) sin duda alguna se obtendrá una predicción excelente.

Para las figuras anteriormente mencionadas en esta sección, cabe mencionar que con la transición entre cada grupo, se aprecia una disminución en la densidad de datos de manera gradual y significativa. Lo cual es esperado, ya que las grabaciones por parte de HUMANA contienen segmentos donde el paciente no está pasando por un episodio epiléptico, lo que se interpreta como ruido en las características de su clase (ictal).

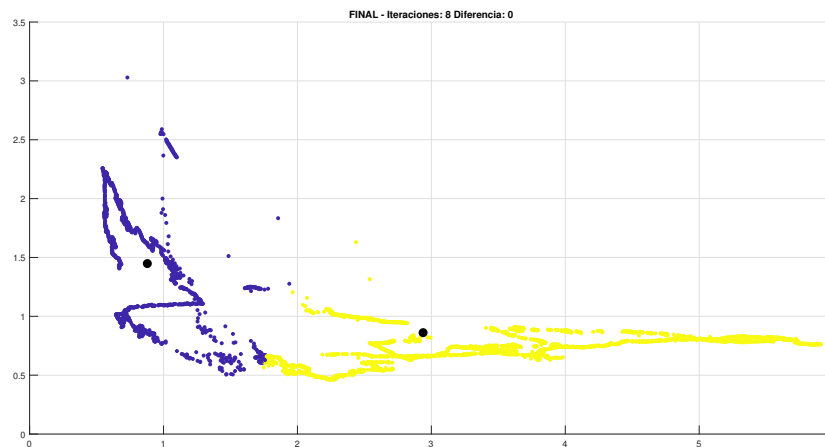


Figura 56: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 2 en el dominio de la frecuencia de actividad ictal y no ictal.

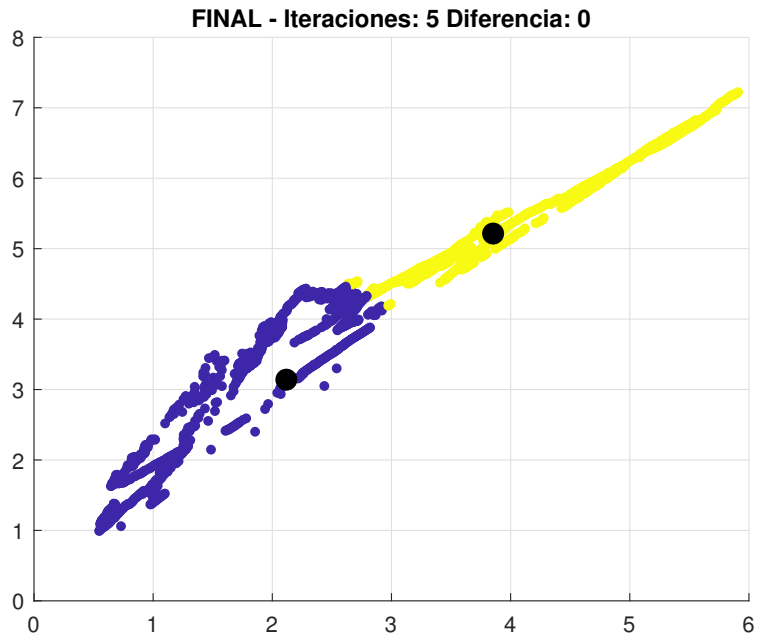


Figura 57: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 3 en el dominio de la frecuencia de actividad ictal y no ictal.

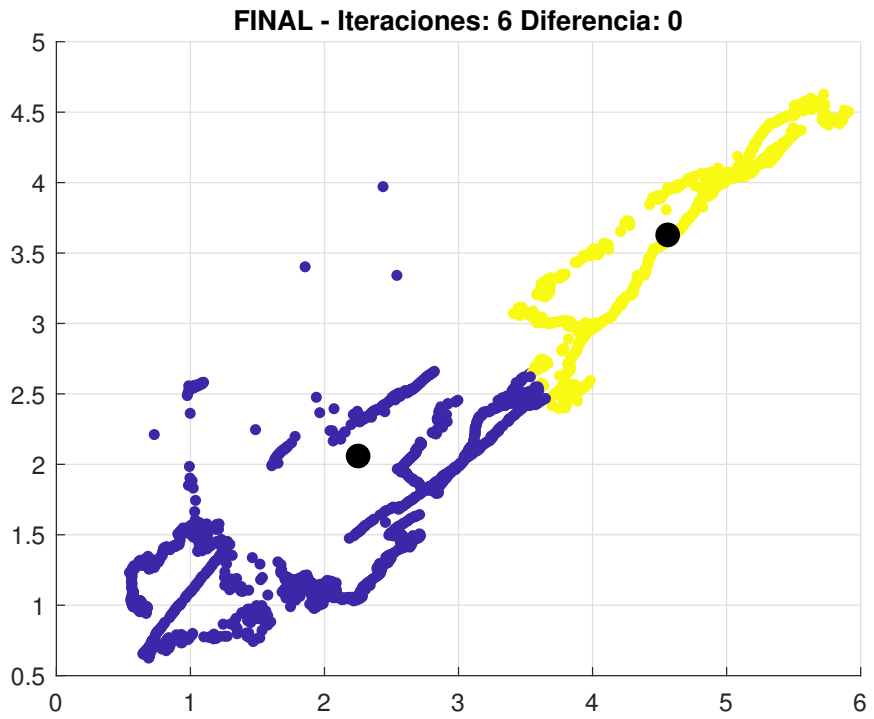


Figura 58: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 4 en el dominio de la frecuencia de actividad ictal y no ictal.

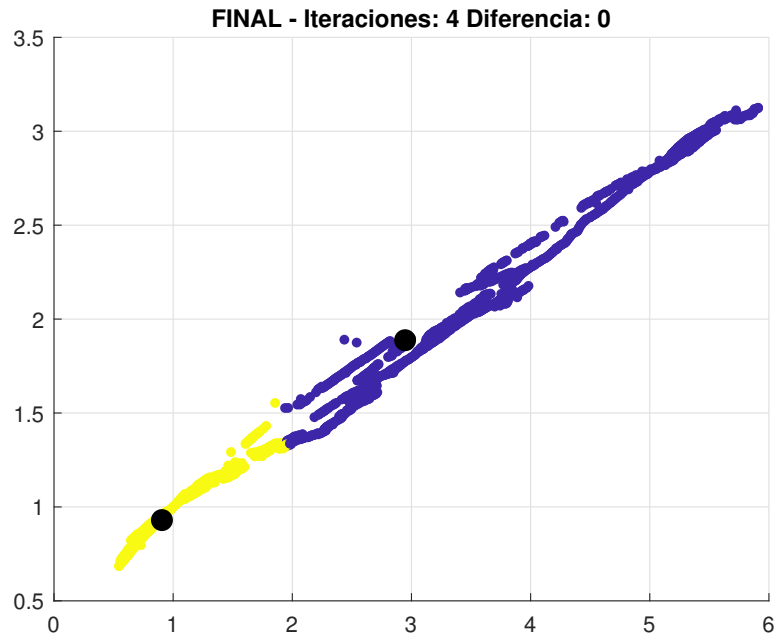


Figura 59: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

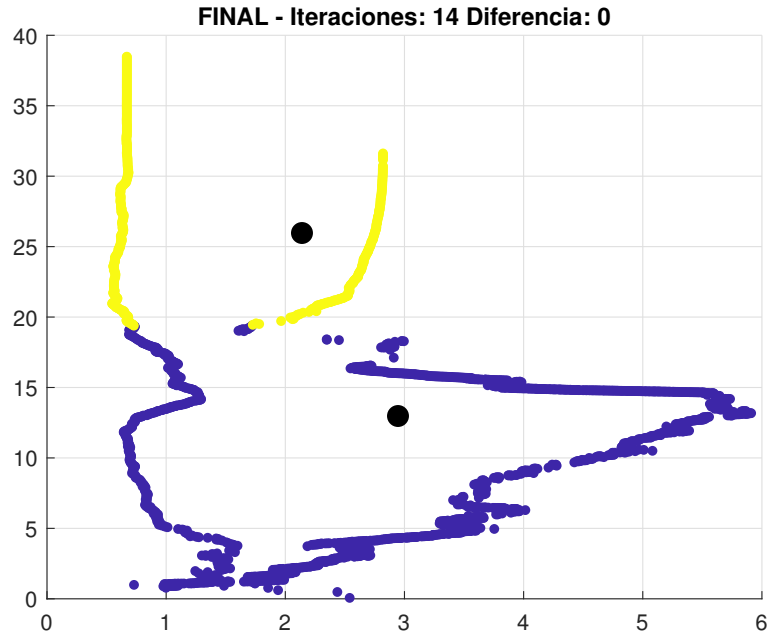


Figura 60: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

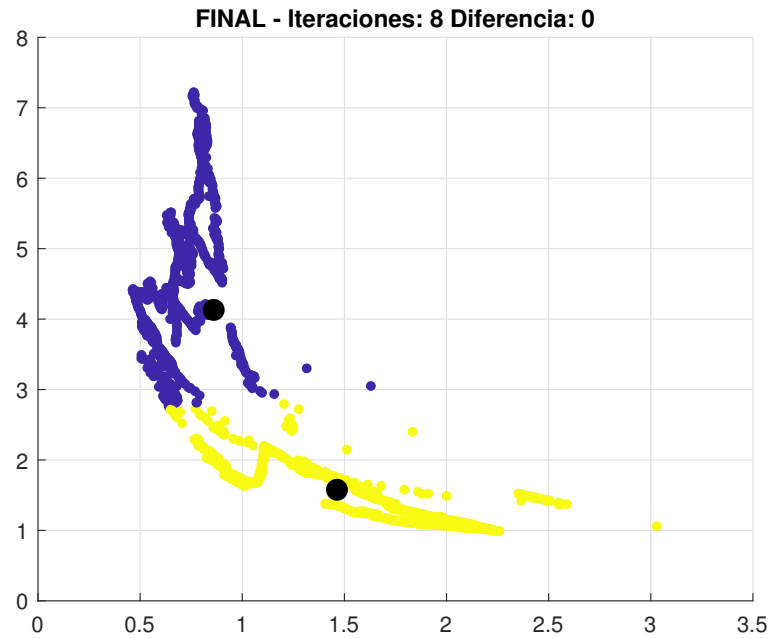


Figura 61: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 3 en el dominio de la frecuencia de actividad ictal y no ictal.

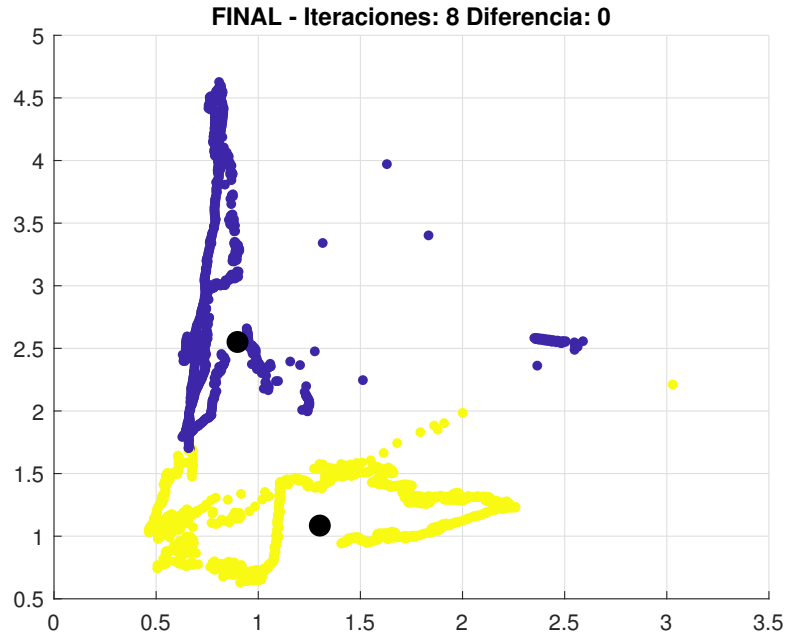


Figura 62: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 4 en el dominio de la frecuencia de actividad ictal y no ictal.

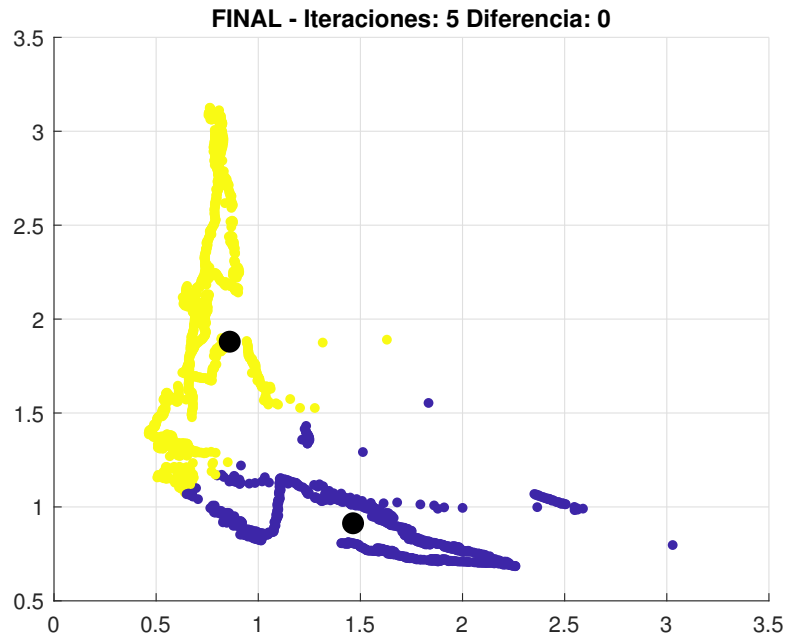


Figura 63: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

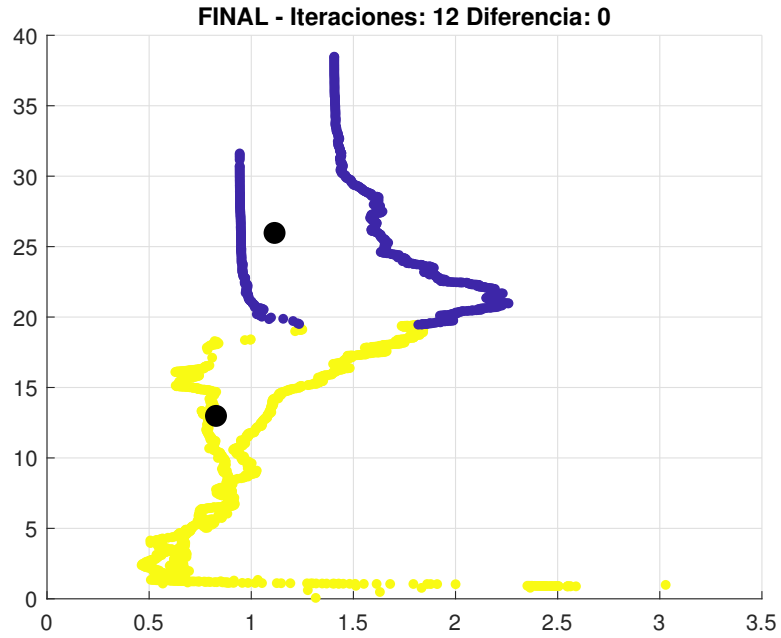


Figura 64: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

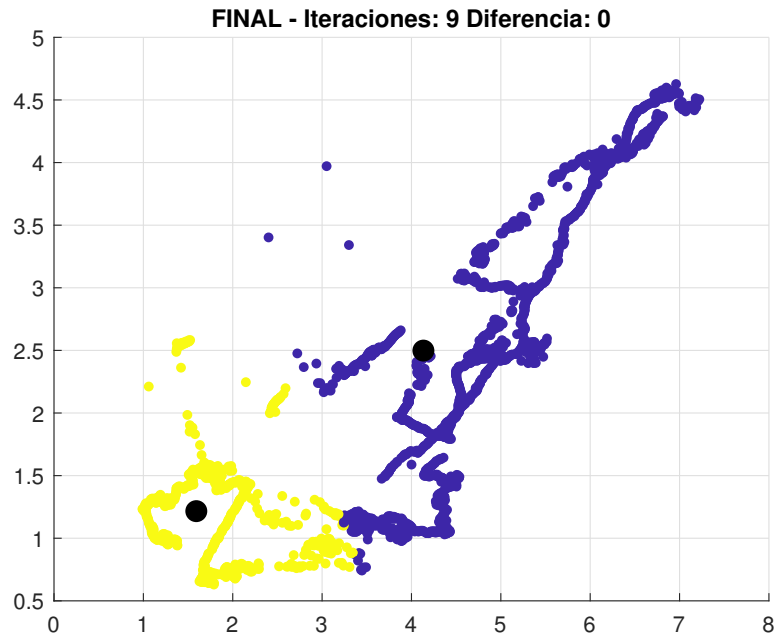


Figura 65: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 4 en el dominio de la frecuencia de actividad ictal y no ictal.

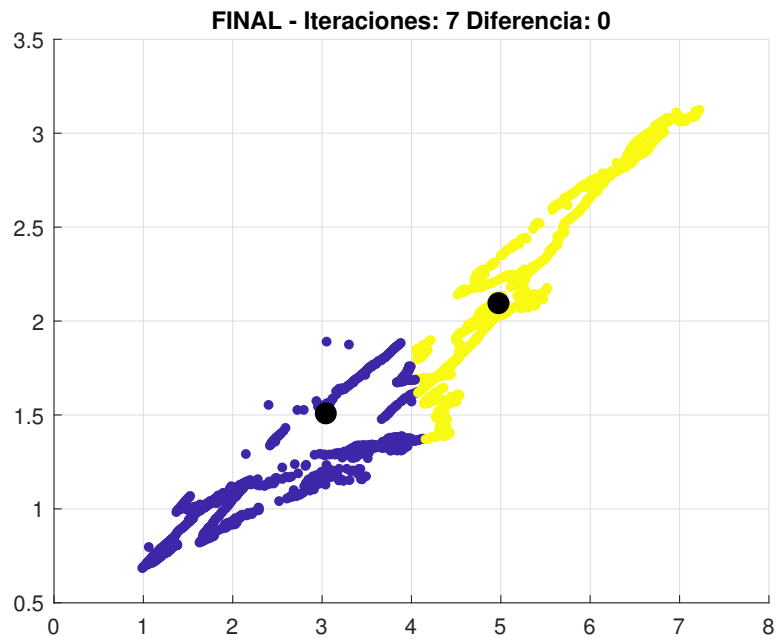


Figura 66: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

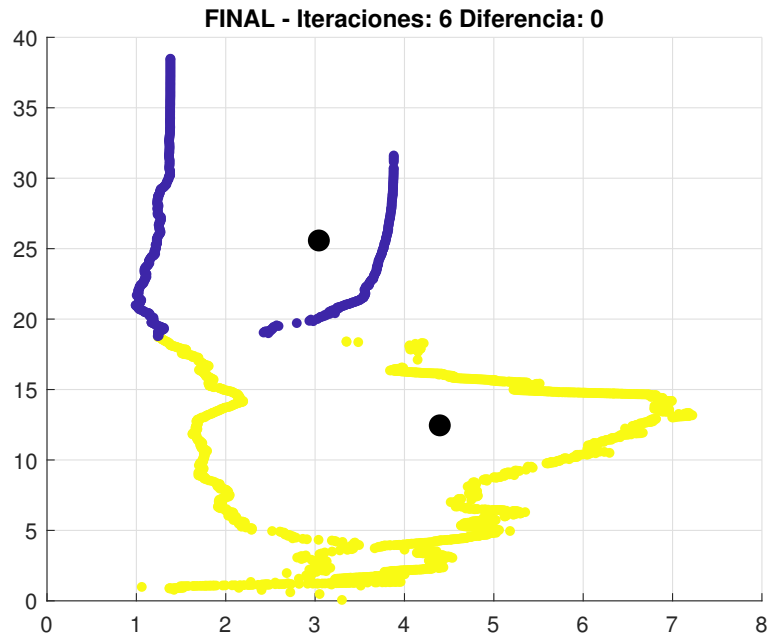


Figura 67: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

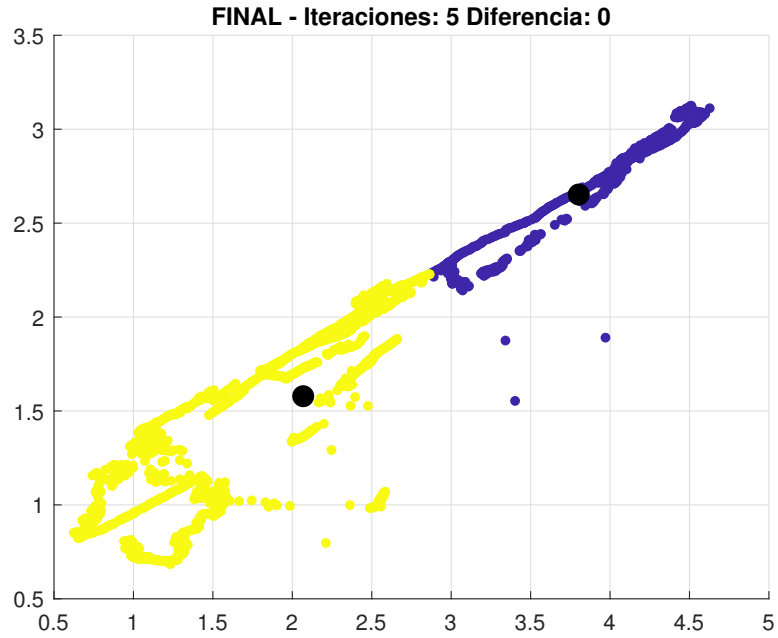


Figura 68: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

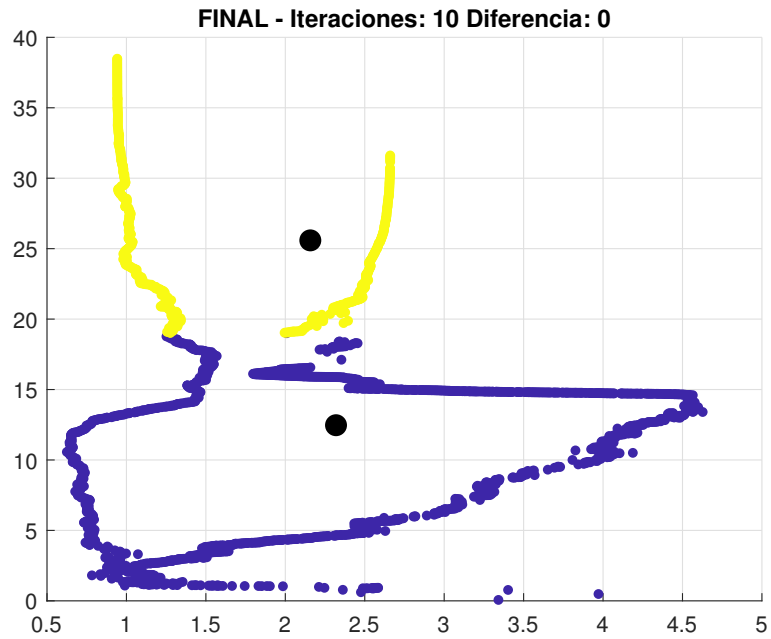


Figura 69: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

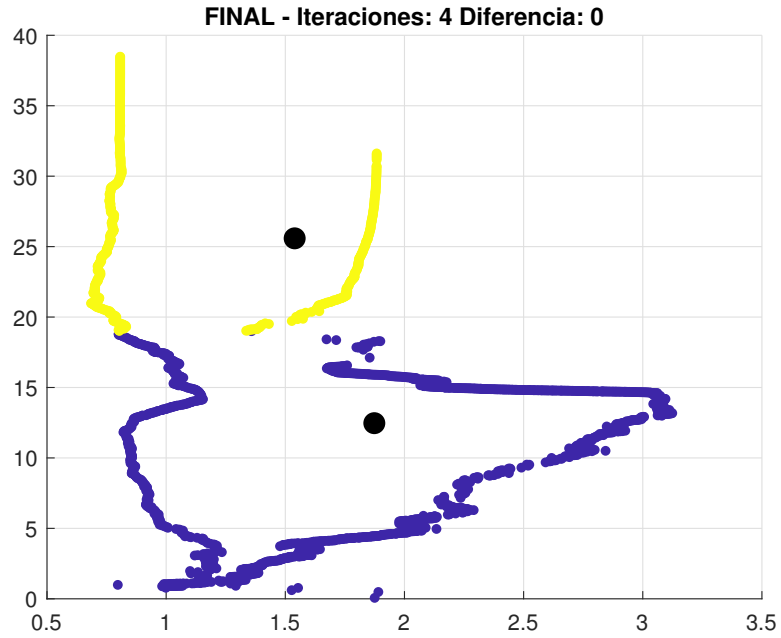


Figura 70: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 5 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

10.2.2. Resultados jerárquico

Los resultados para este algoritmo de agrupación fueron exitosos como se puede observar en las Figuras 71 - 85. Respecto al algoritmo *K-means* se percibe ligeras diferencias, siendo una de estas que la cantidad de épocas percibidas como clase ictal, aumenta considerablemente. Esto sugiere que el análisis de agrupamiento jerárquico, la estructura de árbol permite identificar relaciones de subgrupo dentro de las clases principales, capturando así patrones más complejos y sutiles en los datos. Además, el agrupamiento jerárquico no asume la misma forma esférica de los *clusters* que hace *k-means*, lo que le confiere una mayor flexibilidad para adaptarse a distribuciones de datos más complejas.

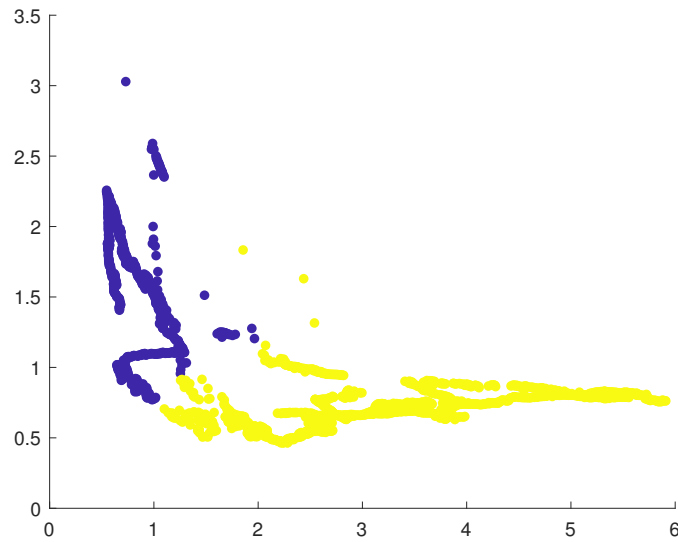


Figura 71: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 2 en el dominio de la frecuencia de actividad ictal y no ictal.

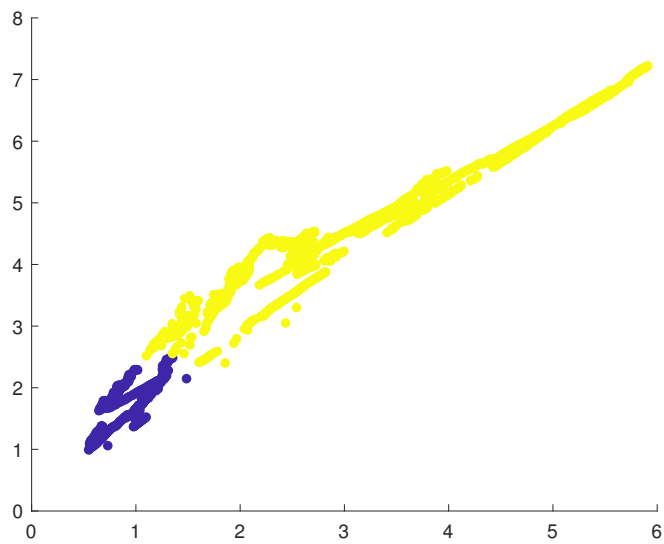


Figura 72: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 3 en el dominio de la frecuencia de actividad ictal y no ictal.

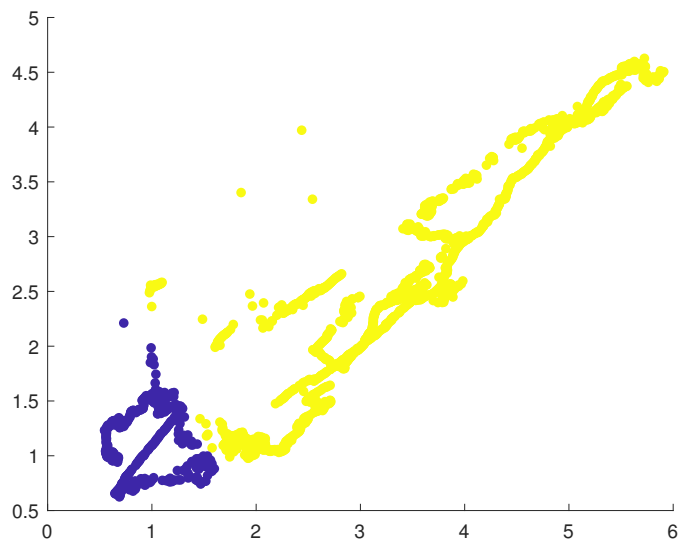


Figura 73: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 4 en el dominio de la frecuencia de actividad ictal y no ictal.

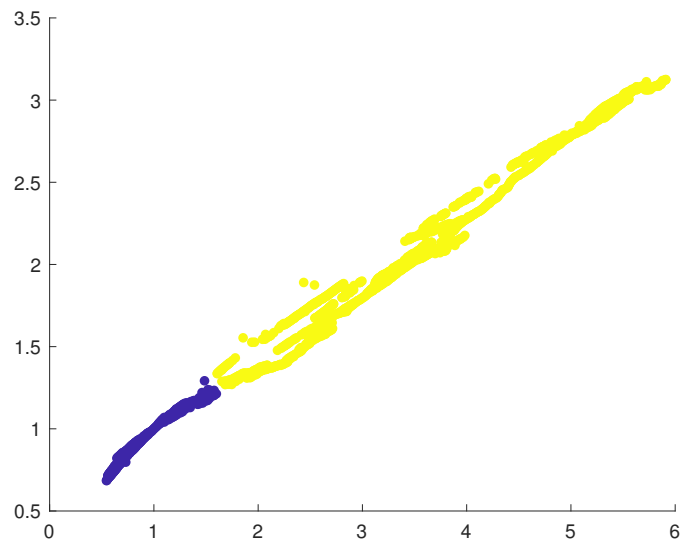


Figura 74: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

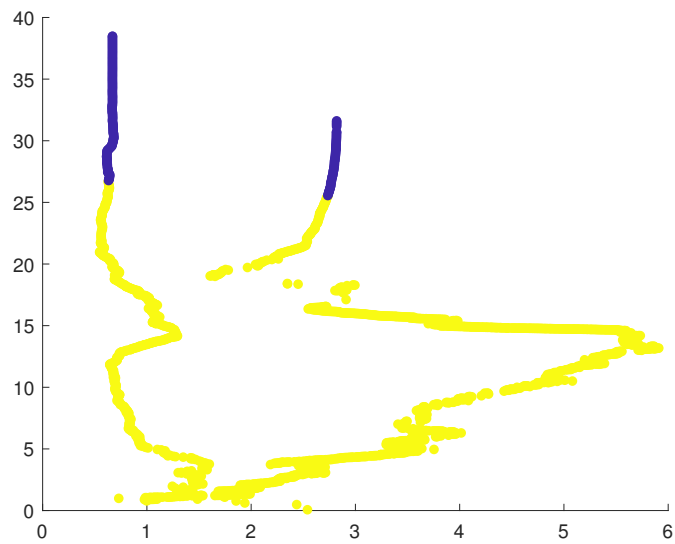


Figura 75: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

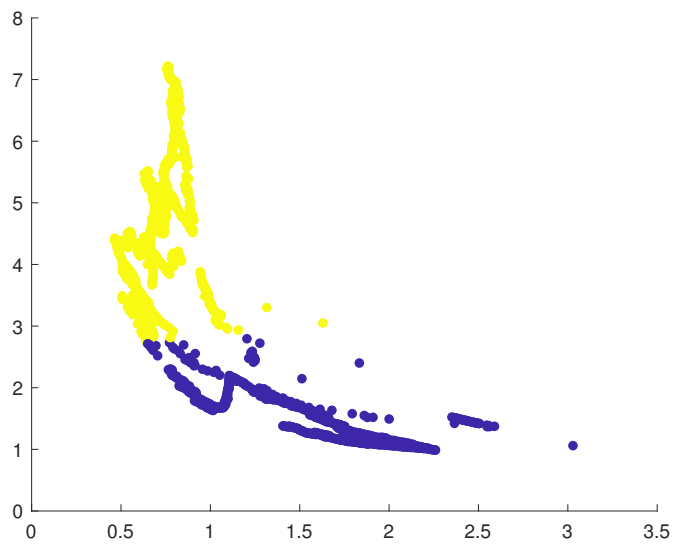


Figura 76: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 3 en el dominio de la frecuencia de actividad ictal y no ictal.

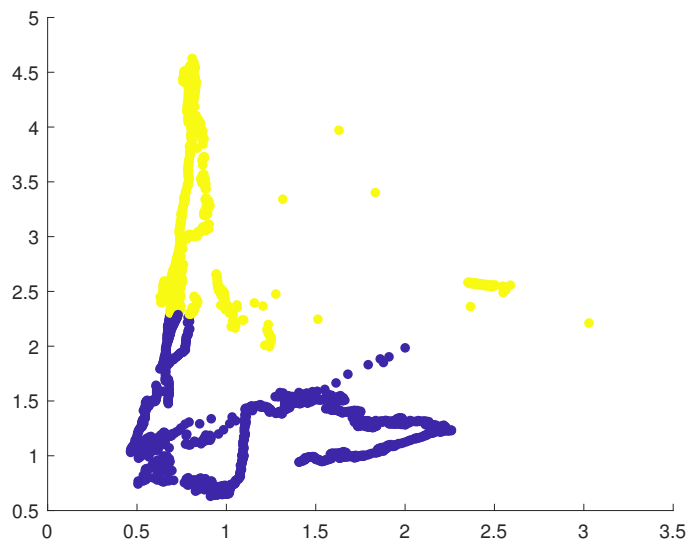


Figura 77: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 4 en el dominio de la frecuencia de actividad ictal y no ictal.

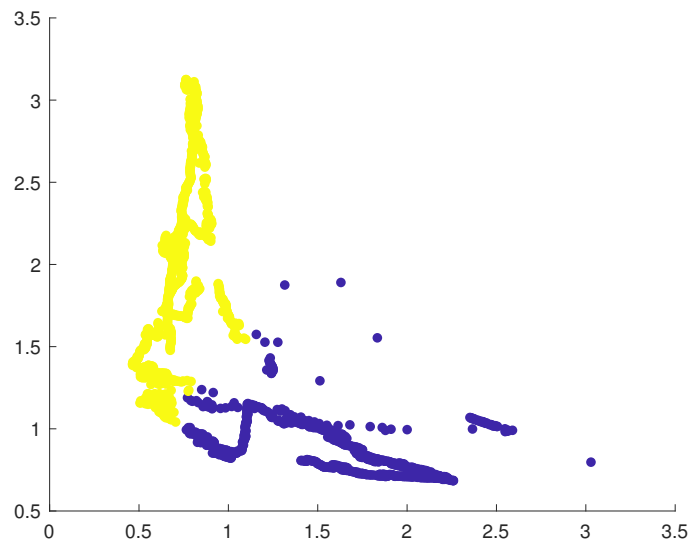


Figura 78: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

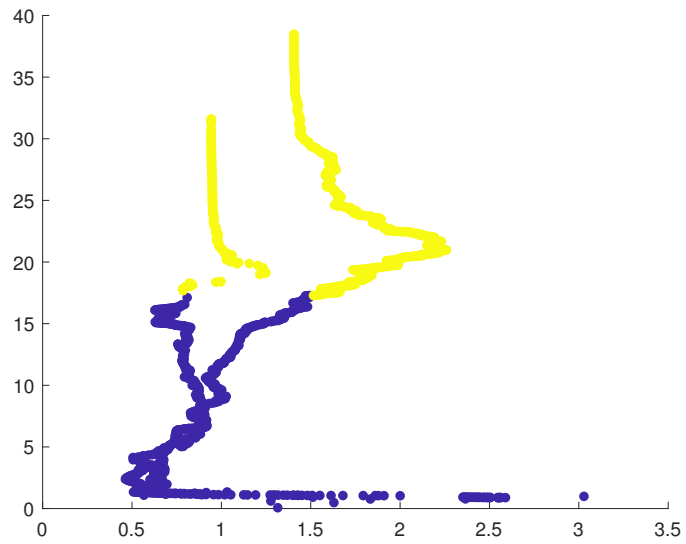


Figura 79: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

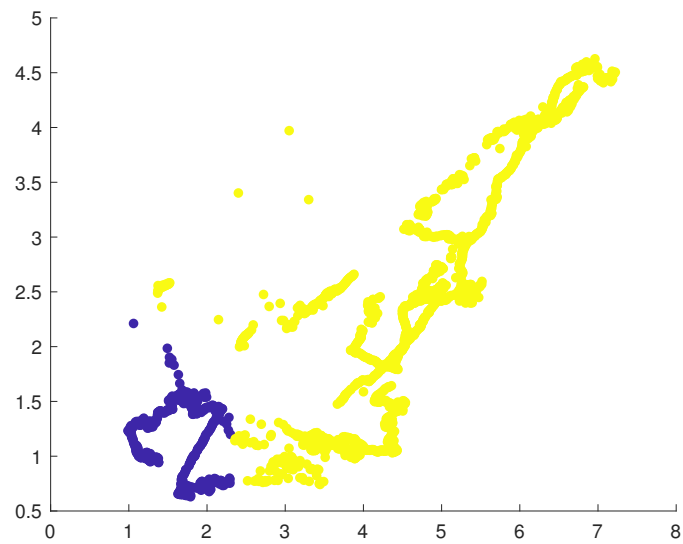


Figura 80: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 4 en el dominio de la frecuencia de actividad ictal y no ictal.

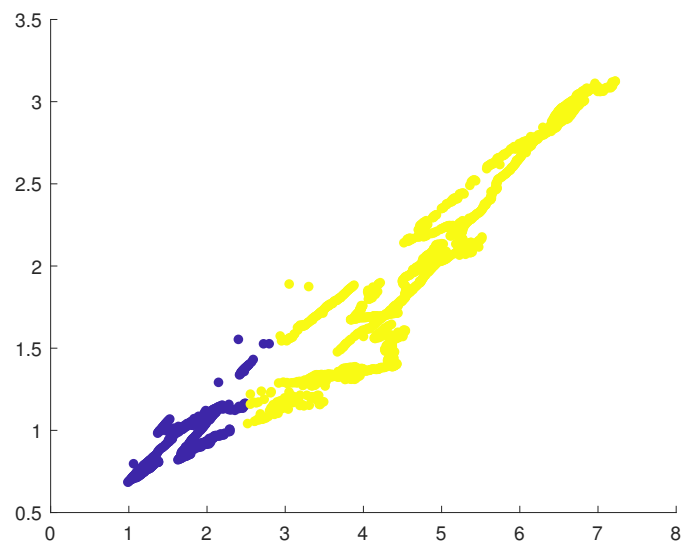


Figura 81: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

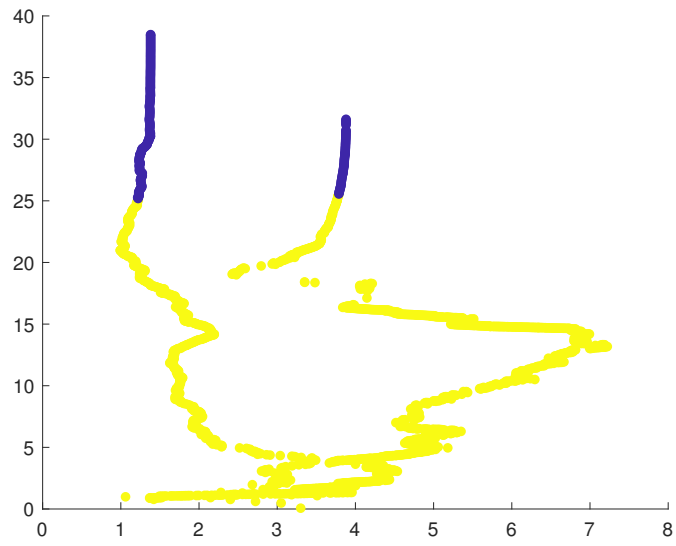


Figura 82: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

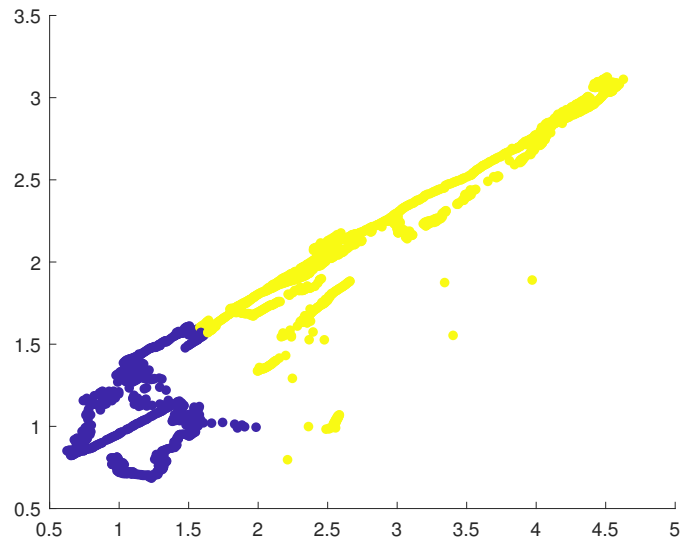


Figura 83: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 5 en el dominio de la frecuencia de actividad ictal y no ictal.

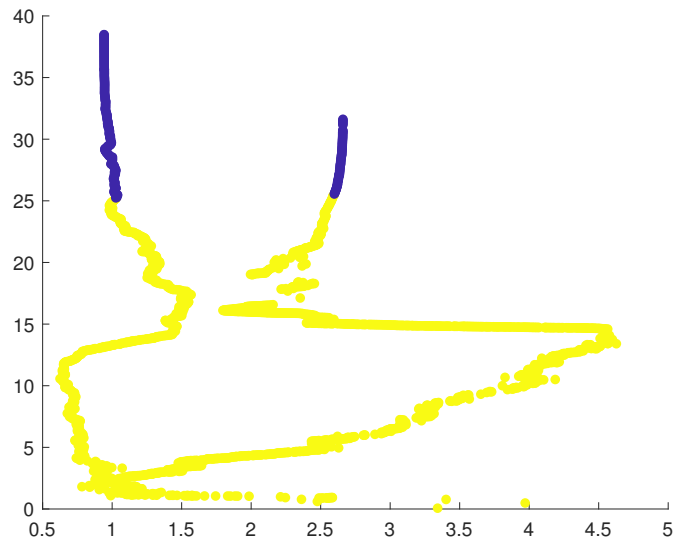


Figura 84: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

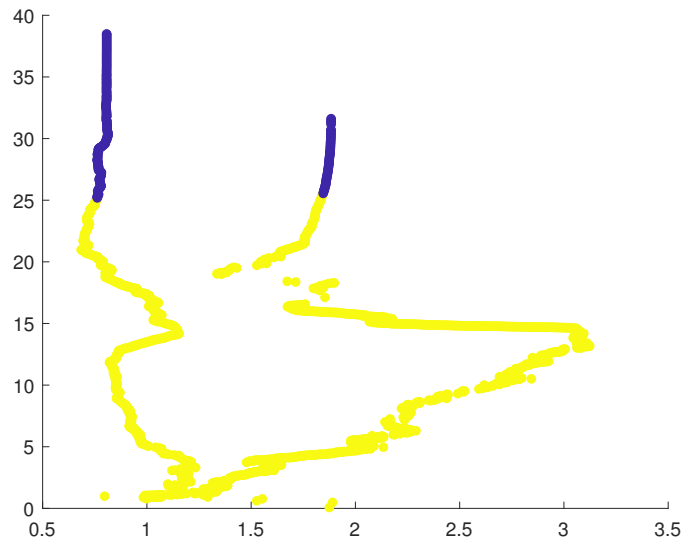


Figura 85: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 5 y 6 en el dominio de la frecuencia de actividad ictal y no ictal.

10.3. Análisis de grupos en el dominio de wavelets

En esta sección, se aborda el análisis de grupos resultantes de la aplicación de técnicas de procesamiento en el dominio de wavelets. Se destacan las siguientes características utilizadas en el análisis:

1. Potencia
2. Desviación
3. Asimetría
4. Media
5. Curtosis
6. Cruces por cero (ZC)

10.3.1. Resultados K-means

La identificación de dos grupos de señales bioeléctricas que están ampliamente separados en el espacio de características es un hallazgo de gran relevancia, como se observa en las Figuras 86 - 92, 94, 95, 97 - 100. Esta distinción evidente indica que las características analizadas, como la potencia, desviación, asimetría, media, curtosis y cruces por cero (ZC), son altamente discriminatorias entre las clases ictal y sana. Esta observación respalda la utilidad de estas características en la detección y diferenciación de estados epilépticos y salud.

La agrupación de datos en dos grupos cercanos, pero con diferentes características de densidad, también presenta un aspecto interesante, como se observar en las Figuras 91, 93 y 96. La concentración de datos en una región más pequeña en un grupo y su expansión en el otro indican que, si bien estos grupos están próximos, aún mantienen diferencias notables. Esta observación puede ser esclarecedora para la comprensión de las variaciones sutiles en las señales bioeléctricas y cómo estas variaciones pueden influir en la clasificación.

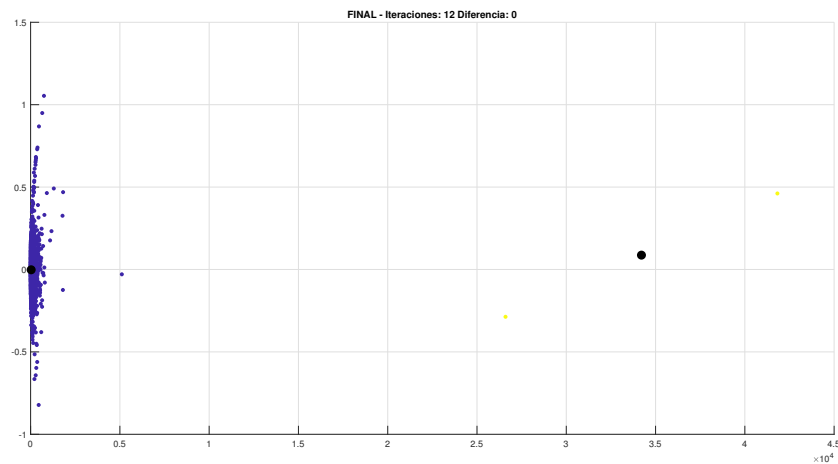


Figura 86: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 2 tipo wavelets de actividad ictal y no ictal.

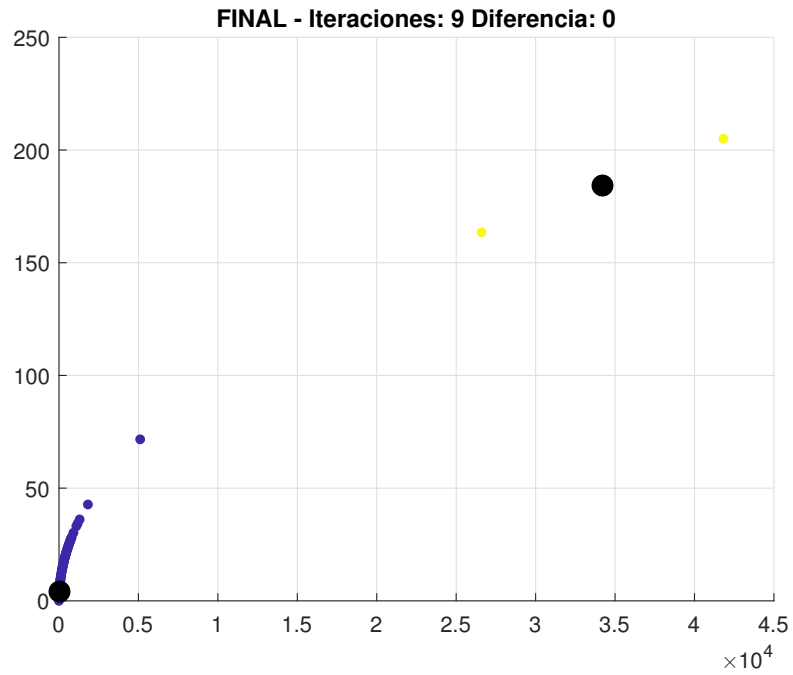


Figura 87: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 3 tipo wavelets de actividad ictal y no ictal.

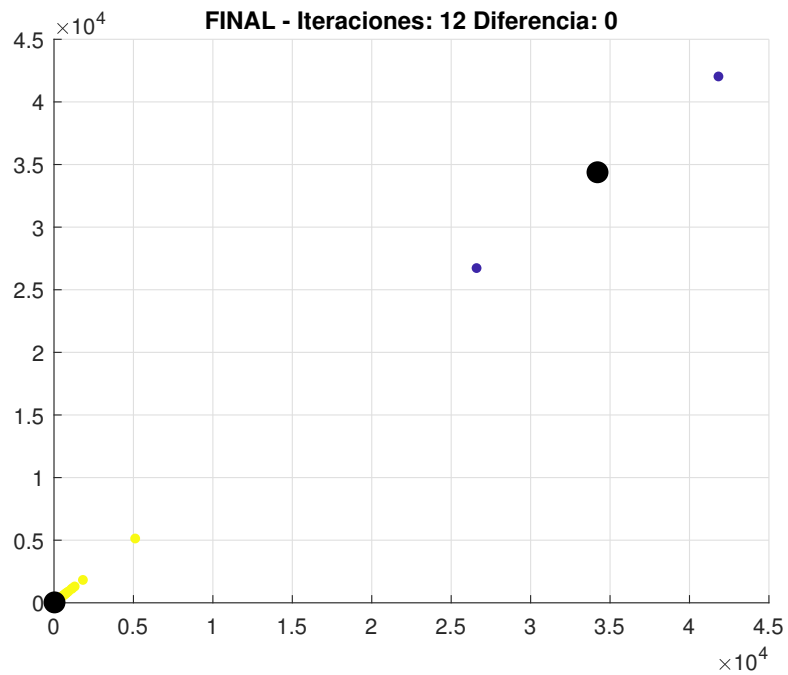


Figura 88: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 4 tipo wavelets de actividad ictal y no ictal.

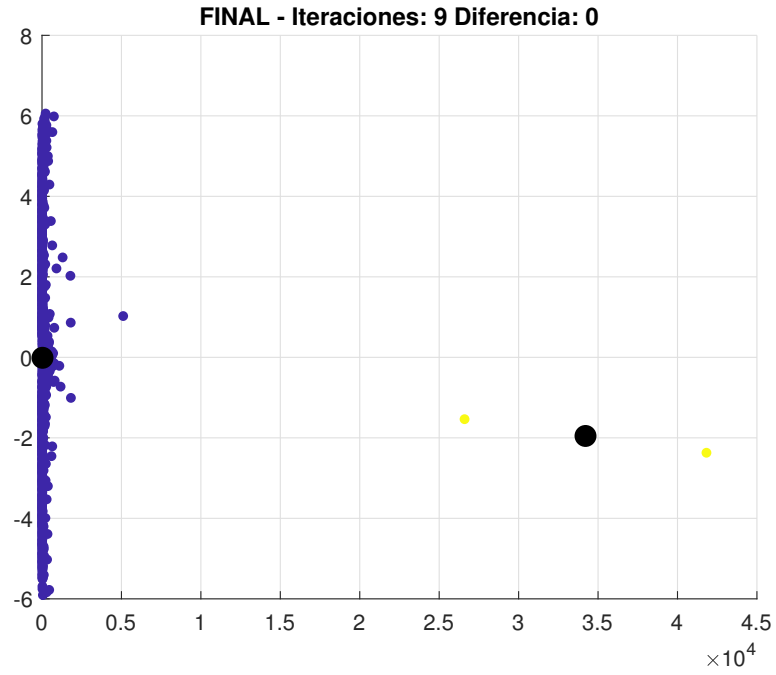


Figura 89: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 5 tipo wavelets de actividad ictal y no ictal.

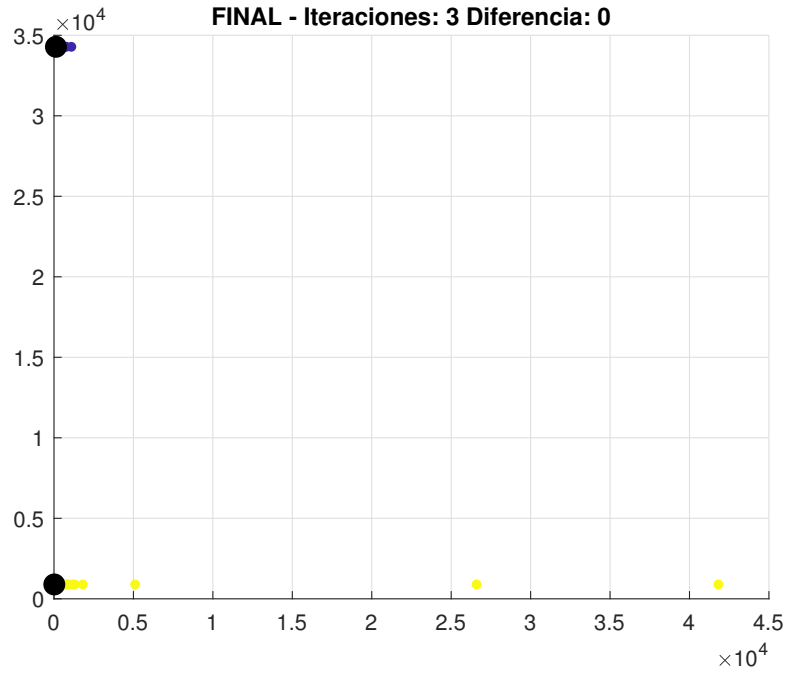


Figura 90: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 1 y 6 tipo wavelets de actividad ictal y no ictal.

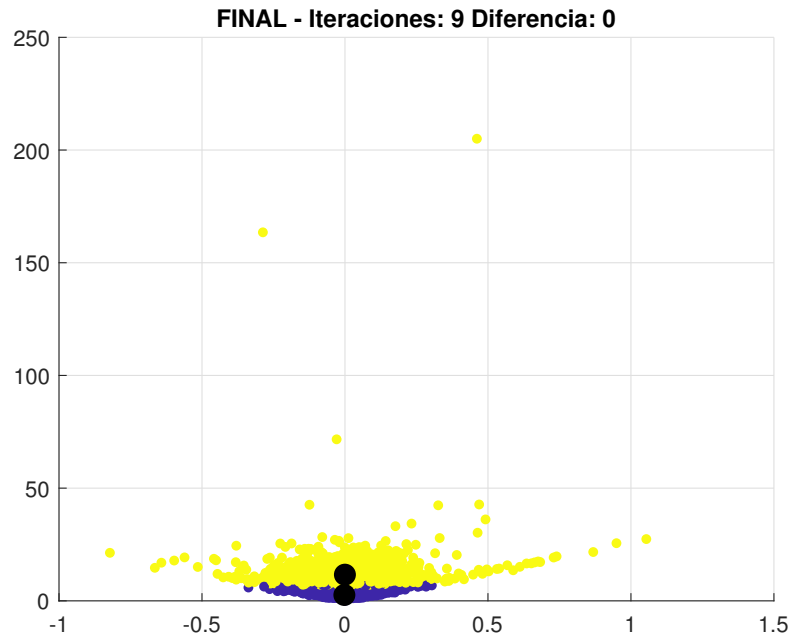


Figura 91: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 3 tipo wavelets de actividad ictal y no ictal.

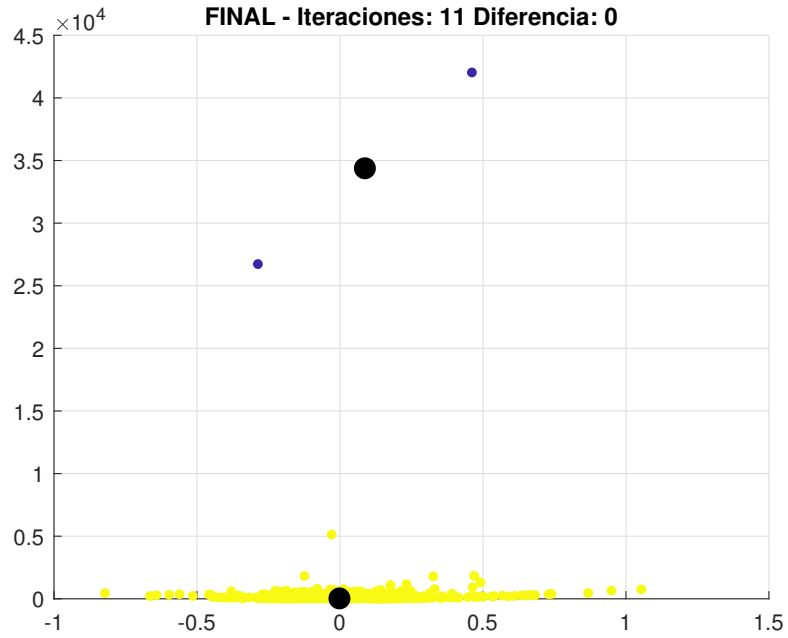


Figura 92: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 4 tipo wavelets de actividad ictal y no ictal.

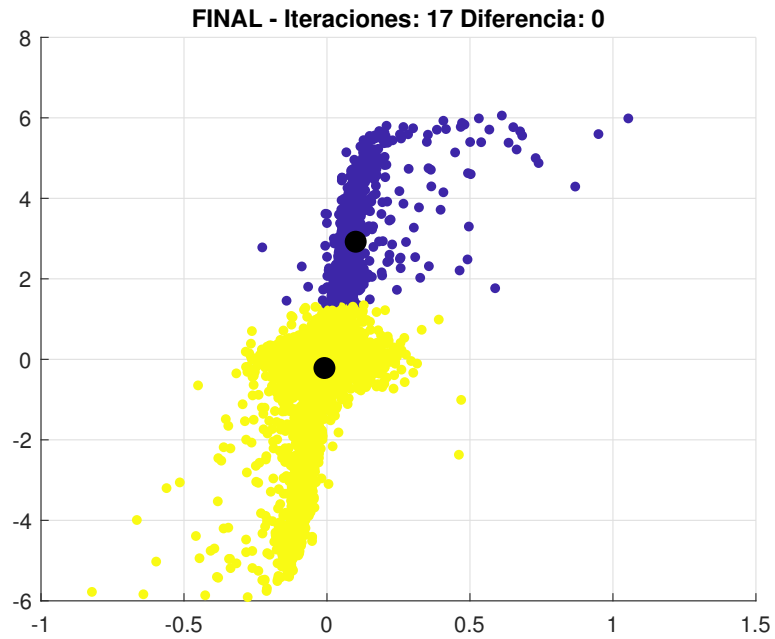


Figura 93: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 5 tipo wavelets de actividad ictal y no ictal.

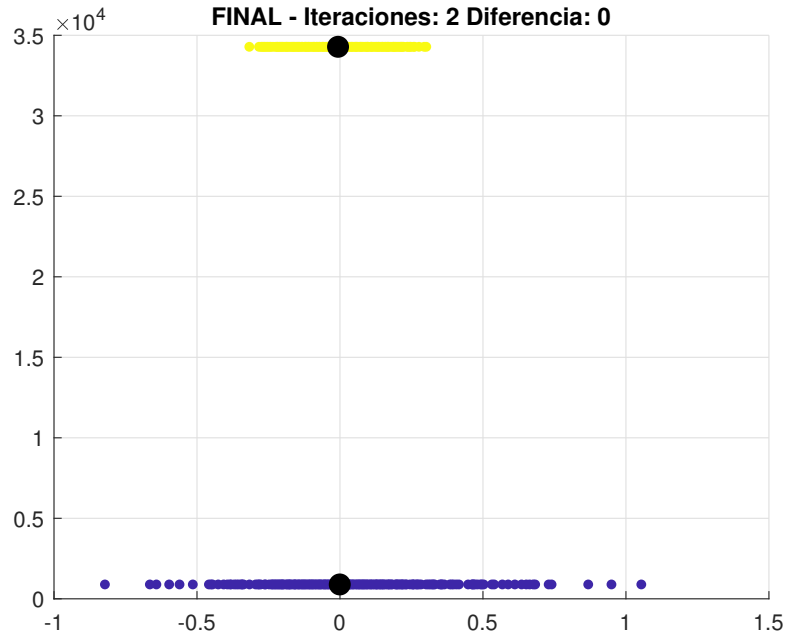


Figura 94: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 2 y 6 tipo wavelets de actividad ictal y no ictal.

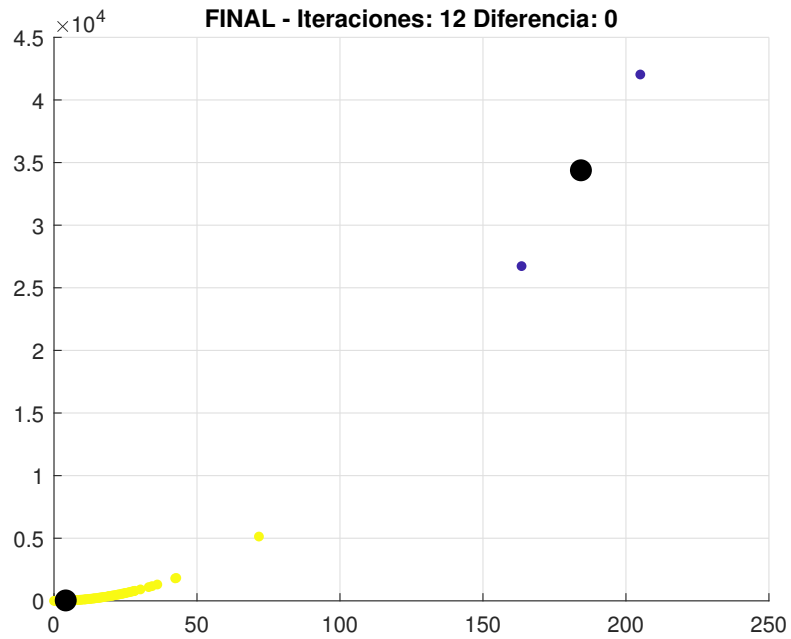


Figura 95: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 4 tipo wavelets de actividad ictal y no ictal.

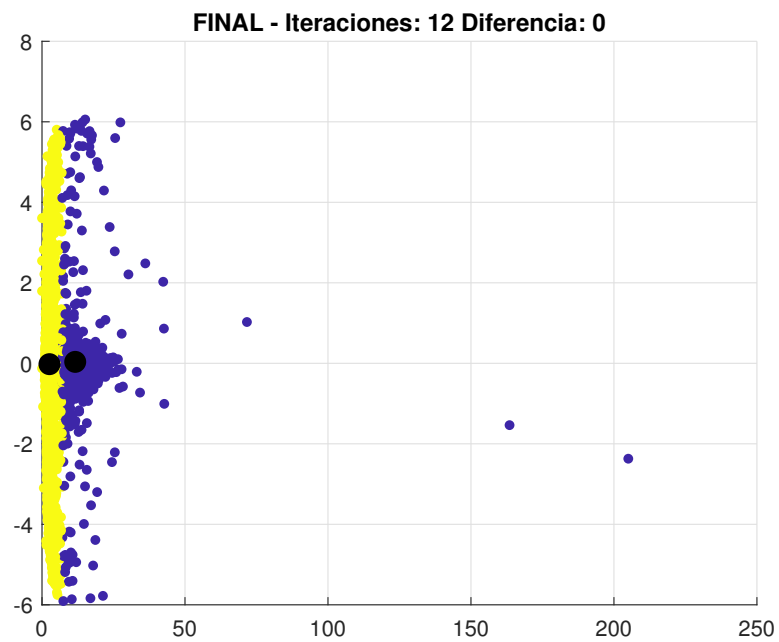


Figura 96: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 5 tipo wavelets de actividad ictal y no ictal.

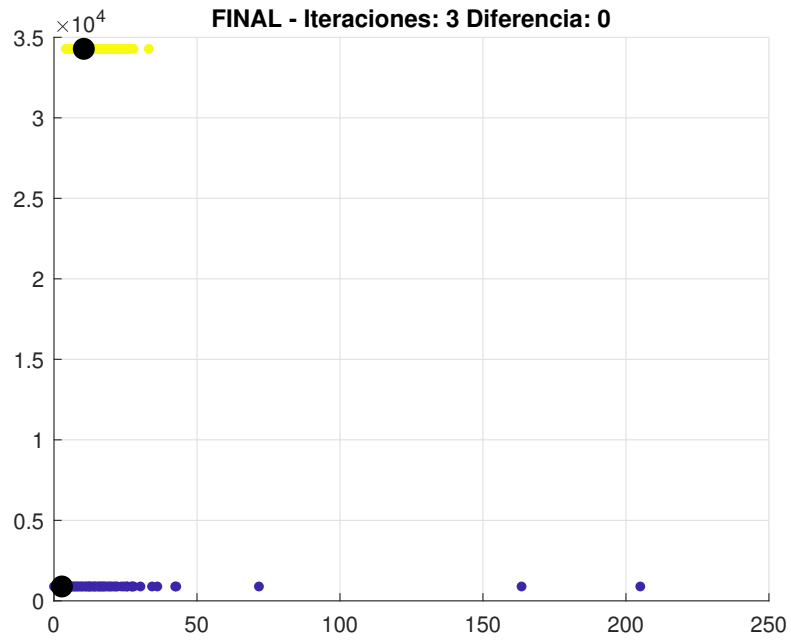


Figura 97: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 3 y 6 tipo wavelets de actividad ictal y no ictal.

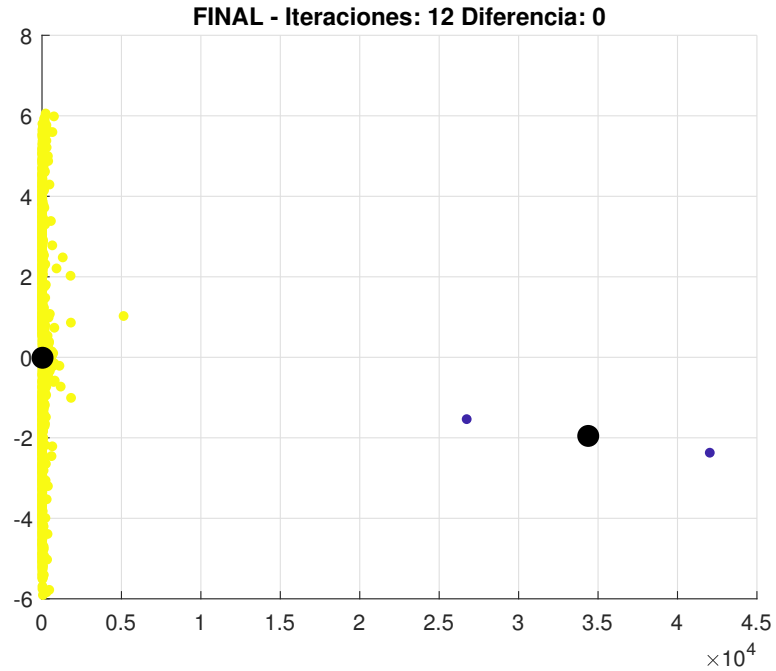


Figura 98: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 5 tipo wavelets de actividad ictal y no ictal.

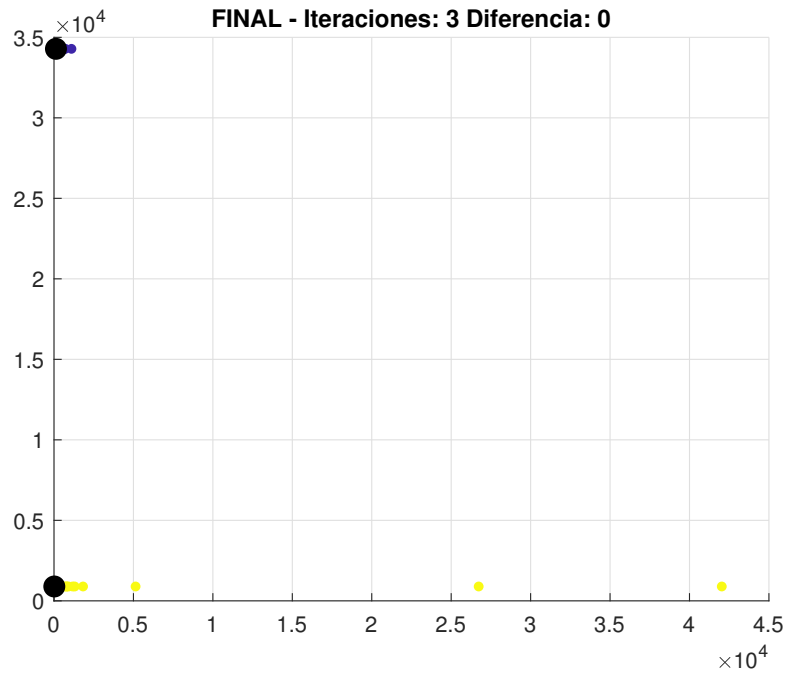


Figura 99: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 4 y 6 tipo wavelets de actividad ictal y no ictal.

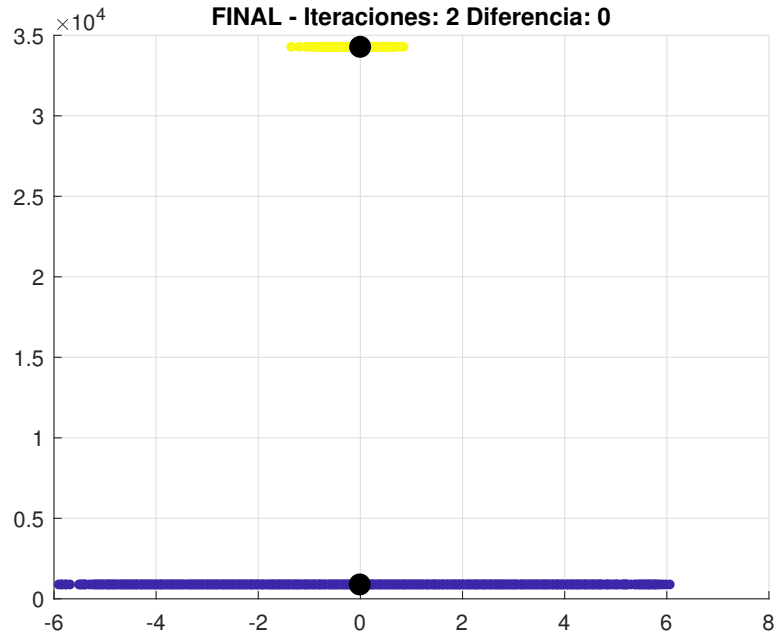


Figura 100: Gráfica de agrupamiento por k-means de señales bioeléctricas de las características 5 y 6 tipo wavelets de actividad ictal y no ictal.

10.3.2. Resultados jerárquicos

Como se puede observar en las Figuras 71 - 115, los resultados fueron exitosos, al ser altamente discriminatorios para cada tipo de clase. Se puede notar también que independientemente del algoritmo de agrupamiento empleado para realizar los modelos, se obtendrán resultados muy similares cuando se utilizan características wavelets.

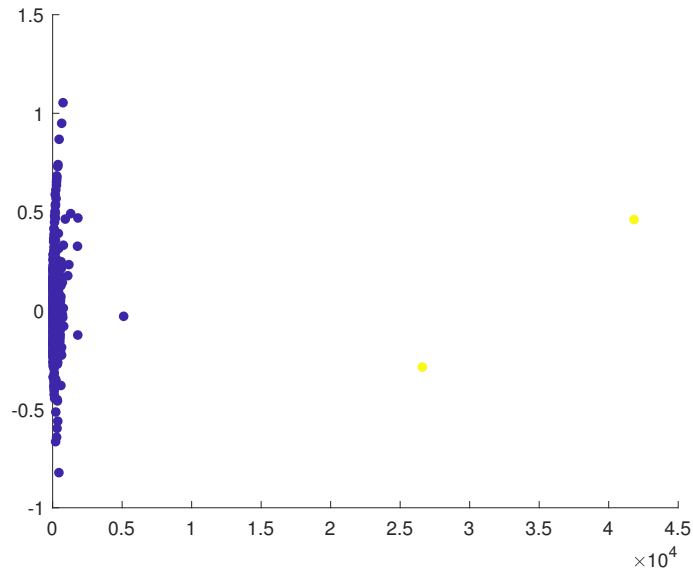


Figura 101: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 2 tipo wavelets de actividad ictal y no ictal.

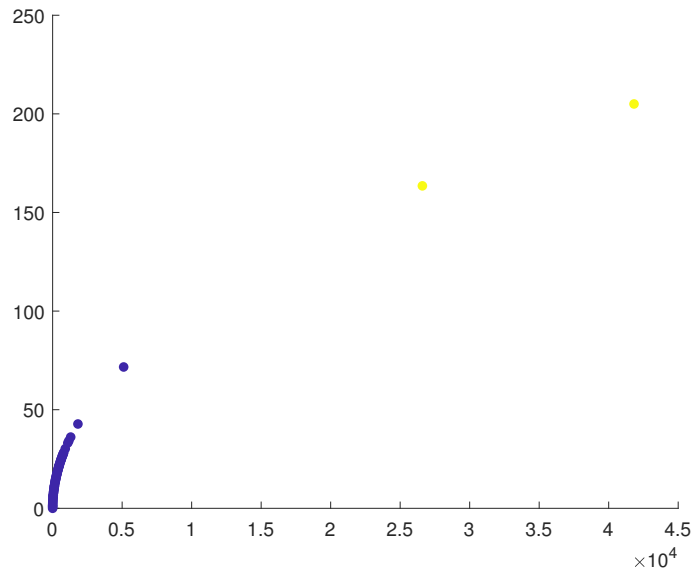


Figura 102: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 3 tipo wavelets de actividad ictal y no ictal.

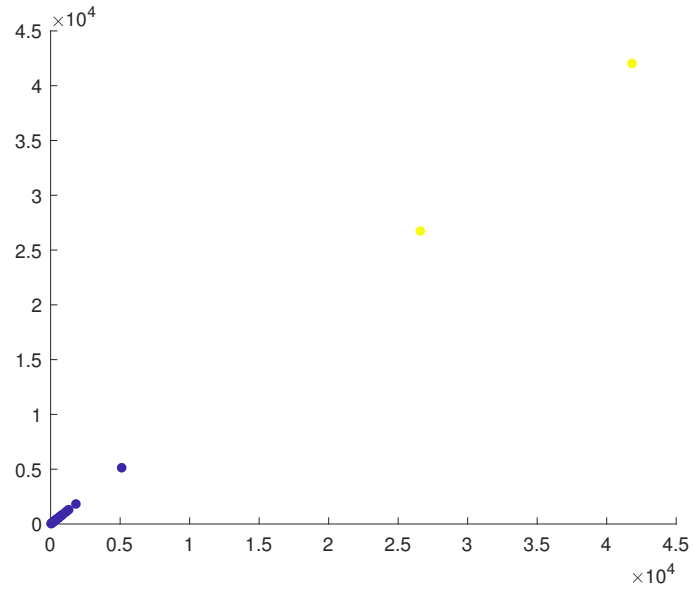


Figura 103: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 4 tipo wavelets de actividad ictal y no ictal.

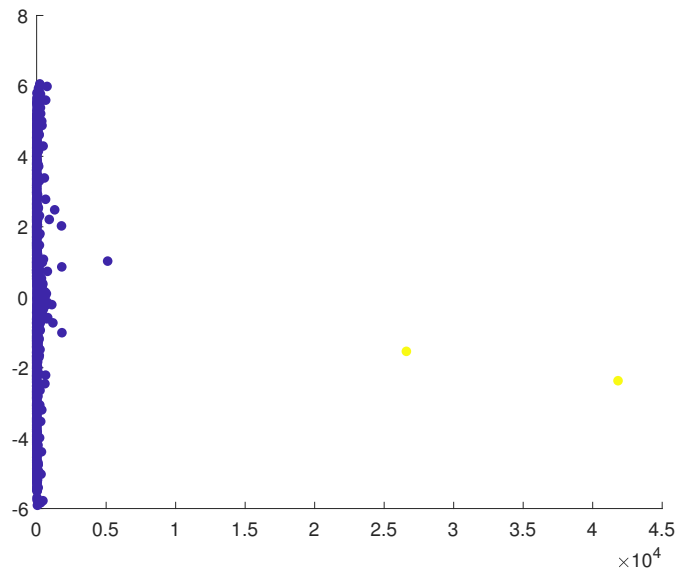


Figura 104: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 5 tipo wavelets de actividad ictal y no ictal.

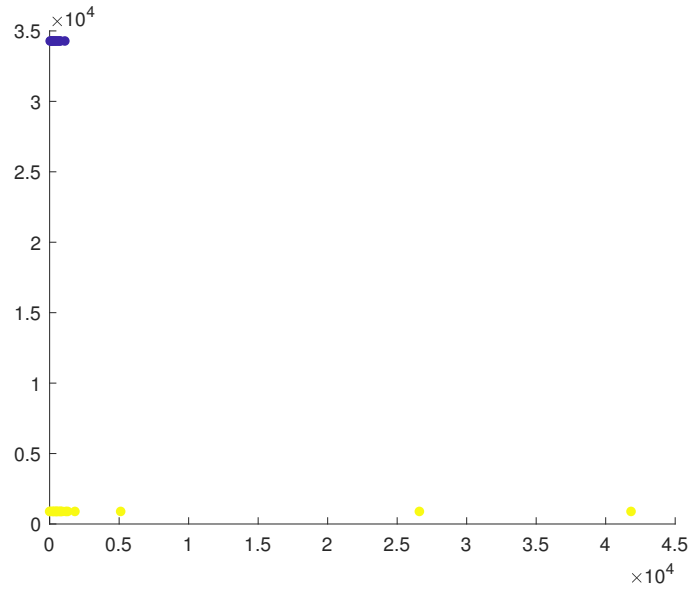


Figura 105: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 1 y 6 tipo wavelets de actividad ictal y no ictal.

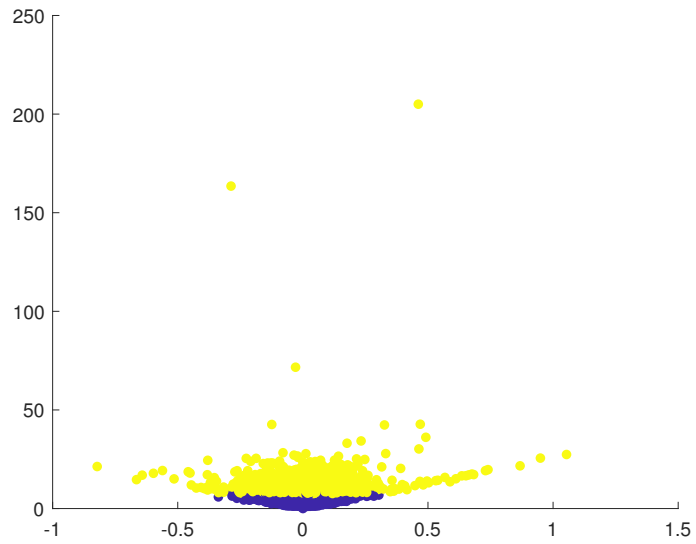


Figura 106: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 3 tipo wavelets de actividad ictal y no ictal.

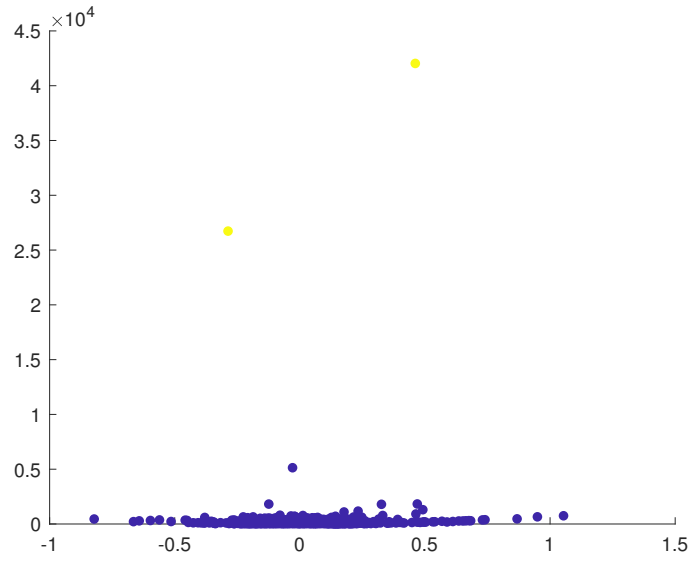


Figura 107: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 4 tipo wavelets de actividad ictal y no ictal.

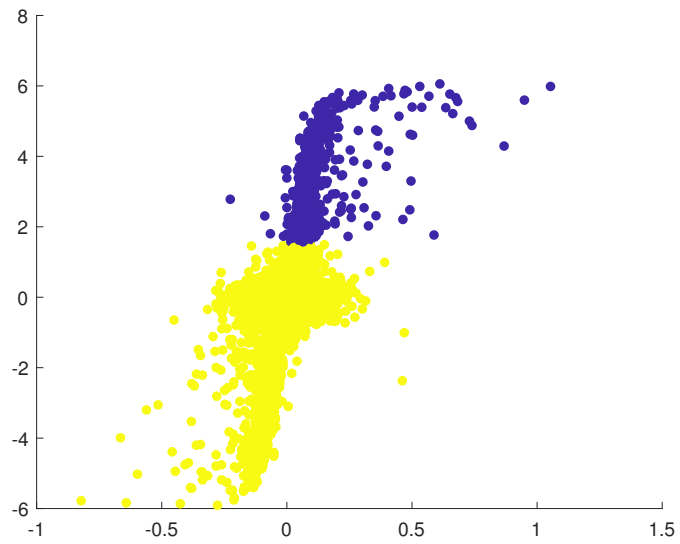


Figura 108: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 5 tipo wavelets de actividad ictal y no ictal.

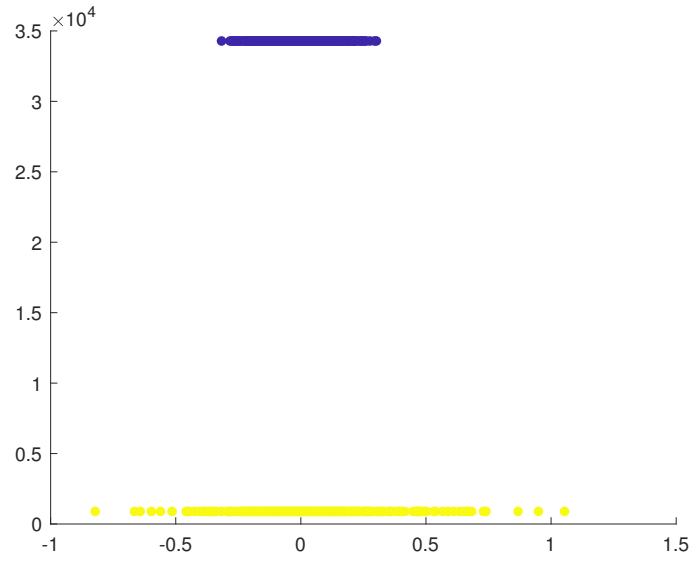


Figura 109: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 2 y 6 tipo wavelets de actividad ictal y no ictal.

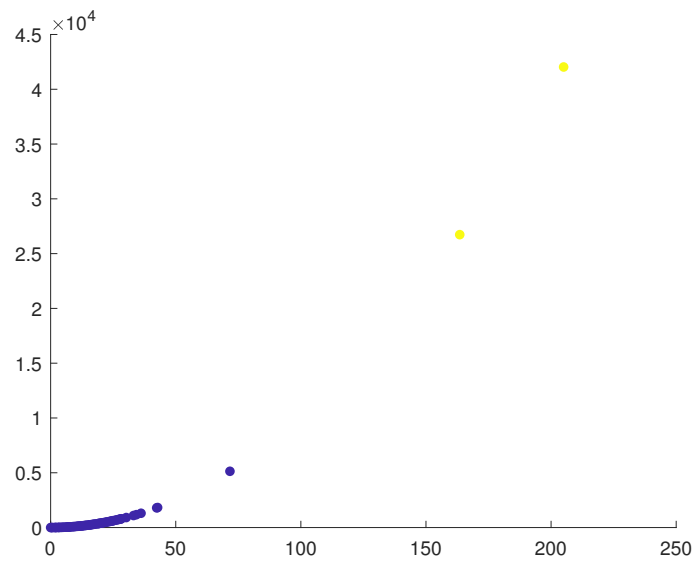


Figura 110: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 4 tipo wavelets de actividad ictal y no ictal.

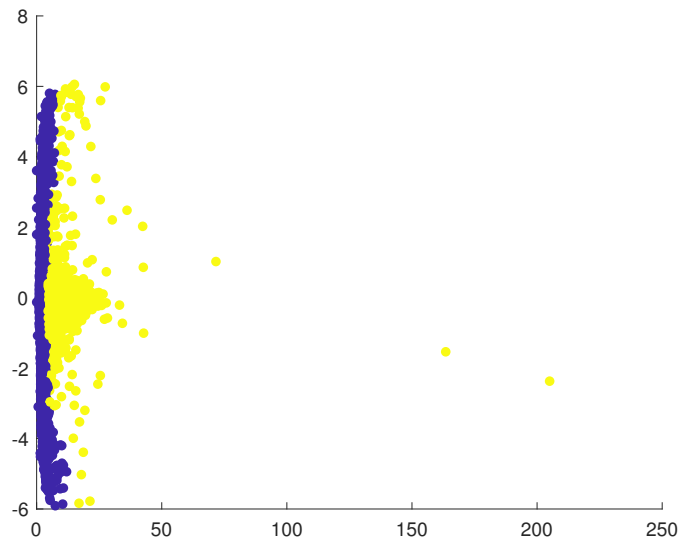


Figura 111: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 5 tipo wavelets de actividad ictal y no ictal.

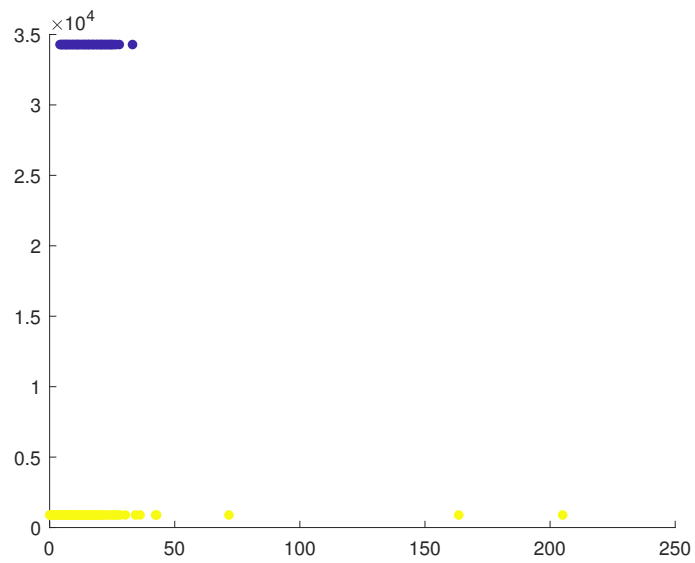


Figura 112: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 3 y 6 tipo wavelets de actividad ictal y no ictal.

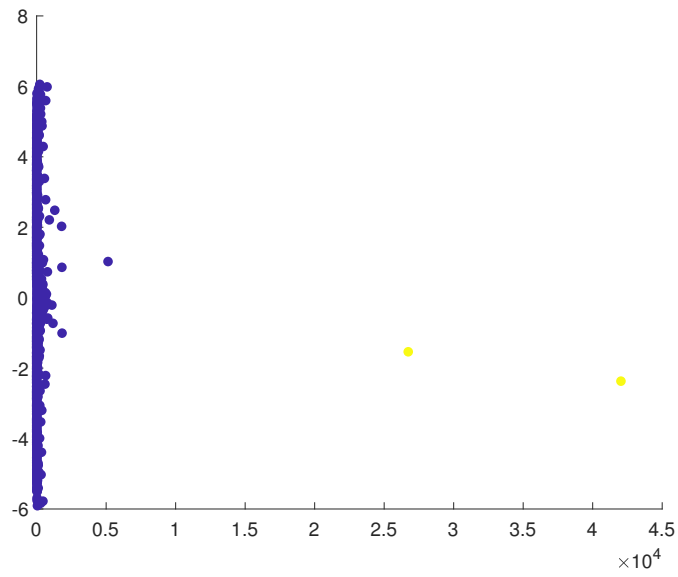


Figura 113: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 5 tipo wavelets de actividad ictal y no ictal.

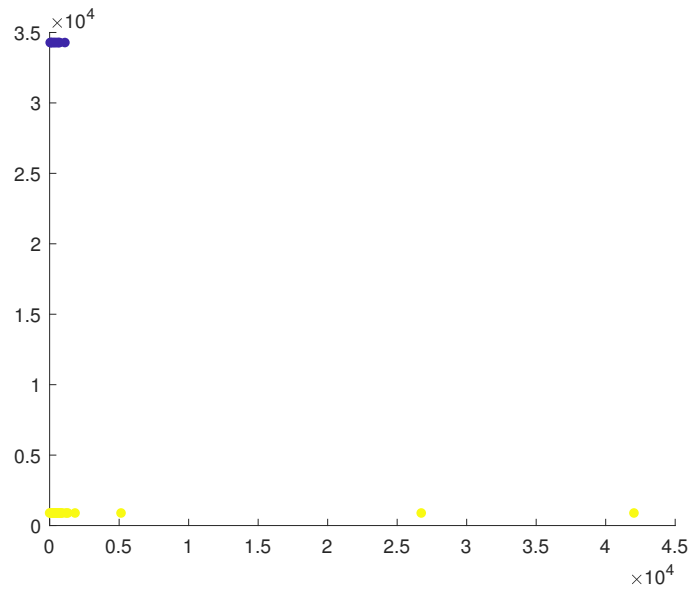


Figura 114: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 4 y 6 tipo wavelets de actividad ictal y no ictal.

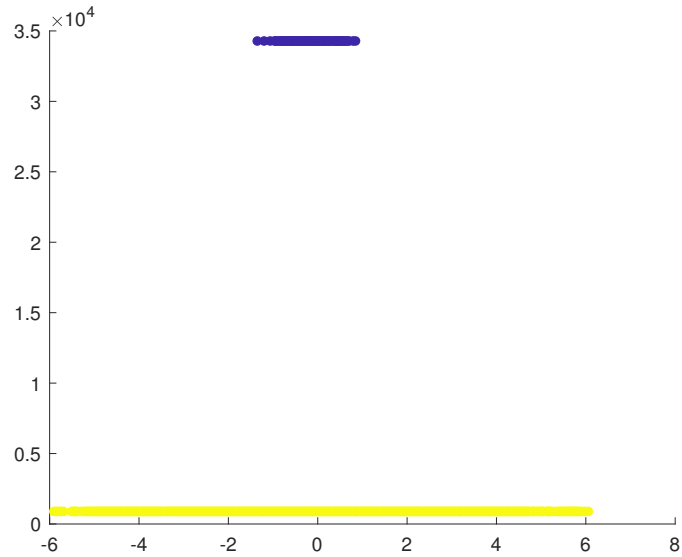


Figura 115: Gráfica de agrupamiento jerárquico de señales bioeléctricas de las características 5 y 6 tipo wavelets de actividad ictal y no ictal.

10.4. Herramienta para la evaluación visual de la tendencia

Se añadió la herramienta “VAT: *A Tool for Visual Assessment of (Cluster) Tendency*” como una herramienta valiosa para la evaluación visual de la tendencia de agrupación de datos en señales bioeléctricas como se observa en la Figura 116. Esta herramienta desempeñó un papel esencial al proporcionar una representación visual clara de la estructura de agrupación de datos en múltiples dimensiones. Su uso se justifica en la necesidad de comprender la tendencia de agrupación en un conjunto de características que puede ser complejo y visualmente abrumador al aplicar técnicas de agrupamiento como K-means o jerárquico.

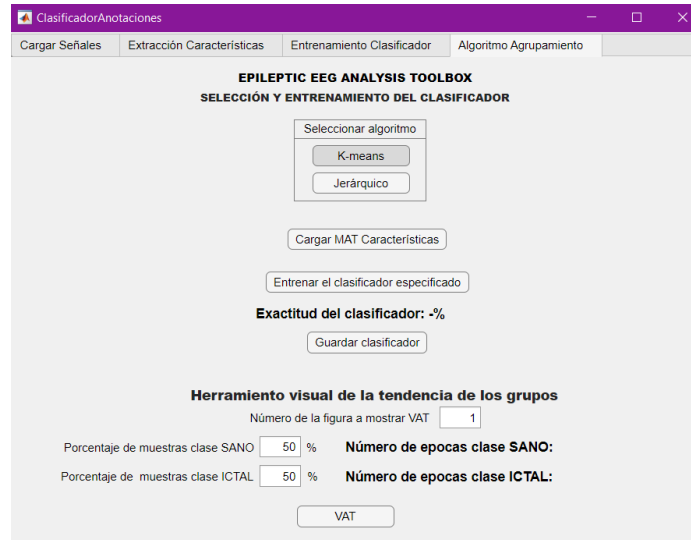


Figura 116: VAT en la herramienta de software para el estudio de la epilepsia.

Para la aplicación de VAT se utilizaron los vectores de características completos, mencionados en el capítulo anterior. Como se puede observar en la Figura 117 no es posible determinar una cantidad de grupos para las características en el dominio del tiempo. Lo cual implica que todos los datos independientemente de su clase (ictal o sano) están muy cerca unos de otros. Permitiendo así validar de forma visual que las características en el dominio del tiempo de las señales EEG no son las más óptimas.

En el caso de las características en el dominio de la frecuencia, revelaron ser más discriminatorias como se puede observar en la Figura 118. Se perciben varios subgrupos los cuales pertenecen a cada tipo de característica y clase. Se puede inferir que las características en el dominio de la frecuencia presentan un mejor comportamiento para realizar predicciones correctas. En cuanto a las características wavelets, son las que mejor discriminación realizan entre clases, como se puede observar en la Figura 119. Dejando bastante evidente la cantidad de clases que existen en el vector de características.

En las Figuras 118 y 119, se puede percibir que a la fecha no se cuenta con la misma cantidad de grabaciones de personas Ictales respecto a grabaciones de personas sanas.

La aplicación de “VAT” permitió una evaluación efectiva de la tendencia de agrupación en los datos de señales bioeléctricas. Los resultados generados por esta herramienta se alinearon con las expectativas teóricas y ayudaron a identificar patrones de agrupación en el conjunto de datos y visualizar el ruido que se puede encontrar entre los tipos de clases. Esta validación refuerza la fiabilidad y utilidad de la herramienta “VAT” en el contexto de este estudio y respalda su contribución a la comprensión de la tendencia de agrupación en señales bioeléctricas. En los anexos se presentan resultados adicionales a los presentados en esta sección.

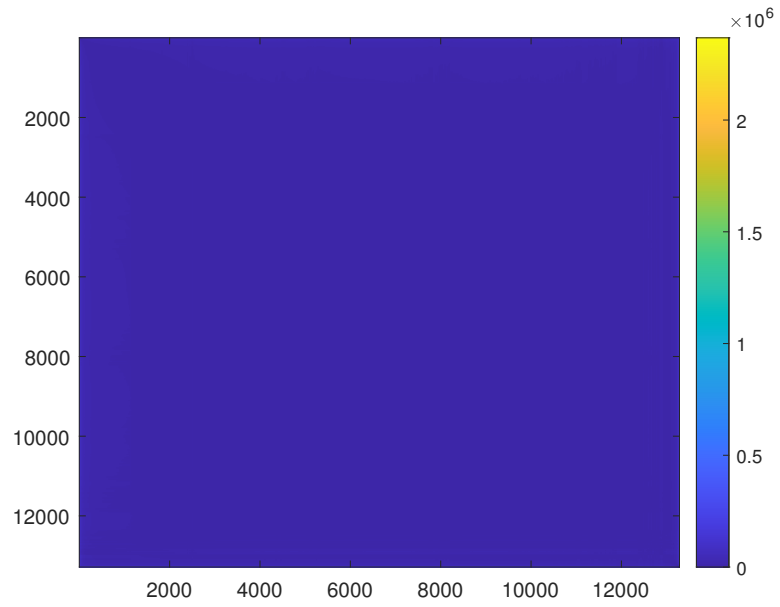


Figura 117: Resultado tras aplicar VAT al vector de características de 5 dimensiones en el dominio tiempo.

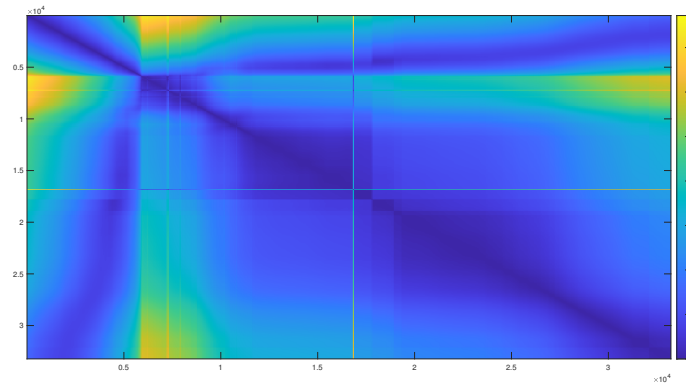


Figura 118: Resultado tras aplicar VAT al vector de características de 6 dimensiones en el dominio frecuencia.

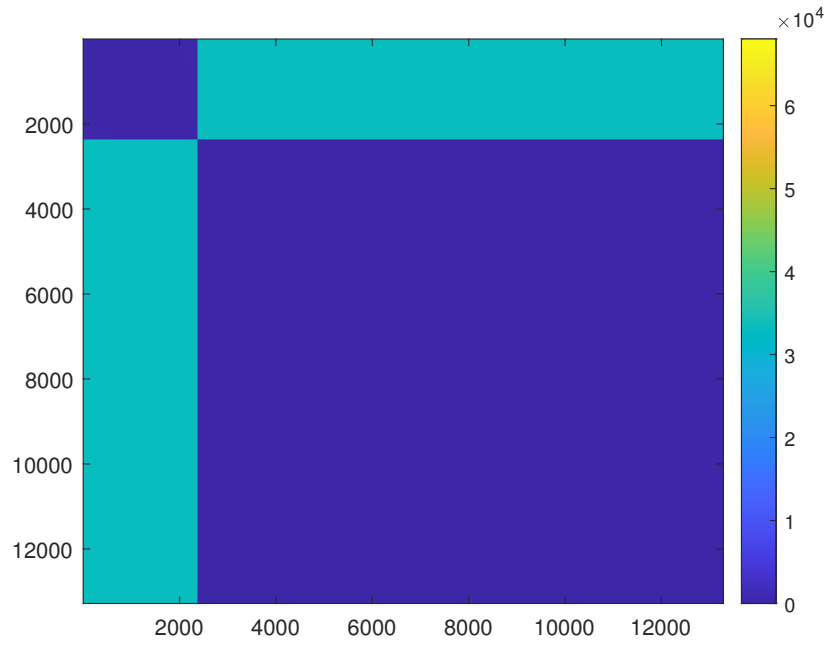


Figura 119: Resultado tras aplicar VAT al vector de características de 6 dimensiones en el dominio wavelets.

Actualización de la herramienta de software para el estudio de la epilepsia

La actualización de la herramienta de software para el estudio de la epilepsia ha resultado en una herramienta más simple de actualizar, ya que los códigos están mejor comentados lo que ayuda definitivamente para la documentación y modificación del mismo. La herramienta ahora cuenta con más programación defensiva, lo cual es de gran ayuda a que la herramienta de software para el estudio de la epilepsia no colapse cuando se produce una advertencia o error inesperado, o por el mal uso de la misma herramienta.

La presente versión de la herramienta de software para el estudio de la epilepsia concluirá siendo implementada en el trabajo de graduación de Diego Méndez titulado “Extensión, validación y migración de una herramienta de software para el estudio de la epilepsia para su uso en el Centro de Epilepsia y Neurocirugía Funcional HUMANA)” [35]. Dicha implementación permitirá obtener una herramienta de software para el estudio de la epilepsia capaz de ser ejecutada en cualquier computadora siendo amigable con el usuario, además de contar con una base de datos local.

Con las actualizaciones hechas, la herramienta tiene las siguientes características:

- Un software más liviano: al no contar con funciones y variables innecesarias, además de, contar con métodos más sofisticados para realizar tareas en específico y funciones nativas de MATLAB que se han descontinuado.
- Facilidad para manipulación de datos: para la importación y exportación de datos, ya que en la actual versión se pueden importar datos del tipo “edf”, “MAT” y de la base de datos local. La exportación de datos se da en formato “MAT”, “XLSX” y “CSV”.

Se implementó una funcionalidad crucial para mejorar la experiencia del usuario: la incorporación de ventanas de avisos. Estas notificaciones están diseñadas para mantener informado al usuario sobre las acciones en curso dentro de la herramienta de software, como

se observa en la Figura 120. Además, ofrecen indicaciones sobre el progreso de procesos extensos, alertan sobre la necesidad de esperar durante operaciones específicas y notifican cualquier problema o inconveniente que pueda surgir durante el uso de la herramienta. Esta adición no solo busca proporcionar una interacción más transparente, sino también brindar al usuario un mayor control y comprensión de las operaciones realizadas en la herramienta de software.



Figura 120: Ventana de avisos.

En el marco del continuo desarrollo y mejora de la herramienta de software, se ha integrado una funcionalidad fundamental para facilitar la comprensión y el depurado del sistema. Ahora, la consola de MATLAB presenta mensajes y anuncios, proporcionando información sobre las acciones que se están llevando a cabo dentro de la herramienta de software. Esta característica no solo contribuye al desarrollo continuo del software, sino que también facilita el proceso de depuración, permitiendo a los desarrolladores identificar y abordar eficientemente cualquier problema que pueda surgir durante la ejecución. Dicha inclusión de mensajes en la consola respalda así un entorno de desarrollo más efectivo y eficiente, observese la Figura 121.

```

EL usuario seleccionó C:\Users\Cris Pat\Desktop\U\Ser
User selected C:\Users\Cris Pat\Desktop\U\Semestre_10\
Procesando...

ans =

      1      409700

      173

Procesado

```

Figura 121: Notas en consola de MATLAB.

En la etapa previa del prototipo, se empleó MATLAB 2022, lo que demanda la actualización de las funciones para garantizar la funcionalidad, eficacia y sostenibilidad de la herramienta. En particular, se llevó a cabo la actualización de la función `insert()` debido a su futura falta de disponibilidad en versiones venideras. Como consecuencia, se ajustó el procedimiento de escritura en la base de datos para adecuarse a la función más reciente `sqlwrite()`.

En la fase anterior, se llevó a cabo una labor preliminar en el sistema de descarga de señales y su correspondiente metadata, como se observa en la Figura 122. En la fase actual, se ha completado satisfactoriamente esta funcionalidad. La incorporación de esta mejora posibilita la obtención de los datos de la señal en un archivo de texto tabulado, que se puede importar fácilmente en aplicaciones como Excel. Los archivos de texto se generan y almacenan en una ubicación seleccionada por el usuario, y se sugiere automáticamente un nombre descriptivo para dichos datos, como se muestra en la Figura 123. El proceso de descarga de pruebas se ha diseñado con la facilidad de uso en mente, de modo que solo se requiere seleccionar la señal y presionar el botón "Descargar".

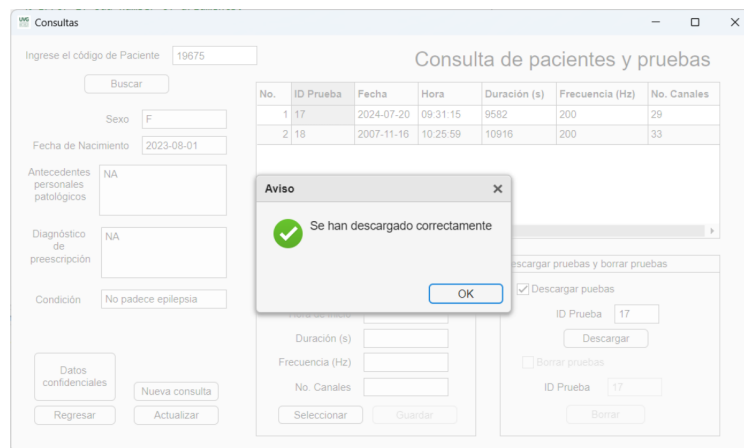


Figura 122: Aviso de descarga finalizada.

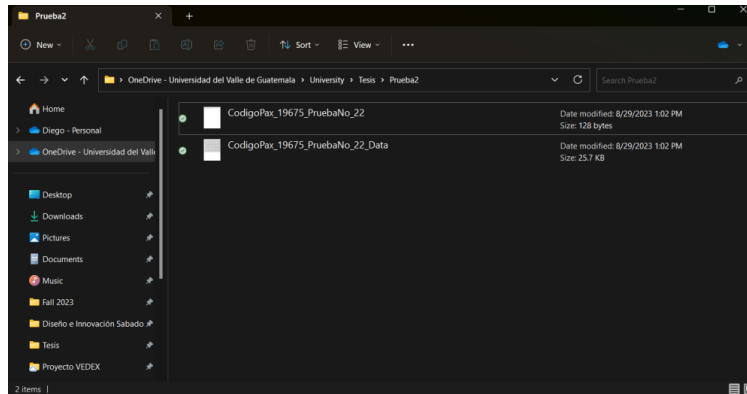


Figura 123: Archivos descargados.

Adicionalmente, se ha diseñado e implementado un procedimiento para la eliminación de datos. Esta adición facilita una mejor gestión de las pruebas almacenadas de cada paciente, proporcionando un recurso valioso para corregir registros erróneos y descartar datos almacenados por error.

En la Figura 124, se muestra la "Ventana Multigráfico Vertical", la cual posibilita la observación simultánea de varios canales para los profesionales. Este proceso se lleva a cabo mediante un desplazamiento vertical de 16 canales, cada uno separado verticalmente por 100 unidades. Cada canal está identificado con su número correspondiente en el eje vertical, siguiendo un enfoque similar al utilizado para el análisis de señales en HUMANA.

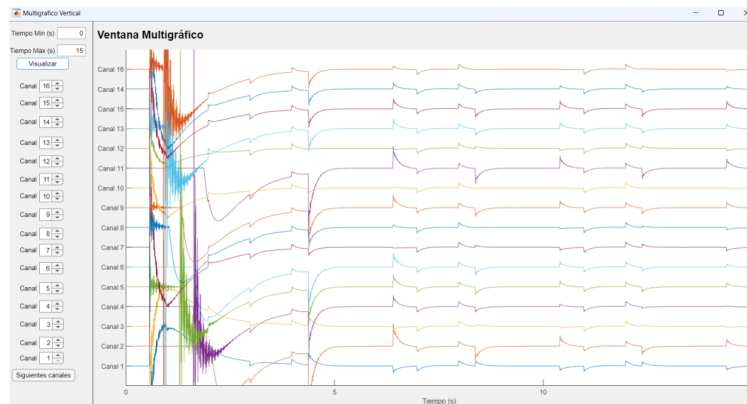


Figura 124: Funcionamiento de ventana multigráfico.

Exploración con algoritmo CHAMELEON

En el capítulo dedicado a la exploración con el algoritmo CHAMELEON, se presenta un análisis detallado de los desafíos enfrentados durante la implementación de esta herramienta en el marco de investigación. Uno de los retos iniciales significativos fue la instalación de las librerías necesarias, ya que, en su versión original, estas estaban diseñadas para sistemas operativos basados en Linux.

Se encontraron diversas dependencias desactualizadas que requerían intervención. El proceso de *debugging* y actualización de estas dependencias fue crucial para lograr una integración exitosa del algoritmo CHAMELEON en el entorno de desarrollo. Este proceso de adaptación y corrección de incompatibilidades permitió superar los obstáculos iniciales y poner en marcha el algoritmo de agrupamiento con las condiciones necesarias para llevar a cabo un análisis efectivo de nuestras señales bioeléctricas.

Considerando lo anterior y su futura implementación en la herramienta de software para el estudio de la epilepsia, se creó un manual para la correcta instalación de las dependencias de librería y las modificaciones pertinentes para que esta pueda funcionar, como se observa en la Figura 125.

```
1 """
2
3     CRISTHOFER PATZAN
4     UNIVERSIDAD DEL VALLE DE GUATEMALA
5     2023
6
7     El siguiente manual son los pasos a seguir para ejecutar el algoritmo
8     CHAMELEON, te recomiendo hacer todo desde un virtual enviroment de
9     python3.
10
11     Se utilizo CentOS, para la compilación, debuggin y uso de la presente
12     libreria.
13 */
14 """
15
16
17
18 #-----
19 #           Instalar CMAKE
20 #-----
21 """/* Debes intsalar cmake desde fuente
22 - Descargar version estable
23 - Desde la carpeta donde este el ".tar" abres una terminal
24 - tar xvf cmake-3.27.9.tar.gz
25 - cd cmake-3.27.9/
26 - ./configure
27 - Si al finalizar este instruccion te aparece un warning de
28   open ssl, entonces instala las dependencias (libssl-dev)
29   Para el caso de centos es:
30   - yum install -y openssl-devel
31   - ejecuta nuevamente:
32   - ./configure
```

Figura 125: Manual de instalación y modificación a dependencias para el uso del algoritmo CHAMELEON.

El algoritmo CHAMELEON demostró su eficacia en la clasificación de grupos mediante la utilización de datos de prueba, revelando resultados altamente efectivos. A pesar de que los datos de prueba no siguieron una distribución de secciones cónicas, el algoritmo exhibió una notable capacidad para clasificarlos de manera precisa y coherente, como se observan en las Figuras 126 y 127. Esta capacidad de adaptación y rendimiento destacado frente a datos con formas diversas resalta la robustez y versatilidad del algoritmo CHAMELEON en el contexto de nuestro estudio. Los resultados obtenidos refuerzan la idoneidad de este enfoque de agrupamiento para la tarea específica de clasificación de señales bioeléctricas, subrayando su potencial aplicabilidad en escenarios complejos y variados.

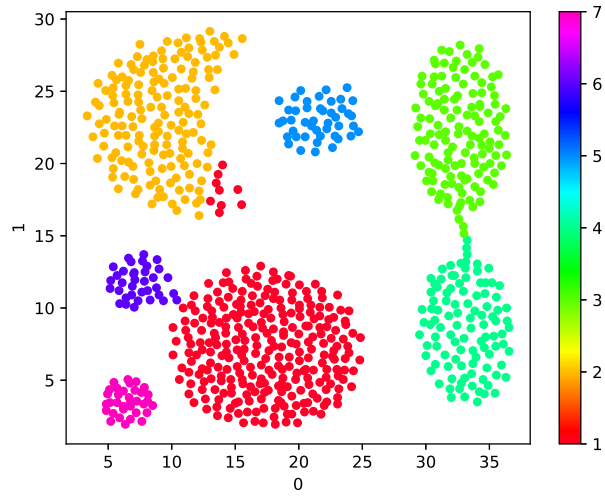


Figura 126: Datos de prueba con secciones cónicas para algoritmo CHAMELEON.

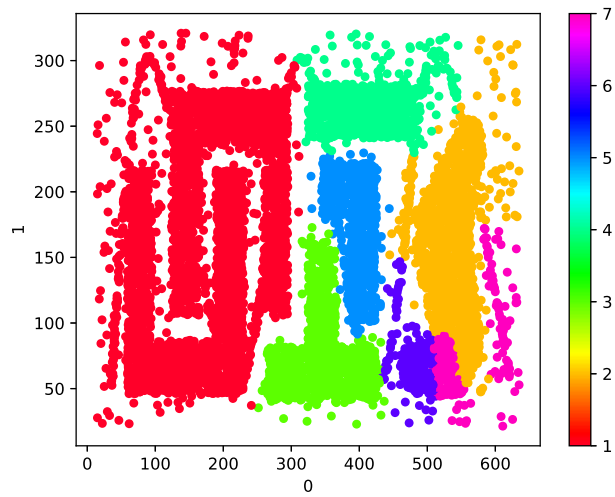


Figura 127: Datos de prueba con secciones no cónicas para algoritmo CHAMELEON.

- La obtención de señales bioeléctricas permitió contar con un conjunto diverso y representativo para su posterior análisis. La recopilación de datos de pacientes con epilepsia atendidos en el Instituto HUMANA y de registros capturados con el equipo de la Universidad del Valle de Guatemala proporcionó una base sólida para la investigación y el desarrollo de la herramienta de software.
- Tras los análisis realizados las características que mejor resultados han presentado para la predicción de clases, son en el dominio tiempo-frecuencia con las transformadas wavelets, además de, ser las que menos costo computacional requieren.
- La aplicación de algoritmos de aprendizaje automático previamente desarrollados para la extracción de características de las señales EEG y EMG fue exitosa. Estos algoritmos demostraron ser efectivos en la identificación de patrones relevantes en las señales bioeléctricas y en la generación de características que sirvieron como entrada para los procesos de detección de segmentos de interés. Aunque se percibe la necesidad de contar con una mayor cantidad de datos de distintas personas por parte de HUMANA.
- Los análisis estadísticos realizados arrojaron resultados positivos en términos del rendimiento de los algoritmos implementados. Estos análisis proporcionaron una base sólida para la toma de decisiones y la optimización de los algoritmos utilizados.
- La revisión y mejora de la herramienta de software desarrollada en fases anteriores ha culminado en una versión notablemente más intuitiva y amigable para el usuario. Las implementaciones realizadas en los algoritmos de clasificación y detección de segmentos de interés representan un avance significativo en la automatización de procesos clínicos relacionados con la epilepsia. También simplifica la tarea de actualizar y depurar la herramienta, permitiendo un desarrollo continuo y eficiente.

- Se recomienda continuar la recopilación de datos de señales bioeléctricas, tanto de pacientes con epilepsia como de sujetos sanos, con el fin de enriquecer la base de datos y mejorar la capacidad de detección de patrones. La inclusión de un conjunto más amplio y diverso de datos permitirá una validación más robusta de los algoritmos desarrollados, añadiendo otro tipo de señales bioeléctricas como electrocardiogramas.
- Considerando el constante avance y desarrollo de algoritmos en el campo de la Inteligencia Artificial (IA) y Aprendizaje Automático (ML), se sugiere la creación de una versión de la herramienta en Python. Dada la prevalencia y la amplia disponibilidad de bibliotecas, herramientas y algoritmos en Python para IA y ML, esta iniciativa permitirá una integración más fluida y la incorporación sencilla de nuevas técnicas y avances en el campo. Python se ha consolidado como un lenguaje de programación ampliamente aceptado en el ámbito de la ciencia de datos, lo que facilitará la colaboración y el desarrollo futuro de la herramienta en un entorno más versátil y adaptable a las innovaciones en este campo en constante evolución.
- Se sugiere realizar una validación clínica más amplia de la herramienta de software desarrollada en entornos médicos reales. La colaboración continua con instituciones médicas, como el Instituto HUMANA, puede proporcionar una plataforma para aplicar la metodología en la práctica médica y evaluar su utilidad en el diagnóstico y tratamiento de pacientes con epilepsia.
- Dado el potencial de los algoritmos de aprendizaje automático para identificar patrones en señales bioeléctricas, se recomienda explorar la aplicación de esta metodología en el estudio de otros trastornos neurológicos. Investigar la detección de patrones relacionados con otras afecciones cerebrales puede ampliar significativamente el impacto y la relevancia de la investigación.

-
- [1] E. Pichon Rivière, “Patogenia y dinanismos de la epilepsia,” 2017.
 - [2] O. organización mundial de la salud, *Epilepsia*, feb. de 2023. dirección: <https://www.who.int/es/news-room/fact-sheets/detail/epilepsy>.
 - [3] C. de Epilepsia y Neurocirugía Funcional, *HUMANA*, *Epilepsia*, mar. de 2021. dirección: <https://humanagt.org/epilepsia/>.
 - [4] 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.
 - [5] M. J. Angulo, “Análisis y Reconocimiento de Patrones de Señales Biomédicas de Pacientes con Epilepsia,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2020.
 - [6] 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.
 - [7] J. E. Mendizabal y L. F. Salguero, “Prevalence of epilepsy in a rural community of Guatemala,” *Epilepsia*, vol. 37, n.º 4, págs. 373-376, 1996.
 - [8] L. Rouhiainen, “Inteligencia artificial,” *Madrid: Alienta Editorial*, 2018.
 - [9] M. Félix y L. Carlos, “Data mining: torturando a los datos hasta que confiesen,” *Coordinador del programa de Data mining*, 2002.
 - [10] E. Wyllie, A. Gupta y D. K. Lachhwani, *The treatment of epilepsy: principles & practice*. Lippincott Williams & Wilkins, 2006.
 - [11] 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.

- [12] R. S. Fisher, H. Cros, J. A. French, N. Higurashi, E. Hirch y F. E. Jansen, *Clasificación operacional de los tipos de crisis por la Liga Internacional contra la Epilepsia: Documento - Posición de la Comisión para Clasificación y Terminología de la ILAE*, 2017. dirección: <https://www.ilae.org/files/ilaeGuideline/Spanish-Traduccion-Fisher-Seizure-types.pdf>.
- [13] D. D. Tinoco y D. F. Gudiño, *Redes neuronales en la Caracterización de Señales bioeléctricas*, 2018. dirección: https://virtual.cuautitlan.unam.mx/intar/?page_id=977.
- [14] 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%B3n_V3_N7.pdf.
- [15] L. Sörnmo y P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic press, 2005, vol. 8.
- [16] P. R. Vallat, *Compute the average bandpower of an EEG signal*, 2018. dirección: <https://raphaelvallat.com/bandpower.html>.
- [17] A. Phinyomark, C. Limsakul y P. Phukpattaranont, "Application of wavelet analysis in EMG feature extraction for pattern classification," *Measurement Science Review*, vol. 11, n.º 2, pág. 45, 2011.
- [18] L. Sörnmo y P. Laguna, "Electrocardiogram (ECG) signal processing," *Wiley encyclopedia of biomedical engineering*, 2006.
- [19] J. A. Zavala-Villeda, "Descripción del electrocardiograma normal y lectura del electrocardiograma," *Revista mexicana de anestesiología*, vol. 40, n.º S1, págs. 210-213, 2017.
- [20] P. Boonyakitanton, A. Lek-uthai, K. Chomtho y J. Songsiri, "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG," 2019. arXiv: 1908.00492 [eess.SP].
- [21] I. Stancin, M. Cifrek y A. Jovic, "A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems," *Sensors (Basel)*, vol. 21, n.º 11, 2021.
- [22] A. S. Al-Fahoum y A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *International Scholarly Research Notices*, vol. 2014, 2014.
- [23] P. M. Bentley y J. McDonnell, "Wavelet transforms: an introduction," *Electronics & communication engineering journal*, vol. 6, n.º 4, págs. 175-186, 1994.
- [24] R. S. Pathak, *The wavelet transform*. Springer Science & Business Media, 2009, vol. 4.
- [25] 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%201a%5C%20gu%5C%C3%5C%ADa-,Definici%5C%C3%5C%B3n%5C%20de%5C%20aprendizaje%5C%20autom%5C%C3%5C%A1tico,de%5C%20grandes%5C%20cantidades%5C%20de%5C%20datos..>
- [26] B. Nguyen Cong, J. L. Rivero Pérez y C. Morell, "Aprendizaje supervisado de funciones de distancia: estado del arte," *Revista Cubana de Ciencias Informáticas*, vol. 9, n.º 2, págs. 14-28, 2015.

- [27] J. C. Tello y S. Informáticos, “Reconocimiento de patrones y el aprendizaje no supervisado,” *Universidad de Alcalá, Madrid*, 2007.
- [28] D. Pascual, F. Pla y S. Sánchez, “Algoritmos de agrupamiento,” *Método Informáticos Avanzados*, págs. 164-174, 2007.
- [29] D. P. González, “Algoritmos de agrupamiento basados en densidad y validación de clusters,” Tesis doct., Universitat Jaume I, 2010.
- [30] G. Karypis, E.-H. Han y V. Kumar, “Chameleon: Hierarchical clustering using dynamic modeling,” *computer*, vol. 32, n.º 8, págs. 68-75, 1999.
- [31] *Data Acquisition, Loggers, Amplifiers, Transducers, Electrodes / BIOPAC*, 2023. dirección: <https://www.biopac.com/>.
- [32] J. C. Bezdek y R. J. Hathaway, “VAT: A tool for visual assessment of (cluster) tendency,” en *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, IEEE, vol. 3, 2002, págs. 2225-2230.
- [33] M. Vela, “Análisis cualitativo y cuantitativo del impacto de los pulsos binaurales en el estado de ánimo, concentración y calidad del sueño de las personas,” Tesis de licenciatura, trabajo en progreso, Universidad Del Valle de Guatemala, 2023.
- [34] O. Fuentes, “Aplicación de técnicas de aprendizaje automático y reconocimiento de patrones a señales bioeléctricas para el estudio del impacto de los pulsos binaurales en el estado ánimo, concentración y calidad de sueño de las personas,” Tesis de licenciatura, trabajo en progreso, Universidad Del Valle de Guatemala, 2023.
- [35] D. Mendez, “Extensión, validación y migración de una herramienta de software para el estudio de la epilepsia para su uso en el Centro de Epilepsia y Neurocirugía Funcional (HUMANA),” Tesis de licenciatura, trabajo en progreso, Universidad Del Valle de Guatemala, 2023.

16.1. Sitios web con herramienta de software y datos recolectados

Aquí, se presentan enlaces a los sitios web específicos donde se puede acceder a la herramienta de software mejorada desarrollada durante el proyecto, así como a los conjuntos de datos recolectados a lo largo de la investigación. Este recurso facilita a los lectores y colaboradores la exploración directa de la herramienta y la revisión de los datos utilizados en el estudio, brindando una visión integral y accesible de los resultados y contribuciones generadas.

- GitHub: https://github.com/pat19218/Tesis_2023
- Banco de datos: Drive

16.2. Pruebas con algoritmo VAT

En la sección dedicada a las pruebas con el algoritmo VAT, se llevó a cabo un exhaustivo análisis de las características de las señales bioeléctricas en distintos dominios: frecuencia, tiempo y wavelets. Cada prueba se diseñó para evaluar la capacidad del algoritmo VAT en la identificación de tendencias de agrupamiento bajo variadas condiciones. Con el propósito de estudiar el impacto de la cantidad de épocas empleadas en el proceso, se realizaron pruebas sucesivas, incrementando gradualmente este parámetro. Este enfoque sistemático permitió una evaluación detallada de la efectividad del algoritmo en la visualización y discernimiento de patrones de agrupamiento en las señales, brindando perspectivas valiosas sobre su rendimiento en diferentes contextos y condiciones.

16.2.1. VAT características en el dominio de frecuencia

Se exploraron detalladamente los resultados obtenidos al aplicar este algoritmo para la visualización de patrones de agrupamiento en las señales bioeléctricas. La representación gráfica generada por el VAT, donde se distinguen los grupos a través de la variación de colores, proporciona una visión integral de la estructura de los datos. Aunque se observa cierta diferenciación entre los grupos, la distinción no es tan pronunciada, lo cual plantea un escenario interesante en términos de la complejidad de los patrones presentes en las señales. Esta característica, aunque implica un mayor desafío en la identificación visual de los grupos, puede brindar *insights* valiosos sobre la naturaleza sutil de las relaciones entre las señales, destacando la importancia de técnicas adicionales para una interpretación más precisa y detallada como se observan en las Figuras 128 - 135.

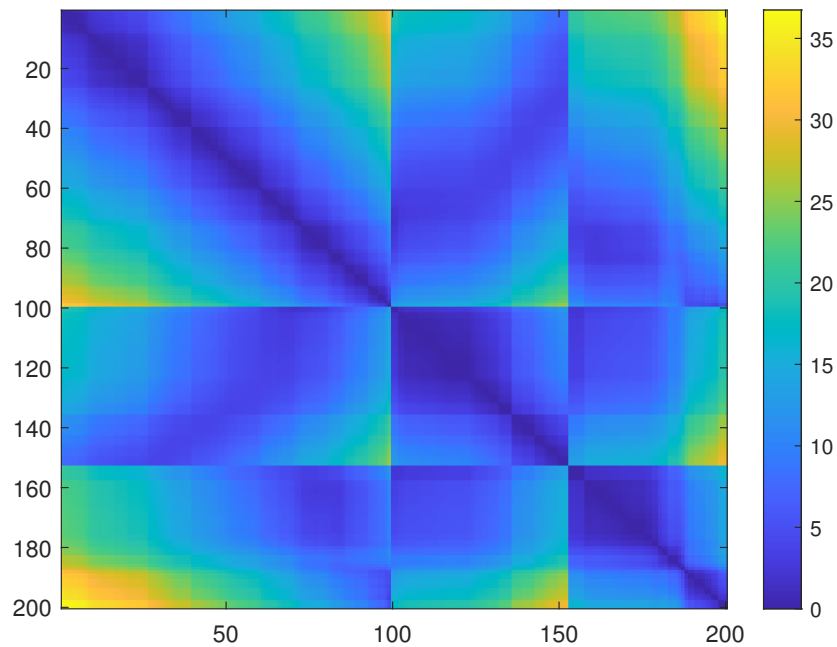


Figura 128: VAT usando 100 épocas de características tipo frecuencia.

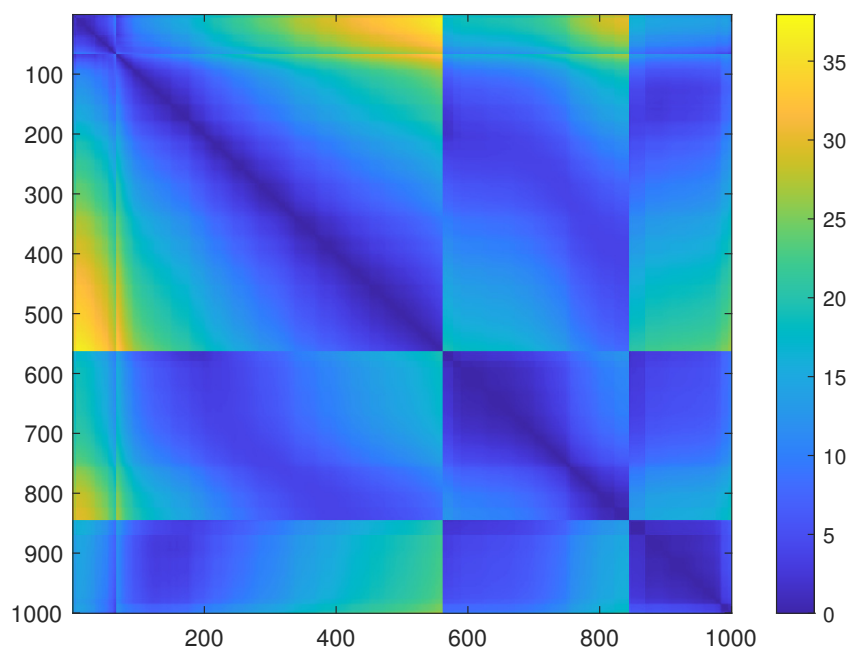


Figura 129: VAT usando 500 épocas de características tipo frecuencia.

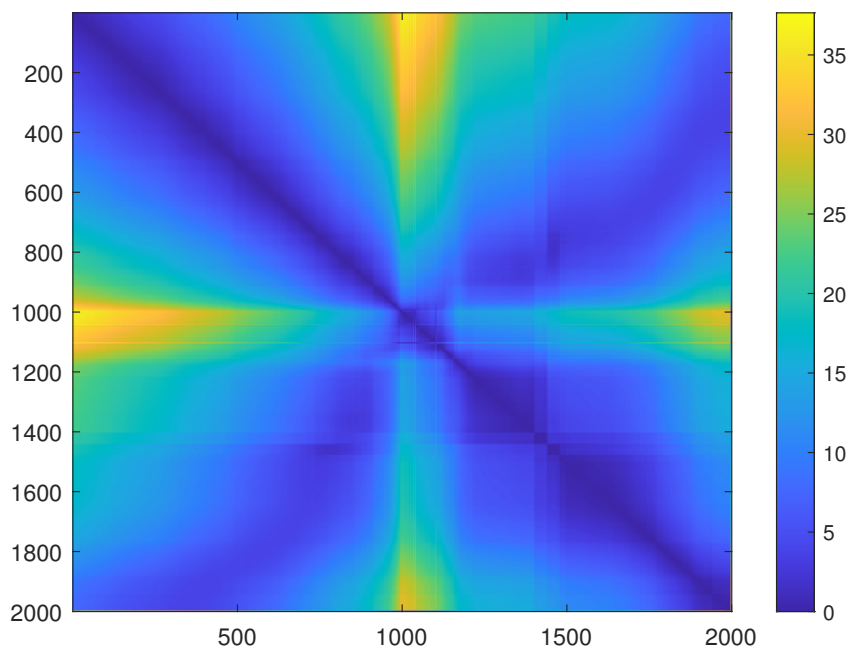


Figura 130: VAT usando 1000 épocas de características tipo frecuencia.

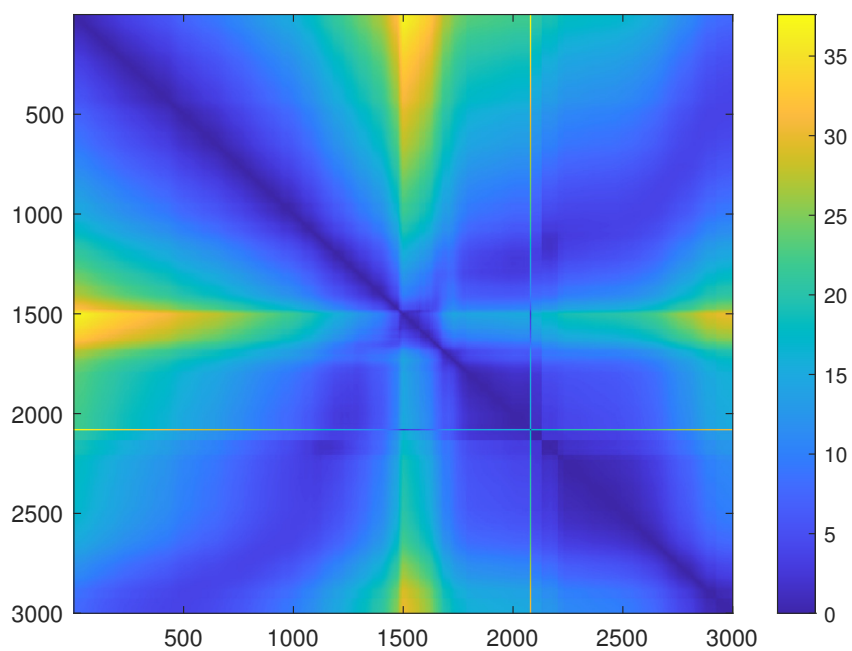


Figura 131: VAT usando 1500 épocas de características tipo frecuencia.

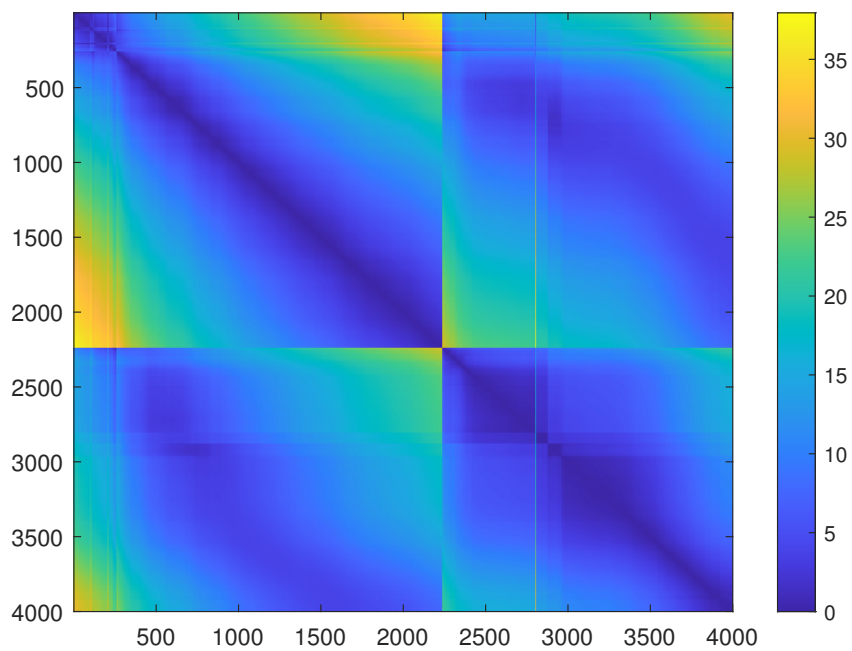


Figura 132: VAT usando 2000 épocas de características tipo frecuencia.

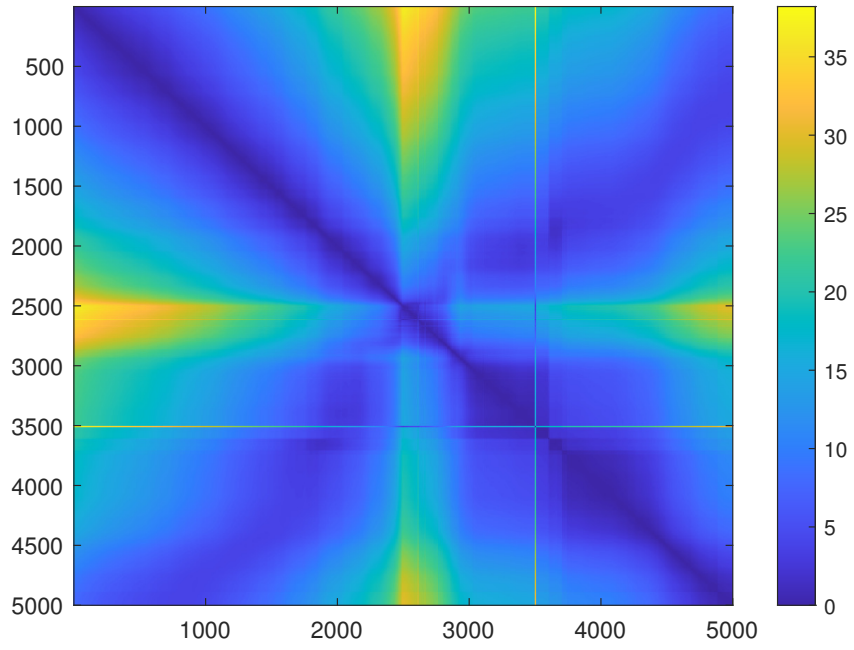


Figura 133: VAT usando 2500 épocas de características tipo frecuencia.

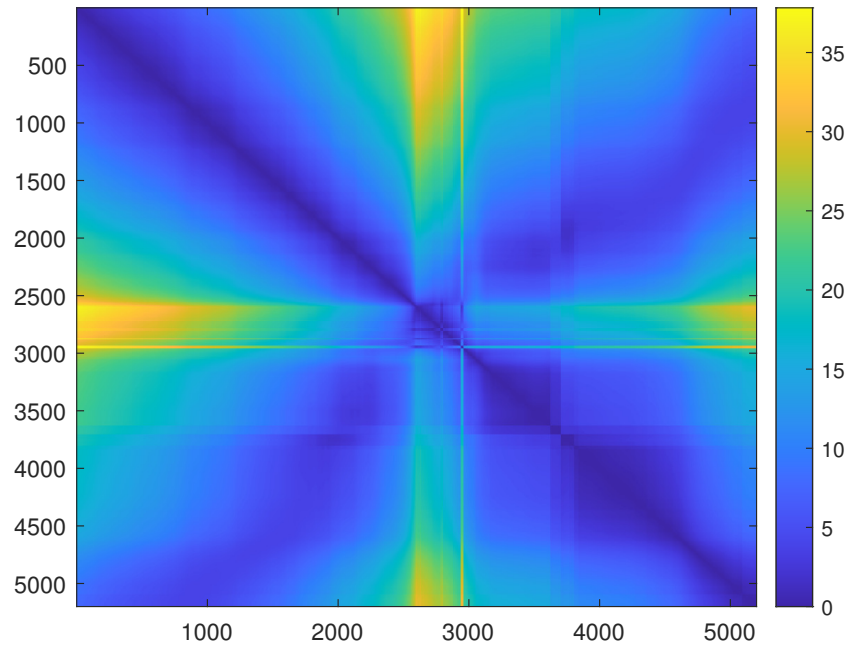


Figura 134: VAT usando 2600 épocas de características tipo frecuencia.

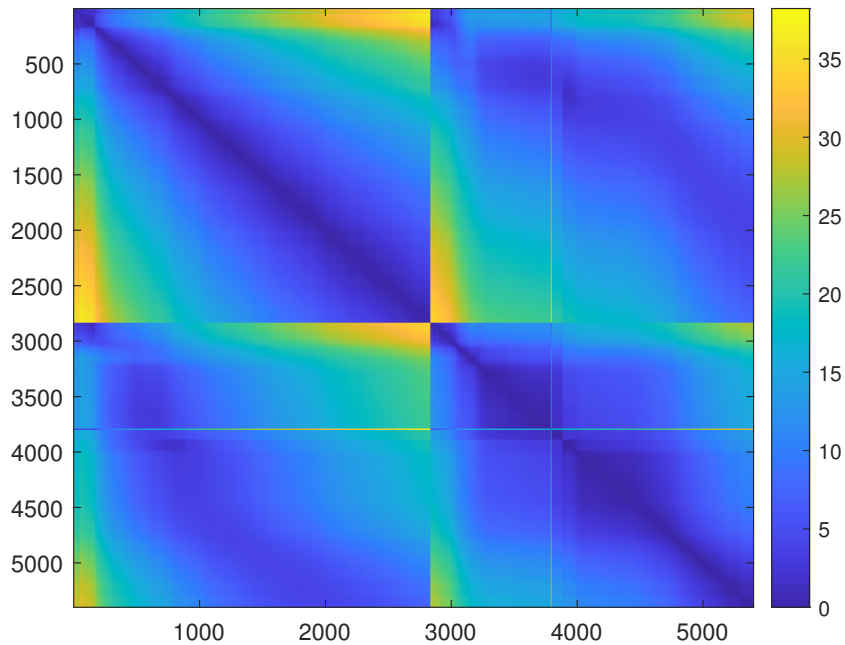


Figura 135: VAT usando 2700 épocas de características tipo frecuencia.

16.2.2. VAT características en el dominio de tiempo continuo

En la exploración de las características en el dominio del tiempo continuo mediante la aplicación del algoritmo VAT, se observa una distribución particular de los datos. La representación visual destaca por la ausencia de una clara diferenciación de grupos, ya que la variación de colores revela una proximidad significativa entre todos los puntos de datos. Esta cercanía en la representación visual sugiere que las características analizadas en este contexto particular pueden no ser tan discriminatorias como se esperaba para identificar patrones distintivos de agrupamiento. Este hallazgo insta a una reflexión más profunda sobre la naturaleza de las características consideradas y la exploración de enfoques alternativos para la identificación de patrones en el dominio del tiempo continuo, como se observa en las Figuras 136 - 143.

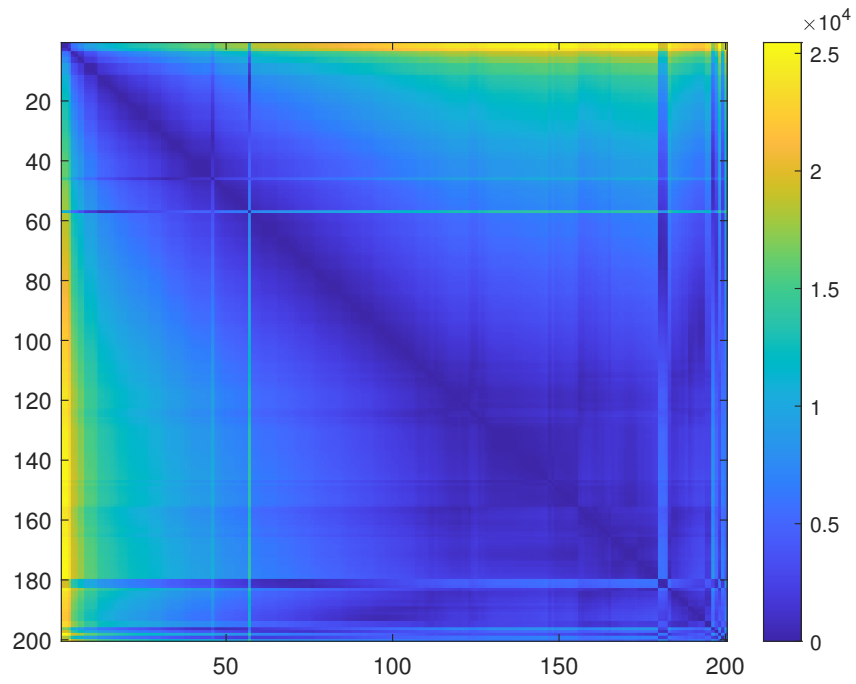


Figura 136: VAT usando 100 épocas de características tipo tiempo continuo.

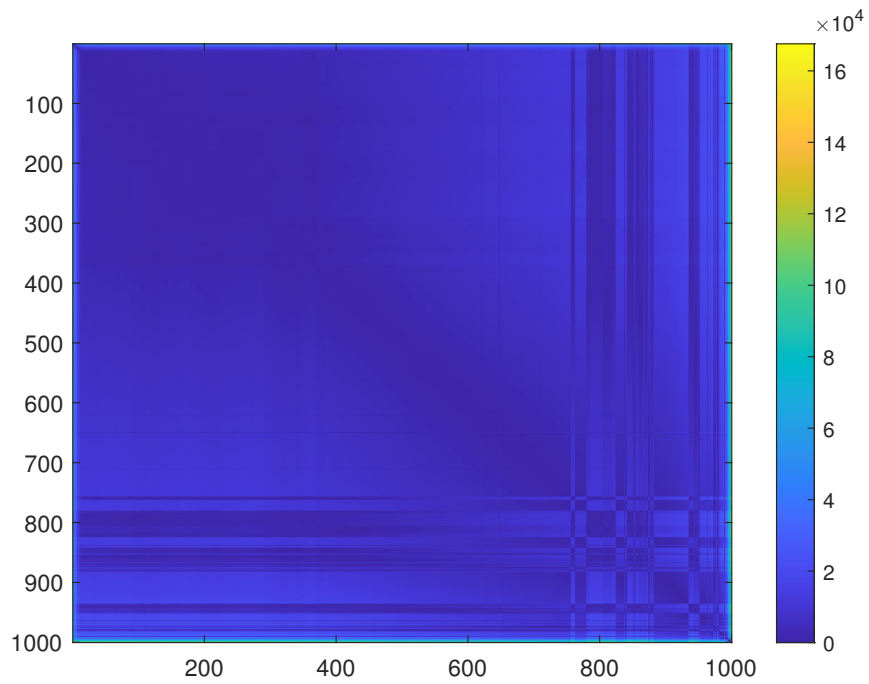


Figura 137: VAT usando 500 épocas de características tipo tiempo continuo.

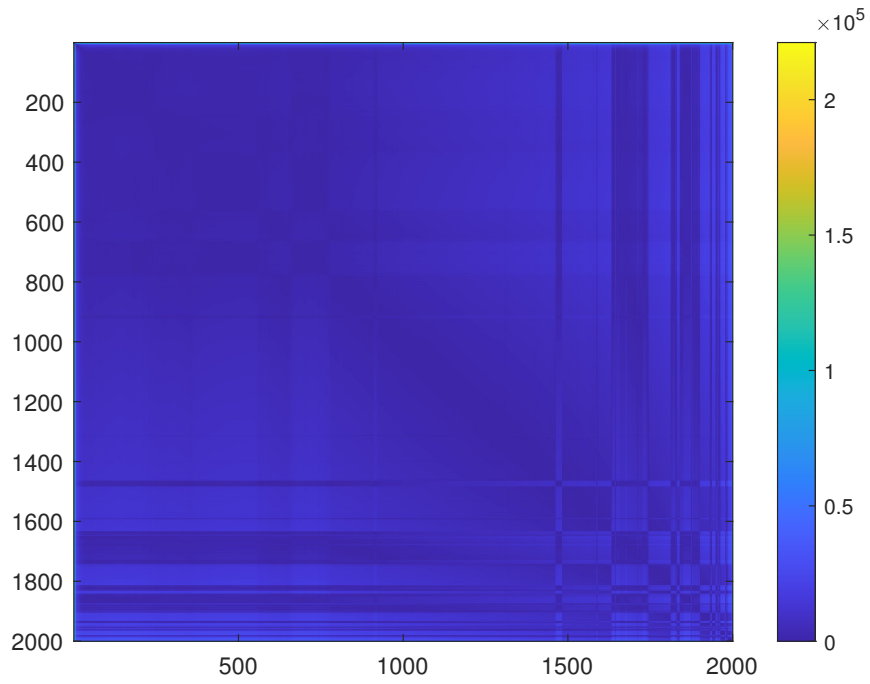


Figura 138: VAT usando 1000 épocas de características tipo tiempo continuo.

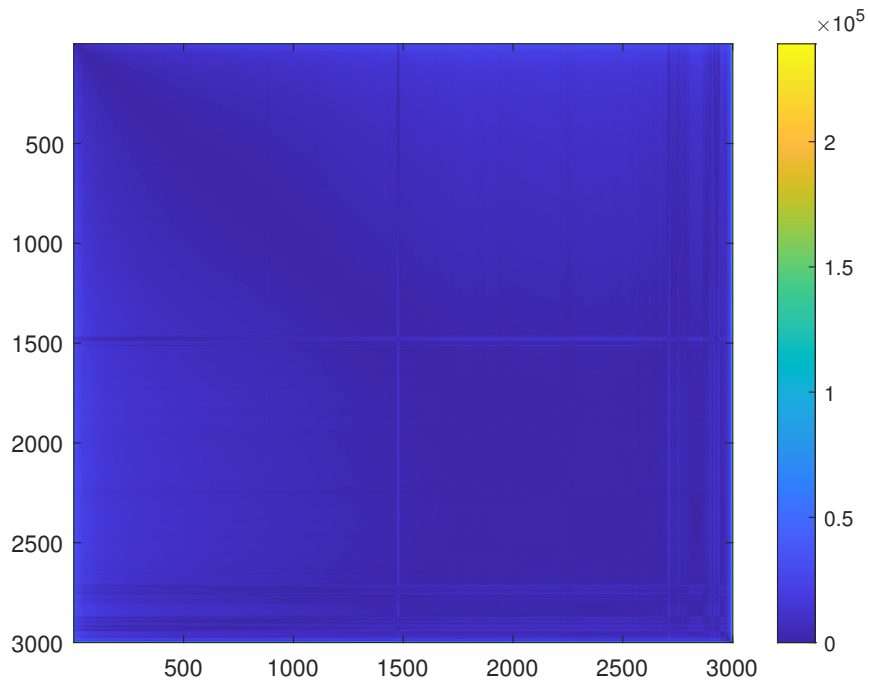


Figura 139: VAT usando 1500 épocas de características tipo tiempo continuo.

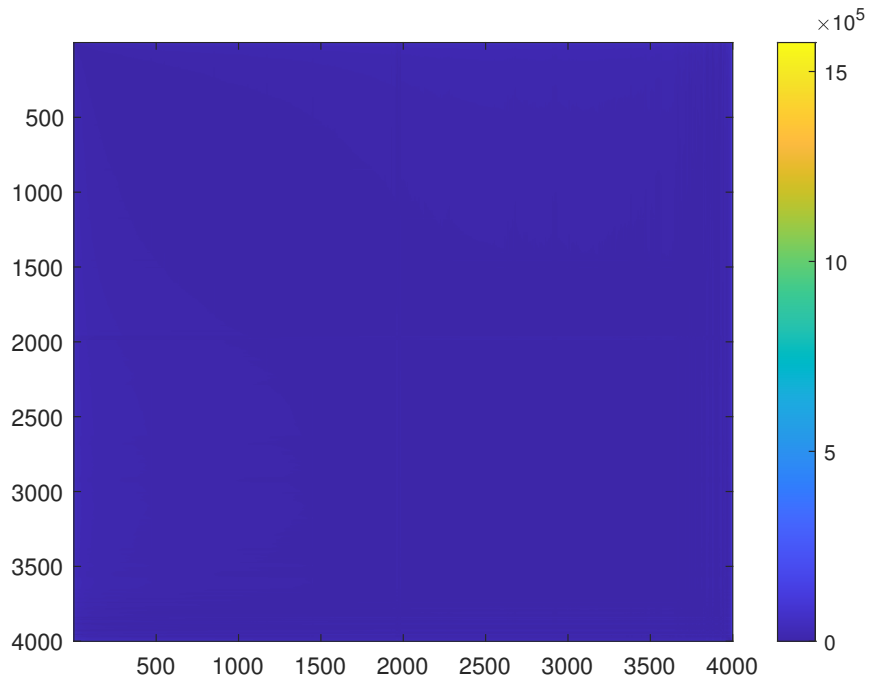


Figura 140: VAT usando 2000 épocas de características tipo tiempo continuo.

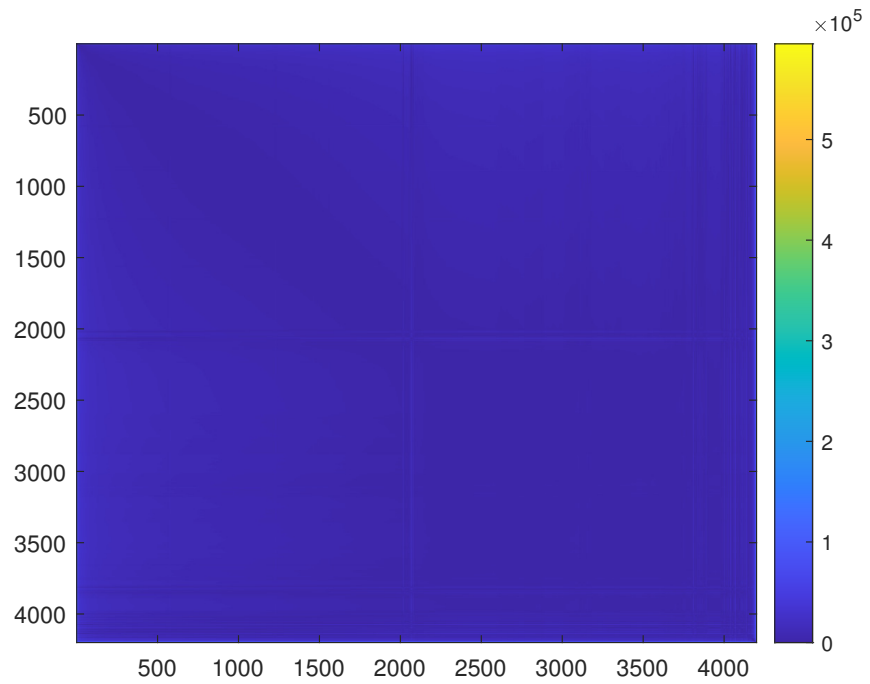


Figura 141: VAT usando 2100 épocas de características tipo tiempo continuo.

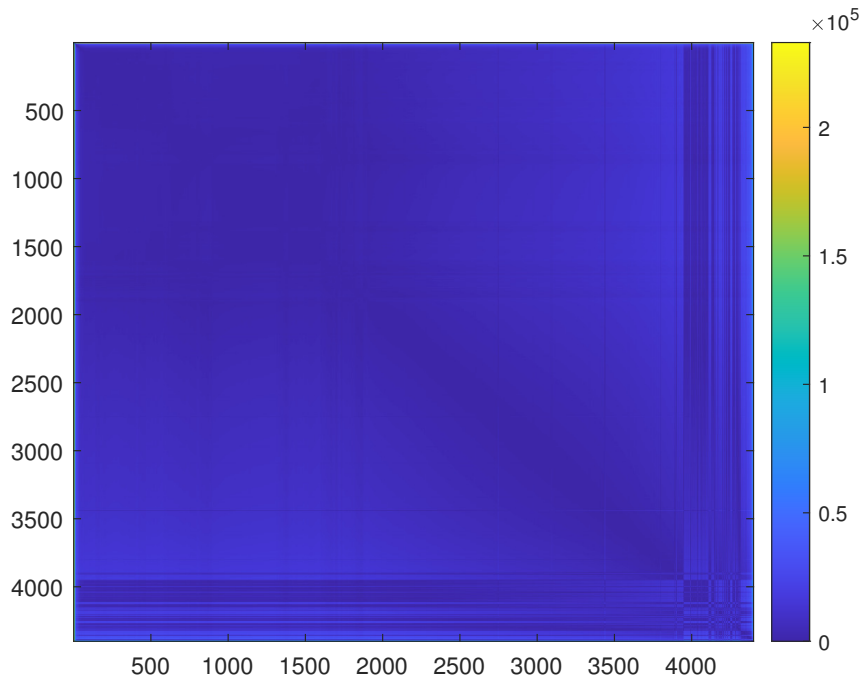


Figura 142: VAT usando 2200 épocas de características tipo tiempo continuo.

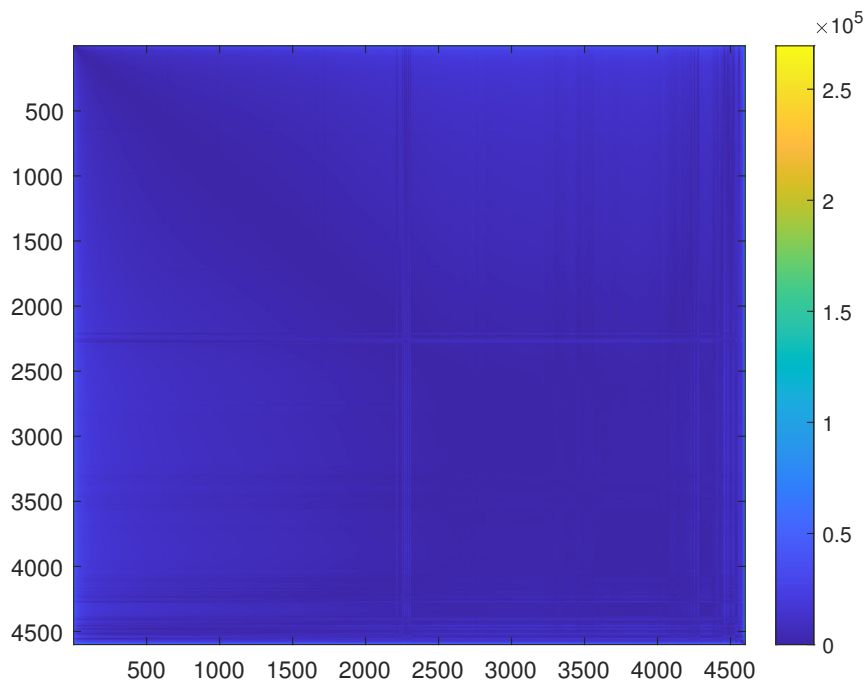


Figura 143: VAT usando 2300 épocas de características tipo tiempo continuo.

16.2.3. VAT características en el dominio de wavelets

La aplicación del algoritmo VAT a las características extraídas en el dominio de wavelets revela resultados notables en la diferenciación de patrones en las señales bioeléctricas. La representación visual muestra una clara distinción entre dos grupos, evidenciada por la marcada variación de colores. Esta observación sugiere que las características en el dominio de wavelets utilizadas en el análisis poseen una capacidad discriminatoria sobresaliente, permitiendo una separación efectiva entre las señales asociadas a estados sanos y ictales. La nitidez en la identificación de estos grupos mediante VAT en el dominio de wavelets subraya la importancia y relevancia de estas características para el análisis y clasificación de señales bioeléctricas. Estos resultados prometedores proporcionan una base sólida para considerar el dominio de wavelets como un enfoque valioso en la identificación de patrones distintivos en señales bioeléctricas relacionadas con la epilepsia. Para ello observar las Figuras 144 - 151.

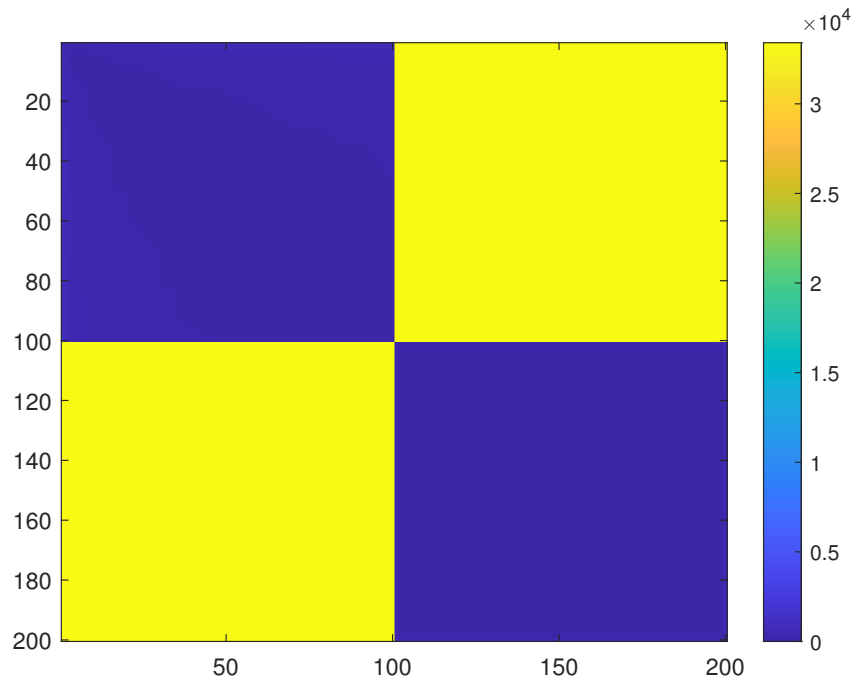


Figura 144: VAT usando 100 épocas de características tipo wavelets.

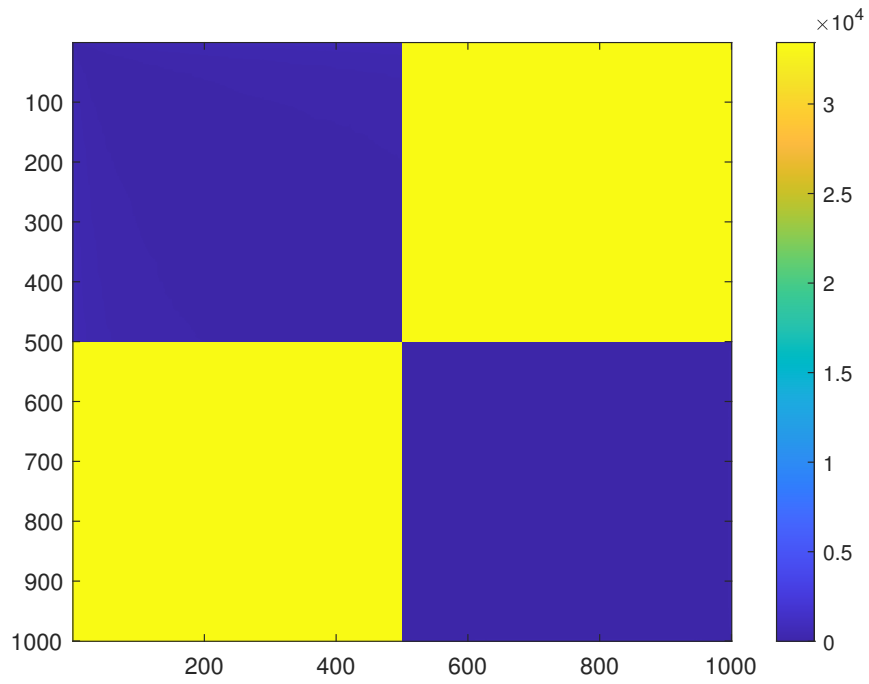


Figura 145: VAT usando 500 épocas de características tipo wavelets.

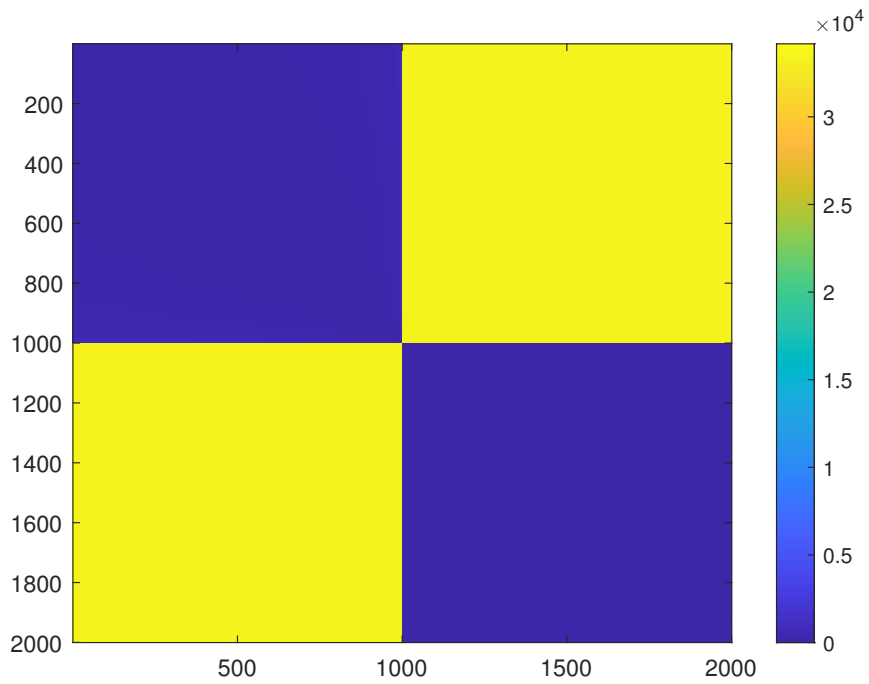


Figura 146: VAT usando 1000 épocas de características tipo wavelets.

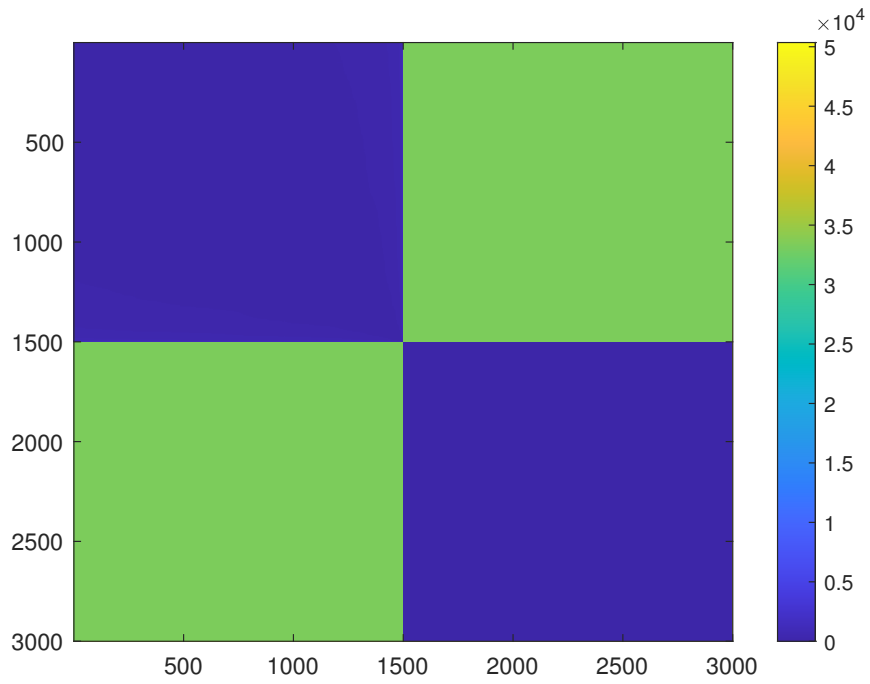


Figura 147: VAT usando 1500 épocas de características tipo wavelets.

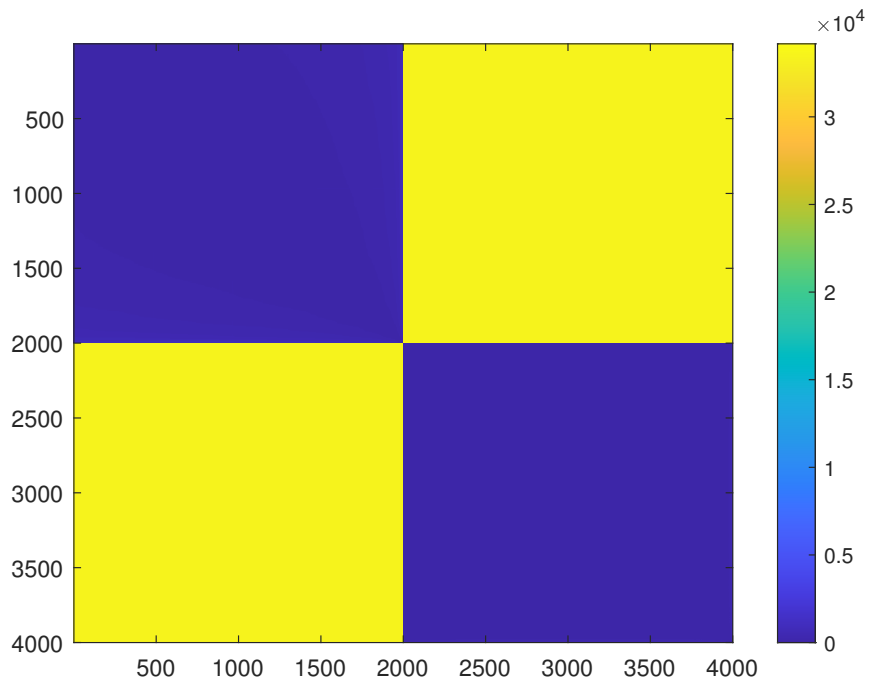


Figura 148: VAT usando 2000 épocas de características tipo wavelets.

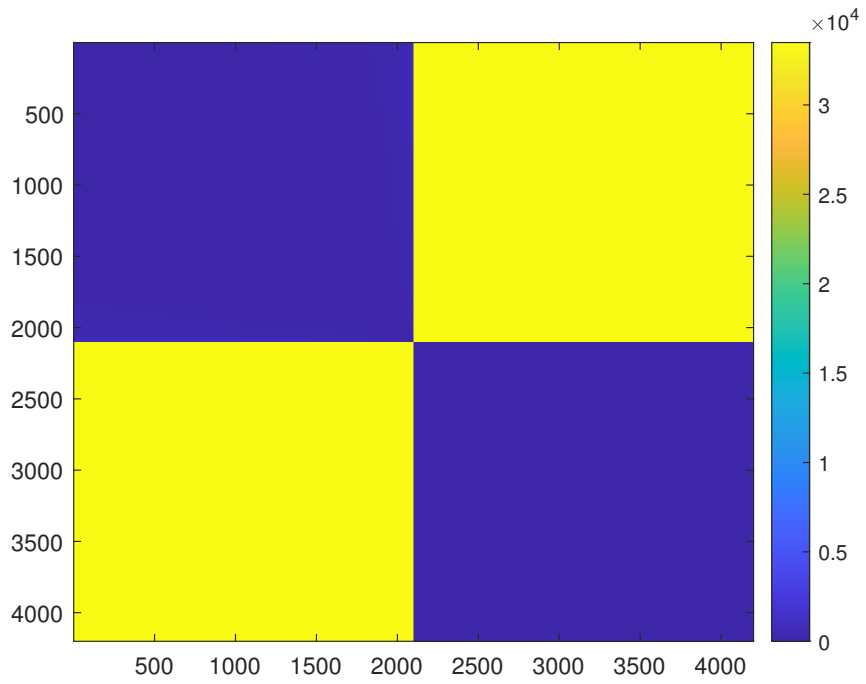


Figura 149: VAT usando 2100 épocas de características tipo wavelets.

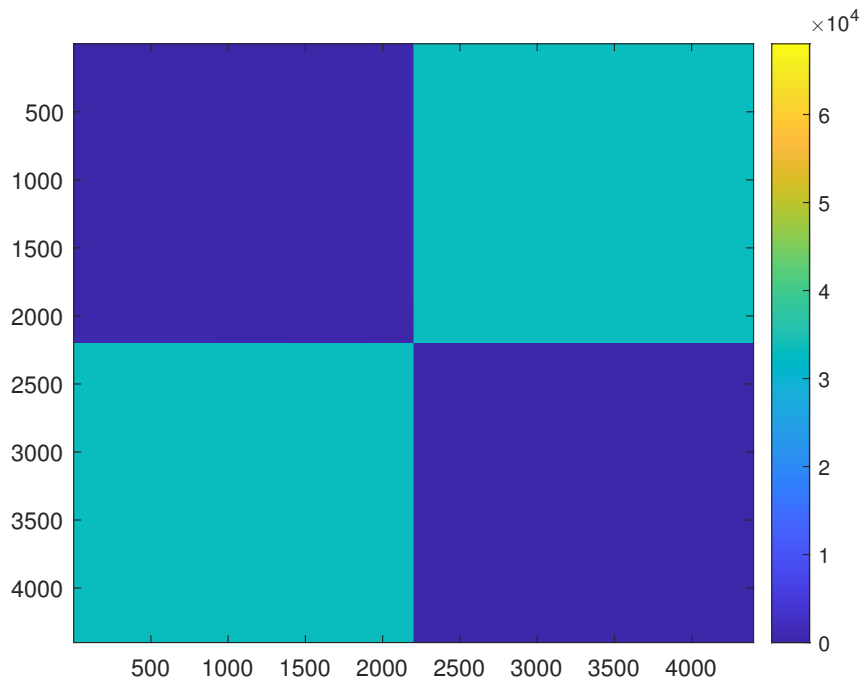


Figura 150: VAT usando 2200 épocas de características tipo wavelets.

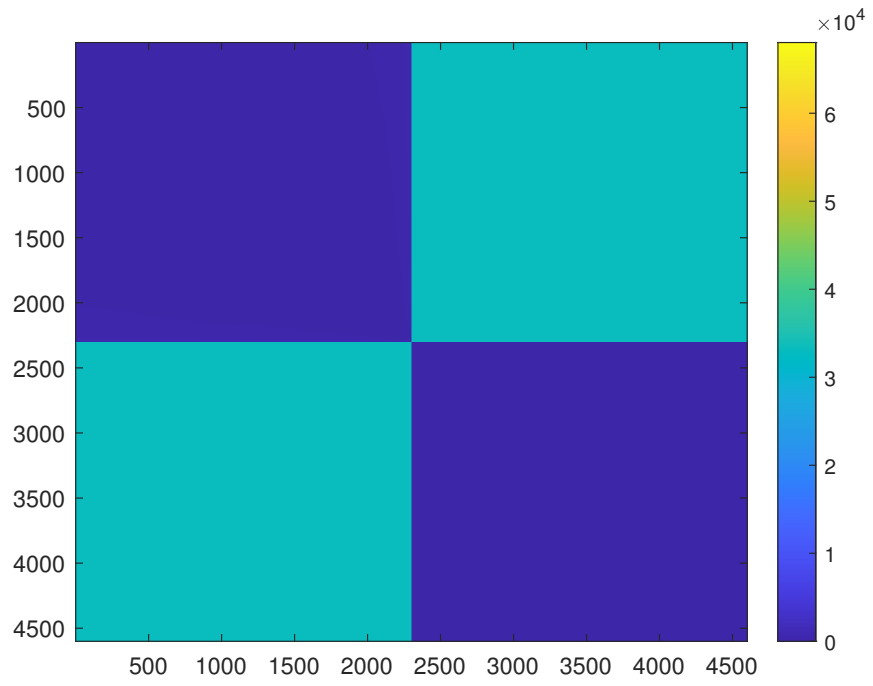


Figura 151: VAT usando 2300 épocas de características tipo wavelets.

Algoritmo: algoritmo es una secuencia finita de instrucciones o reglas bien definidas, diseñadas para realizar una tarea o resolver un problema específico. [3](#)

Frecuencia de muestreo: Se refiere al número de muestras de una señal analógica que se toman en un período de tiempo determinado para convertirla en una señal digital. En otras palabras, es la cantidad de veces que se registran o se toman mediciones de una señal analógica en un segundo. [27](#)

Intersujeto: Es un término que se utiliza en investigaciones y análisis para referirse a las comparaciones o análisis que se realizan entre diferentes sujetos o participantes en un estudio. Se refiere a la variabilidad o diferencias que existen entre individuos distintos en una muestra o grupo de estudio. [28](#)

Intrasujeto: Se refiere a algo que ocurre o se aplica dentro del mismo sujeto o individuo. En el contexto de la investigación científica, especialmente en estudios clínicos o experimentales, el término se utiliza para describir la variabilidad o los efectos dentro de un mismo individuo a lo largo del tiempo o en diferentes condiciones. [28](#)

Kernel: En el contexto de aprendizaje automático y estadísticas, el término “kernel” se refiere a una función matemática que mide la similitud entre pares de datos en un espacio de características. [3](#)

Matrices de confusión: Es una herramienta que se utiliza en el campo de la clasificación en aprendizaje automático para evaluar el rendimiento de un modelo predictivo. Esta matriz presenta de manera sistemática la comparación entre las predicciones de un modelo y las clases reales de un conjunto de datos. [40](#)

Neurociencia: La neurociencia es un campo multidisciplinario que se dedica al estudio del sistema nervioso, que incluye el cerebro, la médula espinal y los nervios. [2](#)

RNA: En el ámbito del aprendizaje automático y la inteligencia artificial, una Red Neuronal Artificial se refiere a un modelo computacional inspirado en la estructura y funcionamiento del cerebro humano. [40](#)

Struct: Es una abreviatura de “estructura” y se refiere a un tipo de datos que se utiliza para almacenar datos relacionados de manera organizada. Una estructura en MATLAB es un contenedor que puede contener diferentes tipos de datos, incluidos escalares, matrices, cadenas de texto y otras estructuras. [27](#)

SVM: Es un algoritmo de aprendizaje supervisado utilizado tanto para tareas de clasificación como de regresión. SVM es particularmente eficaz en espacios de alta dimensión y es ampliamente utilizado en problemas de aprendizaje automático y minería de datos. [40](#)

Wavelets: Son funciones matemáticas que tienen una duración limitada y están localizadas tanto en el tiempo como en la frecuencia. Son utilizadas en el análisis de señales y procesamiento de imágenes para representar y analizar información de manera eficiente. [29](#)