



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para obtener el Título de INGENIERO DE SISTEMAS

APLICACIÓN DE TÉCNICAS DE LA MINERÍA DE DATOS PARA CARACTERIZAR SEÑALES ELECTROENCEFALOGRÁFICAS DE PACIENTES EPILÉPTICOS Y SUJETOS SANOS

Por

Br. Emilio José Zambrano Briceño

Tutor: Dr. Pablo Guillén

Marzo 2010

©2010 Universidad de Los Andes, Mérida, Venezuela

APLICACIÓN DE TÉCNICAS DE LA MINERÍA DE DATOS PARA CARACTERIZAR SEÑALES ELECTROENCEFALOGRÁFICAS DE PACIENTES EPILÉPTICOS Y SUJETOS SANOS

Br. Emilio José Zambrano Briceño

Proyecto de Grado — Investigación de Operaciones, 61 páginas
Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2010

Resumen: El electroencefalograma es uno de los exámenes complementarios más utilizados en el diagnóstico de la Epilepsia. Esto es debido a que en él se refleja el comportamiento de la actividad cerebral, permitiendo así a los especialistas, entender los mecanismos que causan alteraciones de la misma en ciertas patologías como la Epilepsia. El análisis computarizado de registros electroencefalográficos mediante técnicas como las de Minería de Datos, permite la caracterización de los cambios de complejidad que ocurren en la actividad cerebral. Con este propósito, se realiza un estudio con técnicas de Minería de Datos: Medida de Disimilaridad Basada en Compresión (CDM, Compression-based Dissimilarity Measure, por sus siglas en inglés), Agrupamiento Jerárquico y Clasificación con el algoritmo de los k vecinos cercanos, para obtener patrones que permitan cuantificar los cambios de complejidad que ocurren en la patología epiléptica, a través del análisis computarizado del electroencefalograma. La base de datos utilizada en este estudio está constituida por un grupo de cinco pacientes epilépticos con paroxismos (entre los que se encuentra un registro antes y durante una crisis focal registrada accidentalmente en la consulta) y un grupo de cinco sujetos normales (grupo de control). Entre los resultados obtenidos se encontró que los pacientes epilépticos presentan valores de CDM más bajos que los del grupo control, verificando ésto una sincronización (similaridad) de la actividad cerebral de los mismos al comparar entre los canales del EEG. Seguidamente, con los valores de CDM obtenidos para el grupo de pacientes epilépticos y el grupo de sujetos sanos se realiza un agrupamiento jerárquico, separando ésto ampliamente la actividad cerebral entre

canales para cada grupo y cuantificando nuevamente una sincronización entre canales para el grupo de pacientes epilépticos. Una eficiencia del 100% de clasificación entre sujetos sanos y pacientes epilépticos es obtenida con el algoritmo de k vecinos cercanos. Se concluye que índices de complejidad como el presentado en este trabajo para el análisis computarizado del EEG de pacientes epilépticos, podrían ser de gran utilidad para el análisis de la presencia de una crisis, la ubicación de focos epilépticos y a un mejor diagnóstico médico mediante métodos no invasivos.

Palabras clave: Medida de Disimilaridad Basada en Compresión (CDM), Electroencefalograma (EEG), Epilepsia, Minería de Datos (MD).

Este trabajo fue procesado en L^AT_EX.

www.bdigital.ula.ve

Índice general

Índice de Tablas	VII
Índice de Figuras	VIII
1. Introducción	1
1.1. Antecedentes	1
1.2. Planteamiento del Problema	3
1.3. Justificación	3
1.4. Objetivos	4
1.4.1. Objetivo General	4
1.4.2. Objetivos Específicos	4
2. Marco teórico	5
2.1. Epilepsia	5
2.1.1. Definición	5
2.1.2. Clasificación	5
2.1.3. Diagnóstico	8
2.2. Electroencefalograma	10
2.2.1. Descripción	10
2.2.2. Uso general del electroencefalograma en el diagnóstico y tratamiento de las epilepsias	11
2.2.3. Registros electroencefalográficos y base de datos	13
2.3. Minería de Datos	15
2.3.1. Descripción	15

2.3.2.	Tareas de la Minería de Datos	15
3.	Desarrollo de la Metodología	17
3.1.	CDM: Medida de Disimilaridad Basada en Compresión	17
3.1.1.	Matriz de Disimilaridad	19
3.1.2.	Agrupamiento	20
3.1.3.	Clasificación	23
3.2.	Validación de la técnica	24
3.2.1.	Señales	25
3.2.2.	Matriz de Disimilaridad	26
3.2.3.	Agrupamiento	26
3.2.4.	Clasificación	30
4.	Resultados	32
4.1.	Matriz de Disimilaridad	32
4.1.1.	Análisis por canales	37
4.2.	Estudio de formación de grupos	42
4.3.	Clasificación	53
4.4.	Discusión de resultados	53
5.	Conclusiones y Recomendaciones	55
5.1.	Conclusiones	55
5.2.	Recomendaciones	56
	Bibliografía	57
A.	Uso de las funciones en R	59
A.1.	CDM: Medida de disimilaridad Basada en Compresión	59
A.1.1.	Parámetros de entrada	59
A.1.2.	Salida	59
A.2.	Matriz de disimilaridad	60
A.2.1.	Parámetros de entrada	60
A.2.2.	Salida	60

A.3. Agrupamiento Jerárquico	60
A.3.1. Parámetros de entrada	60
A.3.2. Salida	61
A.4. Clasificación	61
A.4.1. Parámetros de entrada	61
A.4.2. Salida	61

www.bdigital.ula.ve

Índice de Tablas

3.1. Ejemplo de Matriz de Disimilaridad	20
3.2. Matriz de Disimilaridad para las señales de validación.	27
3.3. Resultados de la clasificación para las señales de validación.	31
4.1. Matriz de Disimilaridad para el sujeto Normal 2.	33
4.2. Matriz de Disimilaridad para el paciente Epiléptico 4.	34
4.3. Matriz de Disimilaridad para el paciente con crisis.	35
4.4. Disimilaridad de la actividad cerebral de todos los sujetos bajo estudio.	36
4.5. Resultados de la clasificación para las señales EEG.	53

Índice de Figuras

2.1. Sistema Internacional de Registro Electroencefalográfico 10-20	11
2.2. Registros electroencefalográficos de un sujeto sano.	13
2.3. Registros electroencefalográficos de un paciente epiléptico.	14
2.4. Registros electroencefalográficos de un paciente epiléptico en crisis. . .	14
3.1. Ilustración del principio de similaridad por compresión	19
3.2. Ejemplo de Árbol Jerárquico	21
3.3. Ejemplo de clasificación con kNN	24
3.4. Esquema general de los grupos de señales utilizadas para validación de la CDM	25
3.5. Dendrograma de grupos de algunas de las señales de validación (1000 muestras ruidosas).	28
3.6. Dendrograma de grupos de las señales de validación (1000 muestras ruidosas).	28
3.7. Dendrograma de grupos de algunas de las señales de validación (400 muestras ruidosas).	29
3.8. Dendrograma de grupos de las señales de validación (400 muestras ruidosas).	29
3.9. Individuos de entrenamiento para la clasificación con kNN (validación del método)	30
3.10. Individuos de validación para la clasificación con kNN (validación del método)	31
4.1. Análisis de los canales Fp1 y Fp2	37
4.2. Análisis de los canales F7 y F3	38

4.3. Análisis de los canales Fz y F4	38
4.4. Análisis de los canales F8 y T3	39
4.5. Análisis de los canales C3 y Cz	39
4.6. Análisis de los canales C4 y T4	40
4.7. Análisis de los canales T5 y P3	40
4.8. Análisis de los canales Pz y P4	41
4.9. Análisis de los canales T6 y O1	41
4.10. Dendrogramas de los sujetos Normal 1 y Normal 2	42
4.11. Dendrogramas de los sujetos Normal 3 y Normal 5	43
4.12. Dendrograma del sujeto Normal 6.	43
4.13. Dendrogramas de los pacientes Epiléptico 1 y Epiléptico 2	44
4.14. Dendrogramas de los pacientes Epiléptico 3 y Epiléptico 4	44
4.15. Dendrograma del paciente con crisis epiléptica.	45
4.16. Dendrogramas del sujeto Normal 1 con Epiléptico 1 y Normal 1 con Epiléptico 2	45
4.17. Dendrogramas del sujeto Normal 1 con Epiléptico 3 y Normal 1 con Epiléptico 4	46
4.18. Dendrograma del sujeto Normal 1 con el paciente con crisis.	46
4.19. Dendrogramas del sujeto Normal 2 con Epiléptico 1 y Normal 2 con Epiléptico 2	47
4.20. Dendrogramas del sujeto Normal 2 con Epiléptico 3 y Normal 2 con Epiléptico 4	47
4.21. Dendrograma del sujeto Normal 2 con el paciente con crisis.	48
4.22. Dendrogramas del sujeto Normal 3 con Epiléptico 1 y Normal 3 con Epiléptico 2	48
4.23. Dendrogramas del sujeto Normal 3 con Epiléptico 3 y Normal 3 con Epiléptico 4	49
4.24. Dendrograma del sujeto Normal 3 con el paciente con crisis.	49
4.25. Dendrogramas del sujeto Normal 5 con Epiléptico 1 y Normal 5 con Epiléptico 2	50

4.26. Dendrogramas del sujeto Normal 5 con Epiléptico 3 y Normal 5 con Epiléptico 4	50
4.27. Dendrograma del sujeto Normal 5 con el paciente con crisis.	51
4.28. Dendrogramas del sujeto Normal 6 con Epiléptico 1 y Normal 6 con Epiléptico 2	51
4.29. Dendrogramas del sujeto Normal 6 con Epiléptico 3 y Normal 6 con Epiléptico 4	52
4.30. Dendrograma del sujeto Normal 6 con el paciente con crisis.	52

www.bdigital.ula.ve

Capítulo 1

Introducción

Este capítulo comprende aspectos que respaldan la realización del trabajo: antecedentes, definición del problema, justificación, objetivos y metodología a utilizar.

1.1. Antecedentes

Bajo el nombre de Minería de Datos se engloba todo un conjunto de técnicas encaminadas a la extracción de conocimiento potencialmente útil, implícito en las bases de datos. La mayoría de éstas, poseen una gran cantidad de parámetros que deben de ajustarse para que puedan realizar eficazmente su cometido, trayendo esto consigo una serie de inconvenientes.

Recientemente, ha surgido un nuevo paradigma de la Minería de Datos, el cual propone un enfoque de minería libre de parámetros basado en la compresión (Keogh *et al.*, 2004). El trabajo es motivado por los resultados en la bioinformática y la teoría computacional, que no son bien conocidos fuera de estas comunidades. En el estudio anteriormente mencionado, los autores mostraron cómo su enfoque permite realizar tareas clásicas de la Minería de Datos de manera más eficiente que con muchos de los algoritmos tradicionales. Un ejemplo de ello, es la clasificación de series temporales con el algoritmo del los k vecinos cercanos (kNN, k Nearest Neighbors, por sus siglas en inglés) utilizando la CDM como métrica de distancia entre individuos, en donde ésta reveló menores errores porcentuales que el Alineamiento Dinámico del Tiempo (DTW,

Dynamic Time Warping, por sus siglas en inglés) y la distancia euclidiana.

En Venezuela, desde hace algunos años, se ha venido incrementando el interés por parte de los investigadores en el estudio y procesado de registros electroencefalográficos de pacientes epilépticos (Escalona-Morán *et al.*, 2007). En el trabajo realizado por (Coutín *et al.*, 2005), los autores muestran cómo a través de técnicas de la dinámica no lineal, índices de entropía y representación simbólica, se puede caracterizar el comportamiento de la actividad cerebral. Para la Entropía Aproximada, encontraron que esta permite discriminar el nivel de complejidad de la actividad cerebral entre los pacientes epilépticos y los sujetos sanos, presentándose un mayor nivel de regularidad en la actividad cerebral de los enfermos. Los resultados obtenidos de la Dinámica Simbólica, muestran la presencia de un nivel de regularidad asociado con la periodicidad de la actividad paroxística, manifestada en los pacientes epilépticos.

En (Escalona-Morán *et al.*, 2007), dos parámetros de orden introducidos en el contexto de los sistemas caóticos sujetos a ruido externo se emplean para medir los grados de sincronización y de la formación de agrupaciones dinámicas en señales electroencefalográficas. La evolución de estos parámetros muestra la ocurrencia de sincronización intermitente y del agrupamiento en la actividad del cerebro durante una crisis epiléptica. Significativamente, la existencia de un máximo instantáneo de sincronización previo al inicio de una crisis se manifiesta por este procedimiento.

En estudios más recientes (Naive *et al.*, 2009), se emplean las Máquinas de Vectores de Soporte como una herramienta de clasificación. Previo al proceso de clasificación, se aplica una técnica de la Minería de Datos llamada Parámetros de Sincronización, la cual transforma las señales EEG a unos índices estadísticos que extraen patrones e información oculta inherente en las mismas. Una vez obtenidos los parámetros de sincronización, parte de éstos se emplean para el entrenamiento y el resto para la validación de la Máquina. Una eficiencia del 99% de clasificación entre sujetos sanos y pacientes epilépticos es obtenida con la metodología presentada.

1.2. Planteamiento del Problema

La Epilepsia es un trastorno provocado por el aumento de la actividad eléctrica de las neuronas en el cerebro o parte de él. La persona afectada puede sufrir una serie de convulsiones o movimientos corporales incontrolados de forma repetitiva, a esto se le llama “ataque epiléptico”. El análisis electroencefalográfico ha sido una herramienta utilizada por los especialistas durante mucho tiempo para obtener información oculta que caracteriza los desórdenes de la actividad cerebral propios de esta enfermedad. Con este propósito, en el presente trabajo, se realiza un estudio de Minería de Datos con la técnica CDM, para obtener patrones que permitan cuantificar los cambios de complejidad que ocurren en la patología, a través del análisis computarizado del electroencefalograma.

1.3. Justificación

Aproximadamente el 1% de la población mundial sufre de Epilepsia y un 30% de los epilépticos no encuentra mejoría a través de los fármacos que le son suministrados (Hojjat *et al.*, 2003). Para mejorar el tratamiento de esta patología es necesario comprender los mecanismos que afectan a la misma.

El análisis de electroencefalogramas se realiza tradicionalmente mediante inspección visual por especialistas médicos llamados neurólogos. El análisis computarizado de las señales electroencefalográficas a través de técnicas de Minería de Datos, permite extraer información oculta que no puede ser detectada por una simple inspección visual del registro electroencefalográfico. De esta manera, se puede lograr una mejor comprensión de la patología bajo análisis para presentar soluciones más efectivas. Debido a esto, es que, en este Proyecto de Grado, se plantea el uso de la CDM para analizar las series temporales concernientes a electroencefalogramas, ya que este método ha demostrado ser eficaz y eficiente al momento de analizar series temporales (Keogh *et al.*, 2004).

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar e implementar técnicas de la Minería de Datos que permitan analizar el comportamiento de la actividad cerebral de pacientes epilépticos y sujetos sanos.

1.4.2. Objetivos Específicos

- Revisar la literatura concerniente a series temporales, Minería de Datos y Epilepsia.
- Obtener las señales electroencefalográficas de pacientes epilépticos y sujetos sanos para realizar el estudio.
- Desarrollar e implementar en el lenguaje de programación R las técnicas: CDM, Agrupamiento Jerárquico y Clasificación.
- Aplicar las técnicas desarrolladas a las señales EEG de pacientes epilépticos y sujetos sanos.
- Análisis de resultados.

Capítulo 2

Marco teórico

En este capítulo se realiza una descripción general de la Epilepsia, tipos de Epilepsia, diagnóstico de la Epilepsia, el Electroencefalograma y su uso en el diagnóstico de la Epilepsia y registros electroencefalográficos. También se realiza un resumen acerca de la Minería de Datos, sus técnicas y aplicaciones.

2.1. Epilepsia

2.1.1. Definición

- Crisis cerebral consecuencia de una descarga neuronal excesiva. (Organización Mundial de la Salud)
- Manifestación clínica presumiblemente originada por una descarga excesiva de neuronas a nivel cerebral. Esta consiste en fenómenos anormales bruscos y transitorios que incluye alteraciones de la conciencia, motora, sensorial, autonómica, psicológica, que son percibidos por el paciente o un observador. (Liga Internacional contra la Epilepsia).

2.1.2. Clasificación

Una crisis epiléptica es la manifestación clínica de una descarga eléctrica incontrolada en el cerebro. Por ello, según donde se produzca esa descarga (o a las áreas a las que se

propague), así serán las manifestaciones que tenga el paciente. Por ejemplo, si la descarga afecta el área cerebral responsable del movimiento, el paciente notará sacudidas incontroladas de una extremidad. Si afecta al área de la vista, verá fenómenos visuales anormales.

Existen dos grupos de crisis, las llamadas parciales y las generalizadas. La diferencia entre ambas radica en el inicio de éstas.

Crisis Parciales

Son aquellas en las que la actividad epiléptica comienza en una zona determinada del cerebro. Las características de las crisis dependerán de la zona cerebral donde se origine la misma y de la propagación de la actividad epiléptica a otras áreas cerebrales.

Entre las crisis parciales se destacan:

- **Crisis Parciales Simples (CPS):** también denominadas auras. No hay alteración de la conciencia; el paciente queda conectado y se entera de lo que sucede a su alrededor. Suelen durar menos de un minuto. Se producen alteraciones que sólo nota el paciente (olores anormales, sensaciones extrañas, miedo, etc.). Existen varios tipos de auras, entre las cuales quizá la más conocida es el “aura gástrica” o sensación de hormigueo ascendente desde el estómago.
- **Crisis Parciales Complejas (CPC):** existe alteración de la conciencia. Puede ocurrir que una crisis comience como una CPS (un olor anormal, una sensación peculiar...) y luego, debido a la propagación a otras áreas cerebrales, se convierta en una CPC con generalización.
- **Crisis Parciales Secundariamente Generalizadas (CPSG):** la descarga inicial comienza en un grupo neuronal concreto, generalizándose al resto de las estructuras cerebrales, ocasionando una crisis tónico-clónica. Este tipo de crisis puede comenzar como una crisis parcial simple, y progresar hacia una crisis parcial compleja, donde si hay afectación de la conciencia, y derivar en una crisis tónico-clónica. Algunas personas que sufren este tipo de crisis tienen “auras”, estas personas tienen una gran ventaja, ya que antes de tener la crisis, a veces

pueden sentarse o colocarse en una determinada posición para evitar caídas o golpes. Estas crisis pueden empezar súbitamente como crisis tónico-clónica.

Crisis Generalizadas

Son aquellas en las que la actividad eléctrica epileptiforme se manifiesta en todo el cerebro. Se originan movimientos bruscos de todo el cuerpo. Si la actividad epileptiforme comienza en todo el cerebro a la vez, se denomina primariamente generalizada. Entre las crisis generalizadas se distinguen:

- **Crisis Mioclónicas:** contracción brusca (como un susto) de duración muy breve, habitualmente en las extremidades superiores. Son habituales en la epilepsia mioclónica juvenil.
- **Crisis Tónicas:** caracterizadas por la pérdida brusca de la conciencia seguida de una rigidez corporal, más intensa en las extremidades superiores. Suelen ser breves (menos de treinta segundos). Este tipo de crisis son típicas del Síndrome Lennox-Gastaut.
- **Ausencias:** antiguamente denominada "*Petit mal*". Se produce una detención brusca de la actividad motora e interrupción de la conciencia, a veces con movimientos automáticos simultáneos como relamerse los labios o arreglarse la ropa, acompañándose de una mirada fija e inexpresiva, parpadeos y reclinación ligera de la cabeza hacia atrás. Terminan bruscamente, reanudándose la actividad previa. Su duración es de cinco a veinte segundos, recobrándose con extraordinaria rapidez la conciencia. No suelen observarse en niños menores de cuatro años. Son fenómenos de desconexión de breve duración, repentinos y sin pérdida del control postural.
- **Crisis Atónicas:** se produce una pérdida brusca del tono muscular. Son características del Síndrome Lennox-Gastaut.
- **Crisis Tónico-Clónica:** antiguamente denominada "*Grand mal*". Quizás sea la crisis más conocida de todas. Suele ser muy breve (alrededor de dos minutos) pero se hace interminable. Puede resultar desagradable de ver. Comienza de forma

brusca e inesperada; algunos pacientes sienten síntomas (“auras”), otros no. Se distinguen dos fases:

Fase Tónica: comienza con la pérdida brusca del conocimiento, continuada con rigidez en las cuatro extremidades, en ocasiones los labios pueden ponerse morados, es debido a que en esta fase no se respira.

Fase Clónica: sacudidas rítmicas que afectan a los dos brazos y a las dos piernas, o solamente a un lado del cuerpo. Sale espuma (saliva) por la boca, se puede morder la lengua (no siempre) y eliminar orina (no siempre). La crisis termina en un sueño profundo que puede durar de unos minutos a unas horas, dependiendo de la duración que haya tenido la crisis. Al despertar hay confusión, cansancio, desorientación y dolor de cabeza.

2.1.3. Diagnóstico

Como todo problema en Medicina, el diagnóstico comienza con una buena historia clínica. Los datos importantes que cada parte de la historia pueden aportar son los siguientes:

- **Interrogatorio del padecimiento actual:** con el interrogatorio directo al propio paciente, incluido un niño que ya habla, se investigan manifestaciones subjetivas de las que sólo el mismo paciente puede percatarse, tales como: sensaciones del cuerpo, sentirse extraño, ver u oír cosas que no existen, cambios del estado de ánimo, etc., que pueden presentarse como manifestación única (crisis parciales simples) o precediendo a otras en las que exista deterioro de la conciencia (crisis parciales que pasan a complejas o se generalizan secundariamente). En esta situación de que el paciente ya no se da cuenta de nada, la información se obtiene del testigo, quien habiendo visto lo sucedido, podrá describir los cambios que el paciente presenta mientras no se da cuenta, como por ejemplo: caída súbita, sacudidas musculares mioclónicas, aumento del tono muscular, cianosis (ponerse morado) o palidez, dificultad respiratoria, sialorrea (salivación excesiva que escurre de la boca), emisión de orina o heces fecales, movimientos tónico-clónicos de las extremidades, todo esto en caso de crisis generalizadas, ya sean

atónicas, mioclónicas, tónicas, tónico-clónicas o clónicas. En otros casos en los que el paciente se desconecta pero no cae, el testigo puede describir si hay acciones automáticas acompañantes. Inmediatamente después de la crisis generalizada (mayor o parcial compleja), el paciente presenta el período postictal con confusión, somnolencia y cefalea (dolor de cabeza), no así en las ausencias de la niñez y las crisis parciales simples, que habitualmente carecen de alteraciones postictales. Es importante que no se omita en cualquiera de las descripciones la duración del fenómeno en todas sus partes: la crisis suele ser de duración breve que va de pocos segundos a varios minutos, habitualmente dos pero menos de cinco. También es importante conocer su frecuencia, así como la predominancia de las crisis en ciertas horas del día, si son durante el sueño, o el ciclo menstrual o su asociación con diversas circunstancias tales como: ayuno, desvelo, tensión emocional, etc. El médico deberá complementar con la edad de inicio y otros síntomas generales y neurológicos.

- **Antecedentes:** interrogar si hay otros familiares con crisis epilépticas (epilepsias hereditarias), antecedentes del embarazo y condiciones al nacimiento, desarrollo psicomotor o neurológico, enfermedades neurológicas en la infancia, trauma craneoencefálico (golpes fuertes en la cabeza).
- **Exploración:** en la exploración general se pueden detectar malformaciones somáticas, trastornos cutáneos, etc. y en el examen neurológico la presencia de signos que hagan sospechar enfermedades específicas. En la mayoría de los pacientes epilépticos la exploración es normal.
- **Estudios complementarios:** el electroencefalograma se debe hacer en todos los casos. Si la crisis se presenta al momento de realizar el estudio, complementa muy bien el diagnóstico y si el EEG es normal en el momento de presentarse el fenómeno, descarta que sea una crisis epiléptica. En el período interictal el electroencefalograma puede ser anormal o normal, lo cual no excluye el diagnóstico de crisis epiléptica. Otros estudios electrofisiológicos especiales son el mapeo cerebral y el monitoreo video-electroencefalográfico. Los estudios de imagen como la

Tomografía Axial Computada (TAC) y la Resonancia Magnética (RM), se indican cuando el cuadro clínico hace sospechar una lesión neurológica como causa de las crisis y debe realizarse siempre que existan crisis parciales a cualquier edad o en cualquier tipo de crisis si estas aparecen por primera vez después de los veinte años de edad, cuando se consideran de aparición tardía y pueden ser el anuncio de enfermedades progresivas, como por ejemplo, un tumor en el cerebro.

Hasta aquí se ha planteado la manera clínica general de estudiar a un paciente en el que el primer diagnóstico a realizar es si se trata de crisis epilépticas o no; en caso de que sí se apoyen por los datos clínicos el paso siguiente es identificar el tipo de Epilepsia de que se trata y si ésta corresponde al grupo de las epilepsias sintomáticas.

En muchos casos la causa no se puede identificar (epilepsia idiopática) o se trata de una enfermedad que desapareció o no es progresiva pero dejó la Epilepsia como secuela, mientras que en otros, las causas que se pueden detectar son susceptibles de tratamiento específico y es necesario identificarlas oportunamente, como por ejemplo, una fenilcetonuria o un tumor cerebral (epilepsia sintomática).

2.2. Electroencefalograma

2.2.1. Descripción

Es la técnica más clásica de registro de la actividad eléctrica espontánea del sistema nervioso central. Sus fundamentos no han cambiado aunque los aparatos modernos sean más compactos y fiables. Mediante electrodos aplicados al cuero cabelludo se recoge una actividad eléctrica que se genera en la corteza cerebral (potenciales post-sinápticos de las células piramidales). Los electrodos se colocan según un sistema aceptado internacionalmente (Sistema 10-20, ver Figura 2.1) para cubrir todas las áreas corticales (frontales, parietales, temporales y occipitales). La actividad de cada electrodo se deriva de otro de referencia común o en diversos montajes bipolares para definir mejor la actividad de cada área del cerebro en estudio. El paciente se registra en un ambiente relajado y se utilizan diversos artificios para modificar la actividad durante el registro como la apertura y cierre de los ojos, la hiperventilación,

la estimulación luminosa intermitente o el dolor en los enfermos estuporosos.

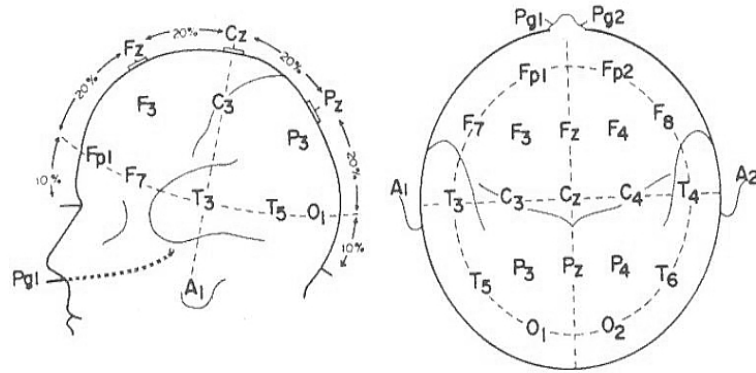


Figura 2.1: Sistema Internacional de Registro Electroencefalográfico 10-20. Este sistema permite posicionar los electrodos para el registro simultáneo de la actividad cerebral en toda la superficie de la cabeza del paciente.

La actividad recogida en el EEG se clasifica según la frecuencia de las oscilaciones (ondas) del trazado; las ondas más lentas de 1-3 Hz se llaman delta, las de 4-7 Hz se llaman theta, las de 8-12 Hz alfa y por encima son denominadas beta (Zarranz, 2003). Normalmente, en el individuo despierto predominan las ondas rápidas, aproximadamente simétricas en las desviaciones de ambos hemisferios; en las regiones occipitales con los ojos cerrados se recoge una actividad regular sinusoidal predominante que se llama ritmo alfa occipital (a unos 8-10 Hz) y que se atenúa al abrir los ojos. La interpretación correcta del EEG requiere una notable experiencia y espíritu crítico, pues no todo lo que se registra y se aparta de lo habitual es patológico, sino que hay considerables variaciones individuales y actividades incluso paroxísticas que carecen de importancia.

2.2.2. Uso general del electroencefalograma en el diagnóstico y tratamiento de las epilepsias

El EEG ha sido y sigue siendo la exploración complementaria que más ha contribuido al conocimiento básico de las Epilepsias, que ha permitido el diagnóstico correcto de

muchos enfermos y que ha servido de soporte objetivo para la definición y clasificación de los síndromes epilépticos.

La descarga neuronal anormal y exagerada, base fisiopatológica de las Epilepsias, puede registrarse en el EEG como grafoelementos, generalmente agudos y aislados (puntas) o agrupados (polipuntas), que pueden estar seguidos de ondas lentas (punta-onda, polipunta-onda) y que pueden ser localizados (foco irritativo) lateralizados sin definir un foco o generalizados. Esta actividad paroxística acompaña prácticamente siempre a la crisis epiléptica clínica (actividad crítica), pero puede aparecer también en los períodos entre crisis (actividad intercrítica) y ésta es la base de la utilidad del registro EEG para el diagnóstico ordinario de las Epilepsias, por cuanto rara vez los enfermos son registrados durante las crisis. El uso del EEG ha sido muchas veces incorrecto y abusivo debido a falta de criterios de valoración.

El estudio EEG en los enfermos con sospecha de Epilepsia tiene dos objetivos principales: apoyar el diagnóstico de la enfermedad y contribuir a la clasificación del síndrome epiléptico. Se basa fundamentalmente en las alteraciones intercríticas y críticas. En las exploraciones EEG de rutina en vigilia exceptuando las ausencias epilépticas, que se provocan por la hiperpnea (respiración acelerada), se pueden utilizar solo las alteraciones intercríticas. En las epilepsias reflejas es posible obtener registro crítico si se conoce el factor precipitante inductor de la crisis. En los casos en que el registro crítico fuera necesario y justificado se requieren métodos de exploración más complejos como la monitorización prolongada video/EEG.

El EEG es muy útil en el estado de mal epiléptico por varios motivos. Es la única técnica capaz de hacer un diagnóstico definitivo del estado de mal no convulsivo, el cual se presenta con un síndrome de confusión, amnesia y conducta automática, que puede ser debido a otras causas no epilépticas como: trastornos psíquicos, efecto de drogas o fármacos, hipoglucemia, etc. En segundo lugar, permite evitar el error de tomar como estado de mal epiléptico a las crisis psicógenas repetidas pseudoconvulsivas que, sin excepción, cursan con EEG normal. Finalmente, el EEG permite verificar la desaparición o persistencia de la actividad paroxística si el enfermo ha sido sedado, relajado o intubado. En estas condiciones el paciente puede no tener convulsiones pero persiste la actividad irritativa cortical. La interpretación del EEG en los estados de

mal y en la UCI (Unidad de Cuidados Intensivos) es particularmente difícil y requiere de mucha experiencia por parte del electroencefalografista.

2.2.3. Registros electroencefalográficos y base de datos

El registro de las señales EEG de cada sujeto se realizó en 19 canales (excepto para el paciente con crisis que no se le registró el canal occipital 2) a partir de electrodos adheridos a la superficie de la corteza cerebral con pasta conductora, en los sitios establecidos por el Sistema Internacional 10-20, midiéndose la diferencia de potencial con respecto a una referencia constituida por ambas orejas cortocircuitadas. La señal se digitalizó a una frecuencia de muestreo de 256 Hz y conversión A/D de 12 bits, filtrada digitalmente entre 0.5 Hz y 30 Hz. Se seleccionaron para el análisis registros de tiempo variable de actividad en reposo de cinco sujetos sanos (Normales) y cinco pacientes epilépticos, uno de los cuales presentó una crisis convulsiva durante el estudio. A continuación, desde la Figura 2.2 a 2.4, se presenta el electroencefalograma de un sujeto sano, uno epiléptico y el epiléptico con crisis, respectivamente.

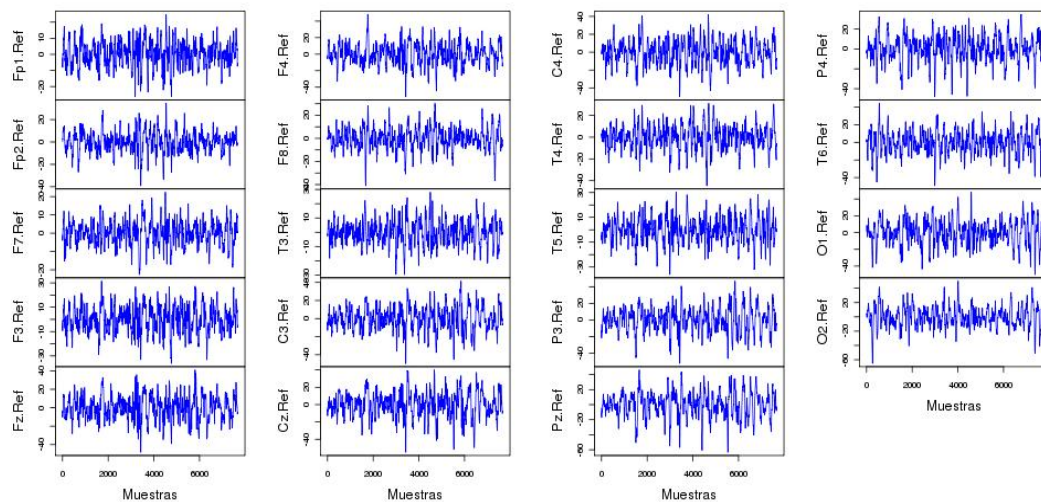


Figura 2.2: Registros electroencefalográficos de un sujeto sano.

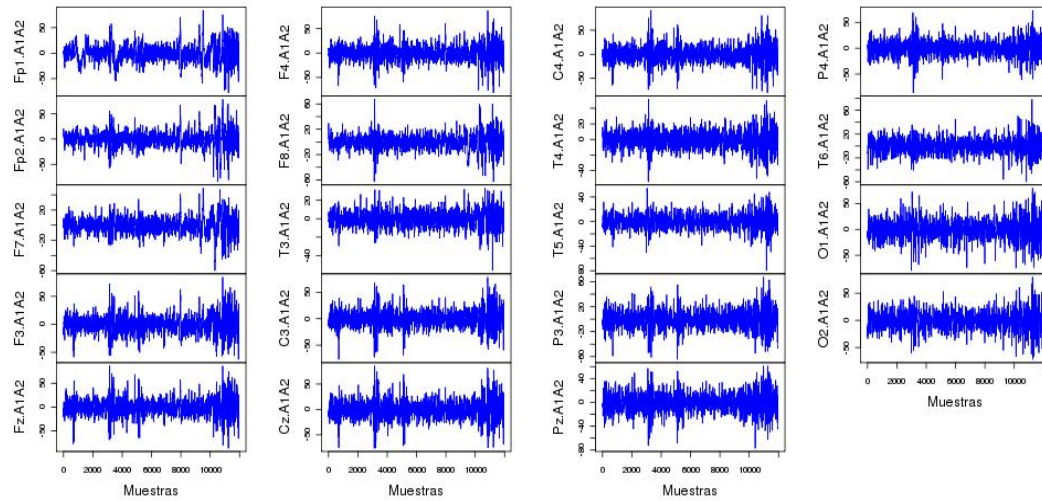


Figura 2.3: Registros electroencefalográficos de un paciente epiléptico.

www.bdigital.ula.ve

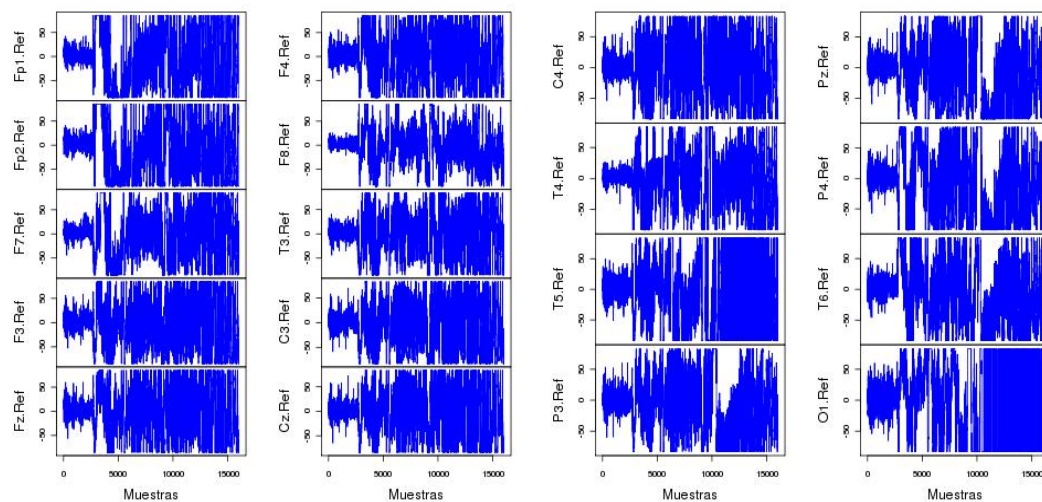


Figura 2.4: Registros electroencefalográficos de un paciente epiléptico en crisis.

2.3. Minería de Datos

2.3.1. Descripción

La Minería de Datos (MD) es un conjunto de metodologías, técnicas y herramientas que ayudan a transformar una gran cantidad de información almacenada en bases de datos en conocimiento, previamente desconocido, que pueda ser de utilidad en la toma de decisiones. Dicho conocimiento generalmente se cristaliza a través de la creación de patrones, reglas o modelos.

La MD emplea todas las técnicas que puedan aportar información útil, desde un sencillo análisis gráfico, pasando por métodos estadísticos mas o menos complejos, complementados con métodos y algoritmos del campo de la inteligencia artificial y el aprendizaje automático que resuelven problemas típicos de agrupamiento, clasificación, predicción de valores, detección de patrones, asociación de atributos, etc. Es, por tanto, un campo multidisciplinar que cubre numerosas áreas y se aborda desde múltiples puntos de vista, como la estadística, la informática (cálculo automático) o la ingeniería.

2.3.2. Tareas de la Minería de Datos

Predictivas

Fundamentalmente se pueden dividir en:

- **Clasificación:** se genera un modelo en donde para cada nueva observación, ésta es clasificada en una de las clases previamente establecidas. Por ejemplo, una empresa aseguradora necesita desarrollar, a partir de toda la información de sus clientes, un modelo clasificador que le indique si un nuevo asegurado, según la información del mismo, supone “poco riesgo” o “mucho riesgo” para la empresa. Dentro de este tipo de algoritmos, nos encontramos con: árboles de decisión, generadores de reglas, redes neuronales, máquinas de soporte vectorial, clasificadores basados en casos y vecindad, entre otros.
- **Predicción:** en este caso, el modelo generado intenta, para cada nueva observación, predecir el valor más probable que pueda tomar cierto parámetro. Por

ejemplo, a partir de una serie de variables de un proceso térmico, queremos predecir la temperatura que va a alcanzar en un cierto momento. Dentro de este tipo de técnicas entran: regresión lineal o no lineal, redes neuronales utilizadas para generar modelos lineales o no lineales, máquinas de soporte vectorial, etc.

Descriptivas

El objetivo de estas técnicas es el descubrimiento de grupos, patrones o reglas, es decir, describir los datos existentes. Dentro de esta serie de técnicas, nos encontramos fundamentalmente con:

- Algoritmos de segmentación (o también llamados de clusterizado): los cuales buscan grupos de datos que tienen características similares.
- Técnicas que describen densidades de probabilidad.
- Métodos que describen las dependencias entre variables (matriz de correlación, gráficos de relación entre variables, etc.).
- Procedimientos que buscan dentro de los datos patrones o reglas, como por ejemplo, las reglas de asociación, que analizan las relaciones entre las variables o las observaciones.

Capítulo 3

Desarrollo de la Metodología

Este capítulo presenta con detalle la metodología para el análisis y estudio de las señales electroencefalográficas tanto del grupo control como el de los pacientes epilépticos. La técnica es utilizada en tres modalidades: Medida de Disimilaridad (Matriz de Disimilaridades), Agrupamiento y Clasificación. La validación de la técnica en cada una de sus modalidades se realizó con señales deterministas provenientes de funciones matemáticas.

3.1. CDM: Medida de Disimilaridad Basada en Compresión

Los fundamentos matemáticos del enfoque de disimilaridad basada en compresión se encuentran en un área de la teoría de la información conocida como Algoritmos de Análisis de Entropía, o Complejidad de Kolmogorov (Min & Vitányi 1997). La Complejidad de Kolmogorov de una secuencia X , $K(X)$, se define como la longitud del programa más corto capaz de producir X en una computadora universal (como una máquina de Turing). Esto puede ser extendido para definir la Complejidad de Kolmogorov condicional, $K(X|Y)$, de una secuencia, X , en relación a otra secuencia, Y . En este caso, $K(X|Y)$ puede ser definida como la longitud del programa más corto que calcula X cuando Y se da como entrada auxiliar al programa. La función $K(XY)$ es la longitud del programa más corto que reproduce Y concatenada a X .

En (Li *et al.*, 2001), los autores consideran la distancia entre dos secuencias X e Y , como sigue

$$d_k(X, Y) = \frac{K(X|Y) + K(Y|X)}{K(XY)} \quad (3.1)$$

la cual satisface la desigualdad triangular, con un pequeño término de error. Una distancia matemáticamente más precisa fué propuesta en (Ming *et al.*, 2004).

El problema de la Complejidad de Kolmogorov, K , es que no hay un procedimiento matemático (algoritmo) exacto para calcularla en un número finito de pasos. No obstante, se ha demostrado que una aproximación matemática puede ser obtenida definiendo la secuencia como una cadena de caracteres, la cual puede ser comprimida por un programa compresor (Ming *et al.*, 2004; Cilibrasi and Vitányi, 2005), tal como los usados en computadoras personales.

En este estudio, se utilizó la métrica propuesta por (Keogh *et al.*, 2004) para aproximar la complejidad de Kolmogorov, la cual dice que dadas dos cadenas X e Y se define la Medida de Disimilaridad Basada en Compresión (CDM) como sigue

$$CDM(X, Y) = \frac{C(XY)}{C(X) + C(Y)} \quad (3.2)$$

Donde:

$C(XY)$ es el tamaño en bytes del archivo comprimido que contiene la concatenación de las cadenas X e Y .

$C(X)$ es el tamaño en bytes del archivo comprimido que contiene la cadena X .

$C(Y)$ es el tamaño en bytes del archivo comprimido que contiene la cadena Y .

La CDM es cercana a uno (1) cuando X e Y no están relacionadas y disminuye a medida que están más vinculadas. Por definición la disimilaridad de una cadena consigo misma es cero (0). La $CDM(X, Y)$ es igual a la $CDM(Y, X)$.

En general, la CDM busca información compartida por las dos cadenas X e Y , esto lo hace a través de los algoritmos de compresión, los cuales se encargan de buscar patrones repetidos dentro de los datos, para que luego se pueda cuantificar la diferencia

existente entre los mismos. Para clarificar esta idea, en la Figura 3.1 se muestra el principio de similaridad por compresión. Los pasos clave en el cálculo de la disimilaridad de X e Y , $CDM(X, Y)$, y la disimilaridad de U y V , $CDM(U, V)$, se muestran en los paneles de arriba y debajo de la Figura 3.1. En este caso el enfoque de disimilaridad basada en compresión podría estimar que $CDM(U, V)$ es mayor que $CDM(X, Y)$.

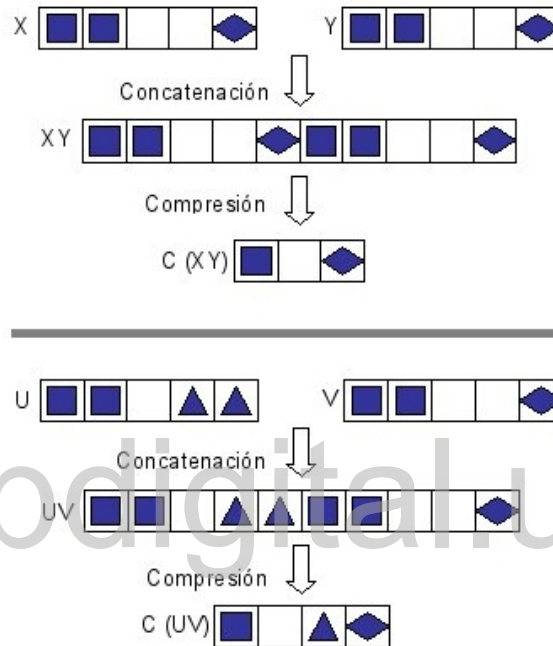


Figura 3.1: Ilustración del principio de similaridad por compresión.

3.1.1.1. Matriz de Disimilaridad

Dado un conjunto de secuencias (a partir de aquí serán llamadas series temporales) podemos calcular la disimilaridad entre todos los pares disponibles de estas. Matrices de Disimilaridad (o distancia) pueden ser construidas para su uso como entrada a algoritmos de agrupamiento o clasificación que requieren de estas, como por ejemplo, el agrupamiento jerárquico. La Tabla 3.1 muestra un ejemplo de una Matriz de Disimilaridad para un conjunto de cinco series temporales (S_1, S_2, S_3, S_4, S_5). Cada valor (celda) en la matriz, representa la disimilaridad, $CDM(i, j)$, entre un

par de series (S_i, S_j) . Por ejemplo, $CDM(4,2)$, es la disimilaridad entre las series S_4 y S_2 .

Series \ Series	S1	S2	S3	S4	S5
S1	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,4}$	$D_{1,5}$
S2	$D_{2,1}$	$D_{2,2}$	$D_{2,3}$	$D_{2,4}$	$D_{2,5}$
S3	$D_{3,1}$	$D_{3,2}$	$D_{3,3}$	$D_{3,4}$	$D_{3,5}$
S4	$D_{4,1}$	$D_{4,2}$	$D_{4,3}$	$D_{4,4}$	$D_{4,5}$
S5	$D_{5,1}$	$D_{5,2}$	$D_{5,3}$	$D_{5,4}$	$D_{5,5}$

Tabla 3.1: Matriz de disimilaridad para un conjunto de cinco series temporales $(S1, S2, S3, S4, S5)$.

La Matriz de Disimilaridad es simétrica con respecto a la diagonal, por lo tanto, basta con calcular una de las triangulares para utilizarla como entrada a otros algoritmos. La diagonal de la matriz es igual a cero.

3.1.2. Agrupamiento

Ya que el objeto del proceso de *clusterizado* es agrupar objetos similares, en este caso series temporales, necesitamos alguna medida para evaluar las diferencias y similitudes entre éstos. La similaridad (similitud) es una medida de correspondencia o semejanza entre los objetos que van a ser agrupados, la estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos. Los objetos con distancias reducidas entre ellos son más parecidos entre sí que aquellos que tienen distancias mayores, y por tanto, se agruparán dentro del mismo grupo (*cluster*). De esta manera, cualquier objeto puede compararse con cualquier otro objeto a través de la medida de similaridad.

Los métodos de agrupamiento se clasifican como sigue:

Métodos Jerárquicos

Utilizan un proceso en el cual se van agrupando los individuos de manera que se maximice una medida de similaridad o se minimice alguna distancia entre ellos. El resultado de un algoritmo de Agrupamiento Jerárquico puede representarse gráficamente mediante un árbol de grupos o dendrograma. El dendrograma presenta el proceso de unión de los grupos, a mayor altura del árbol mayor disimilaridad (distancia) entre los individuos. En la Figura 3.2 se exhibe un ejemplo de un árbol jerárquico cualquiera. Los métodos jerárquicos para construcción de conglomerados pueden ser de dos tipos:

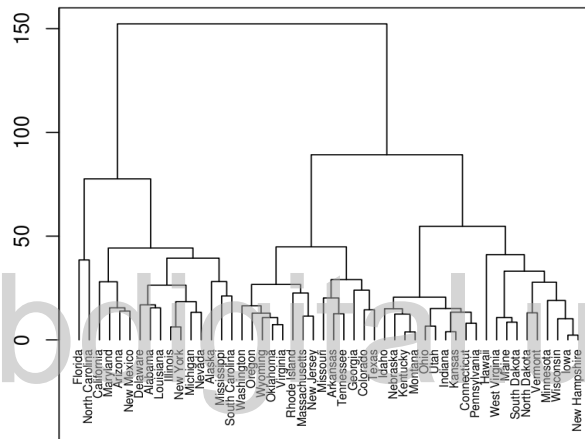


Figura 3.2: Ejemplo de árbol jerárquico.

- Aglomerativos:** se comienza considerando cada elemento como un grupo y en cada paso se unen los dos grupos más cercanos, hasta formar finalmente un único grupo. La unión entre grupos viene dada por una medida de proximidad, en caso de empate, se selecciona aleatoriamente una pareja de grupos (esta selección no afecta el resultado final).
- Divisivos:** se consideran los datos como un solo grupo, y en pasos sucesivos se dividen los grupos de acuerdo a medidas de proximidad entre los mismos, hasta que la cantidad de grupos sea igual a la cantidad de individuos en la base de datos.

Existen diferentes medidas de proximidad entre grupos para formar el Dendrograma. Entre estas tenemos:

- Distancia mínima (*single linkage*): los grupos se unen considerando la menor de las distancias existentes entre los miembros más cercanos de distintos grupos.
- Distancia máxima (*complete linkage*): los grupos se unen considerando la menor de las distancias existentes entre los miembros más lejanos de distintos grupos.
- Distancia entre centros (*centroid*): la distancia entre dos conglomerados es la distancia euclídea entre sus centroides. Se juntan los grupos con menor distancia entre si.
- Distancia promedio: la distancia entre los conglomerados es el promedio de las distancias entre los puntos de un *cluster* y los puntos de otro.
- Método de Ward: los nuevos conglomerados se crean de tal manera que se minimice la suma de cuadrados total de las distancias dentro de cada *cluster*, lo cual permite minimizar la variación Intra Grupal de la estructura formada.

Métodos no jerárquicos (de partición)

Realizan un procedimiento que consiste en elegir una partición de los individuos en K grupos, luego se van intercambiando los miembros de los mismos hasta obtener la mejor partición posible.

En este estudio se centró la atención en el Agrupamiento Jerárquico Aglomerativo, utilizando la CDM como medida de similaridad (o distancia) entre los individuos (series temporales correspondientes a los canales del EEG). La escogencia de este método y no de uno de partición es debido a que se deseaba poder apreciar relaciones existentes (asociadas a la patología) entre las distintas zonas del cerebro sin necesidad de definir los grupos de antemano. Para la construcción de los Dendrogramas se utilizaron todas las medidas de proximidad descritas anteriormente. Además, por razones de visualización, el eje vertical que muestra el valor de distancia (CDM) entre individuos

se reemplazó con el orden de formación de los grupos, sin que ésto altere la manera en que se analiza la similaridad entre las series.

3.1.3. Clasificación

Entre las distintas técnicas de clasificación utilizadas en la Minería de Datos podemos encontrar los llamados clasificadores basados en casos y vecindad.

Los algoritmos de clasificación por vecindad están basados en la búsqueda de un conjunto de los k prototipos más cercanos al patrón a clasificar. Debido a esto, se necesita de la definición de una métrica que ayude a comparar las distancias entre los individuos.

Uno de los algoritmos de clasificación por vecindad es el de k-Vecinos Cercanos (kNN, k Nearest Neighbors, por sus siglas en inglés). La regla de clasificación de este algoritmo consiste en que para un nuevo caso a ser clasificado, éste se ubicará en la clase con más votos en el contexto de los k vecinos más cerca en el conjunto de entrenamiento. Para entender mejor cómo es el funcionamiento del algoritmo kNN a continuación se muestra un ejemplo.

Ejemplo: El patrón que se desea clasificar es el círculo verde (ver Figura 3.3). Se consideran como patrones de entrenamiento a los cuadrados azules y a los triángulos rojos. Para clasificar el círculo, éste se compara con todos los individuos de entrenamiento a través de una métrica de distancia, luego, se define un perímetro alrededor de él, en base a la cantidad de vecinos cercanos establecidos. Si se toma $k = 3$ vecinos (circunferencia continua), el círculo es clasificado con la clase triángulo, ya que hay sólo un cuadrado y 2 triángulos dentro del perímetro que los contiene. Si $k = 5$ (circunferencia punteada), éste es clasificado con la clase cuadrado, ya que hay 2 triángulos y 3 cuadrados dentro del círculo externo.

Para realizar la tarea de clasificación en este trabajo, se utilizó el algoritmo de kNN tomando como métrica de distancia entre los individuos (series temporales) la Medida de Disimilaridad Basada en Compresión. Como individuos de entrenamiento se establecieron las series temporales concernientes a los canales del EEG de un sujeto sano (Normal 1) y el paciente con crisis. La elección de este último y no de un paciente

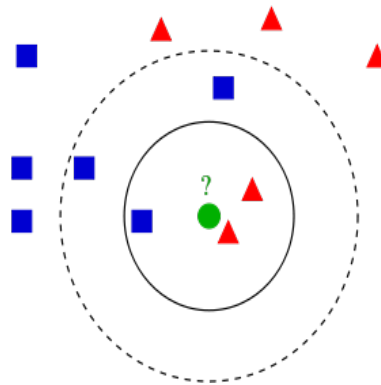


Figura 3.3: Ejemplo de clasificación con el algoritmo de los k vecinos cercanos.

epiléptico es debido a que en la crisis se ponen de manifiesto, en mejor medida, los grafoelementos asociados a la patología. Para la clasificación de un sujeto como sano o enfermo se realizó lo siguiente:

Primero, se calcularon las Matrices de Disimilaridad de los individuos de entrenamiento (EEG del sujeto Normal 1 y EEG del paciente con crisis). Luego, para cada sujeto a clasificar, se calculó la Matriz de Disimilaridad de éste y se comparó canal por canal esta matriz con las de entrenamiento. Esta comparación se realizó con el objetivo de asignarle a cada canal una etiqueta, *enfermo* ó *sano*, en base a los vecinos cercanos que este canal tuviese. Si a todos los canales del EEG del sujeto a clasificar se les asignaba la misma etiqueta, entónces, esta etiqueta le era asignada al sujeto, en caso de no ser así, el sujeto no se podía clasificar. La cantidad de vecinos cercanos se varió de 5 a 13 vecinos. La selección de estos límites es debido a que los pacientes estaban siendo medicados y algunos de los canales de sus EEG tenían un comportamiento normal, lo cual afectaba la clasificación.

3.2. Validación de la técnica

Con la finalidad de validar el método CDM se empleó un conjunto de señales deterministas generadas a partir de funciones matemáticas.

3.2.1. Señales

La base de datos utilizada para validación está compuesta de 16 series temporales, cada una con una longitud de mil muestras (8 de ellas tienen un ruido distribuido uniformemente a lo largo de toda la señal), las cuales fueron generadas a partir de las funciones: seno, coseno, logaritmo neperiano y raíz cuadrada. Para cada aplicación de la técnica (Matriz de Disimilaridad, Agrupamiento y Clasificación), se realizó una pequeña variación del ruido asociado a las series. Un esquema general de las señales se muestra en la Figura 3.4.

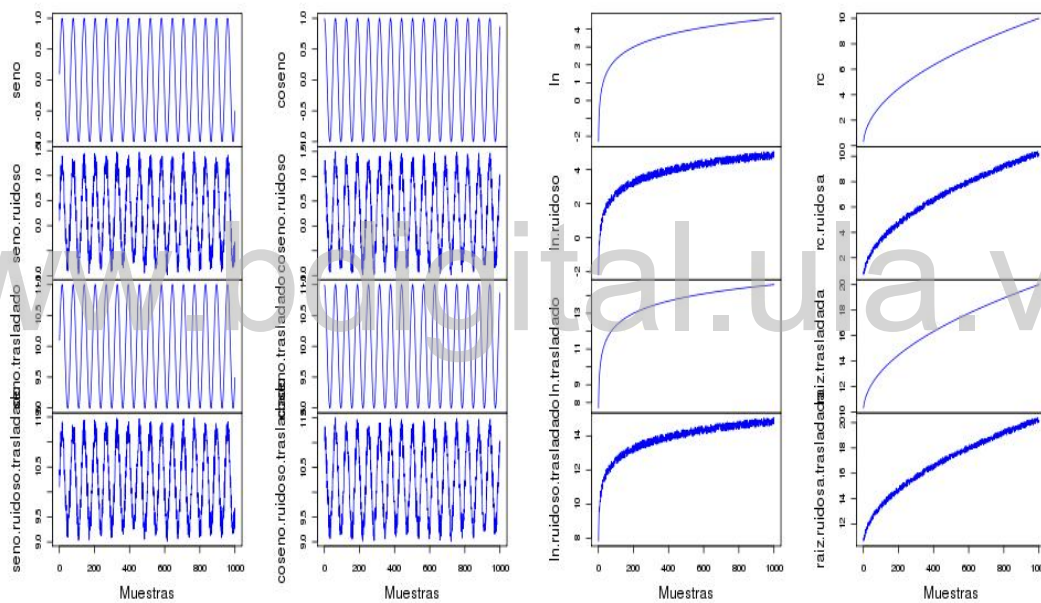


Figura 3.4: Esquema general de los grupos de señales utilizadas para validación. De izquierda a derecha tenemos al seno, coseno, logaritmo neperiano y raíz cuadrada.

La elección de estas series para validar la técnica es respaldada por el hecho de que éstas se parecen entre si, el seno con el coseno y el logaritmo neperiano con la raíz cuadrada, dificultando esto las tareas de agrupamiento y clasificación.

3.2.2. Matriz de Disimilaridad

En la Tabla 3.2 se presenta la Matriz de Disimilaridad correspondiente a la aplicación de la CDM a las señales descritas en la sección 3.2.1.

Para analizar la matriz exhibida anteriormente, podemos seleccionar cualquiera de las señales, por ejemplo, el seno, y observar los valores de disimilaridad que ésta presenta con respecto a las demás. La intuición indica que al comparar la señal seno con el resto, el menor valor de disimilaridad se debe encontrar con alguna de las series pertenecientes al grupo del seno, efectivamente esto es así, ya que el valor más bajo de CDM se encuentra con el seno trasladado (cuadro azul). Otra presunción que puede hacerse al momento del análisis, es que se debe encontrar valores de disimilaridad más bajos con las series del grupo coseno que con las del grupo logaritmo neperiano, al observar los cuadros verdes y amarillos, respectivamente, se puede confirmar que realmente esto sucede.

3.2.3. Agrupamiento

La validación para el agrupamiento se realizó con dos grupos de señales: las descritas en la sección 3.2.1 y con una variación de ellas, la cual consistió en la reducción de la cantidad de muestras ruidosas, disminuyéndose las mismas a 400 muestras. En las Figuras 3.5 y 3.6 se aprecian los dendrogramas correspondientes a las señales con 1000 muestras ruidosas.

En el agrupamiento de las series con ruido de tamaño 1000 se encontró que existen dos grandes grupos en el nivel más alto del dendrograma (ver Figura 3.5): las señales con ruido (izquierda) y las que no tienen ruido (derecha). Esto es debido a que al colocar un ruido a lo largo de una señal, se pierde la forma de la misma. Asimismo, en el dendrograma de la Figura 3.6, se puede apreciar la formación de estos mismos grupos.

En las Figuras 3.7 y 3.8 se puede distinguir la formación de cuatro grupos en los niveles más bajos del árbol, estos son: seno, coseno, raíz cuadrada y logaritmo neperiano. Si se observa el nivel superior del dendrograma, se contempla que existen dos grandes grupos de señales, uno con las funciones seno y coseno y el otro, con las

	s	s r	s t	s r t	c	c r	c t	c r t	ln	ln r	ln t	ln r t	rc	rc r	rc t	rc r t
s	0	0.977	0.874	0.981	0.976	0.979	0.979	0.980	0.980	0.981	0.981	0.981	0.981	0.980	0.977	0.978
s r	0	0	0.972	0.837	0.982	0.978	0.980	0.977	0.984	0.979	0.989	0.981	0.981	0.977	0.979	0.978
s t	0	0	0	0	0.985	0.980	0.978	0.978	0.977	0.979	0.983	0.979	0.978	0.979	0.976	0.977
s r t	0	0	0	0	0.981	0.980	0.983	0.978	0.979	0.979	0.986	0.980	0.981	0.977	0.977	0.977
c	0	0	0	0	0	0.979	0.873	0.978	0.980	0.981	0.980	0.981	0.982	0.978	0.976	0.978
c r	0	0	0	0	0	0	0.982	0.839	0.983	0.980	0.991	0.982	0.986	0.978	0.979	0.979
c t	0	0	0	0	0	0	0	0.978	0.981	0.980	0.979	0.982	0.979	0.980	0.975	0.976
c r t	0	0	0	0	0	0	0	0	0.980	0.979	0.990	0.981	0.981	0.978	0.979	0.977
ln	0	0	0	0	0	0	0	0	0	0.982	0.639	0.981	0.977	0.982	0.970	0.978
ln r	0	0	0	0	0	0	0	0	0	0	0.990	0.627	0.980	0.978	0.977	0.980
ln t	0	0	0	0	0	0	0	0	0	0	0	0.982	0.980	0.982	0.971	0.979
ln r t	0	0	0	0	0	0	0	0	0	0	0	0	0.979	0.979	0.975	0.977
rc	0	0	0	0	0	0	0	0	0	0	0	0	0	0.980	0.627	0.978
rc r	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.982	0.637
rc t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.980
rc r t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 3.2: Matriz de Disimilaridad para las señales de validación.

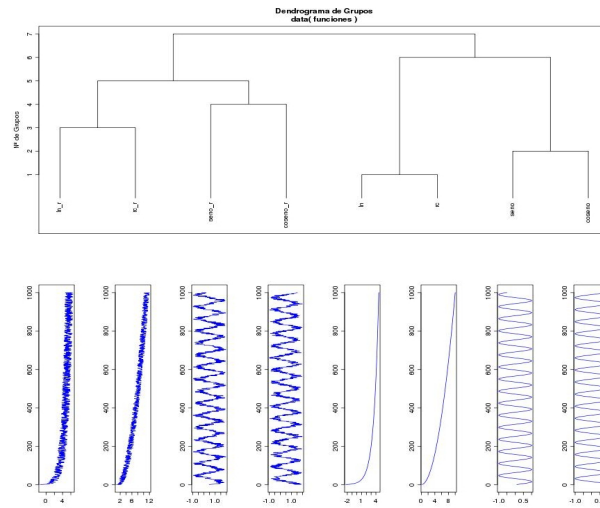


Figura 3.5: Dendrograma de grupos de algunas de las señales de validación (1000 muestras ruidosas).

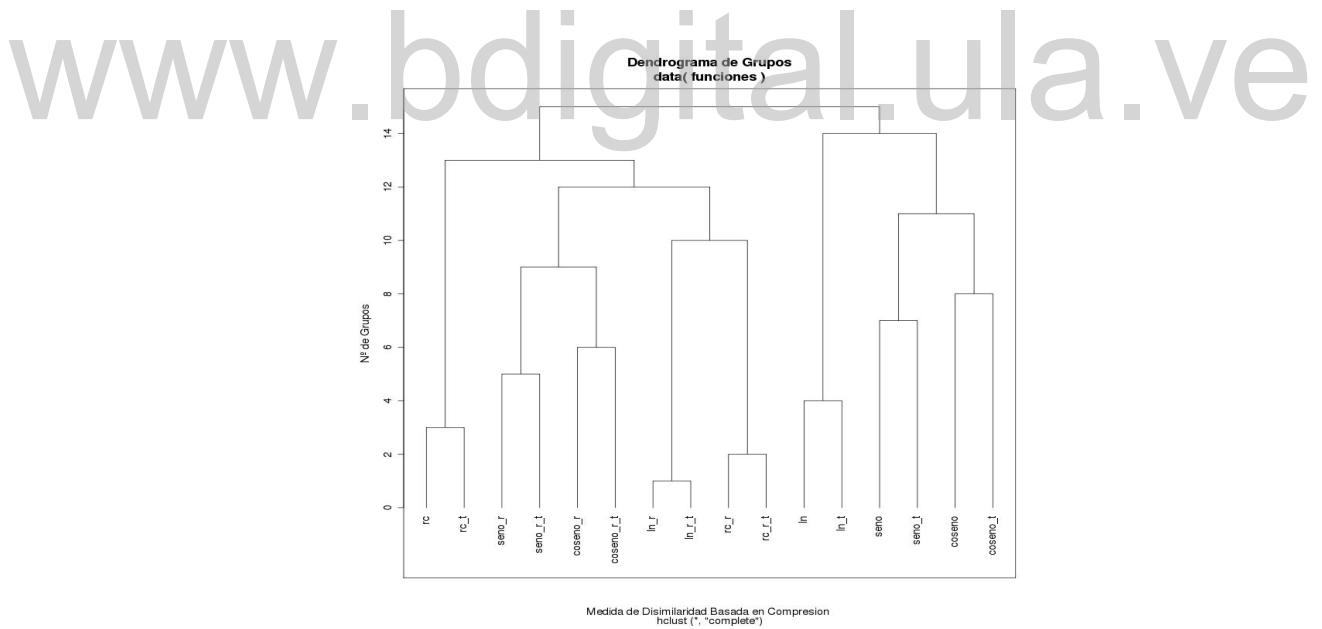


Figura 3.6: Dendrograma de grupos de las señales de validación (1000 muestras ruidosas).

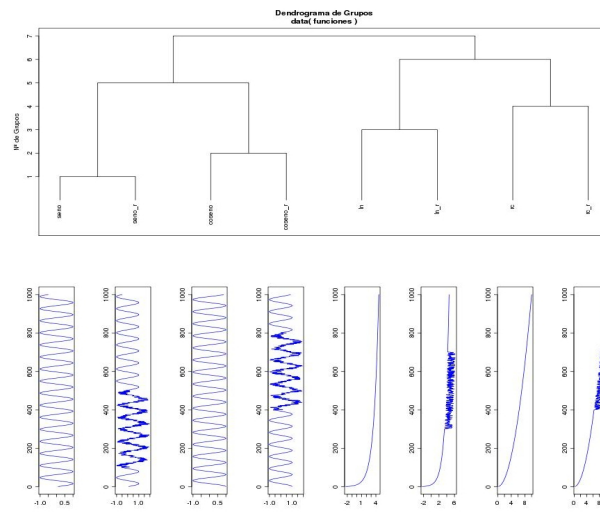


Figura 3.7: Dendrograma de grupos de algunas de las señales de validación (400 muestras ruidosas).

www.bdigital.ula.ve

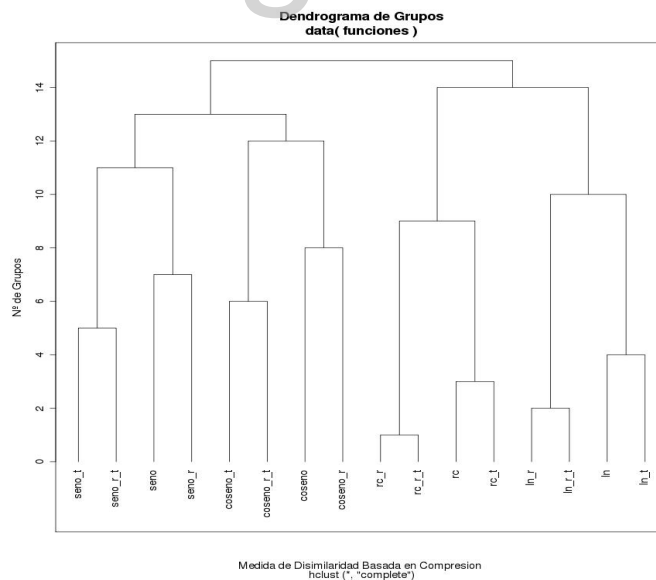


Figura 3.8: Dendrograma de grupos de las señales de validación (400 muestras ruidosas).

series raíz cuadrada y logaritmo neperiano. Estos resultados eran los esperados, puesto que se están logrando definir mediante el agrupamiento jerárquico los grupos de señales de validación. Cabe destacar que estos resultados se obtubieron con un 40 % de ruido presente en las señales.

3.2.4. Clasificación

En las Figuras 3.9 y 3.10 se muestran las series establecidas como individuos de entrenamiento y validación, respectivamente, para la aplicación del clasificador por vecindad kNN.

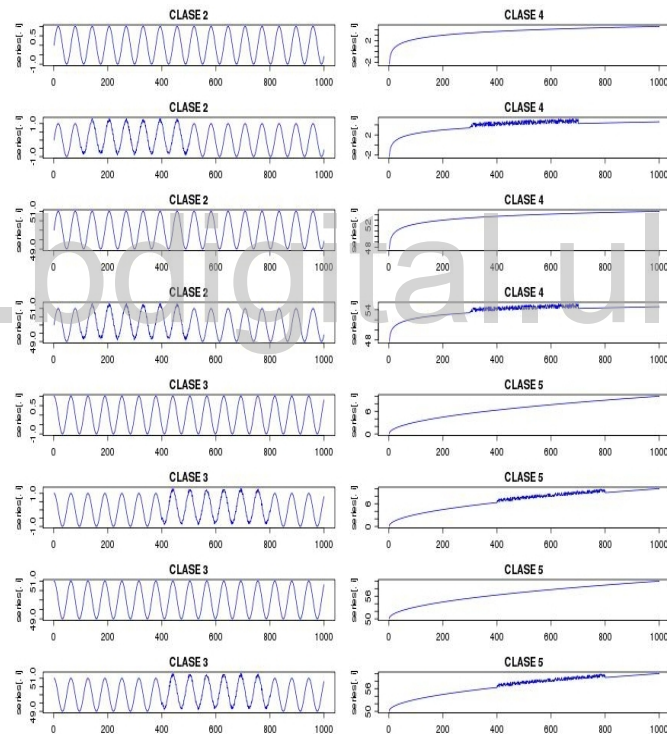


Figura 3.9: Individuos de entrenamiento para la clasificación con kNN.

En general, los individuos de validación tienen menor longitud que los de entrenamiento, esto es debido a que se deseaba comprobar que la CDM es capaz de comparar series de diferente longitud, tal como lo expresan sus autores (Keogh *et al.*, 2004). En la Tabla 3.3, se exponen los resultados de la clasificación utilizando el algoritmo kNN.

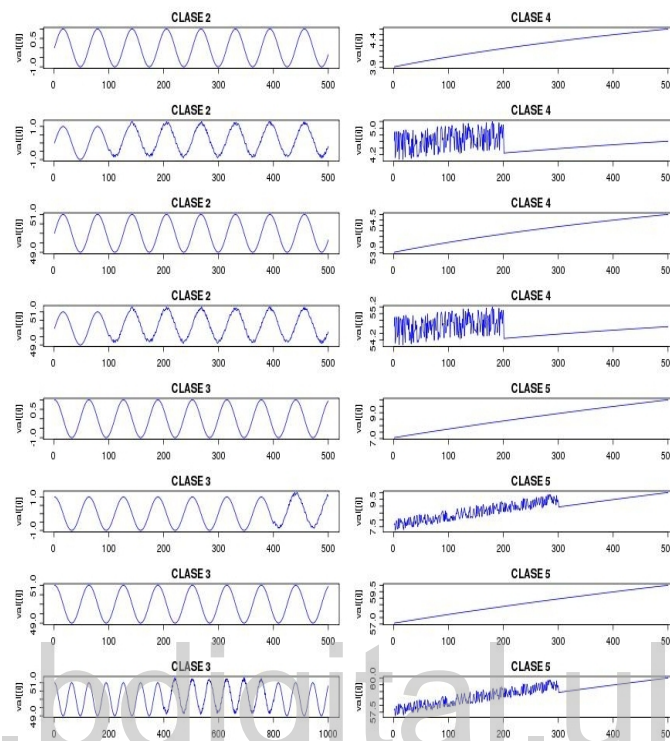


Figura 3.10: Individuos de validación para la clasificación con kNN.

N de Vecinos (k)	Porcentaje de acierto (%)
1	100
3	100
5	100
7	100

Tabla 3.3: Resultados de la clasificación para las señales de validación.

Capítulo 4

Resultados

A continuación se presentan los resultados del análisis de los registros electroencefalográficos con la técnica CDM. Estos resultados consisten en: Matrices de Disimilaridad para sujetos sanos y pacientes epilépticos, gráficos con la comparación entre canales para un sujeto sano, un paciente epiléptico y el paciente con crisis, dendrogramas correspondientes al agrupamiento de los canales de los EEG para todos los individuos en estudio, y por último, la tabla con los resultados de la clasificación con kNN.

4.1. Matriz de Disimilaridad

En las tablas 4.1, 4.2 y 4.3, se muestran las Matrices de Disimilaridad entre canales de un sujeto sano, uno epiléptico y el paciente con crisis, respectivamente. Se presentan solamente estas matrices ya que para toda la base de datos se observó el mismo comportamiento que el reflejado por estos individuos.

En dichas matrices, se puede apreciar que, en general, los sujetos enfermos tienen valores de Disimilaridad más bajos que los sanos, esto se puede percibir al comparar los valores mínimos (color rojo) y valores máximos obtenidos en las mismas. Estos resultados indican que existe una sincronización de la actividad cerebral en los pacientes epilépticos.

Para observar cómo es el comportamiento de la actividad cerebral (en forma global) de todos los sujetos bajo estudio, se realizó el cálculo del promedio de todos los valores

	Fp1.Ref (1)	Fp2.Ref (2)	F7.Ref (3)	F3.Ref (4)	F2.Ref (5)	F4.Ref (6)	F8.Ref (7)	T3.Ref (8)	C3.Ref (9)	C2.Ref (10)	C4.Ref (11)	T4.Ref (12)	T5.Ref (13)	P3.Ref (14)	P2.Ref (15)	P4.Ref (16)	T6.Ref (17)	O1.Ref (18)	O2.Ref (19)
1	0	0.992	0.991	0.991	0.991	0.992	0.992	0.992	0.991	0.992	0.992	0.992	0.992	0.992	0.991	0.992	0.992	0.992	0.992
2	0	0	0.992	0.991	0.991	0.992	0.992	0.992	0.991	0.991	0.991	0.993	0.992	0.991	0.991	0.992	0.992	0.992	0.992
3	0	0	0	0.992	0.991	0.992	0.992	0.990	0.991	0.991	0.991	0.991	0.990	0.991	0.991	0.991	0.989	0.991	0.991
4	0	0	0	0	0.991	0.992	0.991	0.991	0.990	0.991	0.990	0.992	0.991	0.991	0.991	0.991	0.991	0.991	0.992
5	0	0	0	0	0	0.991	0.991	0.992	0.990	0.990	0.990	0.991	0.991	0.990	0.991	0.991	0.991	0.991	0.991
6	0	0	0	0	0	0	0.990	0.990	0.989	0.989	0.989	0.991	0.990	0.989	0.990	0.989	0.990	0.989	0.989
7	0	0	0	0	0	0	0	0.990	0.991	0.992	0.991	0.991	0.990	0.991	0.992	0.992	0.991	0.992	0.992
8	0	0	0	0	0	0	0	0	0.991	0.991	0.990	0.990	0.989	0.991	0.990	0.990	0.989	0.991	0.990
9	0	0	0	0	0	0	0	0	0	0.991	0.991	0.992	0.991	0.990	0.990	0.991	0.991	0.991	0.991
10	0	0	0	0	0	0	0	0	0	0	0.991	0.992	0.990	0.990	0.990	0.990	0.990	0.990	0.991
11	0	0	0	0	0	0	0	0	0	0	0	0.992	0.991	0.991	0.991	0.991	0.991	0.991	0.991
12	0	0	0	0	0	0	0	0	0	0	0	0	0.992	0.994	0.993	0.993	0.991	0.994	0.994
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0.992	0.992	0.993	0.990	0.993	0.992
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.990	0.991	0.991	0.991	0.991
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.990	0.990	0.991	0.991
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.990	0.990	0.991
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.992	0.993
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.990
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 4.1: Matriz de Disimilaridad para el sujeto Normal 2.

	Fp1.A1A2 (1)	Fp2.A1A2 (2)	F7.A1A2 (3)	F3.A1A2 (4)	Fz.A1A2 (5)	F4.A1A2 (6)	F8.A1A2 (7)	T3.A1A2 (8)	C3.A1A2 (9)	Cz.A1A2 (10)	C4.A1A2 (11)	T4.A1A2 (12)	T5.A1A2 (13)	P3.A1A2 (14)	Pz.A1A2 (15)	P4.A1A2 (16)	T6.A1A2 (17)	O1.A1A2 (18)	O2.A1A2 (19)
1	0	0.962	0.966	0.965	0.971	0.969	0.968	0.963	0.965	0.961	0.963	0.965	0.963	0.975	0.962	0.973	0.962	0.963	0.966
2	0	0	0.976	0.968	0.965	0.967	0.972	0.967	0.966	0.967	0.967	0.969	0.966	0.974	0.963	0.967	0.962	0.960	0.966
3	0	0	0	0.977	0.968	0.976	0.969	0.967	0.971	0.969	0.975	0.969	0.970	0.971	0.969	0.970	0.971	0.970	0.975
4	0	0	0	0	0.969	0.968	0.978	0.977	0.965	0.977	0.967	0.975	0.967	0.966	0.962	0.966	0.965	0.966	0.969
5	0	0	0	0	0	0.964	0.973	0.969	0.965	0.963	0.964	0.968	0.969	0.964	0.964	0.967	0.967	0.963	0.979
6	0	0	0	0	0	0	0.970	0.967	0.967	0.965	0.966	0.971	0.964	0.969	0.966	0.969	0.968	0.966	0.971
7	0	0	0	0	0	0	0	0.972	0.973	0.975	0.971	0.970	0.968	0.980	0.972	0.974	0.970	0.967	0.972
8	0	0	0	0	0	0	0	0	0.976	0.966	0.966	0.970	0.967	0.972	0.971	0.970	0.969	0.970	0.972
9	0	0	0	0	0	0	0	0	0	0.961	0.963	0.968	0.965	0.966	0.961	0.966	0.966	0.960	0.967
10	0	0	0	0	0	0	0	0	0	0	0.961	0.971	0.967	0.964	0.962	0.963	0.966	0.962	0.965
11	0	0	0	0	0	0	0	0	0	0	0	0.969	0.966	0.965	0.964	0.968	0.964	0.965	0.966
12	0	0	0	0	0	0	0	0	0	0	0	0	0.971	0.974	0.969	0.972	0.971	0.968	0.971
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0.967	0.965	0.968	0.968	0.960	0.969
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.964	0.968	0.964	0.972	0.966
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.966	0.962	0.964	0.968
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.969	0.964	0.971
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.958	0.965
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.965
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 4.2: Matriz de Disimilaridad para el paciente Epiléptico 4.

	Fp1.Ref (1)	Fp2.Ref (2)	F7.Ref (3)	F3.Ref (4)	Fz.Ref (5)	F4.Ref (6)	F8.Ref (7)	T3.Ref (8)	C3.Ref (9)	Cz.Ref (10)	C4.Ref (11)	T4.Ref (12)	T5.Ref (13)	P3.Ref (14)	Pz.Ref (15)	P4.Ref (16)	T6.Ref (17)	O1.Ref (18)
1	0	0.989	0.991	0.995	0.990	0.990	0.931	0.995	0.990	0.990	0.990	0.989	0.989	0.991	0.990	0.990	0.990	0.988
2	0	0	0.931	0.995	0.927	0.991	0.992	0.996	0.991	0.991	0.991	0.990	0.990	0.991	0.990	0.991	0.991	0.990
3	0	0	0	0.996	0.932	0.991	0.991	0.998	0.990	0.990	0.990	0.990	0.990	0.991	0.990	0.990	0.990	0.989
4	0	0	0	0	0.996	0.995	0.996	0.960	0.996	0.995	0.996	0.995	0.994	0.996	0.996	0.995	0.995	0.997
5	0	0	0	0	0	0.991	0.991	0.998	0.991	0.990	0.990	0.990	0.990	0.992	0.990	0.990	0.991	0.990
6	0	0	0	0	0	0	0.991	0.997	0.991	0.991	0.990	0.990	0.991	0.991	0.990	0.990	0.991	0.990
7	0	0	0	0	0	0	0	0.994	0.991	0.991	0.991	0.990	0.990	0.991	0.990	0.990	0.991	0.990
8	0	0	0	0	0	0	0	0	0.996	0.996	0.995	0.996	0.995	0.995	0.996	0.995	0.995	0.996
9	0	0	0	0	0	0	0	0	0	0.990	0.933	0.933	0.991	0.931	0.990	0.991	0.990	0.924
10	0	0	0	0	0	0	0	0	0	0	0.990	0.990	0.932	0.991	0.990	0.991	0.991	0.990
11	0	0	0	0	0	0	0	0	0	0	0	0.933	0.991	0.931	0.990	0.990	0.991	0.926
12	0	0	0	0	0	0	0	0	0	0	0	0	0.990	0.933	0.990	0.990	0.990	0.927
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0.990	0.989	0.989	0.990	0.988
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.990	0.990	0.990	0.925
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.991	0.930	0.991
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.991	0.990
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.989
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 4.3: Matriz de Disimilaridad para el paciente con crisis.

de Disimilaridad presentes en la matriz de cada uno de los individuos. En la Tabla 4.4 se exhiben los resultados obtenidos.

A pesar de que los pacientes estaban recibiendo fármacos al momento de practicárseles el EEG, se observan valores de Disimilaridad (en todo el cerebro) más bajos en los enfermos que en los sanos. Esto sugiere que existe una regularización de la actividad cerebral asociada a la patología epiléptica.

Sujeto \ CDM	Disimilaridad
N 1	0.991499
N 2	0.991111
N 3	0.989900
N 5	0.976019
N 6	0.969913
E 1	0.972951
E 2	0.962826
E 3	0.973747
E 4	0.967698
E C	0.984924

Tabla 4.4: Disimilaridad de la actividad cerebral de todos los sujetos bajo estudio.

4.1.1. Análisis por canales

Al aplicar la metodología para la obtención del índice CDM para los dos grupos bajo estudio se encontró que este índice presenta un valor más alto para el grupo de sujetos sanos que para el grupo de pacientes epilépticos. En vista de que este comportamiento se obtiene para toda la base de datos, a continuación de la Figura 4.1 a la Figura 4.9 se presenta una comparación de los índices CDM obtenidos entre los canales que constituyen el EEG para un sujeto sano, un paciente epiléptico y el paciente epiléptico en crisis.

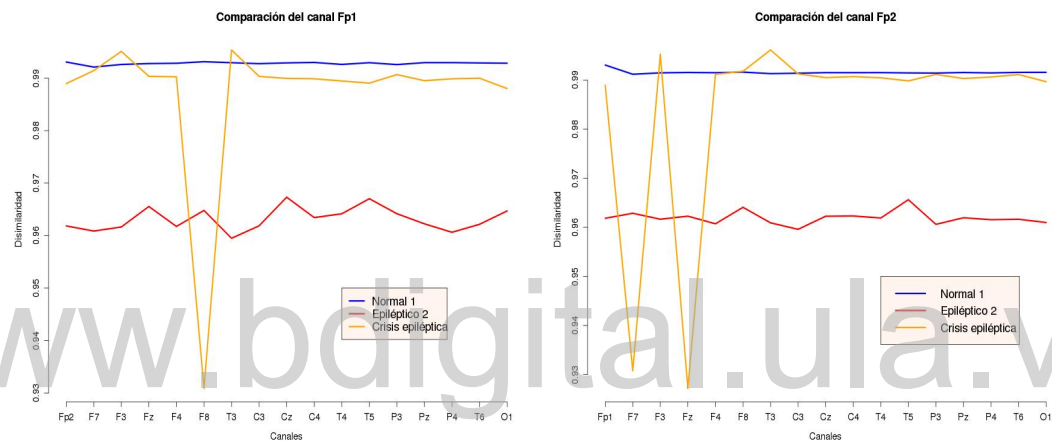


Figura 4.1: Comparación de los canales Fp1 (izquierda) y Fp2 (derecha).

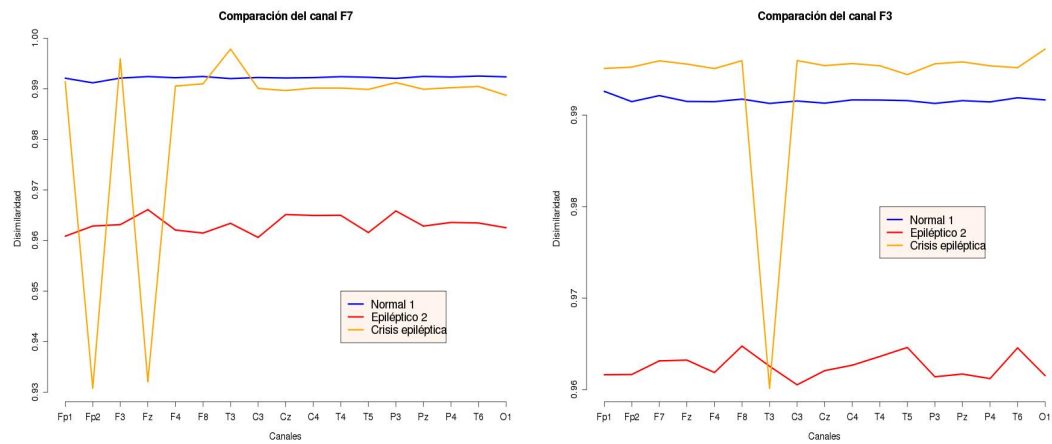


Figura 4.2: Comparación de los canales F7 (izquierda) y F3 (derecha).

www.bdigital.ula.ve

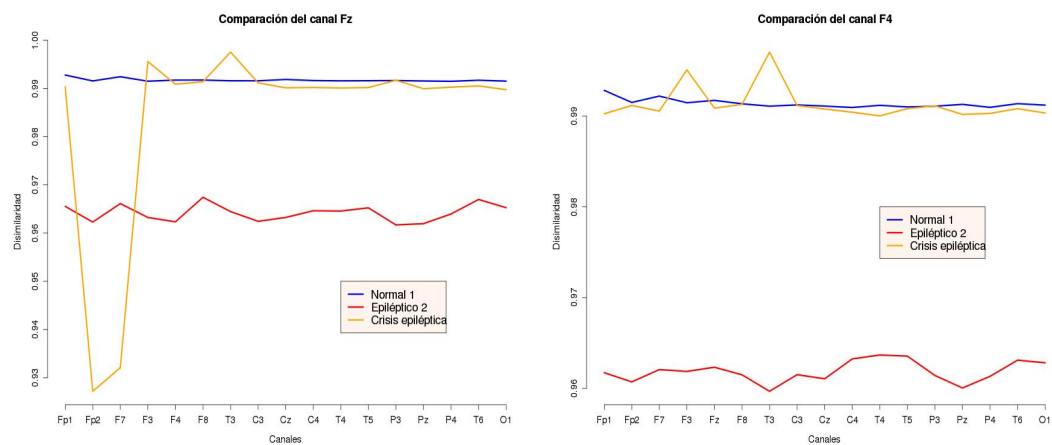


Figura 4.3: Comparación de los canales Fz (izquierda) y F4 (derecha).

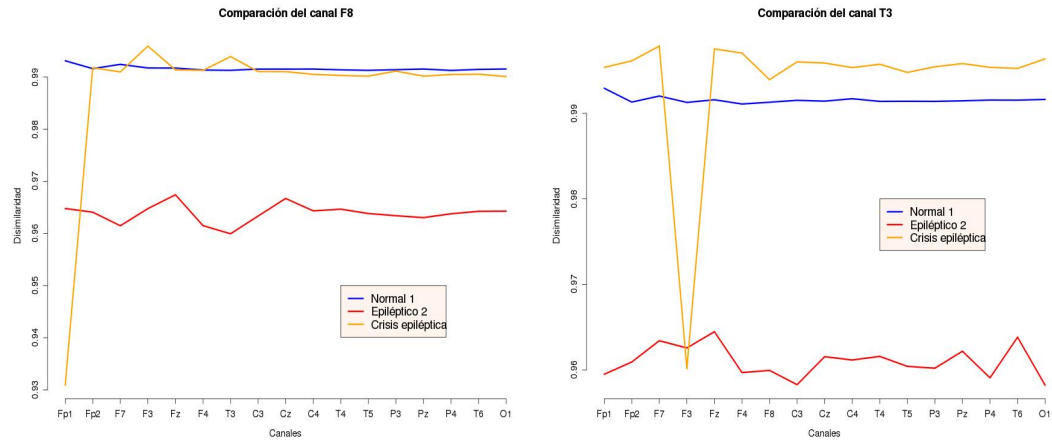


Figura 4.4: Comparación de los canales F8 (izquierda) y T3 (derecha).

www.bdigital.ula.ve

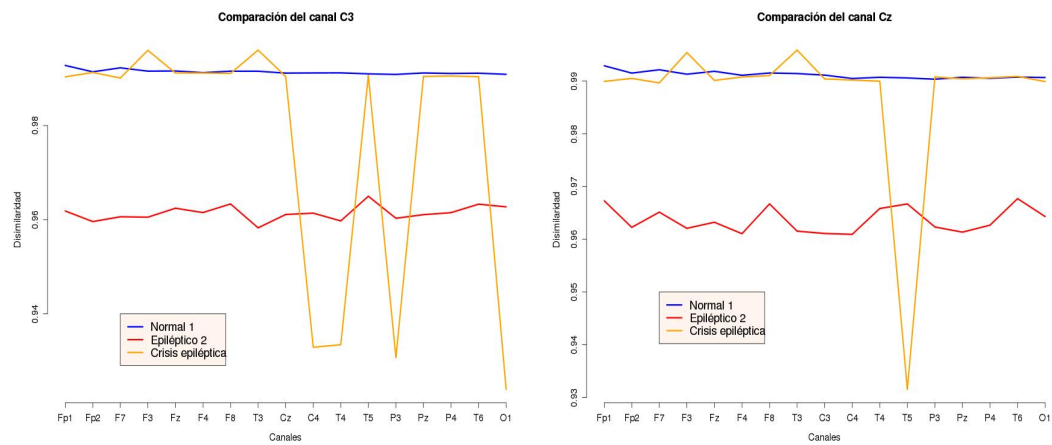


Figura 4.5: Comparación de los canales C3 (izquierda) y Cz (derecha).

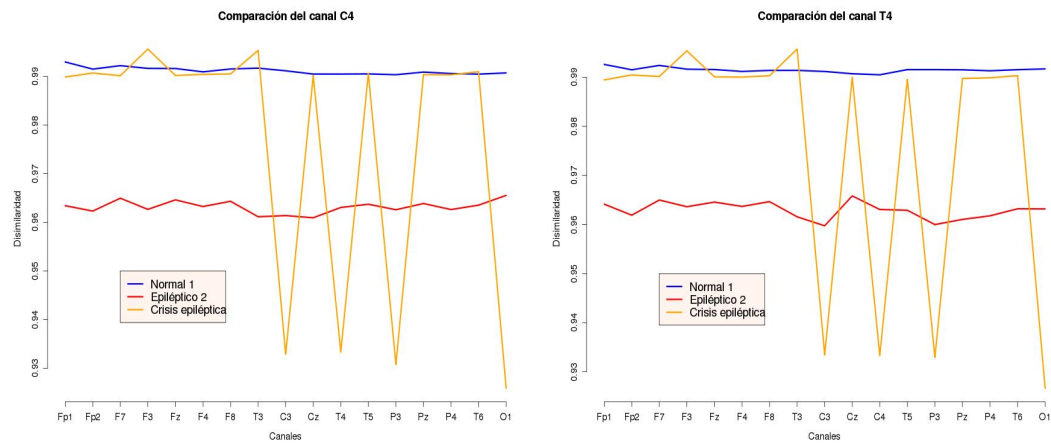


Figura 4.6: Comparación de los canales C4 (izquierda) y T4 (derecha).

www.bdigital.ula.ve

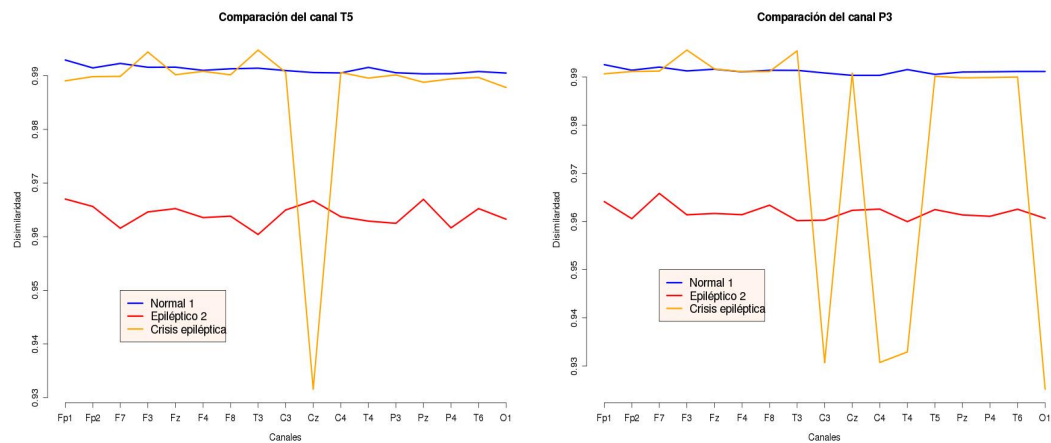


Figura 4.7: Comparación de los canales T5 (izquierda) y P3 (derecha).

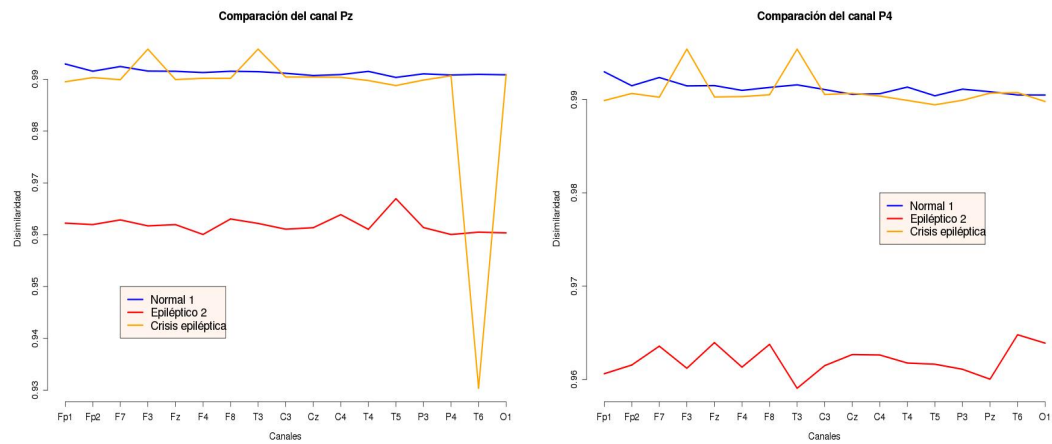


Figura 4.8: Comparación de los canales Pz (izquierda) y P4 (derecha).

www.bdigital.ula.ve

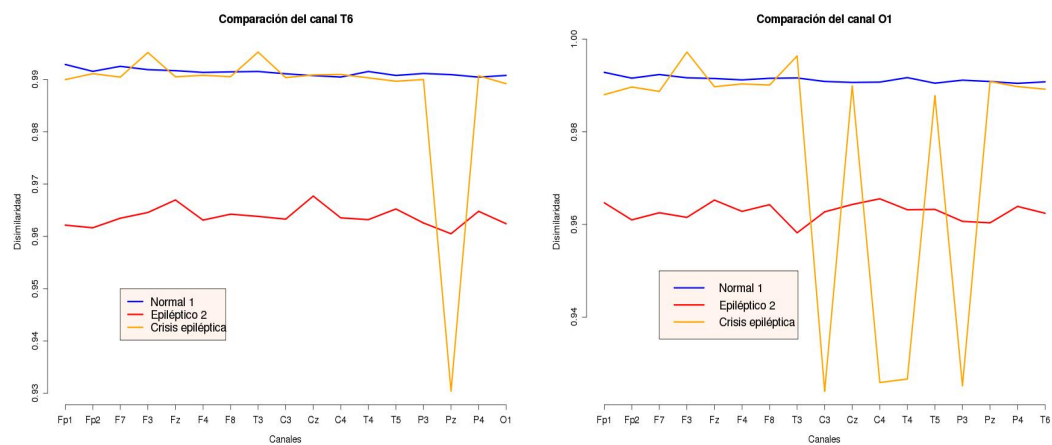


Figura 4.9: Comparación de los canales T6 (izquierda) y O1 (derecha).

4.2. Estudio de formación de grupos

De la Figura 4.10 a la Figura 4.15, se aprecian los dendrogramas correspondientes al agrupamiento de los canales del EEG de cada uno de los sujetos bajo estudio. Comenzando en la Figura 4.16 hasta la 4.30, se contemplan los resultados del proceso de creación de grupos al combinar los canales del EEG de cada uno de los sujetos sanos con todos los canales del EEG de los pacientes epilépticos.

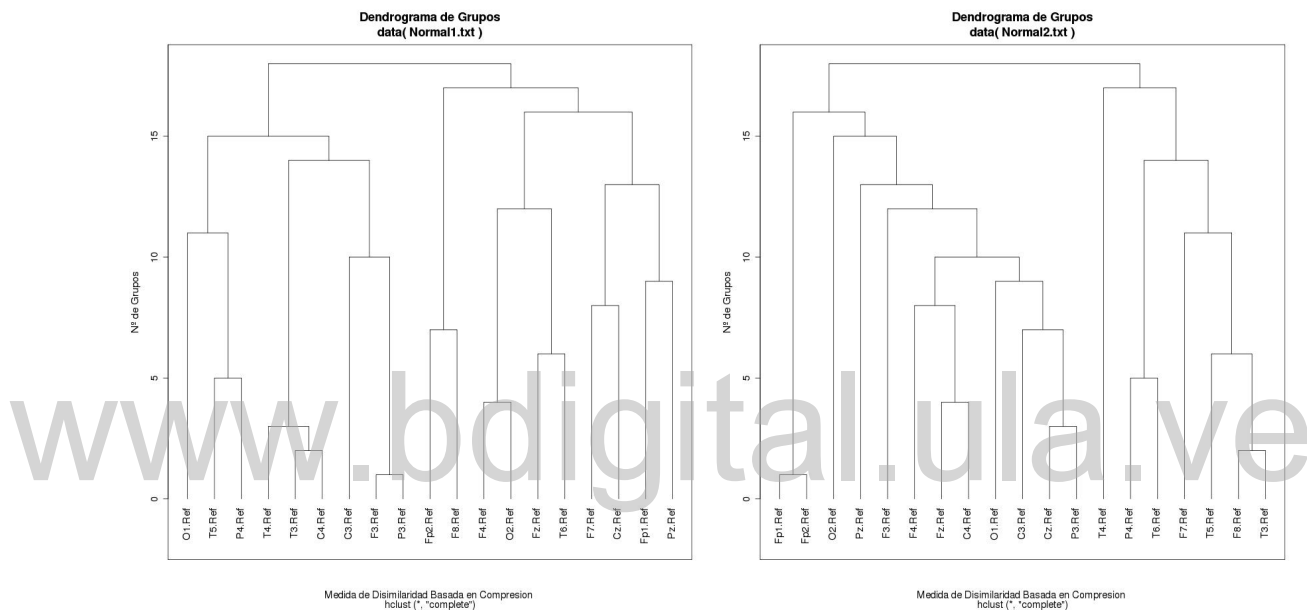


Figura 4.10: Dendrogramas de los sujetos Normal 1 (izquierda) y Normal 2 (derecha).

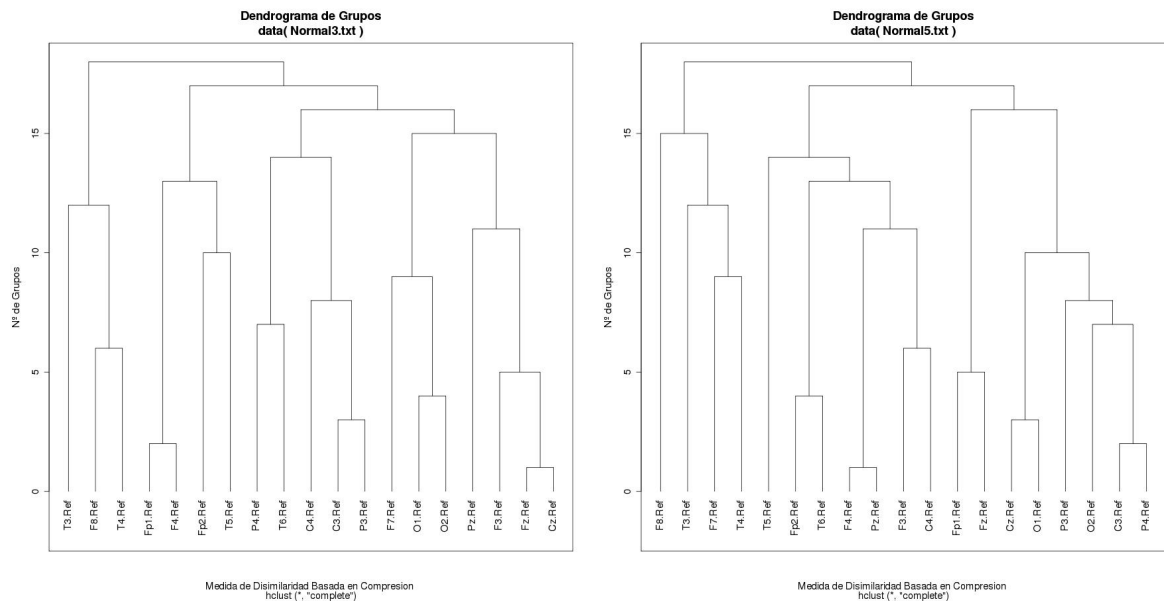


Figura 4.11: Dendrogramas de los sujetos Normal 3 (izquierda) y Normal 5 (derecha).

www.bdigital.ula.ve

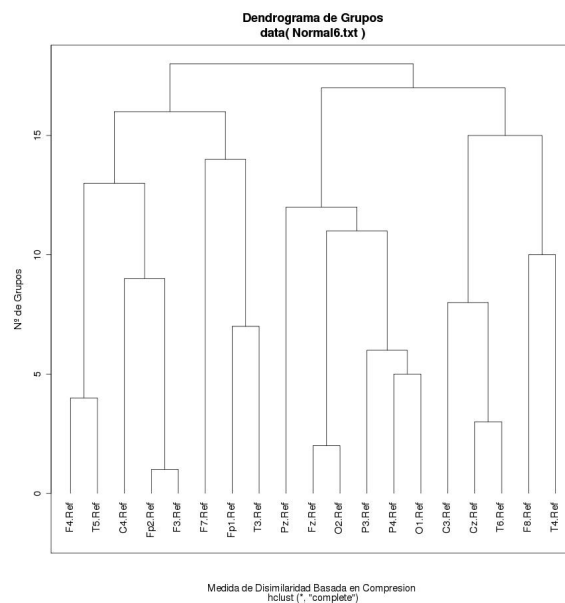


Figura 4.12: Dendrograma del sujeto Normal 6.

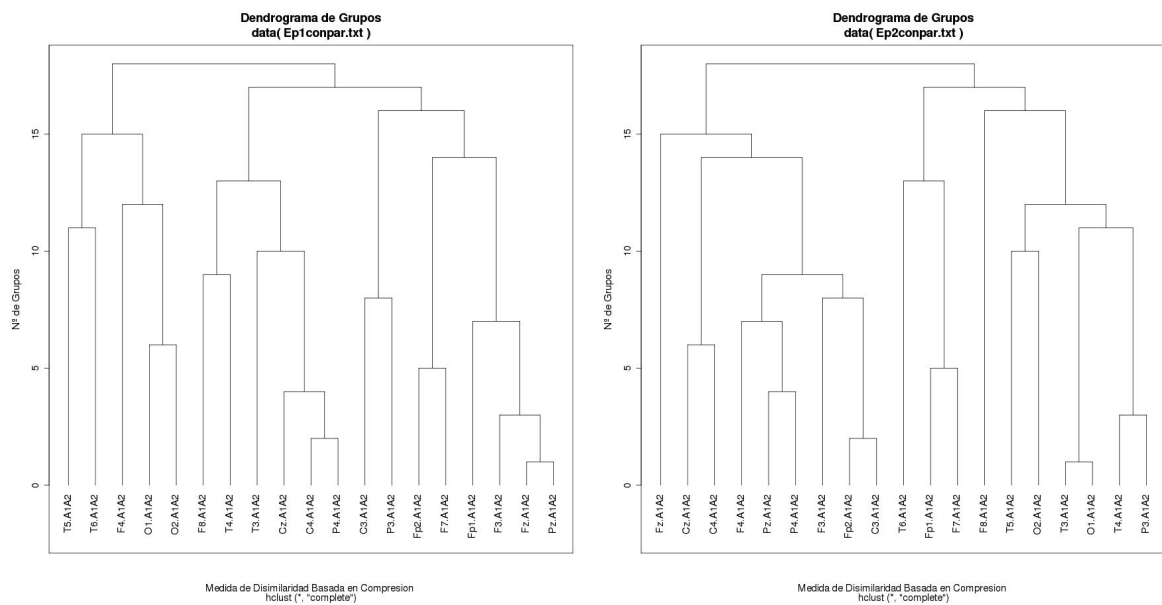


Figura 4.13: Dendrogramas de los pacientes Epiléptico 1 (izquierda) y Epiléptico 2 (derecha).

www.bdigital.ula.ve

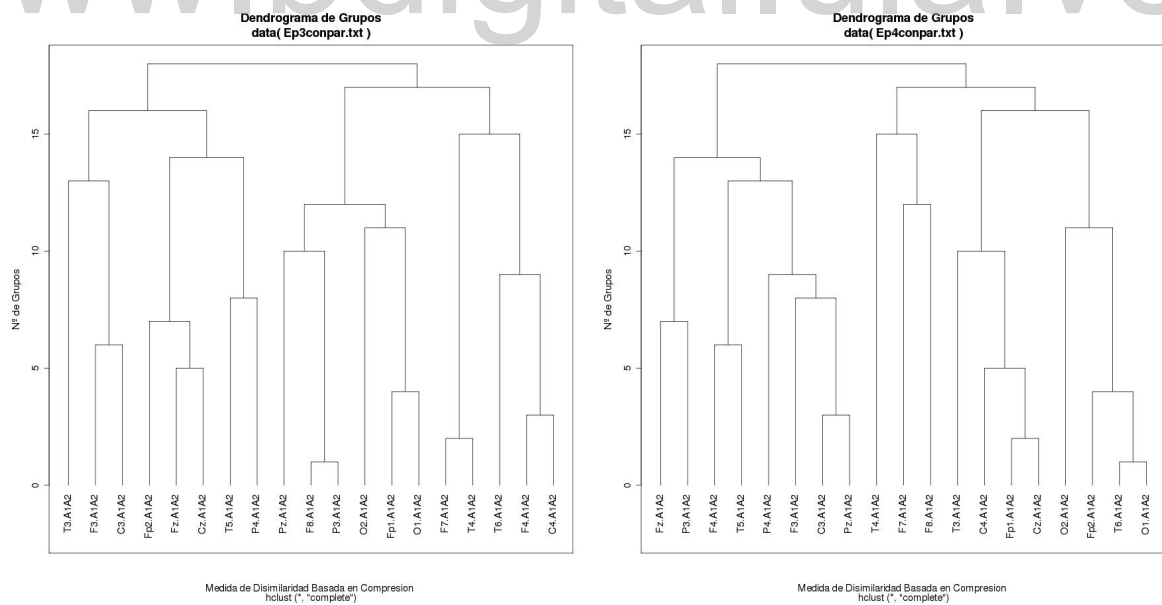


Figura 4.14: Dendrogramas de los pacientes Epiléptico 3 (izquierda) y Epiléptico 4 (derecha).

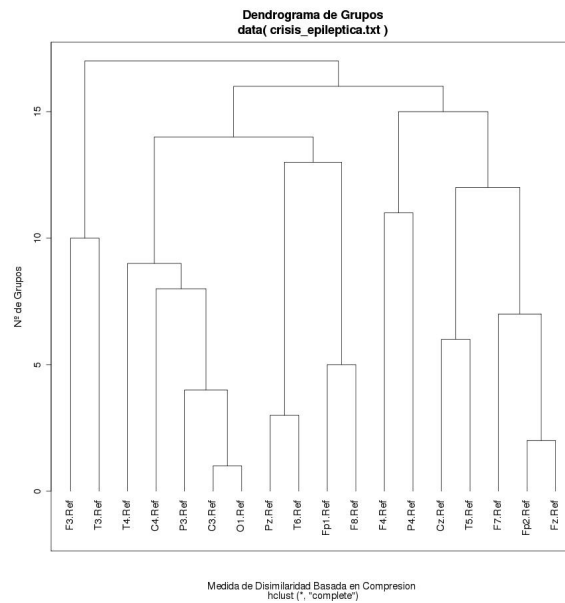


Figura 4.15: Dendrograma del paciente con crisis epiléptica.

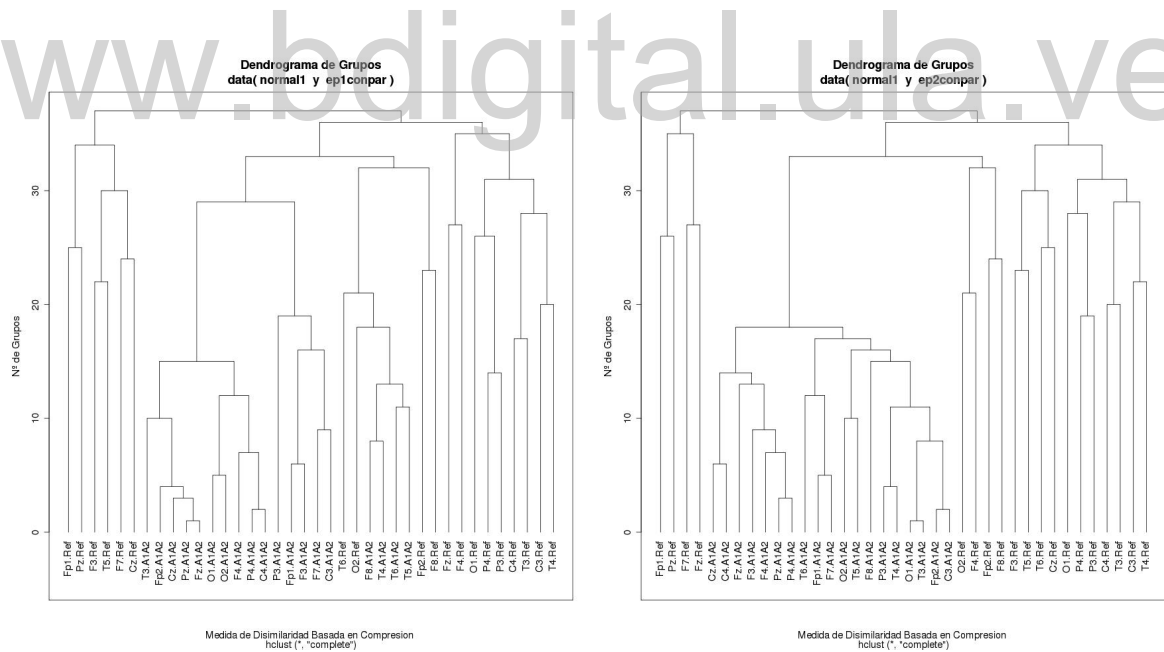


Figura 4.16: Dendrogramas del sujeto Normal 1 con Epiléptico 1 (izquierda) y Normal 1 con Epiléptico 2 (derecha).

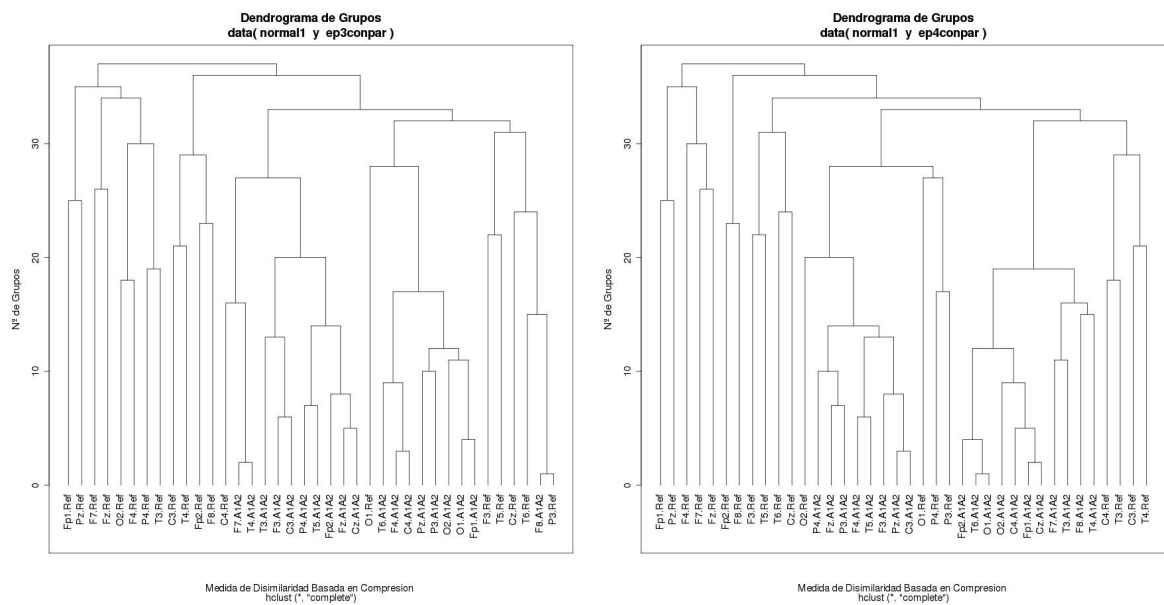


Figura 4.17: Dendrogramas del sujeto Normal 1 con Epiléptico 3 (izquierda) y Normal 1 con Epiléptico 4 (derecha).

www.bdigital.ula.ve

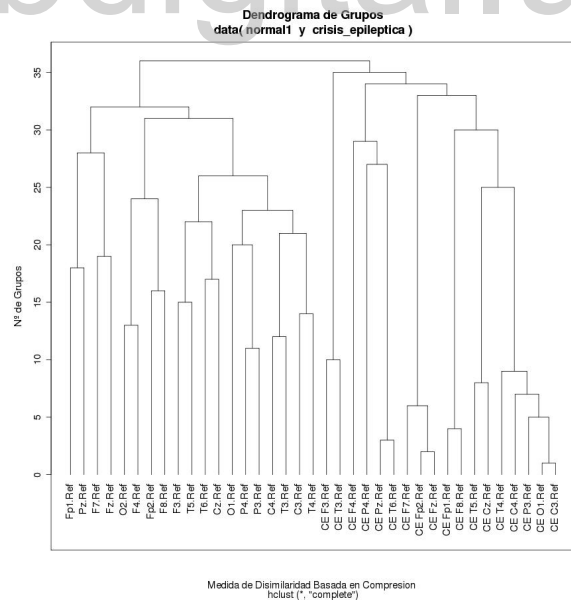


Figura 4.18: Dendrograma del sujeto Normal 1 con el paciente con crisis.

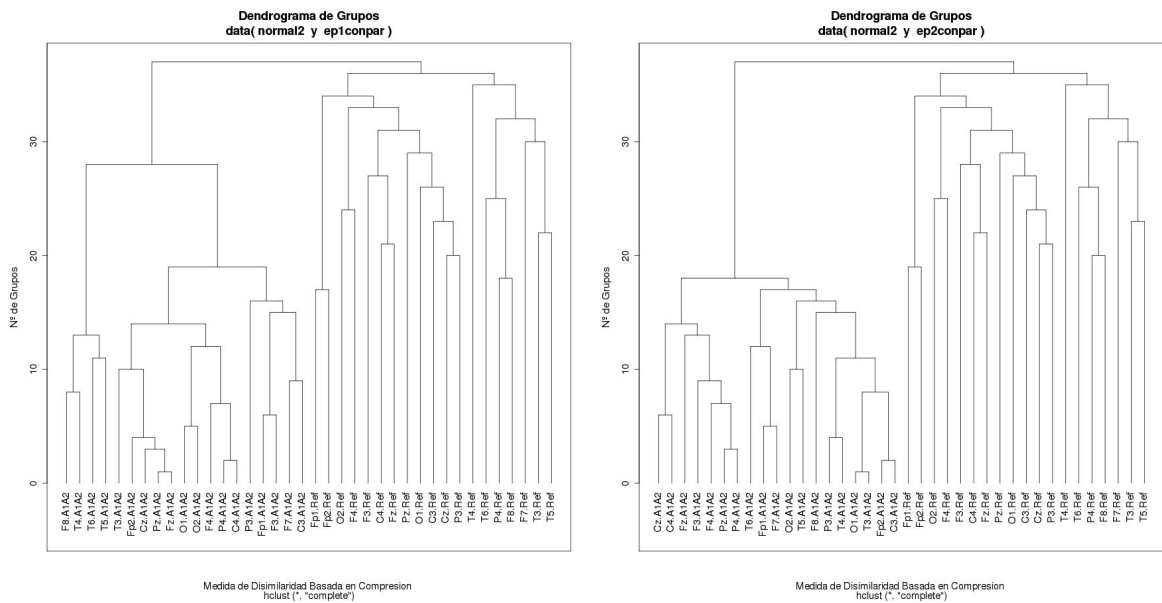


Figura 4.19: Dendrogramas del sujeto Normal 2 con Epiléptico 1 (izquierda) y Normal 2 con Epiléptico 2 (derecha).

www.bdigital.ula.ve

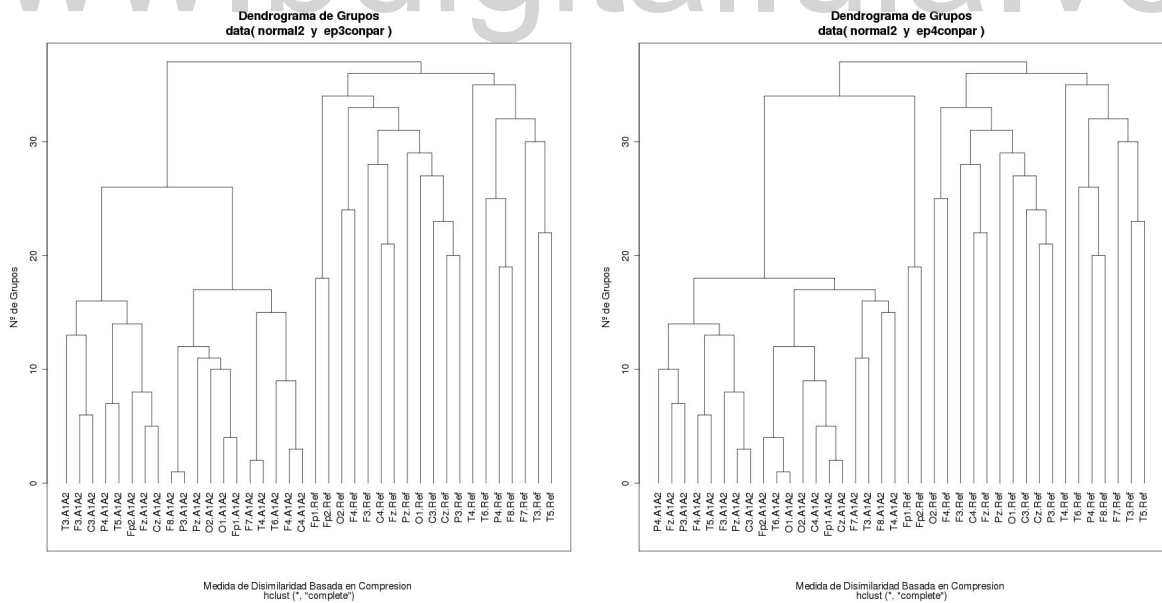


Figura 4.20: Dendrogramas del sujeto Normal 2 con Epiléptico 3 (izquierda) y Normal 2 con Epiléptico 4 (derecha).

Normal 2 con crisis epileptica

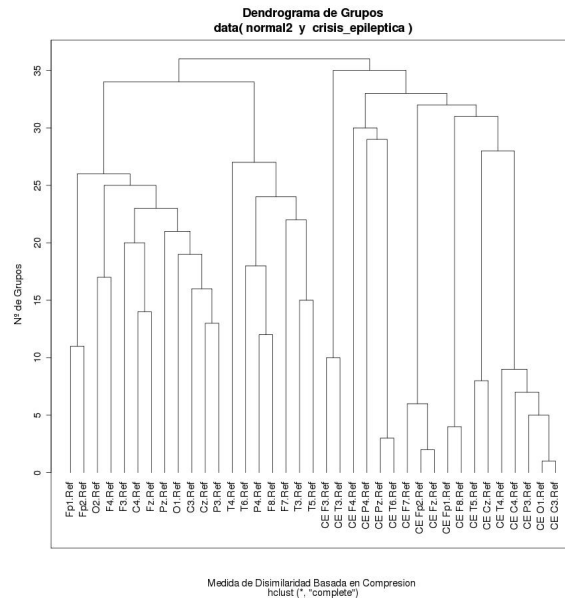


Figura 4.21: Dendrograma del sujeto Normal 2 con el paciente con crisis.

www.bdigital.ula.ve

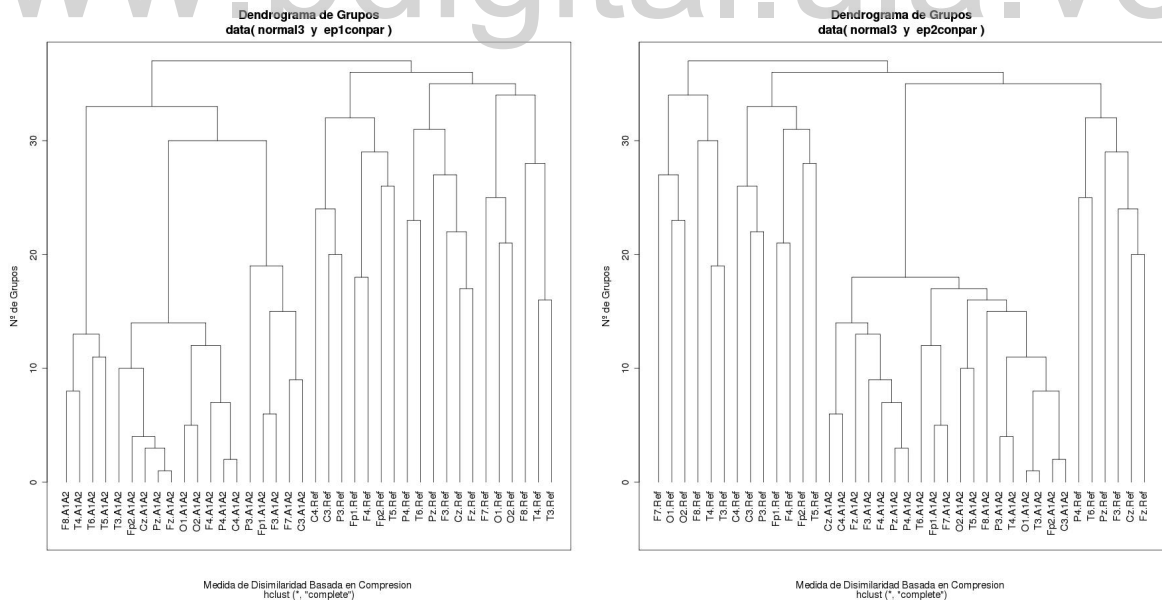


Figura 4.22: Dendrogramas del sujeto Normal 3 con Epiléptico 1 (izquierda) y Normal 3 con Epiléptico 2 (derecha).

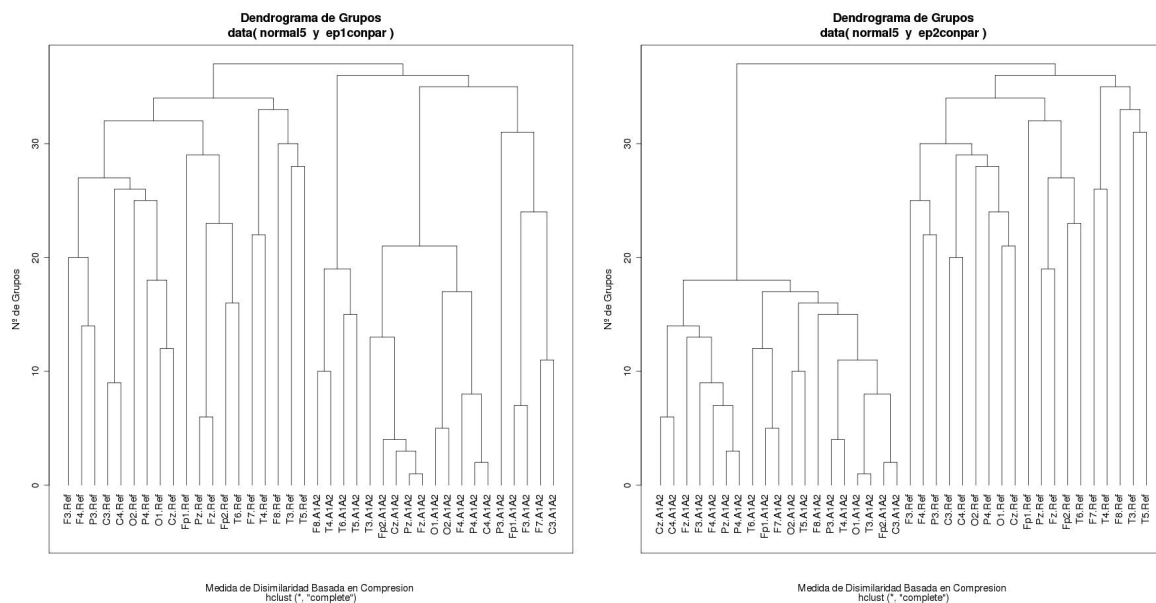


Figura 4.25: Dendrogramas del sujeto Normal 5 con Epiléptico 1 (izquierda) y Normal 5 con Epiléptico 2 (derecha).

www.bdigital.ula.ve

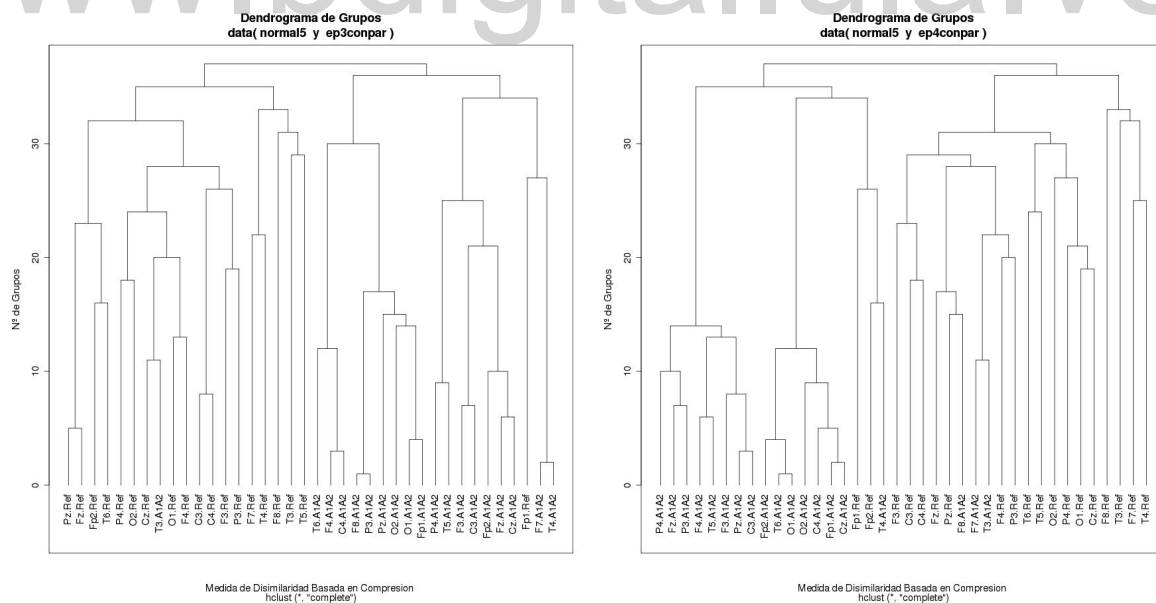


Figura 4.26: Dendrogramas del sujeto Normal 5 con Epiléptico 3 (izquierda) y Normal 5 con Epiléptico 4 (derecha).

Normal 5 con crisis epileptica

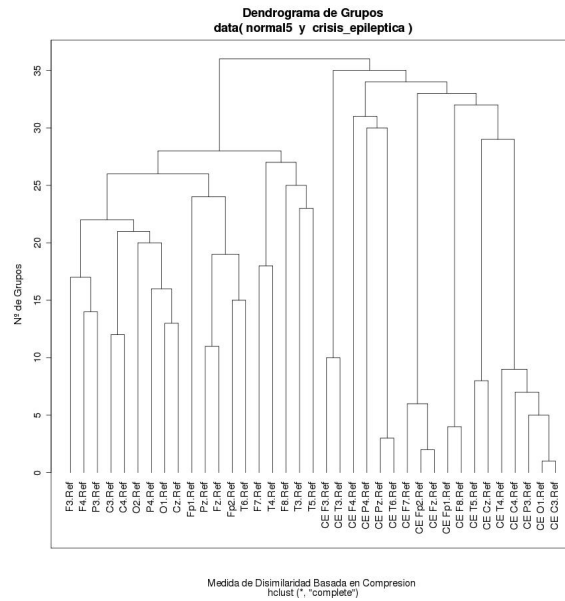


Figura 4.27: Dendrograma del sujeto Normal 5 con el paciente con crisis.

www.bdigital.ula.ve

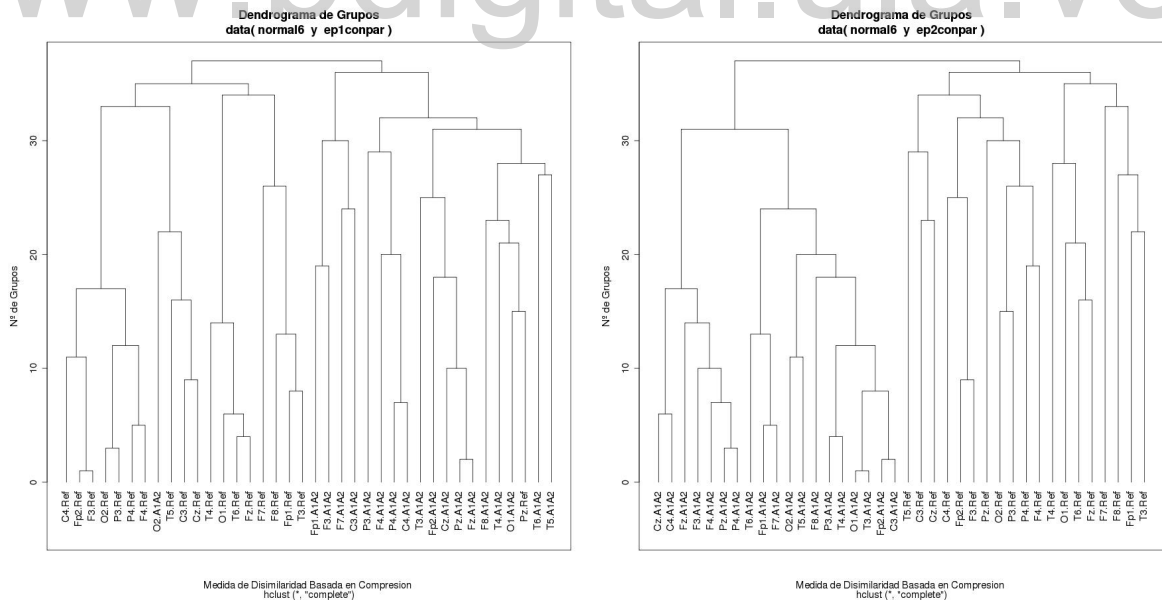


Figura 4.28: Dendrogramas del sujeto Normal 6 con Epiléptico 1 (izquierda) y Normal 6 con Epiléptico 2 (derecha).

4.3. Clasificación

Los resultados de la aplicación del algoritmo kNN con CDM como métrica de distancia entre los individuos se resumen en la Tabla 4.5, en donde se aprecian los porcentajes de instancias correctamente clasificadas para diferentes cantidades de vecinos cercanos.

Número de Vecinos (k)	Sanos	Epilépticos
5	100 %	100 %
7	100 %	100 %
9	100 %	100 %
11	100 %	100 %
13	100 %	100 %

Tabla 4.5: Resultados de la clasificación para las señales EEG.

En la Tabla 4.5, se puede notar que para todos los valores de k considerados se obtiene un porcentaje de instancias correctamente clasificadas del 100 %. Lo que indica que, sin importar el valor de k que se elija (dentro de este rango), el algoritmo de kNN realiza eficientemente la distinción entre un EEG de un sujeto sano y un EEG de un paciente epiléptico.

4.4. Discusión de resultados

En el análisis de los EEG mediante la Matriz de Disimilaridad se encontró que los pacientes epilépticos presentaban valores de CDM más bajos que los sujetos sanos, asimismo, este tipo de comportamiento se pudo apreciar en los promedios de Disimilaridad calculados para cada EEG. Este comportamiento de los pacientes evidencia una sincronización de la actividad cerebral asociada a la patología.

Al efectuar una inspección visual de los valores de CDM en la comparación entre canales, se puede apreciar que los pacientes epilépticos presentan valores de CDM más bajos que los sujetos sanos, evidenciando esto un comportamiento de similaridad (sincronización) de la actividad cerebral. Asimismo, se puede apreciar que para el paciente en crisis existen zonas de disimilaridad y similaridad (sincronización), cuantificada por

una valor bajo de la CDM, al comparar entre canales. La fluctuación de la CDM permite caracterizar las crisis en vista de la predicción de la misma. Clínicamente, la crisis epiléptica focal presentada durante la realización del EEG corresponde a un tipo de crisis uncinada del lóbulo fronto temporal, dada por la sensación de olor putrefacto de minutos de duración sin pérdida del nivel de consciencia, observándose relación directa entre los canales graficados (F7, F8, Fp1, Fp2, F3, F4, Fz, T3, T4, T5, T6) y el área anatomofuncional afectada, las zonas de disimilaridad y similaridad se pueden explicar debido a la acción contrarreguladora protectora de homeostasis del área cerebral del foco epiléptico que está descargando.

En los dendrogramas donde se combinan los EEG de un sujeto normal con uno enfermo, se puede apreciar que, en general, se forman dos grandes grupos, el correspondiente al sujeto sano y el del paciente epiléptico. En los casos en que canales de un sujeto enfermo se agrupan con los de los de un sano, se debe a que los pacientes estaban bajo tratamiento médico al momento de practicárseles el electroencefalograma. Además, existe una diferencia de jerarquía entre los dos grupos observados, sanos y enfermos, en donde estos últimos presentan alturas del árbol menor. Lo cual indica una sincronización entre de la actividad cerebral entre canales.

En la clasificación se obtuvo un excelente resultado, 100 % de instancias correctamente clasificadas. Este rendimiento es gracias a que en el sujeto epiléptico con crisis, se puede apreciar en buena medida los paroxismos inherentes a la Epilepsia, lográndose así, distinguir entre los sujetos sanos y los enfermos.

Capítulo 5

Conclusiones y Recomendaciones

Este capítulo exhibe las conclusiones a las que se llegaron a lo largo del desarrollo de este estudio y señala algunas recomendaciones para continuar aportando mejoras para el diagnóstico de la enfermedad, como también extendiendo el universo de aplicación de la técnica.

5.1. Conclusiones

La aplicación del índice CDM presentada en este trabajo permitió caracterizar y diferenciar la actividad cerebral entre el grupo de pacientes epilépticos y el grupo de sujetos sanos.

La sincronización temporal y frontal del índice CDM en el paciente epiléptico en crisis se correlaciona clínicamente con una crisis focal.

Un agrupamiento jerárquico de los índices CDM obtenidos para los grupos en estudio permitió separar ampliamente la actividad cerebral que caracteriza cada grupo. Adicionalmente, el agrupamiento para los pacientes epilépticos evidenció una sincronización entre los canales y para un paciente epiléptico en crisis mostró zonas de sincronización y desincronización, clínicamente relacionadas con la epilepsia focal.

El uso de técnicas computacionales, como las presentadas en este trabajo, permite excluir la subjetividad del médico neurólogo interpretador de las señales del EEG en el momento de analizar visualmente las señales electroencefalográficas, dando información

vital y precisa en el diagnóstico y decisión terapéutica.

El índice CDM es una técnica prometedora en el campo de la epilepsia (médica-quirúrgica) al permitir ubicar anatómicamente el foco epiléptico durante la descarga cerebral anormal, por lo que es necesario ampliar la muestra y correlacionar con otras variedades de epilepsia.

En vista de los resultados obtenidos, la técnica CDM podría ser útil en la monitorización y análisis de registros electroencefalográficos, que permitiría cuantificar y caracterizar los cambios de complejidad de la actividad cerebral de pacientes epilépticos

5.2. Recomendaciones

El trabajo realizado en este Proyecto de Grado es una investigación inicial en la aplicación de la técnica CDM para analizar registros electroencefalográficos de pacientes epilépticos. Para estudios futuros se recomienda lo siguiente:

- Realizar el estudio con mayor cantidad de sujetos, tanto epilépticos como sanos.
- Tener información del tipo de Epilepsia que posean los sujetos enfermos y de los medicamentos que le estén siendo suministrados. De ser posible, realizar el estudio con sujetos a los que no se les estén suministrando fármacos.
- Combinar la técnica CDM con otros métodos de clasificación. Por ejemplo, Máquinas de Soporte Vectorial y Redes Neuronales.
- Utilizar la CDM en conjunto con otras técnicas para la predicción de crisis epilépticas.
- Realizar un sistema experto que permita el análisis de Electroencefalogramas a través de técnicas como las presentadas en este estudio.

Bibliografía

- Cilibrasi, R. and Vitányi, P. (2005). Clustering by compression. *IEEE T. Inform. Theory*, pages 1523–1545.
- Coutín, E., Arias, S., Guillén, P., and Coutín, P. (2005). Análisis de registros electroencefalográficos en pacientes epilépticos utilizando dinámica no lineal. In *Memorias del Congreso ASME USB 2005*, Caracas, Venezuela.
- Escalona-Morán, M., Consenza, M., Guillén, P., and Coutín, P. (2007). Synchronization and clustering in electroencephalographic signals. *Chaos Solitons and Fractals*, 13(6):820–825.
- Hojjat, A., Ziqin, Z., and Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods* 123, pages 69–87.
- Kandel, E., Schwartz, J., and Jessell, T. (1991). *Principles of Neural Science*. Elsevier Science Publishing Co., New York, third edition.
- Keogh, E., Lonardi, E., and Ratanamahatana, C. (2004). Towards parameter-free data mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Seattle, USA.
- Li, M., Badger, J., Chen, X., Kwong, S., Kearney, P., and Zhang, H. (2001). An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics*, pages 149–154.
- Min, L. and Vitányi, P. (1997). *A Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, second edition.

Ming, L., Xin, C., l. Xin, Bin, M., and Vitányi, P. (2004). The similarity metric. *IEEE Transactions on information theory*, 50(12):3250–3264.

Naive, H., Sanz, J., Malavé, V., Barrera, M., Spinetti, H., and Guillén, P. (2009). Máquinas de aprendizaje para clasificar señales electroencefalográficas. *Avances en Sistemas e Informática*, 6(1).

Niedermeyer, E. (2005). *Electroencephalography, basic principles, clinical applications, and related fields*. Williams and Wilkins, Philadelphia, USA, fifth edition.

Zarranz, J. (2003). *Neurología*. Elsevier España S.A., Madrid, España, tercera edition.

www.bdigital.ula.ve

Apéndice A

Uso de las funciones en R

En esta sección se explica el uso de las funciones que fueron programadas para el desarrollo de este proyecto.

A.1. CDM: Medida de disimilaridad Basada en Compresión

Esta función permite realizar el cálculo de la CDM entre dos series temporales.

A.1.1. Parámetros de entrada

- *serie1*: vector que contiene una de las series a comparar.
- *serie2*: vector que contiene una de las series a comparar.
- *modo*: este parámetro indica cuál de los compresores se utilizará en los cálculos.

A.1.2. Salida

Número real con el valor de CDM entre las series comparadas.

A.2. Matriz de disimilaridad

Esta función permite realizar el cálculo de la CDM entre un grupo de n series temporales.

A.2.1. Parámetros de entrada

- *series*: es un data frame, lista ó matriz que contiene las series temporales a comparar.
- *modo*: este parámetro indica cuál de los compresores se utilizará en los cálculos.

A.2.2. Salida

Matriz que contiene los valores de CDM entre cada par de series. Solamente se calcula la triangular superior de la matriz.

A.3. Agrupamiento Jerárquico

Esta función permite realizar el agrupamiento de series temporales.

A.3.1. Parámetros de entrada

- *series*: es un data frame, lista ó matriz que contiene las series temporales a comparar.
- *metodo*: indica el método de proximidad entre grupos para realizar los *clusters*,
- *modo*: este parámetro indica cuál de los compresores se utilizará en los cálculos.
- *data*: nombre de los datos que serán agrupados.
- *grafico*: indica si se graficarán los dendrogramas o no.
- *cex*: tamaño de los ejes del dendrograma.

A.3.2. Salida

Devuelve una lista con un objeto de la clase *cluster*(*hclust*) y los gráficos de los dendrogramas.

A.4. Clasificación

Esta función permite realizar la clasificación de series temporales mediante el algoritmo del vecino más cercano.

A.4.1. Parámetros de entrada

- *individuo*: es un vector que contiene la serie temporal a clasificar.
- *modo*: este parámetro indica cuál de los compresores se utilizará en los cálculos.
- *series*: data frame, matriz o lista que contiene las series temporales consideradas como individuos de entrenamiento.
- *clases*: vector que contiene las clases asociadas a las series de entrenamiento.
- *kNN*: número de vecinos cercanos a considerar.

A.4.2. Salida

Devuelve una lista con: la clase asignada al individuo, un vector con los nombres y las disimilaridades de los vecinos cercanos y otro vector con las clases de los vecinos.