



INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

DISEÑO Y APLICACIÓN DE
HERRAMIENTAS
COMPUTACIONALES PARA
ELECTROENCEFALOGRAFÍA
CUANTITATIVA

TESIS

QUE PARA OBTENER EL GRADO DE:

Maestra(o) en Sistemas Computacionales

Presenta:

Yoselyn Nohemí Ortega Gijón

Director:

Alejandro del Rey Torres Rodríguez

Co-Director:

Carlos Arturo Hernández Gracidas

Misantla, Veracruz Agosto del 2018



**INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA
DIVISIÓN DE ESTUDIOS PROFESIONALES
AUTORIZACIÓN DE IMPRESIÓN DE TRABAJO DE TITULACIÓN MAESTRÍA**

FECHA: 07 de Agosto de 2018.

ASUNTO: **AUTORIZACIÓN DE IMPRESIÓN
DE TESIS.**

A QUIEN CORRESPONDA:

Por medio de la presente se hace constar que el (la) C:

YOSELYN NOHEMÍ ORTEGA GIJÓN

estudiante de la maestría en SISTEMAS COMPUTACIONALES con No. de Control 162T0082 ha cumplido satisfactoriamente con lo estipulado por el **Lineamiento de Posgrado para la obtención del grado de Maestría** mediante **Tesis.**

Por tal motivo se **Autoriza** la impresión del **Tema** titulado:

**DISEÑO Y APLICACIÓN DE HERRAMIENTAS COMPUTACIONALES PARA
ELECTROENCEFALOGRAFÍA CUANTITATIVA**

Dándose un plazo no mayor de un mes de la expedición de la presente a la solicitud del examen para la obtención del grado de maestría.

ATENTAMENTE

**DR. Alejandro del Rey Torres Rodríguez
Presidente**



**M.I.A. Roberto Ángel Meléndez Armenta
Secretario**

**M.S.C. Eddy Sánchez de la Cruz
Vocal**

Archivo.

Dedicatoria

A mi Madre Delfina Gijón por su apoyo en cada decisión que tomaba, por acompañarme en cada paso que doy, por cada consejo que me ha dado siempre velando por mi bien, por mostrarme que el amor de una familia es lo mas importante, la quiero muchísimo y Gracias por todo.

A mi hermano Tomas por siempre estar a mi lado haciéndome reir en momentos duros y ayudarme cuando siempre lo necesite.

A mis amigos, que gracias a su apoyo, y conocimientos hicieron de esta experiencia una de las mas especiales.

Agradecimientos

A mis formadores, personas de gran sabiduría quienes se han esforzado por ayudarme a llegar al punto en el que me encuentro.

Al CONACyT y al Dr. Carlos Arturo Hernández Gracidas, responsable técnico del proyecto SALUD 2016-1 272345, por la colaboración y apoyo en este trabajo de investigación.

Al Dr. Alejandro del Rey Torres quien se ha tomado el arduo trabajo de transmitirme sus diversos conocimientos, especialmente del campo y de los temas que corresponden a mi profesión. Por siempre estar dispuesto a compartir sus conocimientos.

Resumen

En este trabajo se presentan los resultados obtenidos de la colaboración en un proyecto de investigación sobre salud mental. El objetivo del proyecto de investigación es desarrollar y aplicar modelos computacionales, para el análisis e interpretación cuantitativa de datos obtenidos por electroencefalografía. En esta tesis se elaboró una propuesta de modelos y algoritmos, que fueron implementados en entornos de programación libre. La propuesta de modelos consta de algoritmos de análisis de señales en minería de datos, considerando las señales electroencefalográficas como una serie de tiempo multivariada. Por otra parte, también se propone un modelo de red bayesiana para caracterizar la estructura espacial de las señales electroencefalográficas, con el propósito de analizar la estructura de correlaciones entre diferentes canales de la señal de acuerdo con su ubicación en el montaje electroencefalográfico. Para estos algoritmos y modelos se han planteado implementaciones en entornos de programación de carácter libre, con la finalidad de facilitar la transferencia de estas tecnologías a plataformas digitales de información de salud mental. El entorno de programación elegido es R, en el cual existen algunas librerías para análisis de señales y minería de datos que fueron utilizadas. Sin embargo, algunos algoritmos fueron programados durante este trabajo, y en todos los casos se realizó una adaptación para ajustar los programas a las características de formato electrónico que producen los dispositivos electroencefalográficos de uso comercial. Es destacable la elaboración de interfaces gráficas dinámicas que permitirán el procesamiento de señales electroencefalográficas para dispositivos como interfaces cerebro-computadora, entre otros.

Abstract

This work is related to the participation on a research project on Mental Health. The research project is intended to develop and apply computational models to quantitative analysis of electroencephalographic data. This proposal is build on with algorithms of signal analysis and data mining, considering electroencephalographic signals as a multivariate time series. On the other hand, it is presented a bayesian network model to characterize the spatial structure if the electroencephalographic signals, intending to analyze the correlations between differents channels, according to its spatial location. This algorithms and models were programed on free software enviroment in order to facilitate its transfer to digital plataforms of mental health. The free software is R, in which there are some libraries, intended to signal analysis and data mining, used in this work. However some algorithms were programed, and in all cases it was necessary to adapt these programs to the electronic format produced by electroencephalologic device, commercial in nature. it is remarkable the development of grap-hic dynamical interfaces, which may be able the processing of electroencephalographic for brain-computer-interfaces, among others.

Índice general

Autorización de impresión	II
Dedicatoria	III
Agradecimientos	IV
Resumen	V
Abstract	VI
Artículos publicados	X
Índice de figuras	XI
Índice de tablas	XIII
1. Generalidades	14
1.1. Introducción	14
1.2. Planteamiento del problema	16
1.2.1. Justificación	17
1.3. Objetivos	18
1.3.1. Objetivo general	18
1.3.2. Objetivos específicos	18
1.4. Hipótesis	19
1.5. Propuesta de solución	19
1.6. Alcances y limitaciones	19
1.6.1. Alcances	19
1.6.2. Limitaciones	20
1.7. Estructura de la tesis	20
2. Marco teórico	21
2.1. Señales biomédicas	21
2.2. Electroencefalograma	22
2.2.1. Definición	22

2.2.2.	Desarrollo histórico	23
2.2.3.	Características de la señal	24
2.3.	Análisis de señales	27
2.3.1.	Ruidos y artefactos	27
2.3.2.	Transformada discreta de Fourier	28
2.3.3.	Filtrado de Tikhonov	28
2.4.	Reconocimiento de patrones	29
2.4.1.	Análisis de Componentes Principales (PCA)	30
2.5.	Minería de datos	32
2.5.1.	Análisis Discriminante	33
2.5.2.	Regresión logística	33
2.5.3.	Red Bayesiana	35
3.	Estado del arte	36
3.1.	Reconocimiento de patrones	36
3.2.	Minería de datos	38
4.	Presentación de herramientas computacionales	41
4.1.	Desarrollo de la Aplicación para Filtrado por regularización de Tikhonov	41
4.1.1.	Características de los datos	41
4.1.2.	Algoritmo de Tikhonov	42
4.2.	Desarrollo de la Aplicación para el Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG	45
4.2.1.	Características de los datos	46
4.2.2.	Análisis de Componentes Principales	46
4.2.3.	Regresión Logística	46
4.3.	Desarrollo de la aplicación para el Análisis Longitudinal y Topográfico de la Conectividad Funcional Cerebral. Detección de patrones espaciales y Temporales de Conectividad Funcional Cerebral.	48
4.3.1.	Características de los datos	48
4.3.2.	Análisis de Componentes Principales	49
4.3.3.	Regresión Logística	49
4.4.	Desarrollo de la Aplicación para el análisis de correlaciones basado en modelos gráficos gaussianos.	51
4.4.1.	Características de los datos	51
4.4.2.	Análisis de Componentes Principales	51
4.4.3.	Red Bayesiana	51
4.5.	Validación	54
4.5.1.	Validación del Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG	54

4.5.2.	Validación del Análisis Longitudinal y Topográfico de la Conectividad Funcional Cerebral. Detección de patrones espaciales y Temporales de Conectividad Funcional Cerebral.	57
4.5.3.	Validación del análisis de correlaciones basado en modelos graficos gaussianos.	65
5.	Análisis de resultados	69
5.1.	Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG	69
5.2.	Análisis Longitudinal y Topográfico de la Conectividad Funcional Cerebral. Detección de patrones espaciales y Temporales de Conectividad Funcional Cerebral	69
5.3.	Análisis de correlaciones basado en modelos gráficos gaussianos.	69
5.4.	Comparación con el Estado del Arte.	71
6.	Conclusiones y Trabajo a futuro	73
6.1.	Conclusiones	73
6.2.	Trabajo a futuro	73
	Referencias	74
	Anexos	76
6.3.	Extracción de datos .edf	76
6.3.1.	Gráfica de las señales	76
6.3.2.	Ejemplo para el canal FP1-F7	76
6.4.	Extracción de Bandas	77
6.4.1.	Gráficas de las bandas	79
6.5.	Modelos gráficos	81
6.5.1.	Red de Bayes	81

Artículos publicados

Los artículos publicados durante el proceso de realización de este trabajo de investigación son los siguientes:

- 1.- Yoselyn Nohemí Ortega Gijón, Alejandro del Rey Torres Rodríguez, María Luisa Córdoba Tlaxcalteco, Carlos Arturo Hernández Gracidas. (2018, Junio). Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG. Impulso Tecnológico CIMVER 2017, 1, pp. 313-314.
- 2.- Yoselyn Nohemí Ortega Gijón, Alejandro del Rey Torres Rodríguez, María Luisa Córdoba Tlaxcalteco, Carlos Arturo Hernández Gracidas. (2017, Octubre). Análisis discriminante longitudinal de la Conectividad Funcional Cerebral usando componentes principales y regresión logística.. COLOQUIO DE INVESTIGACIÓN MULTIDISCIPLINARIA JOURNAL CIM 2017, 5, pp. 592-598.
- 3.- María Luisa Córdoba Tlaxcalteco, Alejandro del Rey Torres Rodríguez, Yoselyn Nohemí Ortega Gijón, Carlos Arturo Hernández Gracidas. (2018, Febrero 22). Red Bayesiana para la distribución espacial de la Conectividad Funcional Cerebral, con EEG. Ingenieros 2017, 1, pp. 53-58.

Índice de figuras

2.1. Señales biomédicas.	22
2.2. Montaje del Sistema Internacional 10-20.	26
2.3. Curva Logística.	34
2.4. Detalle de la RB, nodo X_i con sus padres X_j y X_k	35
4.1. Diagrama de flujo para el desarrollo de la aplicación para el filtrado por regularización de Tikhonov.	42
4.2. Parte real de los valores propios de la matriz A_α para diferentes valores de α	43
4.3. Gráfica que muestra el efecto de regularización en el filtrado de Tikhonov	44
4.4. Diagrama de flujo para el desarrollo de la aplicación para el Análisis Longitudinal de Conglomerados para CFC	45
4.5. Diagrama de flujo para el desarrollo de la aplicación para el Análisis Longitudinal y Topográfico de la CFC.	48
4.6. Representación de la RB para los canales EEG.	52
4.7. Gráfica de logitos de fr de 2 pacientes (mujer y hombre), con el logaritmo del número de toma en el eje horizontal.	54
4.8. Agrupamiento de los individuos de acuerdo a la regresión logística (Coeficientes)	55
4.9. Gráfica de la dinámica del grupo 1	56
4.10. Gráfica de la dinámica del grupo 2	56
4.11. Gráfica topográfica de la dinámica de fr en Mujer	58
4.12. Gráfica topográfica de la dinámica de fr en Hombre	58
4.13. Gráfica de distribución espacio temporal en una Mujer por canal.	59
4.14. Gráfica de distribución espacio temporal en un Hombre por canal.	59
4.15. Montaje del Sistema Internacional 10-20.	59
4.16. Gráfica topográfica de la dinámica de fr en Mujer.	60
4.17. Gráfica topográfica de la dinámica de fr en Hombre.	60
4.18. Gráfica de distribución espacio temporal en una Mujer por canal.	61
4.19. Gráfica de distribución espacio temporal en un Hombre por canal.	61
4.20. Montaje del Sistema Internacional 10-20.	61
4.21. Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en una Mujer	62

4.22. Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en un Hombre	62
4.23. Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en Mujeres	63
4.24. Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en Hombres	63
4.25. Gráfica topográfica de la regresión logística (Coeficientes) en Mujeres.	64
4.26. Gráfica topográfica de la regresión logística (Coeficientes) en Hombres.	64
4.27. Distribución Normal por canal	66
4.28. Distribución de residuales de la red bayesiana.	66
4.29. Estimaciones de coeficientes de correlación	67
4.30. Limites Mujeres	68
4.31. Limites Hombres	68
6.1. Gráfica del canal FP1-F7	77
6.2. Bandas del canal FP1-F7	78
6.3. Banda Delta del canal FP1-F7	79
6.4. Banda Theta del canal FP1-F7	79
6.5. Banda Alpha del canal FP1-F7	80
6.6. Banda Beta del canal FP1-F7	80
6.7. Banda Gamma del canal FP1-F7	81

Índice de tablas

1.1. AVISA perdidos en México por algunos padecimientos neurológicos.	17
5.1. Comparativa del Estado del Arte de algoritmos.	71
5.2. Comparativa del Estado del Arte de algoritmos.	72

Capítulo 1

Generalidades

1.1. Introducción

El proyecto de investigación denominado “SALUD-1-2016-232345” es un proyecto resultado de la colaboración entre el Instituto Tecnológico de Ciudad Victoria (ITCV), el Instituto Tecnológico Superior de Misantla (ITSM), y otras instituciones. Comenzó en noviembre 2016 con el financiamiento del “Fondo sectorial de investigación en Salud y Seguridad Social”, en el que concurren el CONACYT, La Secretaría de Salud (SS), el Instituto Mexicano del Seguro Social (IMSS), el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE). Su título es “Análisis de trazos EEG para caracterizar la densidad del espectro de potencia y de diversos rangos de frecuencias en seres humanos separados por quinquenio de la vida”, y su objetivo general es contribuir a definir la normalidad en el sistema nervioso, entender las etapas de su maduración, y finalmente, si la forma de los trazos electroencefalográficos varía conforme a la edad de las personas.

Entre sus metas se encuentran el desarrollo de repositorios de datos neurológicos, y la puesta en marcha de servidores de aplicaciones para análisis individual y colectivo, de datos biofísicos. La proponente de este trabajo ha colaborado en el proyecto de investigación, diseñando y elaborando las aplicaciones informáticas para el análisis cuantitativo de los datos neurológicos obtenidos en el proyecto. Dichas aplicaciones tienen la finalidad de contribuir en la difusión de las técnicas de Análisis de señales, Reconocimiento de patrones, y Minería de datos que son utilizados en la Neurología cuantitativa. En el presente trabajo se encuentran los resultados obtenidos en este sentido.

Por siglos, el ser humano ha intentado descifrar los enigmas del funcionamiento de su cerebro. Esta curiosidad nos ha llevado a avances tales como ser capaces de interactuar con la computadora con el simple hecho de pensarlo, con las llamadas interfaces cerebro-computadora. No obstante, hace solo unos siglos, muchas enfermedades que hoy están mejor documentadas, eran diagnosticadas, erróneamente, como posesiones diabólicas o se catalogaba a los enfermos como lunáticos, pues se creía que su mal estaba relacionado con las fases de la luna. Incluso hoy en día, es muy frecuente decir que una persona está de “mal humor”, cuando pudiera estar mostrando síntomas de depresión. Esto da una idea de la necesidad

de desarrollar técnicas actualizadas para el diagnóstico y tratamiento de padecimientos del sistema nervioso. Muchos de los estudios para detectar padecimientos neurológicos requieren de técnicas invasivas con un alto nivel de riesgo y efectos secundarios. Ejemplos de ellos son punciones lumbares, radiografías y angiografías. Por otra parte, los estudios en electroencefalografía tienen la ventaja de ser técnicas no invasivas, es decir, que no requieren de punción, sedación o la ingestA de medicamentos, basándose en el análisis de las propiedades eléctricas de las células biológicas. En particular, la electroencefalografía se presenta como una herramienta que permite al médico especialista analizar y diagnosticar padecimientos neurológicos midiendo la actividad eléctrica del encéfalo; esta técnica ha sido objeto de un gran avance tecnológico, pasando desde aquellos equipos extremadamente sensibles y delicados donde el operador requería de un amplio conocimiento técnico para su calibración y posterior realización de los estudios, hasta los equipos digitales de la actualidad, los cuales permiten la aplicación de algoritmos de eliminación de artefactos o ruidos por movimiento, ambiente, etc., así como los que incluyen herramientas automáticas de detección de actividad neurológica anormal como es el caso de la actividad paroxística.(Freeman y Quiroga, 2012)

En este contexto de desarrollo de la moderna electroencefalografía, más cercana a un mayor número de usuarios, el proyecto de investigación pretende hallar evidencias de que al agrupar una muestra de seres humanos en quinquenios de la vida y analizar sus trazos EEG con respecto a sus espectros de potencia y los diversos rangos de frecuencias, mediante la aplicación de técnicas computacionales, será posible caracterizar la normalidad, las etapas de maduración neurológica y la no estacionalidad de estos grupos. Entre sus objetivos específicos se encuentra elaborar las herramientas computacionales de análisis de la actividad eléctrica cerebral espontánea y provocada de cada uno de los segmentos de estudio establecidos en el protocolo de registro. El trabajo de Tesis toma el propósito de elaborar dichas herramientas computacionales. El aspecto neurológico que se aborda como objeto de estudio es la Conectividad Funcional Cerebral. Esto se hace a través de tres enfoques. En el primero, tomando una medida global de conectividad. El segundo, especializando el primer análisis a las diferentes partes del encéfalo. Y por último, haciendo uso de un modelo gráfico probabilista que tiene la aportación de brindar análisis de significación estadística para la conectividad. Es importante aclarar que en este trabajo se aborda el diseño y el desarrollo inicial de las herramientas computacionales de cada uno de los enfoques citados anteriormente. Este desarrollo inicial permite la validación de la eficacia de las aplicaciones para el análisis de conectividad. Con el propósito de facilitar su transferencia a un sistema distribuido de uso público, se ha enfatizado el desarrollo de aplicaciones libres (no comerciales) en el entorno de programación “R”. Algunos algoritmos son elaboraciones propias de esta Tesis, y los gráficos han sido diseñados ad hoc para su posterior uso por especialistas neurólogos.

Este trabajo también incluye el desarrollo de una herramienta de filtrado multicanal, basado en el algoritmo de regularización de Tikhonov. La organización de la tesis es como sigue. En el primer capítulo se hacen los planteamientos generales del problema, el enfoque metodológico para abordarlo, y se concluye estableciendo los alcances y las limitaciones. En el segundo capítulo se presentan, en forma general, el marco teórico donde se enuncian los elementos computacionales y matemáticos de la Electroencefalografía cuantitativa, que van

a ser los componentes de las herramientas computacionales desarrolladas. El capítulo tres es una discusión del estado del arte de las aplicaciones de reconocimientos de patrones y minería de datos aplicadas a Electroencefalografía. El capítulo cuatro presenta los desarrollos de las aplicaciones computacionales. El capítulo cinco contiene de los resultados obtenidos por las herramientas desarrolladas, utilizadas en un conjunto de datos EEG. El siguiente capítulo, que es el último, tiene una discusión de los trabajos a futuro que pueden desarrollarse a partir de esta tesis.

1.2. Planteamiento del problema

La Electroencefalografía (EEG) es el estudio del comportamiento eléctrico de la corteza cerebral, a través del registro de las diferencias de potencial eléctrico en diferentes zonas de la cabeza humana. La EEG es una herramienta de diagnóstico para la Neurología, y típicamente se realiza por un especialista entrenado en la interpretación visual de gráficos que registran la actividad eléctrica en ciertos periodos de tiempo. La EEG cuantitativa o mapeo cerebral (desarrollada a partir de los años 70's) es el conjunto de técnicas computacionales que sirven para procesar e interpretar las señales EEG, a través de algoritmos de reconocimiento de patrones y el análisis de señales. El análisis cuantitativo del EEG requiere del desarrollo tecnológico de dispositivos electrónicos que registren, almacenen, y transmitan los registros biofísicos cerebrales. Esto lleva paralelamente el desarrollo de software para el análisis de las señales. De manera reciente también se ha gestado software para realizar algunas comparaciones entre señales de diferentes sujetos. Estas aplicaciones, hasta ahora, están vinculadas al uso de electroencefalografos comerciales, de tal forma que es imposible utilizarlas sin haber adquirido el equipo comercial. Aunado a esto, el análisis tradicional, el análisis médico, es visual y no cuantitativo. Uno de los objetivos del proyecto "SALUD-1-272345" es construir repositorios de señales y aplicaciones con código abierto para la mayor difusión y uso de Electroencefalografía cuantitativa. Esto requiere el desarrollo de programas informáticos para administrar bases de datos, y realizar cómputo cuantitativo. El desarrollo de las técnicas de análisis de señales, reconocimiento de patrones, y minería de datos para EEG enfrenta, sin embargo, dos obstáculos. Primero el carácter no estacionario de las señales con origen biológico. Y también, el aspecto multidimensional asociado al registro de señales biofísicas en múltiples canales sobre el mismo sujeto. La mayoría de los trabajos de análisis en el aspecto neurológico en electroencefalografía están orientados a personas enfermas. Otro aspecto de la problemática consiste en que muchos trabajos de análisis de bioseñales se basan en la noción de señal estacionaria, lo cual no es congruente con la intención de detectar cambios en el tiempo. Por lo tanto, en este trabajo se aborda el desarrollo de aplicaciones computacionales que, por un lado, no dependan de la hipótesis de invarianza en el tiempo de las características estadísticas de la señal (no estacionariedad). Y por otro, las técnicas a utilizar contemplan el carácter multidimensional de señales correlacionadas entre sí.

1.2.1. Justificación

De acuerdo a INEGI, la esperanza de vida al nacimiento en México subió de 55.9 años en 1950 a 75.3 años en 2010. Según datos del Banco Mundial, la esperanza de vida en México en 2014 era de 76.72 años. El aumento de la esperanza de vida tiene, como contraste, la tasa de enfermedades neurológicas se ha mantenido estable en los últimos años. Como un ejemplo, en 2006 la epilepsia atacaba a 50 millones de personas en el mundo, lo cual da una muestra de lo relevante que son estos padecimientos incluso hoy en día (Organization and others, 2006). En la siguiente tabla se muestran los datos sobre la pérdida estimada de años de vida saludable en México, por anormalidades neurológicas.

Padecimiento	Hombres	Mujeres
Enfermedades cerebrovasculares	2.4	2.8
Encefalopatía neonatal	1.4	1.4
Epilepsia	1.4	1.6

Tabla 1.1: AVISA perdidos en México por algunos padecimientos neurológicos. Principales padecimientos neurologicos en México, y su impacto en años perdidos de vida saludable.

Una estadística que es muy representativa de la relevancia que tienen los padecimientos neurológicos en la población en general son los años de vida saludables (AVISA), siendo de particular interés para este tipo de padecimientos los AVISA perdidos; es decir, los años de vida en estado de salud adecuada perdidos a causa de una muerte prematura o de una enfermedad discapacitante. Las estadísticas respecto a padecimientos neurológicos muestran que estos no tienen índices de mortalidad altos, sin embargo, se observa que un paciente que llega a padecer este tipo de enfermedades pasará una gran cantidad de años viviendo con un cierto nivel de discapacidad, lo que se traduce en un gran número de AVISA perdidos; la Tabla 1 ejemplifica esto para algunos padecimientos neurológicos considerados.

Dado que, en términos generales, la esperanza de vida en nuestro país ha aumentado significativamente, y que los índices de padecimientos neuronales prácticamente se mantienen iguales, tenemos como resultado que cada vez una mayor cantidad de gente se encuentre incapacitada, sufriendo tanto de estos padecimientos como de sus secuelas. Esto, además de la gran cantidad de AVISA perdidos que genera, representa un fuerte desembolso para las familias que deben mantenerlos en caso de no contar con un seguro médico que los ampare. Incluso, si la persona tiene la suerte de contar con un seguro médico, el gasto que esto representa para el gobierno al tener que pagar por los tratamientos, medicamentos y equipos médicos para la atención de los pacientes es tanto o más relevante.

Un primer paso para disminuir estos costos radica en modelar la normalidad neurológica por cada quinquenio de la vida estudiado y, a futuro, emplear esta información para detectar los problemas neurológicos mediante el análisis de los espectros de potencia y los rangos de frecuencia que ayuden a detectar patrones para, así, atenderlos de forma oportuna. Modelar la normalidad neurológica nos permitiría determinar que un paciente requiere de atención

y estudios especializados, dado que sus condiciones neurológicas no se encuentran dentro de los parámetros esperados para su etapa de la vida; esto ayudaría a que, incluso sin haber determinado cuál es la condición que lo aqueja pueda comenzar a atenderse, con la posibilidad de atenderlo oportunamente y así brindarle una mejor calidad de vida y, adicionalmente, que se evite un fuerte desembolso en la atención de los mismos una vez que ya no puedan ser revertidos o, mínimamente, atenuados, lo cual se espera permita disminuir los AVISA perdidos debido a padecimientos neurológicos. La presente propuesta busca dar los primeros pasos hacia esta caracterización.

Estos estudios no suelen ser utilizados para dar un seguimiento de rutina de su salud, pues, en su mayoría, se acostumbra que los estudios EEG se empleen únicamente para corroborar o descartar padecimientos neuronales, siendo que la persona ya ha presentado síntomas clínicos de un padecimiento de este tipo. Con un conjunto de datos de EEG de este tipo sería algo nuevo en el estudio de los padecimientos neuronales. La EEG cuantitativa permite encontrar anomalías cerebrales, que no se pueden detectar por inspección visual de gráficos.

Algunas de estas anomalías son: Depresión, ansiedad, trastornos de la atención, autismo, demencias, enfermedades degenerativas, y otras. El éxito del uso de EEG cuantitativa depende de la capacidad tecnológica para realizar el registro y el almacenamiento de los datos que constituyen las señales, así como la existencia de base de datos donde estos registros puedan ser comparados y clasificados. La neuropsicología, por ejemplo, utiliza la EEG cuantitativa para realizar evaluación del desarrollo cognoscitivo de niños y niñas, a fin de prevenir o diagnosticar desviaciones del proceso normal de maduración infantil.

En este contexto, las herramientas de Electroencefalografía cuantitativa y la minería de datos asociada, representan el salto del análisis individual, al análisis neurológico colectivo, permitiendo descubrir características de diferenciación o asociación, así como la construcción de conglomerados de individuos semejantes o diferentes, desde el punto de vista electroencefalográfico.

1.3. Objetivos

1.3.1. Objetivo general

Elaborar herramientas computacionales, que son implementaciones de algoritmos de minería de datos, análisis de señales, y modelos cuantitativos sobre actividad neurológica derivada de electroencefalografía. ¹

1.3.2. Objetivos específicos

- Diseñar y/o seleccionar modelos cuantitativos para la detección de patrones en Electroencefalografía multicanal.

¹En el presente documento, por indicación del director de tesis, los términos diseño, aplicación y desarrollo de técnicas o herramientas computacionales, se refieren a la implementación de algoritmos tradicionales.

- Seleccionar algoritmos de cómputo científico para la estimación o ajuste de parámetros de modelos para Electroencefalografía cuantitativa multicanal.
- Elaborar e implementar las herramientas computacionales en entornos de programación libre.
- Validar las herramientas computacionales mediante su aplicación a conjuntos de datos de electroencefalografía.

1.4. Hipótesis

Es factible elaborar herramientas computacionales de análisis y visualización de patrones de señales de electroencefalografía multicanal, mediante minería de datos y análisis de señales, relevantes para especialistas en neurología.

1.5. Propuesta de solución

Se plantea la elaboración de herramientas computacionales, que constan de la implementación de algoritmos cuantitativos y salidas gráficas de información, en los siguientes casos:

- Filtrado multicanal basado en la regularización de Tikhonov. Para la eliminación de ruidos y artefactos, así como corregir el mal condicionamiento de la matriz de la transformada de Fourier multicanal discreta.
- Análisis de conglomerados longitudinal, que sirve para identificar grupos de individuos de acuerdo con el patrón de cambio en el tiempo de las señales.
- Análisis discriminante espacial y temporal, que es similar al anterior con la inclusión de información de ubicación de las señales.
- Herramienta para la visualización de la estructura de correlaciones entre canales de señales, basada en modelos gráficos computacionales. Sirve para identificar y cuantificar la conectividad entre diferentes partes funcionales del cerebro.

1.6. Alcances y limitaciones

1.6.1. Alcances

-Las herramientas computacionales elaboradas en este trabajo son utilizadas en datos electroencefalográficos en formatos estándar, como CSV, EDF y otros. Esto libera del uso de formatos comerciales, a menudo encriptados.

-Las herramientas fueron diseñadas usando la segmentación de las señales en el tiempo, para contemplar el carácter no estacionario de la señal biofísica, en contraposición a la hipótesis habitual de que la señal es estacionaria.

-La programación de las herramientas en entorno libre hace factible la asociación o vinculación de las mismas con otras aplicaciones informáticas, como administradores de base de datos, o dispositivos electrónicos.

1.6.2. Limitaciones

-La transferencia de esta tecnología a los usuarios finales requiere, entre otras cosas, de la evaluación de especialistas en salud mental.

-Queda pendiente la implementación de las herramientas en una plataforma tipo web.

-El uso de estas herramientas en tiempo real, por ejemplo Interfaces BCI, requiere más evaluación de su eficacia en términos de costo computacional.

1.7. Estructura de la tesis

La primera parte contiene el planteamiento general de la tesis. La segunda parte es el marco teórico, donde vienen descritas y referidas todos los elementos de las herramientas computacionales. La tercera parte es una discusión del estado del arte. A continuación se describen a detalle los elementos de las herramientas computacionales. El capítulo 5 contiene el uso de estas herramientas en bases de datos, y los resultados obtenidos en esta actividad. Para concluir, el capítulo 6 contiene las conclusiones del trabajo. En el marco teórico se muestran más a detalle los métodos y materiales que se van a ocupar en el presente trabajo, explicando su funcionamiento y el modo de aplicación para la resolución. En este capítulo se presenta el estado del arte, donde se visualizan los artículos y trabajos relacionados en los que se basa la presente investigación.

Capítulo 2

Marco teórico

2.1. Señales biomédicas

Una señal es una representación de información de un solo valor en función de una variable independiente (por ejemplo, tiempo). El tipo específico de información que se representa puede tener valores reales o complejos.

Las fuentes de señales biomédicas son infinitas en variedad e identificaremos solo algunas como ejemplos de los diferentes tipos de estas señales. Una gran clase de señales biomédicas comprende aquellas que son de naturaleza eléctrica. El electrocardiograma (ECG), que se puede registrar colocando electrodos en el cofre, es una de esas señales. Registra el potencial eléctrico en el electrodo (o la diferencia de potencial entre dos electrodos) inducido por la presencia de actividad eléctrica variable en el tiempo en el músculo cardíaco asociada con la generación y propagación de potenciales de acción. Cada latido del corazón produce una secuencia de ondas eléctricas (ondas P, Q, R, S y T) y al examinar la forma de las ondas del ECG un médico puede obtener información considerable sobre si las contracciones de el corazón está ocurriendo normalmente.(Bruce, 2001).

Otro ejemplo de señal biomédica es el electromiograma (EMG). Las señales EMG se registran colocando electrodos en, sobre o cerca de un músculo y amplificando el potencial eléctrico (o la diferencia de potencial entre dos electrodos) que resulta de la generación y propagación de potenciales de acción a lo largo de las fibras musculares. Un EMG de unidades múltiples (MUEMG) registra los potenciales de muchas fibras musculares a la vez. La presencia de actividad MUEMG indica que un músculo se contrae activamente y muchos investigadores han intentado desarrollar técnicas para filtrar la señal MUEMG para producir otra señal que sea proporcional a la fuerza que el músculo está generando.

La Figura 2.1, muestra ejemplos de las señales biomédicas. (a) Velocidad instantánea media del flujo sanguíneo en la arteria cerebral. Marcador de tiempo = 1 s; (b) Electromiograma (EMG) de dos cables firmemente debajo de la lengua por una boquilla. Marcador de tiempo = 0.2 s; (c) Se obtuvo el ángulo de rotación de la rodilla desde un sensor de ángulo (Datos cortesía de James Abbas); (d) Un electrocardiograma (ECG) grabación (datos cortesía de Abhijit Patwardhan); (e) Tasa de carga instantánea (beat/min, recíproco del intervalo latido

a latido) durante 100 latidos cardíacos consecutivos.

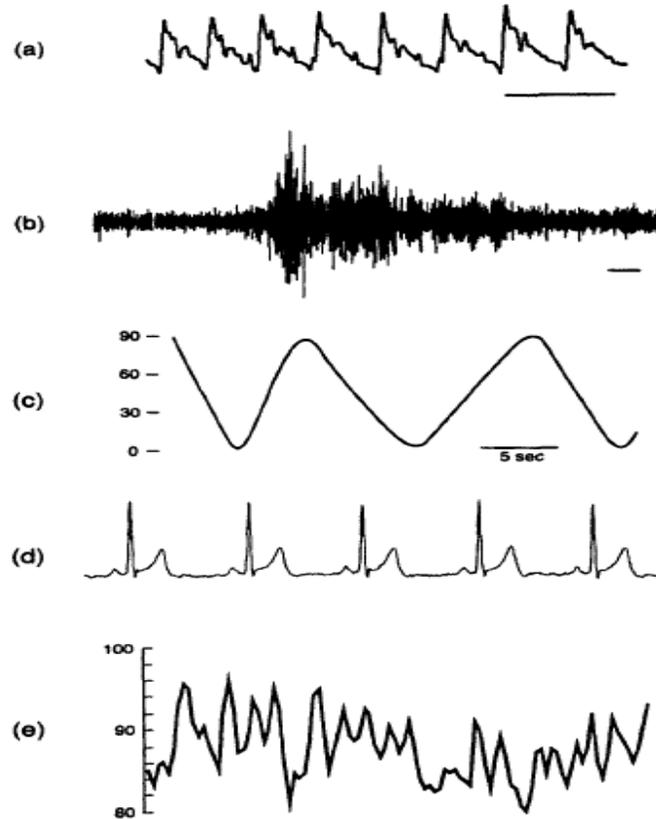


Figura 2.1: Señales biomédicas.

(a) Velocidad del flujo sanguíneo, (b) Electromiograma, (c) Ángulo de rotación de la rodilla, (d) Electrocardiograma, (e) Latidos cardíacos consecutivos. Fuente:(Bruce, 2001)

Las señales biomédicas pueden estar corrompidas por varios tipos de ruidos. Ejemplos típicos son: Interferencia de línea de potencia, ruido por contacto de electrodo, artefactos, contracción muscular, ruido de instrumentación generado por dispositivos electrónicos.

2.2. Electroencefalograma

2.2.1. Definición

El EEG mide la actividad eléctrica del cerebro en diferentes sitios de la cabeza, generalmente utilizando electrodos colocados en el cuero cabelludo. Sus principales ventajas sobre otras técnicas de registro son su alta resolución temporal y el hecho de que puede grabarse de forma no invasiva (es decir, sin la necesidad de una cirugía). Debido a su costo relativamente

bajo, las grabaciones de EEG son ampliamente utilizadas en entornos clínicos y laboratorios de investigación. Esto hace que el EEG sea una herramienta muy accesible y útil, que es particularmente interesante para el análisis de procesos cerebrales de alto nivel que surgen de la actividad grupal de poblaciones de células grandes. Dichos procesos pueden estar bien localizados en el tiempo o pueden correlacionarse con patrones variables en el tiempo, como las oscilaciones cerebrales, que están más allá de la resolución temporal de las técnicas de imagen como imágenes de resonancia magnética funcional (fMRI). La advertencia de los EEG no invasivos es el hecho de que reflejan la actividad promedio de un gran número de fuentes lejos de los sitios de registro y, por lo tanto, no tienen una resolución espacial óptima.

Aunque la forma de registrar las señales EEG no cambió tanto como en el caso de las grabaciones microscópicas y macroscópicas (aunque en capítulos posteriores, describiremos las pautas básicas para los diseños de electrodos que mejoran el análisis espacial de los EEG), ha habido cambios significativos. avances en la metodología para analizar datos de EEG. De hecho, las grabaciones de EEG han sido un desafío final para la mayoría de los métodos de procesamiento de señales debido a su alta complejidad, baja relación señal / ruido, no linealidad y no estacionariedad. Como describiremos en este libro, el desarrollo e implementación de nuevos algoritmos que están específicamente diseñados para señales complejas como los EEG nos permitirá obtener mucha más información de la que ha sido accesible con los métodos anteriores y la inspección visual convencional del grabaciones, como lo hacen los electroencefalographers entrenados. Estos métodos abren una nueva ventana al estudio de los procesos cognitivos de alto nivel en humanos con métodos no invasivos técnicas y sin grandes gastos.(Freeman y Quiroga, 2012).

2.2.2. Desarrollo histórico

La historia de las grabaciones de EEG humanas se remonta a Hans Berger (1873-1941), profesor de psiquiatría en la Universidad de Jena, Alemania. Siguiendo el trabajo de Richard Caton (1842-1926), un cirujano de Liverpool que registró con éxito la actividad eléctrica de los hemisferios cerebrales expuestos de monos y conejos en 1875, Hans Berger fue el primero en registrar la actividad eléctrica del cuero cabelludo humano en 1924. Después de 5 años recolectando datos y reexaminando sus resultados, finalmente publicó en 1929 "Über das Elektroencephalogramm des Menschen". En este trabajo seminal, Berger ya informó la presencia de oscilaciones cerebrales de aproximadamente 10 ciclos por segundo, lo que llamó ondas alfa , visto con el sujeto en un estado relajado con los ojos cerrados. Al abrir los ojos, estas ondas desaparecieron (bloqueo alfa) y se observaron oscilaciones de frecuencias más altas (ondas beta). También se observó un tipo similar de oscilaciones beta con los ojos cerrados cuando los sujetos realizaban tareas aritméticas mentales.

La importancia del trabajo de Berger no fue reconocida hasta 1934 cuando Lord Edgar Adrian (1889-1977), en Cambridge, confirmó sus resultados. A partir de ese momento, la técnica de EEG provocó una revolución en el modo de estudiar los procesos cerebrales normales y patológicos. Para mencionar algunos de los principales logros, en los años 30 Gray Walter, primero en Londres y luego en el Burden Neurological Institute en Bristol, informa-

ron oscilaciones lentas (ondas delta) sobre los tumores cerebrales hemisféricos e introdujeron el concepto de topografía EEG para localizar el cerebro lesiones Inmediatamente después, la investigación de EEG se extendió a los Estados Unidos. En Harvard, Hallowell Davis, Frederic Gibbs, Erna Gibbs y William Lennox comenzaron a estudiar los patrones paroxísticos de EEG relacionados con la epilepsia. Estos patrones anormales, como los picos o las ondas de pico, se siguen utilizando para ayudar al diagnóstico de la epilepsia. La década de 1940 vio el comienzo de los estudios del sueño. Al final de esta década, se realizaron las primeras grabaciones intracraneales humanas. En nuestros días, este tipo de grabaciones se utilizan principalmente en pacientes que son candidatos a la cirugía de epilepsia con el fin de determinar el origen de las convulsiones. En la década de 1950, Wilder Penfield y Herbert Jasper, en la Universidad McGill de Montreal, utilizaron estimulación eléctrica, con cirugías cerebrales abiertas bajo anestesia local, para localizar áreas involucradas en diferentes procesos cerebrales. En la misma década, George Dawson, en Londres, introdujo un gran avance en el campo, y desarrolló una técnica de suma para visualizar las respuestas medias de EEG a los estímulos. Más adelante, un gran avance fue la introducción de computadoras en el análisis de señales EEG, especialmente con el uso de la transformada rápida de Fourier desarrollada por Cooley y Tukey (1965).

Una desaceleración significativa en la investigación EEG resultó como una consecuencia de la introducción de otras metodologías para medir la actividad cerebral, tales como grabaciones de neuronas individuales en la década de 1950 y especialmente la aparición de técnicas de imagen y magnetoencefalografía en la década de 1980. En nuestros días, las grabaciones de EEG generalmente se usan para diagnósticos clínicos, como lesiones en la cabeza, tumores cerebrales y epilepsia. Los neurocientíficos también estudian diferentes tipos de actividad EEG durante el comportamiento controlado en sujetos humanos y animales.

Hay tres ventajas principales de EEG sobre otros métodos: (1) no es invasivo y, por lo tanto, es posible hacer experimentos con sujetos humanos normales; (2) tiene una resolución de tiempo muy alta, del orden de milisegundos - que permite seguir la dinámica temporal de los procesos cerebrales; y (3) es relativamente barato. Además de eso, ha habido avances significativos en el desarrollo de métodos para estudiar señales complejas, y la mayoría de ellos solo están comenzando a usarse en grabaciones de EEG. Como veremos en capítulos posteriores, algunos de estos métodos ofrecen una nueva perspectiva para estudiar los EEG y los procesos cerebrales en general.(Freeman y Quiroga, 2012).

2.2.3. Características de la señal

Toma de datos

Las grabaciones de EEG de cuero cabelludo se realizan usando electrodos de alta conductividad (es decir, con una impedancia nanométrica inferior a 5 kW) colocados en la parte superior de la cabeza. Los electrodos se distribuyen en ubicaciones específicas.

Los electrodos se colocan en la cabeza con gel para aumentar la conductividad con el cráneo. Los electrodos solían colocarse manualmente de uno en uno, pero ahora, por lo general, ya vienen colocados en una tapa, que puede ajustarse fácilmente alrededor de la

cabeza del sujeto. Esto es más práctico y consume menos tiempo, teniendo en cuenta las posibilidades de los equipos modernos,

El EEG se puede registrar con referencia a un electrodo pasivo común - grabaciones monopolares (referenciales) - o se puede registrar diferencialmente entre pares de electrodos contiguos - grabaciones bipolares. En el caso posterior, hay varias formas de elegir los pares de electrodos a estas combinaciones se les llama montajes. Los montajes son conocidos como el conjunto de pares de electrodos que se encuentran conectados a cada canal de amplificación de señal. Un canal se define como la diferencia de potencia de dos electrodos que se conectan.

según los montajes diseñados para visualizar la propagación de la actividad en diferentes direcciones. Algunos montajes particulares pueden ser muy útiles para visualizar las fuentes de diferentes patrones de EEG.

El sistema internacional de posicionamiento 10-20 (Figura 2.2), es un estándar internacional definido por la FISE (Federación Internacional de Sociedades de Electroencefalografía) que describe en qué ubicación del cráneo deben ser colocado los electrodos, para la medición mediante un Encefalograma. Dicho sistema está basado en la relación entre la localización del electrodo y el área de la corteza cerebral ubicada bajo el, los números "10z "20" se refieren a que la distancia entre los electrodos vecinos puede ser de 10 % o 20 % de la distancia total entre la zona frontal y posterior, o de derecha a izquierda del cráneo, cada zona tiene asignada una letra para identificar el lóbulo y un número para identificar en qué hemisferio se encuentra, si el número es par, las posiciones de los electrodos pertenecen al hemisferio derecho del cerebro, si el número es impar pertenecen al hemisferio izquierdo. Las letras mayúsculas denotan las diferentes ubicaciones de electrodos, F para frontal, C para central, P para parietal, T para temporal y O para occipital.(Freeman y Quiroga, 2012).

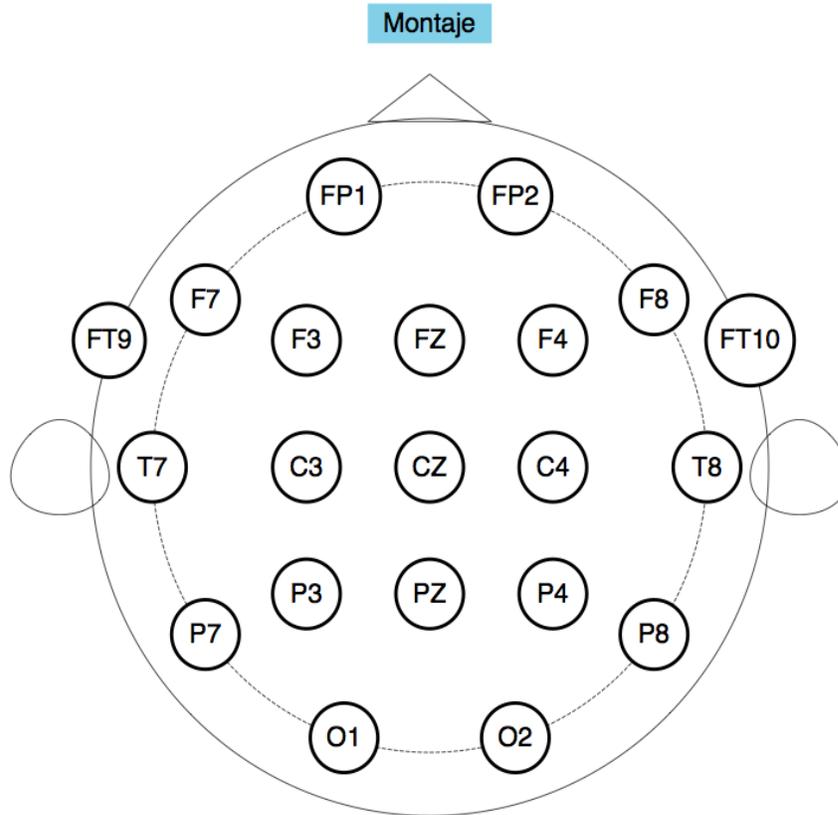


Figura 2.2: Montaje del Sistema Internacional 10-20.
Fuente:(Nomenclature, 1991)

Frecuencia

Las señales EEG se graban con una frecuencia de muestreo de 100 Hz o superior. Los modernos sistemas de adquisición pueden lidiar fácilmente con altas tasas de muestreo, y ahora es habitual registrar EEG con una frecuencia de muestreo de 500 Hz o más para permitir el estudio de oscilaciones de alta frecuencia o transiciones rápidas entre los diferentes electrodos.(Freeman y Quiroga, 2012)

Una señal EEG se graba digitalmente y se guarda en un formato definido en los archivos de una PC, existen diferentes formatos en que pueden ser guardados dependiendo del artefacto con que se hayan tomado, algunos son EEG, CSV, EDF, Mat. Esta señal finaliza siendo una matriz cuyas dimensiones depende del número de canales (N filas) y la duración junto con la frecuencia de muestreo del electroencefalograma (N columnas), donde se puede continuar realizando el análisis computacional.

El análisis EEG plantea una serie de desafíos que lo convierten en un campo de aplicación interesante:

1) EEG viene en grandes bases de datos. P.ej. una grabación de toda la noche del sueño humano da como resultado ocho horas de datos multicanal muestreados con hasta 256 Hz.

2) Las señales EEG son muy ruidosas. Mientras que la actividad de fondo eléctrico del cerebro humano está en el rango de 1-200 mV, los potenciales evocados (EP) tienen una amplitud de solo 1-30 mV.

3) Las señales EEG tienen una gran variación temporal. Mientras que la localización espacial de EEG ya está bien investigada, aún se necesita un gran esfuerzo para tener en cuenta la variación temporal entre sujetos.

4) El análisis de los datos de EEG requiere el uso de la gama completa de técnicas. Hay tareas para análisis de señales, minería de datos, reconocimiento de patrones, etc.

2.3. Análisis de señales

2.3.1. Ruidos y artefactos

Al recibir la señal cruda de los electrodos antes de realizar cualquier tipo de análisis matemático, es necesario que se realice algún tipo de pre procesamiento sobre dicha señal, esto abarca desde, reorganizar los datos que se obtienen, como extraer algún punto de interés sin realizar alguna otra modificación o remover algunos elementos de un electrodo con mala calidad, hasta un filtrado o transformación que facilite el análisis posterior modificando los datos crudos del EEG. La razón detrás del pre procesado es que toda señal EEG viene acompañada por un ruido que entorpece el procesamiento, este ruido suele ser parte de la señal por lo que eliminar parte del ruido puede derivar en eliminar alguna porción de datos correctos.

Debido a su muy baja amplitud, las señales EEG también se contaminan fácilmente por fuentes externas. Estos "artefactos", inherentes a las grabaciones de EEG del cuero cabelludo, se producen por movimientos de la cabeza, parpadeo, electrocardiograma, actividad muscular, etc. Los artefactos musculares pueden, por ejemplo, ser generados por una postura tensa, que generalmente se correlaciona con la actividad de alta frecuencia. En principio, estos artefactos pueden eliminarse utilizando filtros digitales estándar, pero desafortunadamente, en algunos casos, esto no es posible porque se superponen con las frecuencias de interés. Los movimientos de la cabeza se correlacionan con la actividad de baja frecuencia, y en este caso, el EEG típicamente muestra una línea base fluctiva. Un punto importante es que los artefactos generalmente limitan la longitud de las grabaciones de EEG que pueden considerarse estacionarias (es decir, segmentos en los que las principales características de la señal, como su media, varianza y espectro de potencia no cambian). De hecho, la duración relativamente corta de las grabaciones estacionarias de EEG es uno de los principales desafíos para su análisis. (Freeman y Quiroga, 2012)

El objetivo del preproceso es eliminar el ruido y preparar la señal para un procesamiento posterior. Las operaciones incluyen el filtrado de la señal y la eliminación de ciertos artefactos.

2.3.2. Transformada discreta de Fourier

Jean Baptiste Joseph Fourier, un brillante matemático francés, realizó su mayor trabajo en la conducción de calor, con la forma de descomponer señales periódicas en sumas ponderadas de funciones seno y coseno. Esta descomposición, en nuestros días conocida como serie de Fourier, ha sido su mayor contribución a la ciencia, en gran parte trascendiendo su aplicación original a la conducción de calor.

Siguiendo las ideas de Fourier, las señales como las registradas desde los electrodos de superficie del EEG del cuero cabelludo pueden representarse en el dominio del tiempo o alternativamente en términos de su descomposición en senos y cosenos en el dominio de la frecuencia, dando la idea de usar representaciones de frecuencia. Básicamente, trata de obtener una imagen más simple de algunas de las características básicas de la señal, que generalmente no son obvias a partir de representaciones de tiempo ruidosas y complejas.

Las señales digitales tienen una longitud infinita y se muestrean con una determinada frecuencia de muestreo. Esta longitud y muestreo finitos introducen varios problemas que discutiremos en esta y las siguientes secciones. Consideremos una señal discreta $x[n]$ $n = 1, \dots, N$, que se ha derivado de una señal continua $x(t)$ muestreando a intervalos de tiempo iguales Dt (es decir, con una frecuencia de muestreo). Obviamente, la longitud de la señal es $T = N * Dt$. La Transformada discreta de Fourier se define como:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{2i\pi kn/N} \quad (2.1)$$

El cálculo de la Transformada discreta de Fourier requiere N^2 multiplicaciones complejas, porque para cada una de las N frecuencias discretas k tenemos que calcular una suma de N multiplicaciones con exponenciales complejos. Esto puede tomar demasiado tiempo para N grande, pero, afortunadamente, es posible reducir drásticamente la velocidad de cálculo mediante el uso del algoritmo de Transformada rápida de Fourier (FFT, Cooley y Tukey 1965). La introducción de la FFT ha revolucionado el análisis de las señales digitales y, en particular, ha impulsado el estudio de los EEG en el dominio de la frecuencia. (Freeman y Quiroga, 2012)

2.3.3. Filtrado de Tikhonov

En esta sección se analizará el desarrollo de una aplicación para utilizar el filtrado de Tikhonov en el problema de eliminar ruidos y artefactos en señales multicanal, de manera conjunta. (Lampe y Voss, 2013).

Para explicarlo, sea X una matriz con los datos de las señales, de tal forma que el número de canales es el número de filas de la matriz, y el número de muestras es el número

de columnas de la matriz.

Así, x_{lj} es el dato electroencefalográfico en el canal l y la muestra j . La transformada de Fourier discreta se define como:

$$d = XA \quad (2.2)$$

Donde la matriz A se define como:

$$A_{tu} = \frac{1}{N} e^{2i\pi kj/N} \quad (2.3)$$

A está mal condicionada. El filtrado de T , consiste en reemplazar la matriz A en la ecuación tal, por parámetro de regularización:

$$d_\alpha = X(A + \alpha^2 I) \quad (2.4)$$

La selección óptima de α se realiza mediante el criterio de la curva "L", que consiste en una regularización que hace que los valores propios mas pequeños de A , que corresponden a las altas frecuencias de sus vectores propios, aumenten en magnitud, regularizando la variabilidad de la señal. Es destacable que en esta propuesta el filtrado se realiza con todos los canales a un tiempo, por primera vez.

El filtrado consiste en resolver el sistema:

$$d_\alpha = X_\alpha A_\alpha, A_\alpha = A + \alpha^2 I \quad (2.5)$$

X_α es la señal filtrada.

2.4. Reconocimiento de patrones

El Reconocimiento de Patrones (RP) es una disciplina de las Ciencias Computacionales cuyo fin es extraer características, a partir de un conjunto de datos, que permitan discernir, mediante métodos automáticos, entre aquellas condiciones o comportamientos comunes y aquellas consideradas anómalas para esos datos; estas características, una vez determinadas, pueden ser usadas para representar tales datos para su reconocimiento, recuperación, filtrado o interpretación.

El RP ha mostrado su utilidad al ser aplicado en una inmensa variedad de contextos, donde se hace necesario identificar dichos patrones para automatizar procesos, optimizar recursos, detectar fallas, etc. Las técnicas automáticas para el Reconocimiento de Patrones han sido usadas en una gran variedad de ámbitos experimentales con resultados prometedores. Muchas de estas mismas técnicas, incluso, son empleadas habitualmente en una amplia gama de aplicaciones del mundo real como son la Visión Computacional y el Procesamiento de Lenguaje Natural, por mencionar algunos. (Alba y Cid, 2006)

2.4.1. Análisis de Componentes Principales (PCA)

Es una forma de identificar patrones en los datos y expresarlos de tal manera que resalten sus similitudes y diferencias. Dado que los patrones en los datos pueden ser difíciles de encontrar en datos de alta dimensión, donde el lujo de la representación gráfica no está disponible, la PCA es una poderosa herramienta para analizar datos. La otra ventaja principal de PCA es que una vez que ha encontrado estos patrones en los datos, comprime los datos, es decir, reduciendo el número de dimensiones, sin mucha pérdida de información. Para realizar este objetivo, ocupa diferentes propiedades matemáticas básicas de fondo que serán necesarias para comprender el proceso del Análisis de Componentes Principales. (Smith, 2002).

Covarianza

Muchos conjuntos de datos tienen más de una dimensión, y el objetivo del análisis estadístico de estos conjuntos de datos suele ser ver si hay alguna relación entre las dimensiones.

La covarianza es un valor que indica el grado de variación conjunta de dos variables aleatorias respecto a sus medias. Es el dato básico para determinar si existe una dependencia entre ambas variables y además es el dato necesario para estimar otros parámetros básicos, como el coeficiente de correlación lineal o la recta de regresión.

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))) \quad (2.6)$$

Para cada elemento de datos, multiplique la diferencia entre X valor y la media de X, por la diferencia entre el valor Y y la media de Y. Sume todos estos y divida entre (n-1). Si el valor es positivo, eso indica que ambas dimensiones aumentan juntas, lo que significa que, en general, a medida que aumenta X, también lo hace Y. Si el valor es negativo, a medida que aumenta una dimensión, la otra disminuye. En el último caso, si la covarianza es cero, indica que las dos dimensiones son independientes entre sí.

Recuerde que la covarianza siempre se mide entre 2 dimensiones. Si tenemos un conjunto de datos con más de 2 dimensiones, se puede calcular más de una medida de covarianza. Por ejemplo, a partir de un conjunto de datos tridimensionales (dimensiones, X, Y, Z), puede calcular $\text{cov}(x, y)$, $\text{cov}(X, Z)$, y $\text{cov}(Y, Z)$. De hecho, para un conjunto de datos tridimensionales, puede calcular diferentes valores de covarianza. smith2

Una forma útil de obtener todos los valores de covarianza posibles entre todas las diferentes dimensiones es calcularlos todos y ponerlos en una matriz. Entonces, la definición para la matriz de covarianza para un conjunto de datos con dimensiones es:

$$C^{n \times n} = (c_{ij}, c_{ij} = \text{cov}(Dim_i, Dim_j)) \quad (2.7)$$

Donde hay una matriz C con n filas y n columnas, y Dim_x es la dimensión xth. Todo lo que esta fórmula dice es que si tienes un conjunto de datos n-dimensionales, entonces la

matriz tiene n filas y columnas (por lo que es cuadrada) y cada entrada en la matriz es el resultado de calcular la covarianza entre dos dimensiones separadas.

Ejemplo de la matriz de covarianza para un conjunto de datos, los valores son los siguientes:

$$C = \begin{pmatrix} cov(x, x) & cov(x, y) & cov(x, z) \\ cov(y, x) & cov(y, y) & cov(y, z) \\ cov(z, x) & cov(z, y) & cov(z, z) \end{pmatrix}$$

Algunos puntos a tener en cuenta: en la diagonal principal, el valor de covarianza está entre una de las dimensiones y ella misma. Estas son las variaciones para esa dimensión. El otro punto es que, dado que $cov(a,b)=cov(b,a)$, la matriz es simétrica con respecto a la diagonal principal.

AutoVectores

Los vectores propios solo se pueden encontrar para matrices cuadradas. Y, no todas las matrices cuadradas tienen vectores propios. Y, dada una matriz $n * n$ que tiene vectores propios, hay n de ellos.

Todos los vectores propios de una matriz son perpendiculares, es decir. en ángulo recto entre sí, no importa cuántas dimensiones tenga. Por cierto, otra palabra para perpendicular, en el habla matemática, es ortogonal. Esto es importante porque significa que puede expresar los datos en términos de estos vectores propios perpendiculares, en lugar de expresarlos en términos de ejes X y Y .

Encontrar los vectores propios, desafortunadamente, es fácil si tienes una matriz bastante pequeña, como no más grande que $3 * 3$. Después de eso, la forma habitual de encontrar los vectores propios es mediante algún método iterativo complicado.(Smith, 2002)

AutoValores

Los valores propios están estrechamente relacionados con los vectores propios.

Entonces puede ver que los vectores propios y valores propios siempre vienen en pares. Cuando calculas sus vectores propios, normalmente obtienes los valores propios asociados a los vectores propios.(Smith, 2002)

Procedimiento

Dada una matriz X , la matriz de covarianzas muestrales es:

$$M = \frac{1}{n-1}XX^t \tag{2.8}$$

La diagonalización de M es una factorización (única) tal que:

$$M = UDU^{-1} \tag{2.9}$$

Donde D es la matriz diagonal con los valores propios λ_i , que se obtienen como raíces de la ecuación:

$$\det(M - \lambda I) = 0 \quad (2.10)$$

Y las columnas de U son soluciones de:

$$(M - \lambda_j I)U_j = 0 \quad (2.11)$$

2.5. Minería de datos

Una definición de Minería de datos, DM se refiere al proceso general de descubrimiento de conocimiento útil a partir de datos, pero también limpieza y preprocesamiento de datos, reducción y proyección de datos, incorporación de conocimientos previos y también la validación e interpretación correctas de los resultados. Se considera que está relacionada con los campos de aprendizaje automático, estadísticas, reconocimiento de patrones, inteligencia artificial, bases de datos y visualización, entre otros. Sin embargo, el objetivo es la extracción de conocimiento a partir de datos en el contexto de grandes bases de datos.

Las técnicas de minería de datos provienen de la inteligencia artificial y de la propia estadística. Se trata de algoritmos que se aplican sobre un conjunto de datos con el objetivo de obtener resultados. Las técnicas más representativas que se utilizan en la minería de datos son. (Flexer, 2000)

CLUSTERING (Agrupamiento). Agrupan datos dentro de un número de clases preestablecidas o no, partiendo de criterios de distancia o similitud, de manera que las clases sean similares entre sí y distintas con las otras clases. Su utilización ha proporcionado significativos resultados en lo que respecta a los clasificadores o reconocedores de patrones, como en el modelado de sistemas. Este método debido a su naturaleza flexible se puede combinar fácilmente con otro tipo de técnica de minería de datos, dando como resultado un sistema híbrido.

Un problema relacionado con el análisis de cluster es la selección de factores en tareas de clasificación, debido a que no todas las variables tienen la misma importancia a la hora de agrupar los objetos. Otro problema de gran importancia y que actualmente despierta un gran interés es la fusión de conocimiento, ya que existen múltiples fuentes de información sobre un mismo tema, los cuales no utilizan una categorización homogénea de los objetos. Para poder solucionar estos inconvenientes es necesario fusionar la información a la hora de recopilar, comparar o resumir los datos.

REGRESIÓN LINEAL. Es una de las técnicas más utilizadas para la formación de relaciones entre datos. Se trata de un sistema rápido y eficaz pero que a su vez cuenta con insuficiencias en los espacios multidisciplinares donde puedan relacionarse más de 2 variables.

2.5.1. Análisis Discriminante

Es una técnica estadística multivariante cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos para, en el caso de que existan, explicar en qué sentido se dan y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados. (De la Garza García, Morales Serrano, y González Cavazos, 2013)

Pasos necesarios para aplicar el Análisis discriminante.

Podrían simplificarse en tres pasos:

1. Selección de las variables discriminantes

Puede hacerse a través de la identificación de variables importantes y, por otro lado, de la aplicación de pruebas estadísticas. Ambas vías son complementarias y no excluyentes. La selección estadística consiste en la combinación de un proceso de análisis descriptivo (cálculo de las medias y desviaciones típicas), correlación entre las variables y diferencias significativas entre los grupos. La correlación se comprobará dentro del grupo y entre grupos.

2. Selección de la función discriminante

La función discriminante no es otra cosa que un factor o una nueva variable, combinación lineal de las anteriores. Podremos obtener más de una función discriminante pero solo una permitirá hacer predicciones de las mejores clasificaciones futuras. Se utilizará el coeficiente de “autovalor” y los coeficientes estandarizados de las funciones discriminantes.

3. Clasificación de los sujetos u objetos

Para poder asignar un sujeto a un grupo u otro tenemos que diseñar una regla de decisión.

2.5.2. Regresión logística

La regresión logística es un tipo especial de regresión que se utiliza para explicar y predecir una variable categórica binaria (dos grupos) en función de varias variables independientes que a su vez pueden ser cuantitativas o cualitativa. Permite modelizar la probabilidad de que ocurra un evento dado una serie de variables independientes. (Anderson, 2005)

Esto requiere: 1. Una o varias variables continuas, que cambian a lo largo del tiempo. 2. Registros de las variables en tres o más instantes del tiempo, en el mismo sujeto. 3. Una unidad de tiempo relevante al fenómeno que se estudia, tomando en cuenta su cadencia y velocidad de cambio.

Modelo:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \text{tiempo} + \text{error} \quad (2.12)$$

$$fr = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (2.13)$$

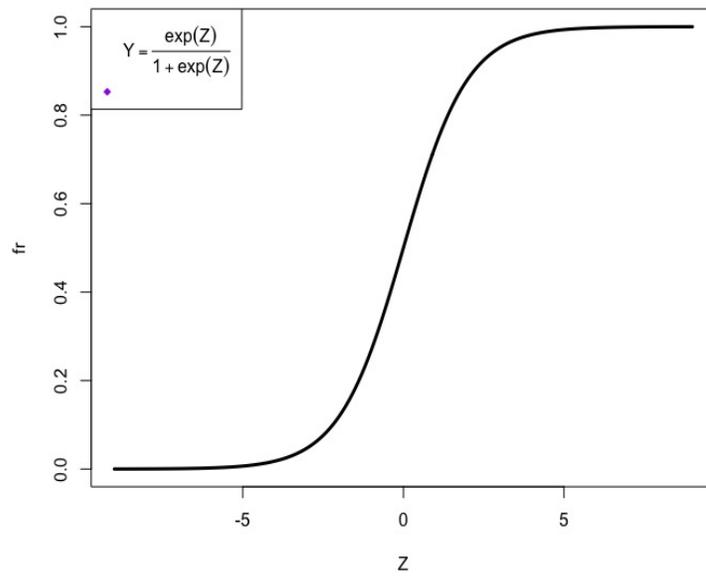


Figura 2.3: Curva Logística.
Fuente:(Anderson, 2005)

2.5.3. Red Bayesiana

La Red Bayesiana es una clase de modelos gráficos que permiten una representación concisa de dependencias probabilísticas entre un conjunto dado de variables $X = \{ X_1 \dots X_p \}$, en forma de un grafo acíclico dirigido. Cada nodo corresponde a una variable aleatoria. Si el arco dirigido que va del nodo X_j al nodo X_i pertenece a la RB, se va a escribir $X_j \rightarrow X_i$. Para cada nodo X_i de la RB, $\prod X_i$ denota al subconjunto de los “padres” de X_i . Un padre de X_i es cualquier nodo donde empieza un arco dirigido que acaba en X_i (ver Fig. 2).

$$\prod X_i = \{ X_j \in X \mid X_j \rightarrow X_i \} \quad (2.14)$$

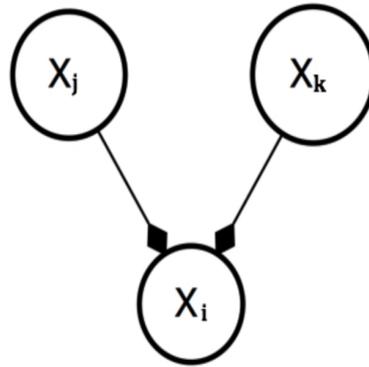


Figura 2.4: Detalle de la RB, nodo X_i con sus padres X_j y X_k .

Por el teorema de Bayes, la distribución conjunta de $X = \{ X_1 \dots X_p \}$ está determinada por la estructura de dependencia entre cada nodo de la RB, y sus padres:

$$P_X(X) = \prod_{i=1}^p P_{X_i}(X_i \mid \prod X_i) \quad (2.15)$$

Capítulo 3

Estado del arte

En esta sección se presentan los artículos y avances relacionados con esta investigación para hallar evidencias de que al analizar trazos EEG con respecto a sus espectros de potencia y los diversos rangos de frecuencias, mediante la aplicación de técnicas computacionales, será posible caracterizar la normalidad, las etapas de maduración neurológica y la no estacionalidad de estos grupos. Se dividió en 2 secciones que son: reconocimiento de patrones y Minería de datos, para presentar como se han ocupado las diferentes áreas en la electroencefalografía y sus avances.

3.1. Reconocimiento de patrones

Se han realizado la predicción de ataques convulsivos para su uso en dispositivos de epilepsia terapéutica para activar la intervención para prevenir las convulsiones antes de comenzar. Los estudios sobre la predicción de las convulsiones varían ampliamente en sus enfoques teóricos del problema, la validación de los resultados y la cantidad de datos analizados. Se han aplicado DWT(Transformada Wavelet) para el análisis tiempo-frecuencia de señales EEG para la clasificación utilizando coeficientes wavelet. Las señales EEG se descomponen en sub-bandas de frecuencia usando DWT. Para después extraer un conjunto de características estadísticas de estas sub-bandas para representar la distribución de los coeficientes wavelet. PCA, ICA(Análisis de Componentes Independientes) y LDA(Análisis Discriminante Lineal) se utilizan para reducir la dimensión de los datos.Las características se utilizan como una entrada a una máquina de vector de apoyo (SVM) con dos salidas discretas: convulsiones epilépticas o no epilépticas. El mejor rendimiento se logra en la extracción de características LDA, el proceso de entrenamiento se lleva a cabo utilizando el núcleo RBF a PCA + SVM, ICA + SVM, y LDA + SVM. Después del entrenamiento, utilizan tres métodos de extracción de características diferentes. Usando las características de PCA, ICA y LDA se extraen de los conjuntos de características originales. Además, el número de vectores de soporte (SV) disminuye debido a la extracción de características. La tasa de clasificación con extracción de características LDA es la más alta (100%) y la ACI llegó segunda (99,5%). El PCA tuvo el porcentaje de clasificación correcto más bajo (98,75%) en comparación con los homólogos

de LDA e ICA. También la simulación muestra que SVM por extracción de características utilizando PCA, ICA o LDA siempre puede tener mejor rendimiento que sin extracción de características (98 %), como presentan A. Subasi y M. Ismail Gursoy (Subasi y Gursoy, 2010).

Se ha comparado el rendimiento de PCA e ICA en aplicaciones con electroencefalograma (EEG) de modelado. Una aplicación es la reducción de la dimensionalidad, otra es el modelado de una interesante característica de las señales EEG denominada evento relacionado con el potencial (ERP). Se ocupa PCA e ICA para reducir la dimensionalidad. Utilizán el algoritmo JADE para estas 2 aplicaciones con un paquete Matlab1 para realizar el algoritmo JADE. Este paquete fue diseñado para facilitar el análisis EEG y para visualizar gráficamente las fuentes estimadas. Con PCA, se encontro sólo un componente independiente necesario para analizar el pico P300. Sin embargo, este análisis no separa los fenómenos P3a y P3b. Todo el procedimiento es automático y fácil de configurar. ICA es muy prometedor para analizar las señales biomédicas multidimensionales y mucho más eficiente que el PCA en el contexto de análisis EEG, al poder separar los 2 fenómenos, como se muestra en el artículo de C. Bugli y P. Lambert (Bugli y Lambert, 2007).

X. Liu et al. Investigan la complejidad de las señales de EEG y distinguir las señales de EEG con ritmos diferentes. Métodos: Los datos de la EEG se utilizan en el modelo de masa natural. El modelo con diferentes parámetros produce las señales de EEG con ritmos diferentes, tales como señales EEG normales y picos epileptiformes. La permutación-entropía modificada se utiliza para extraer las características de complejidad de las señales EEG normales y los picos epileptiformes y se compara con el algoritmo de permutación-entropía y el algoritmo de permutación-entropía de Rényi. El análisis estadístico se utiliza para caracterizar las diferencias entre las señales. Resultados: Las señales correspondientes a $A = 3,25$ mV y $A = 3,6$ mV son fáciles de distinguir según los valores medios y las desviaciones estándar de la permutación-entropía modificada. Los resultados estadísticos revelan que las diferencias de grupo de los valores de permutación-entropía modificados para las señales correspondientes a $A = 3,25$ mV y $A = 3,4$ mV son significativas, manifestadas como $F < 1$ y $p > 0,05$. En comparación con la entropía de permutación y la entropía de permutación de Rényi, los valores F de la permutación-entropía modificada son los mayores y los valores p de la permutación-entropía modificada son los más pequeños. Por lo tanto, concluyen que la permutación-entropía modificada es una medida de complejidad efectiva para cuantificar las señales de EEG y la permutación-entropía modificada es más eficaz para distinguir las señales de EEG con ritmos diferentes en comparación con la entropía de permutación y la entropía-permutación de Rényi. (Liu, Wang, Gao, y Gao, 2017)

Existe un algoritmo robusto de análisis de componentes principales (PCA) para reducir la dimensión de las características del EEG para la estimación de la vigilancia. El rendimiento se compara con el estándar PCA, L1-norma PCA, escasa PCA y robusto PCA en reducción de la dimensión de la característica en un conjunto de datos de EEG de veintitrés sujetos. Para evaluar el rendimiento de estos algoritmos, utilizan características de entropía diferencial suavizada como las características de EEG relacionadas con la vigilancia. Los resultados experimentales demuestran que la robustez y el rendimiento del PCA robusto son mejores que otros algoritmos para la estimación de vigilancia fuera de línea y en línea. El RMSE

promedio (error cuadrático medio) de la estimación de vigilancia fue de 0,158 cuando se aplicó PCA robusto para reducir la dimensionalidad de las características, mientras que el RMSE medio fue de 0,172 cuando se usó PCA estándar en la misma tarea. Este algoritmo lo presentan Li-Chen Shi et al. (Shi, Duan, y Lu, 2013)

Kavita Mahajan. Investigaron el procesamiento y análisis de electroencefalograma (EEG) en un marco propuesto que lo llevaron a cabo con DWT para la descomposición de la señal en sus sub-bandas de frecuencia y un conjunto de características estadísticas extrajeron de las subbandas para representar la distribución de los coeficientes wavelet. La reducción de la dimensión de los datos se hace con la ayuda de análisis de componentes principales y análisis de componentes independientes. Entonces estas características se utilizaron como una entrada a una red neuronal para la clasificación de los datos como normal o de otra manera. Presentaron y compararon el desempeño del proceso de clasificación debido a diferentes métodos para mostrar el excelente proceso de clasificación. Estos hallazgos se presentan como un ejemplo de un método para el entrenamiento y para probar un método de predicción normal y anormal en los datos de pacientes epilépticos individuales pequeños. Método de extracción de características, Exactitud (%), Sensibilidad (%), Especificidad (%), PCA 93,63 % 62,93 % 98,83 %. ICA 96,75 % 96,75 % 96,75 %.(Mahajan, Vargantwar, y Rajput, 2011)

Seungjin et al. Realizaron un sistema del dominio de la interfase informática cerebral(BCI) basado en EEG. El sistema consta de dos procedimientos: (1) extracción de características; (2) clasificación. Presentaron dos métodos para la clasificación de patrones EEG, el primero inicia con PCA y las características de los componentes principales extraídas separadamente de los canales C3 y C4 se concatenan, luego se introducen en el correspondiente HMM(modelo de Hidden Markov) (que modela el movimiento hacia la izquierda o hacia la derecha) para el entrenamiento. Por otro lado, en la segunda metodología, las características de los componentes principales de cada canal se introducen en dos HMM por separado, lo que da como resultado cuatro HMM en total. El SVM se emplea en el segundo método para tomar una decisión final a partir de las puntuaciones de probabilidad calculadas por HMMs. La comparación de rendimiento para características de PCA, datos de EEG sin procesar y parámetros de Hjorth reflejó los siguientes resultados, metodología 1: PCA 75.70, Raw 60.63, Hjorth 56.88, metodología 2 : PCA 78.15, Raw 64.38, Hjorth 66.50, por lo tanto la metodología 2 mostró un rendimiento ligeramente mejor en comparación con la metodología 1.(Lee y Choi, 2003)

3.2. Minería de datos

Schaltenbrand informa sobre el trabajo que intenta automatizar basado en la estadificación del sueño. Usan perceptrones multicapa (MLP) para clasificar las grabaciones de sueño durante toda la noche de acuerdo con las reglas de R & K. Usan parámetros espectrales como potencia relativa y total, relación de potencias y frecuencias medias de densidades de EEG (dos canales, uno central, un occipital), señales EOG y EMG como entradas a la red. Para cada segmento de dos segundos se calcula dicho vector de características de la dimensión 17. Aunque los autores afirman que el cálculo de estas características solo es significativo en in-

tervalos cortos (dos segundos) en los que las señales pueden decirse estacionarias, promedian 15 de los vectores de características para obtener una estimación del vector de características correspondiente al Ventanas de 30 segundos en las que se basan las reglas de R & K. Usando estos vectores de características, un MLP puede alcanzar una precisión de aproximadamente 80 % en 12 nuevas grabaciones de sueño que nunca antes se habían visto durante toda la noche.(Schaltenbrand, Lengelle, y Macher, 1993).

Empleando una versión simplificada de ART2, NeoART, el sistema puede rechazar vectores de entrada que son muy diferentes de los utilizados durante el entrenamiento. Por lo tanto, se pueden detectar artefactos y épocas generalmente ambiguas de las grabaciones de sueño y las decisiones finales se pueden entregar a expertos humanos. NeoART es un algoritmo estrechamente relacionado con el algoritmo clásico de LBG [Linde et al. 1980] para la cuantificación de vectores, así como para el enfoque de clustering k-means desarrollado en la literatura de análisis de cluster a partir de [MacQueen 1967]. Todos estos algoritmos representan un conjunto de datos a través de un conjunto más pequeño de vectores de libros de códigos prototípicos o centros de clusters. Se rechaza un nuevo patrón de entrada si no cae dentro de ninguna de las hiper-esferas de un radio dado alrededor de los centros obtenidos por NeoART.(Carpenter y Grossberg, 1990).

(Roberts y Tarassenko 1992) no quieren aplicar un conjunto fijo de reglas para clasificar automáticamente las etapas del sueño, sino que apuntan a dar alguna indicación de la dinámica del sueño en los humanos. Los autores desean desarrollar un sistema de análisis alternativo que no esté basado en las reglas de R & K, especialmente porque estos últimos se han definido para sujetos sanos normales y tienden a descomponerse cuando se aplican a patrones de sueño anormales. Su base de datos consiste en nueve registros de sueño de toda la noche de adultos jóvenes y sanos, de los cuales se utilizaron solo un canal de EEG y EMG para su posterior análisis. La señal de EEG se parametrizó usando un filtro de Kalman de décimo orden para cada segmento de un segundo que proporciona diez coeficientes de predicción lineales variables en el tiempo como entrada para la red. La red era un mapa de autoorganización (SOM) según [Kohonen 1984] con un mapa de salida bidimensional de $10 * 10$ unidades. Se puede decir que SOM hace clustering y al mismo tiempo preserva la ordenación espacial de los datos de entrada reflejados por un ordenamiento de los centroides del clúster en un espacio de salida de una o dos dimensiones. La última propiedad está estrechamente relacionada con la escala multidimensional (MDS) en las estadísticas. Se puede encontrar más información sobre la relación de los SOM con la agrupación y el MDS en (Blakrishnan et al. 1994), (Flexer 1997), (Bishop et al. 1998) y (Flexer 1999). Utilizando los coeficientes de Kalman para cada segmento de un segundo como vectores de entrada para el SOM descuidando su orden en el tiempo, los autores observaron ocho grupos claramente visibles en el mapa de salida. Aún más importante, fueron capaces de distinguir tres trayectorias típicas entre estos grupos.(Roberts y Tarassenko, 1992)

(Kubat et al. 1994) introducen dos métodos de aprendizaje automático, la inducción de árboles de decisión (ID3) y la cuantificación del vector de aprendizaje (LVQ), al problema de la clasificación del sueño en los bebés. Su trabajo está relacionado con la estadificación del sueño según las reglas de (Guilleminault & Souquet). ID3 por (Quilan, 1982) es un algoritmo

para la inducción automática (aprendizaje) de un árbol de decisiones a partir de un conjunto de ejemplos. Un árbol de decisión se puede ver como una forma jerárquica de describir una partición de un espacio de datos. Una ventaja sobre otros enfoques de clasificación es que la terminología de los árboles es gráfica y fácil de comprender. Es importante tener en cuenta que cada rama introduce una nueva partición del conjunto de datos en función de los valores de una sola variable. Por lo tanto, ID3 también logra algo así como una clasificación de la importancia de las variables, ya que se usan variables más discriminantes en las ramas ubicadas en lo alto de la jerarquía. LVQ introducido por [Kohonen 1990] es un método no paramétrico de clasificación. Un libro de códigos inicializado aleatoriamente similar a un conjunto de centros de clúster divide el conjunto de datos y se mejora iterativamente. Durante cada pasada a través del conjunto de datos del conjunto de códigos de datos, los vectores se desplazan hacia ejemplos cercanos de su propia clase, y se alejan de los de otras clases. Vea los capítulos respectivos en (Ripley 1996) para la vista estadística en ambos árboles de decisión y LVQ. (Kubat, Pfurtscheller, y Flotzinger, 1994)

Capítulo 4

Presentación de herramientas computacionales¹

4.1. Desarrollo de la Aplicación para Filtrado por regularización de Tikhonov

El uso de la transformada de Fourier para obtener filtrado multicanal del EEG, presenta dos importantes problemas. El primero, la no estacionariedad de la señal biofísica. Esto se remedia segmentando la señal en partes que sean homogéneas. Cada una de éstas partes se llama ventana, de tal suerte que la transformación de Fourier se hace ventana por ventana. El segundo problema es el mal condicionamiento de la transformada de Fourier Discreta, vista esta como una transformación lineal. Para remediar este defecto, que puede llegar a introducir inestabilidades numéricas en los cálculos, se presenta el filtrado de Tikhonov multicanal. Hasta donde se sabe, este es el primer uso de este filtrado en el preprocesamiento de señales EEG, y la hipótesis planteada es que su uso elimina inestabilidades en los posteriores algoritmos lineales de clasificación o discriminación.

4.1.1. Características de los datos

La base de datos consta de grabaciones de EEG de sujetos pediátricos con convulsiones intratables (Gotman, Kobayashi, Bagshaw, Bénar, y Dubeau, 2006). Los sujetos fueron monitorizados durante varios días. Las grabaciones están agrupadas en 22 casos (5 varones, edad de 3 a 22 años, y 17 mujeres, edad de 1.5 a 19 años). Cada caso contiene registro de diferencias de potencial eléctrico (en μV), con frecuencia de muestreo de 256 muestras por segundo, y una resolución de 16 bits. Las tomas se realizaron mientras el sujeto estaba en estado de reposo. Se utilizó el sistema internacional 10-20 de posicionamiento de electrodos. La base de datos consta en total de 484 archivos con extensión EDF.

¹En el presente documento, por indicación del director de tesis, los términos diseño, aplicación y desarrollo de técnicas o herramientas computacionales, se refieren a la implementación de algoritmos tradicionales.

4.1.2. Algoritmo de Tikhonov

A continuación el diagrama de flujo 1 (Figura. 4.1) muestra el procedimiento que se llevó a cabo para el desarrollo de la aplicación y a continuación se explica cada proceso mas a detalle.

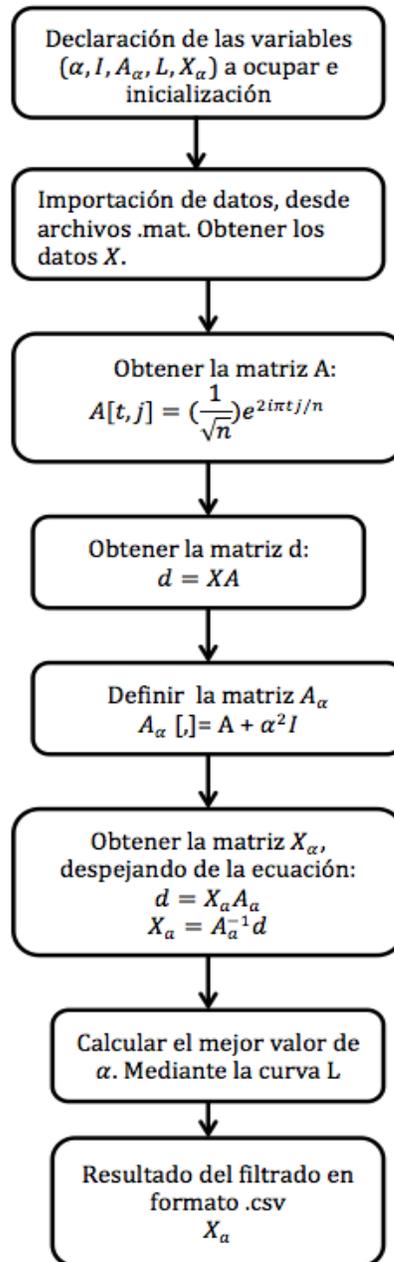


Figura 4.1: Diagrama de flujo para el desarrollo de la aplicación para el filtrado por regularización de Tikhonov.

En la Figura 4.2 se muestra el efecto del parametro α sobre el valor de la parte real de los valores propios de la matriz A_α . Se aprecia que al mayor valor del α , el condicionamiento del parametro de la matriz mejora.

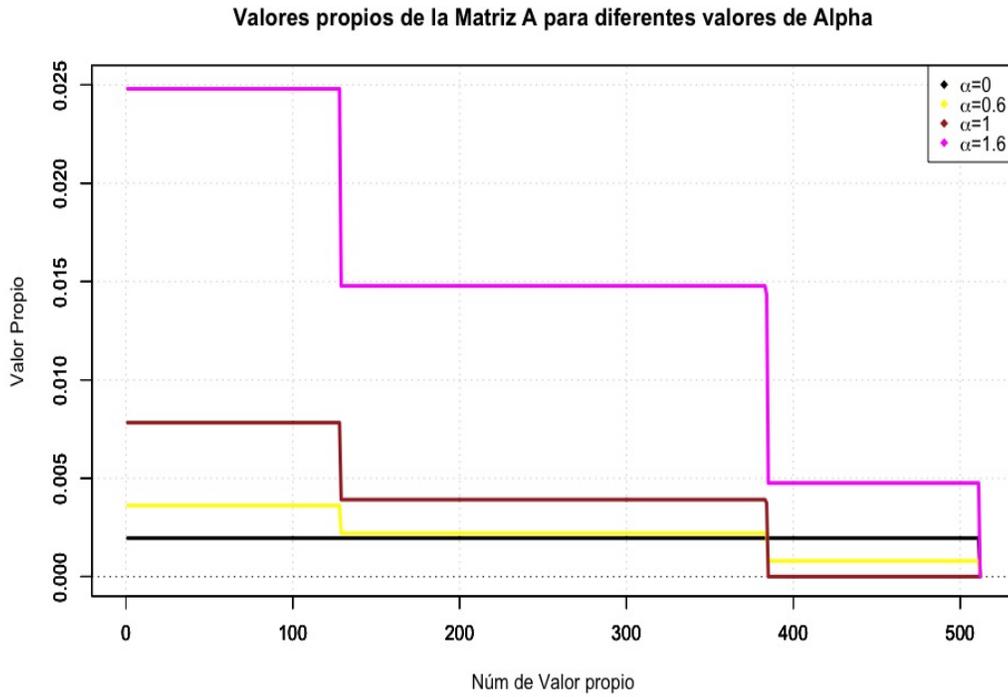


Figura 4.2: Parte real de los valores propios de la matriz A_α para diferentes valores de α .

La Figura 4.3 visualiza el efecto de la selección del parámetro α sobre la señal original. Se aprecia que el mejor valor del α es el 0.6, ya que no pierde las frecuencias principales de la señal, mejorando el condicionamiento.

Figura 4.3: Gráfica que muestra el efecto de regularización en el filtrado de Tikhonov

4.2. Desarrollo de la Aplicación para el Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG

A continuación el diagrama de flujo 1 (Figura 4.4) muestra el procedimiento que se llevó a cabo para el desarrollo de la aplicación y a continuación se explica cada proceso mas a detalle.

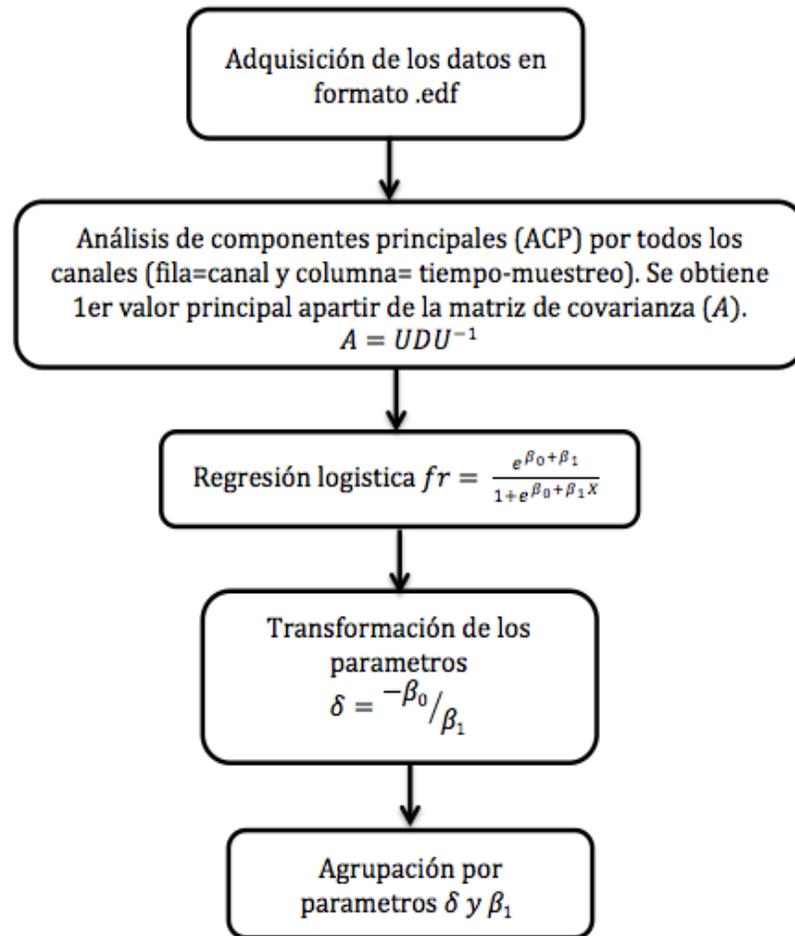


Figura 4.4: Diagrama de flujo para el desarrollo de la aplicación para el Análisis Longitudinal de Conglomerados para CFC

4.2.1. Características de los datos

La base de datos consta de grabaciones de EEG de sujetos pediátricos con convulsiones intratables (Gotman y cols., 2006). Los sujetos fueron monitorizados durante varios días. Las grabaciones están agrupadas en 22 casos (5 varones, edad de 3 a 22 años, y 17 mujeres, edad de 1.5 a 19 años). Cada caso contiene registro de diferencias de potencial eléctrico (en μV), con frecuencia de muestreo de 256 muestras por segundo, y una resolución de 16 bits. Las tomas se realizaron mientras el sujeto estaba en estado de reposo. Se utilizó el sistema internacional 10-20 de posicionamiento de electrodos. La base de datos consta en total de 484 archivos con extensión EDF.

4.2.2. Análisis de Componentes Principales

Es una técnica utilizada principalmente para reducción de la dimensión. En el análisis cuantitativo EEG multicanal se utiliza para segmentación, clasificación y detección de eventos convulsivos, por lo cual se asume que es funcional para el análisis y detección de patrones de cambio de la CFC. (Berka, 2009).

Los componentes principales se calculan a partir de la señal por eigen-factorización de la matriz de covarianza de la señal multicanal. Si X es la matriz de los 22 canales del EEG (512 filas, 22 columnas), la matriz de covarianza ($n=512$) es:

$$\frac{1}{n-1}xx^T \quad (4.1)$$

Sea λ el conjunto formado por cada primer valor propio de cada toma (son 18 tomas en este ejemplo numérico), entonces la fracción de variación explicada por el primero de ellos es:

$$0 \leq fr = \frac{\lambda_1}{\sum \lambda_i} \leq 1 \quad (4.2)$$

4.2.3. Regresión Logística

El logito de fr se define (Sprott, 2008) por:

$$Y = \log\left(\frac{fr}{1-fr}\right) \quad (4.3)$$

En el ajuste lineal del logito de fr , se usa como variable regresora X al logaritmo natural del número de toma (tiempo).

$$Y = \beta_0 + \beta_1 * \chi + \epsilon \quad (4.4)$$

Donde la distribución de los errores ϵ se supone normal con media 0 y varianza constante. Los parámetros β_0, β_1 son estimados por máxima verosimilitud (Sprott, 2008). Por lo tanto, la fracción estimada de variación explicada por el primer valor propio es:

$$\widetilde{fr} = \frac{e^{\widetilde{\beta}_0 + \widetilde{\beta}_1 * \chi}}{1 + e^{\widetilde{\beta}_0 + \widetilde{\beta}_1 * \chi}} \quad (4.5)$$

Siendo $\widetilde{\beta}_0, \widetilde{\beta}_1$ los estimadores de máxima verosimilitud de los parámetros. Para mostrar el desempeño de esta técnica, se evaluaron los estimadores de fracción de variación \widetilde{fr} , de acuerdo al sexo de los pacientes, buscando diferencias en los patrones de cambio de las CFC mediante análisis de conglomerados logístico. Se adopta la transformación de parámetros:

$$\delta = -\frac{\widetilde{\beta}_0}{\widetilde{\beta}_1} \quad (4.6)$$

$$Y = \frac{\exp(Z)}{1 + \exp(Z)} \quad (4.7)$$

4.3. Desarrollo de la aplicación para el Análisis Longitudinal y Topográfico de la Conectividad Funcional Cerebral. Detección de patrones espaciales y Temporales de Conectividad Funcional Cerebral.

El proceso para el desarrollo de la aplicación se muestra en la Figura 4.5, el diagrama expone los pasos para la elaboración de la aplicación, y a continuación se explica cada proceso mas a detalle.

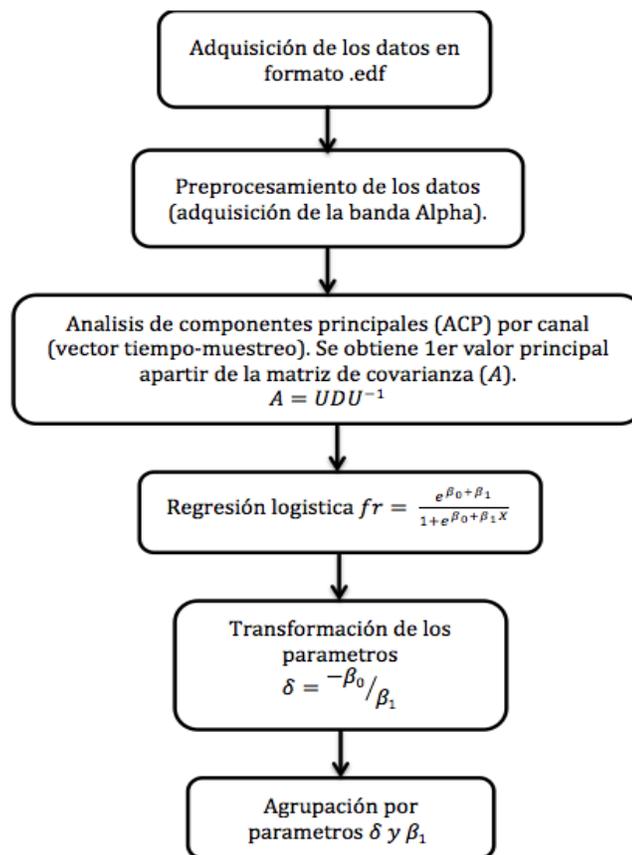


Figura 4.5: Diagrama de flujo para el desarrollo de la aplicación para el Análisis Longitudinal y Topográfico de la CFC.

4.3.1. Características de los datos

La base de datos consta de grabaciones de EEG de sujetos pediátricos con convulsiones intratables (Gotman y cols., 2006). Los sujetos fueron monitorizados durante varios días.

Las grabaciones están agrupadas en 22 casos (5 varones, edad de 3 a 22 años, y 17 mujeres, edad de 1.5 a 19 años). Cada caso contiene registro de diferencias de potencial eléctrico (en μV), con frecuencia de muestreo de 256 muestras por segundo, y una resolución de 16 bits. Las tomas se realizaron mientras el sujeto estaba en estado de reposo. Se utilizó el sistema internacional 10-20 de posicionamiento de electrodos. La base de datos consta en total de 484 archivos con extensión EDF.

4.3.2. Análisis de Componentes Principales

La Figura 1 se muestra el montaje monopolar de electrodos utilizado. Las letras F, T, C, P y O representan los lóbulos frontal, temporal, central, parietal y occipital (Nomenclature, 1991), respectivamente. Los números pares (2, 4, 6, 8) se refieren a las posiciones de los electrodos en el hemisferio derecho, mientras que los números impares (1,3, 5,7) se refieren a las del hemisferio izquierdo. Por "z" (zero) se entiende a un electrodo colocado en la línea media.

$$\Sigma = \frac{1}{n-1}xx^T \quad (4.8)$$

La señal completa está formada por 22 segmentos, llamados episodios. Cada episodio contiene los 22 canales de EEG. Así que, cada canal por episodio tiene asociada una matriz de covarianza Σ . El primer valor propio de cada matriz de covarianza se denota por la letra λ . Sea λ_i el conjunto formado por cada primer valor propio de cada episodio. Para el episodio i , la fracción de variación explicada por su primer valor propio es:

$$0 \leq fr = \frac{\lambda_i}{\sum_{j=1}^{18} \lambda_j} \leq 1 \quad (4.9)$$

Esta será la variable respuesta usada para el agrupamiento.

4.3.3. Regresión Logística

El logito de fr (Spratt, 2008), se define por:

$$Z = \log\left(\frac{fr}{1-fr}\right) \quad (4.10)$$

En el ajuste lineal de Z , se usa como variable regresora X , al logaritmo natural del número de episodio (tiempo):

$$Z = \beta_0 + \beta_1 * X + \epsilon \quad (4.11)$$

Donde X es el logaritmo natural del número de episodio, y la distribución de los errores ϵ se supone normal con media 0 y varianza constante. Los parámetros β_0 , β_1 son estimados por máxima verosimilitud (Spratt, 2008). Por lo tanto, la fracción estimada de variación explicada por el primer valor propio es:

$$\widetilde{fr} = \frac{e^{\widetilde{\beta}_0 + \widetilde{\beta}_1 * \chi}}{1 + e^{\widetilde{\beta}_0 + \widetilde{\beta}_1 * \chi}} \quad (4.12)$$

Siendo $\widetilde{\beta}_0$ y $\widetilde{\beta}_1$ los estimadores de máxima verosimilitud de los parámetros. Para mostrar el desempeño de esta técnica, se evaluaron los estimadores de fracción de variación (\widetilde{fr}), de acuerdo al sexo de los pacientes, buscando diferencias en los patrones de cambio de las CFC mediante análisis discriminante logístico. Se adopta la transformación de parámetros:

$$\delta = -\frac{\widetilde{\beta}_0}{\widetilde{\beta}_1} \quad (4.13)$$

La función de separación o discriminación se calcula a partir de los centroides en cada grupo.

$$Y = \frac{\exp(Z)}{1 + \exp(Z)} \quad (4.14)$$

4.4. Desarrollo de la Aplicación para el análisis de correlaciones basado en modelos gráficos gaussianos.

4.4.1. Características de los datos

La base de datos consta de grabaciones de EEG de sujetos pediátricos con convulsiones intratables (Gotman y cols., 2006). Los sujetos fueron monitorizados durante varios días. Las grabaciones están agrupadas en 22 casos (5 varones, edad de 3 a 22 años, y 17 mujeres, edad de 1.5 a 19 años). Cada caso contiene registro de diferencias de potencial eléctrico (en μV), con frecuencia de muestreo de 256 muestras por segundo, y una resolución de 16 bits. Las tomas se realizaron mientras el sujeto estaba en estado de reposo. Se utilizó el sistema internacional 10-20 de posicionamiento de electrodos. La base de datos consta en total de 484 archivos con extensión EDF.

4.4.2. Análisis de Componentes Principales

Para cada uno de los 22 canales EEG se filtra la señal a la banda Alfa(8-13 Hz), por medio de la FFT(Fast Fourier Transform). Dada la señal Alfa Z , se calcula la matriz de covarianza

$$M = \frac{1}{n-1} Z Z^t \quad (Z^t \text{ es el transpuesto del vector } Z, \text{ y } n \text{ es su longitud}).$$

Se realiza el Análisis de Componentes Principales(PCA), sobre la matriz de covarianza M , obteniendo λ_i el conjunto de valores propios de la matriz de covarianza M , ordenados de mayor a menor. Para el canal i , la variable continua X_i , es la transformación logito de la fracción fr de variación explicada por el primer valor principal λ_1 .

$$0 \leq fr = \frac{\lambda_1}{\sum \lambda_i} \leq 1 \quad (4.15)$$

$$X_i = \left(\frac{fr}{1 - fr} \right) \quad (4.16)$$

4.4.3. Red Bayesiana

La metodología para formar la RB con el montaje es como sigue. Cada canal representa la diferencia de potencial entre dos electrodos adyacentes. Por ejemplo, el canal “FP1-F7” representa la diferencia de potencial entre el electrodo FP1 y el electrodo F7. El siguiente canal en el montaje, “F7-C3”, representa la diferencia de potencial entre F7 y C3, y así sucesivamente. Se realiza una identificación entre los nodos de la RB, y los canales del montaje: Cada canal es un nodo de la RB. Se elige a los nodos O2P4 y O1P3, como los nodos sin padre, conectando el resto a partir de ellos. Así, se obtiene una gráfica acíclica dirigida, comenzando de la zona Occipital, hacia la zona Frontal-Parietal del cerebro, ver Figura 4.6. El modelado de la RB para la CFC tiene dos propósitos. En primer lugar, realizar un modelo

gráfico sobre el conjunto de interrelaciones entre canales de EEG. En segundo lugar, hacer estimaciones cuantitativas de estas relaciones.

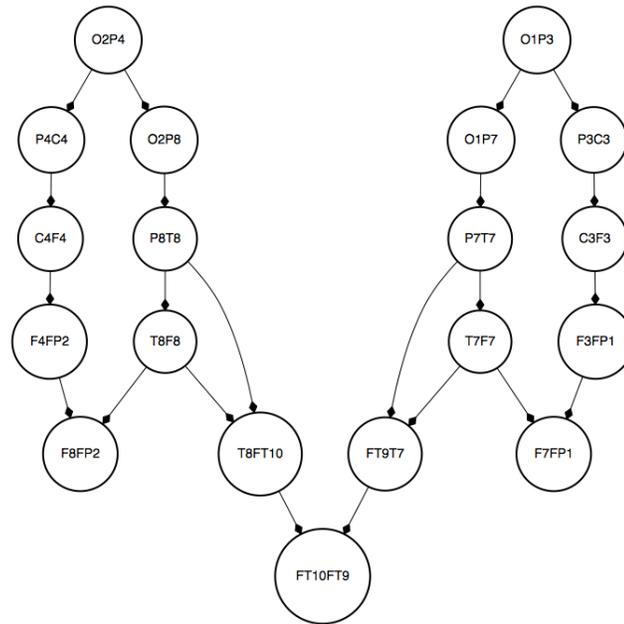


Figura 4.6: Representación de la RB para los canales EEG.

Fuente:(María Luisa Córdoba Tlaxcalteco y del Rey Torres Rodríguez, 2017)

En este trabajo se supone un modelo de dependencia lineal entre nodos. Por ejemplo, si $\prod X_i = \{ X_j, X_k \}$, la distribución de X_i , dados los padres $\prod X_i$ (ver Figura 4.6), se define por:

$$X_i = \beta_0 + \beta_1 * X_j + \beta_2 * X_k + error \quad (4.17)$$

Donde β_0 es el intercepto, y β_1 y β_2 son los parámetros que miden la influencia de X_j y X_k sobre X_i , respectivamente. El término *error* denota una variable aleatoria con distribución gaussiana. El cálculo de estimaciones sobre los parámetros de la RB se llama Aprendizaje de la RB (Li y cols., 2011). Para este fin, en el presente trabajo se utiliza el Algoritmo de Causalidad inductiva, implementado en R a través del paquete bnlearn. (Bishop, 2006).

Se comprobó que los datos contienen una distribución normal multivariada, la imagen tal presenta la distribución de cada uno de los nodos de la red bayesiana.

Al tener una distribución normal multivariada se obtuvo la desviación estándar, las matrices de covarianza y de correlación, mediante las siguientes formulas.

$$\sigma_Y^2 = \sigma^2 + \beta^T \sum \beta \quad (4.18)$$

$$\text{Cov}(X_i : Y) = \sum_j \beta_j \sum_{ij} \quad (4.19)$$

$$r_{X,Y} = \frac{\text{Cov}(XY)}{\sigma_X \sigma_Y} \quad (4.20)$$

4.5. Validación

En esta sección se presentan la validación de las herramientas computacionales, mediante su uso en conjunto de datos electroencefalograficos descritos.

4.5.1. Validación del Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG

En esta sección se muestra el uso de la nueva técnica de Análisis Longitudinal de Conglomerados (ALC) en señales Electroencefalográficas (EEG) multicanal, ordenadas a través del tiempo. El ALC hace uso del Análisis de Componentes principales, la Regresión Logística, y transformación de parámetros. Se presentan los resultados del ALC aplicado a una base de datos EEG. El comportamiento dinámico de los conglomerados se muestra en gráficas 2D y animadas.

En la Figura 4.7 se muestra, como ejemplo, los cambios en el logito de fr en un paciente hombre y uno mujer. Esta gráfica sirve para la bondad de ajuste del modelo lineal en cada paciente. En este caso existe un buen ajuste para todos los pacientes.

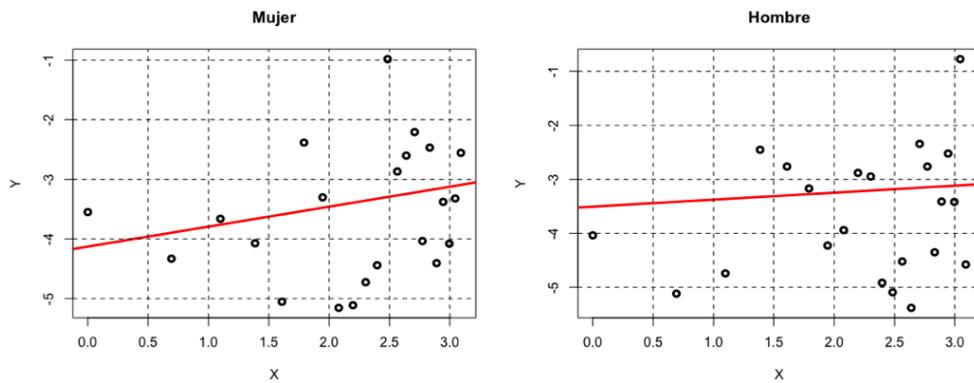


Figura 4.7: Gráfica de logitos de fr de 2 pacientes (mujer y hombre), con el logaritmo del número de toma en el eje horizontal.

Fuente:(Yoselyn Nohemi Ortega Gijón y Tlaxcalteco, 2017)

Finalmente, se muestra la gráfica (Figura 4.8) de los coeficientes lineales (δ, β_1) , por sexo.

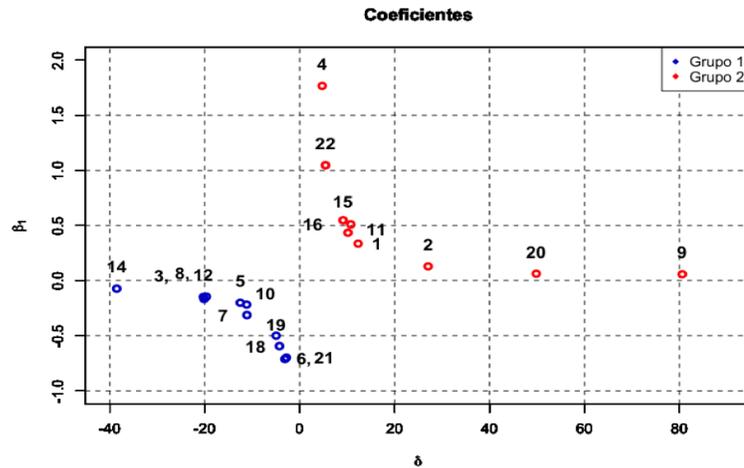


Figura 4.8: Agrupamiento de los individuos de acuerdo a la regresión logística (Coeficientes)
Fuente:(Yoselyn Nohemi Ortega Gijón y Tlaxcalteco, 2017)

Es notable el agrupamiento por grupos que muestra los puntos que representan a los pacientes. Las Figuras 4.9 y 4.10 muestran los sucesivos valores de f_r sobre la gráfica función logística.

Figura 4.9: Gráfica de la dinámica del grupo 1
Fuente:(Yoselyn Nohemi Ortega Gijón y Tlaxcalteco, 2017)

Figura 4.10: Gráfica de la dinámica del grupo 2
Fuente:(Yoselyn Nohemi Ortega Gijón y Tlaxcalteco, 2017)

4.5.2. Validación del Análisis Longitudinal y Topográfico de la Conectividad Funcional Cerebral. Detección de patrones espaciales y Temporales de Conectividad Funcional Cerebral.

Se presentan los resultados de evaluar la herramienta que detecta los patrones de cambio en tiempo y ubicación de la Conectividad Funcional en humanos, a partir de señales Electroencefalográficas (EEG) multicanal. Los elementos de la técnica son: El Análisis de Componentes Principales (ACP), la Regresión Logística (RL) y el agrupamiento. La técnica fue aplicada a una base de datos EEG, con el propósito de evaluar diferencias entre hombres y mujeres. Los resultados esperados son gráficas 2D, topograficas y dinámicas.

Las Figuras 4.11,4.12,4.13,4.14 muestran los sucesivos valores de fr sobre la gráfica función logística en una mujer y un hombre respectivamente.

Las Figuras 4.13,4.14 muestran la dinámica del parametro fr en cada uno de los canales EEG. Es destacado que los canales con diferencias más notables son FP1_F7,T7_P7, P7_O1,FP1_F3,FP2_F4 entre ambos pacientes. Esto corresponde mayormente al hemisferio izquierdo, centro de las facultades intelectuales superiores como el habla, la logica entre otras similares.

Se agrupo por sexo y las Figuras 4.16,4.17,4.18,4.19 muestran los sucesivos valores de fr sobre la gráfica función logística en el grupo de mujeres y el grupo de hombres respectivamente.

Las Figuras 4.18,4.19 muestran la dinámica del parametro fr en cada uno de los canales EEG. Es notable que por sexo las mujeres tienen mayor conectividad en los canales P4_O2,T8_P8,P8_O2, FT10_T8, entre otros. Esto corresponde al hemisferio derecho, centro de las facultades visual y espaciales, sensaciones, sentimientos y otras similares. Mientras que los hombres tiene mayor conectividad en el hemisferio izquierdo, que corresponde al centro de las facultades intelectuales, como la logica, el habla, entre otras.

Figura 4.11: Gráfica topográfica de la dinámica de fr en Mujer

Figura 4.12: Gráfica topográfica de la dinámica de fr en Hombre

Paciente1-Paciente2

Figura 4.13: Gráfica de distribución espacio temporal en una Mujer por canal.

Figura 4.14: Gráfica de distribución espacio temporal en un Hombre por canal.

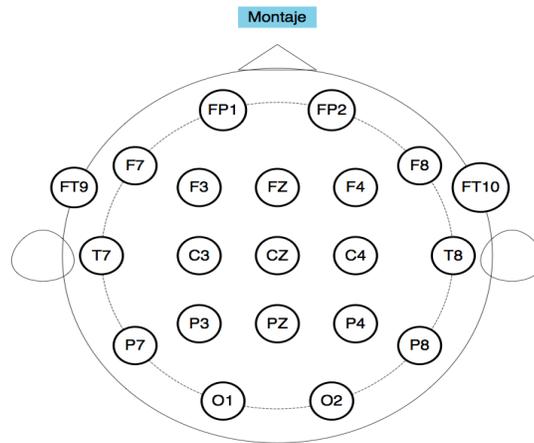


Figura 4.15: Montaje del Sistema Internacional 10-20.
Fuente:(Nomenclature, 1991)

Hombres-Mujeres

Figura 4.16: Gráfica topográfica de la dinámica de fr en Mujer.

Figura 4.17: Gráfica topográfica de la dinámica de fr en Hombre.

Figura 4.18: Gráfica de distribución espacio temporal en una Mujer por canal.

Figura 4.19: Gráfica de distribución espacio temporal en un Hombre por canal.

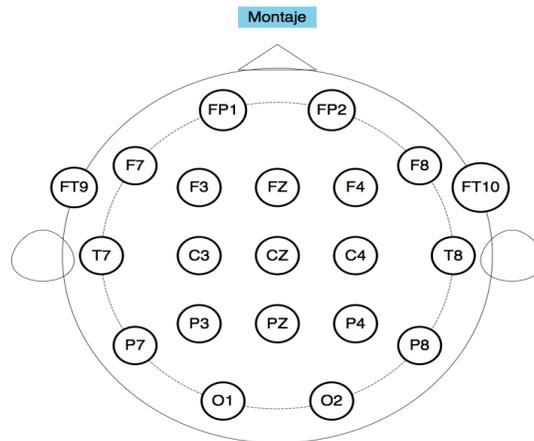


Figura 4.20: Montaje del Sistema Internacional 10-20.
Fuente:(Nomenclature, 1991)

Finalmente, se muestra la gráfica 4.21 y 4.22 de los coeficientes lineales (δ, β_1) , por paciente (Mujer-Hombre) y las Figuras 4.23 y 4.24 muestran los coeficientes lineales (δ, β_1) , por sexo. La mayor diferenciación se encuentra en los hombres, mientras que los parametros en las mujeres se encuentran más uniformes.

Paciente1-Paciente2

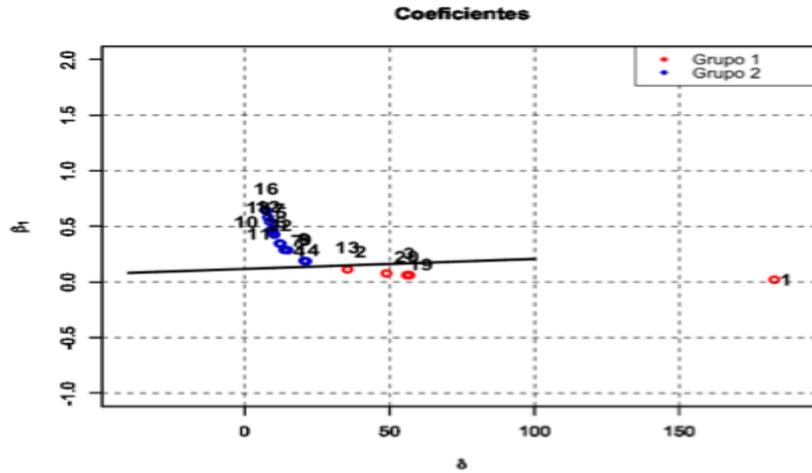


Figura 4.21: Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en una Mujer

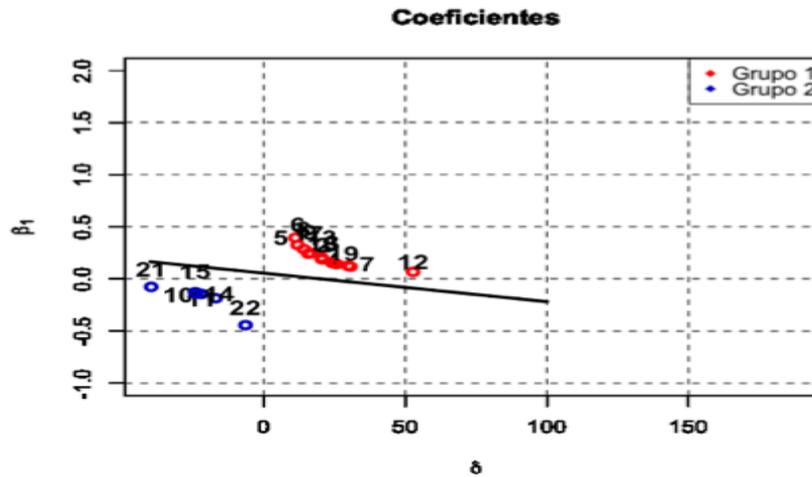


Figura 4.22: Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en un Hombre

Mujeres-Hombres

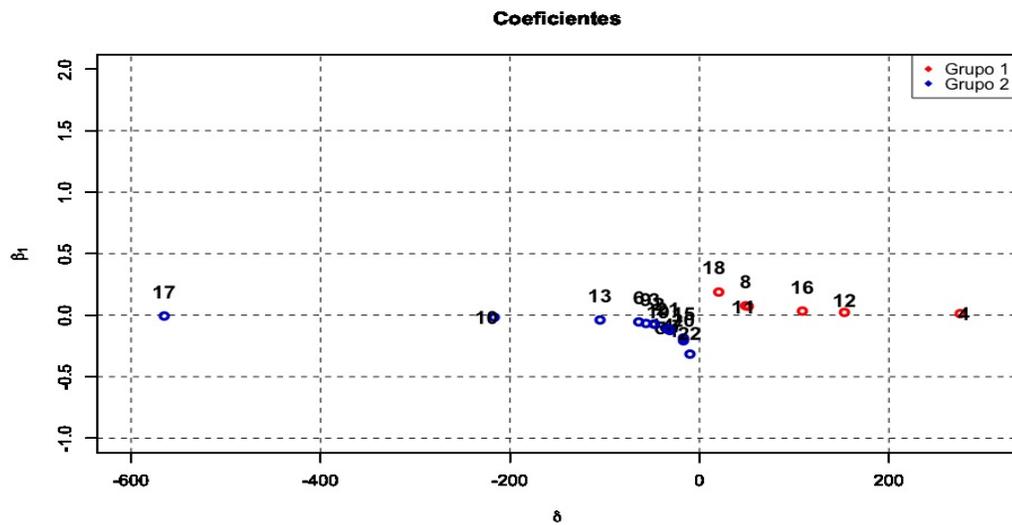


Figura 4.23: Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en Mujeres

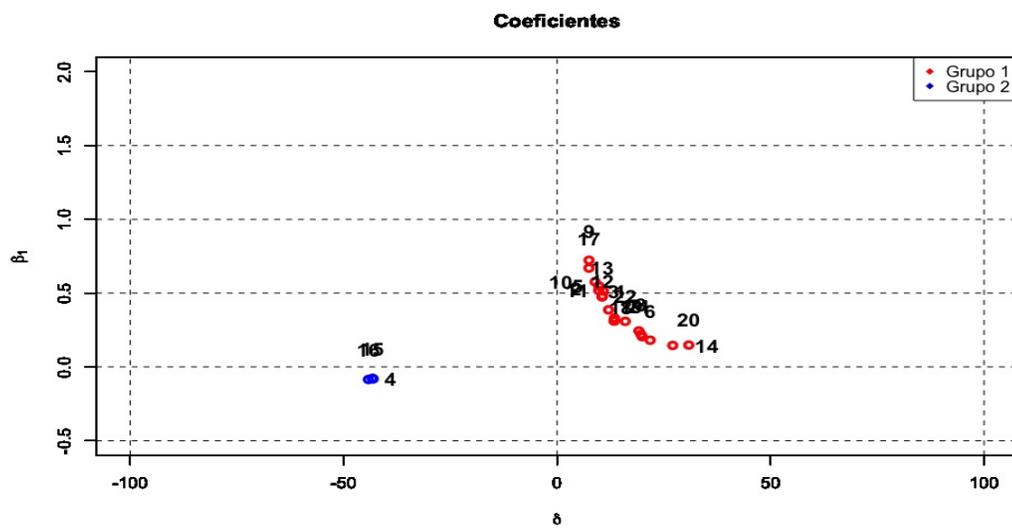


Figura 4.24: Agrupamiento de los canales de acuerdo a la regresión logística (Coeficientes) en Hombres

4.5.3. Validación del análisis de correlaciones basado en modelos graficos gaussianos.

Validación del uso del modelo de Red Bayesiana (RB) lineal con Electroencefalografía (EEG), para detectar la distribución topográfica de la CFC. Se utiliza la RB para analizar las señales EEG multicanal que pertenecen a una base de datos. En este caso, el objetivo de la RB es evaluar gráficamente la existencia de dimorfismo sexual en la distribución espacial de la CFC.

Se comprobó que los datos siguen una distribución normal multivariada, la Figura 4.28 presenta la distribución de cada uno de los nodos de la red bayesiana. Muestra evidencia de que la red tiene distribución normal.

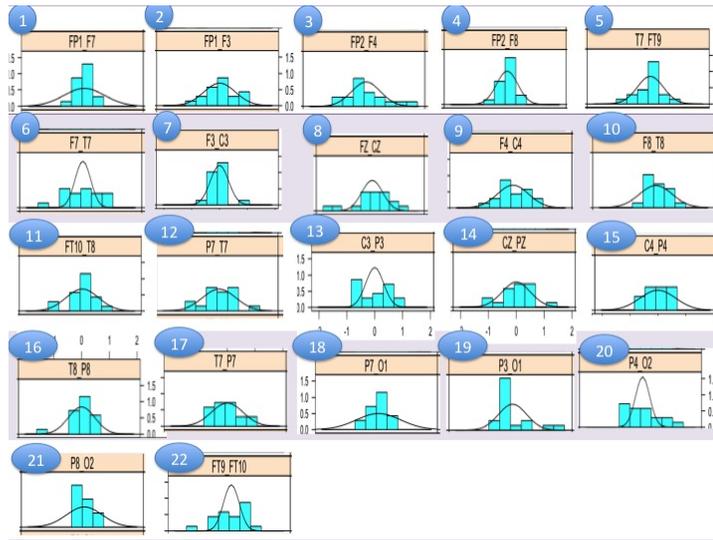


Figura 4.27: Distribución Normal por canal

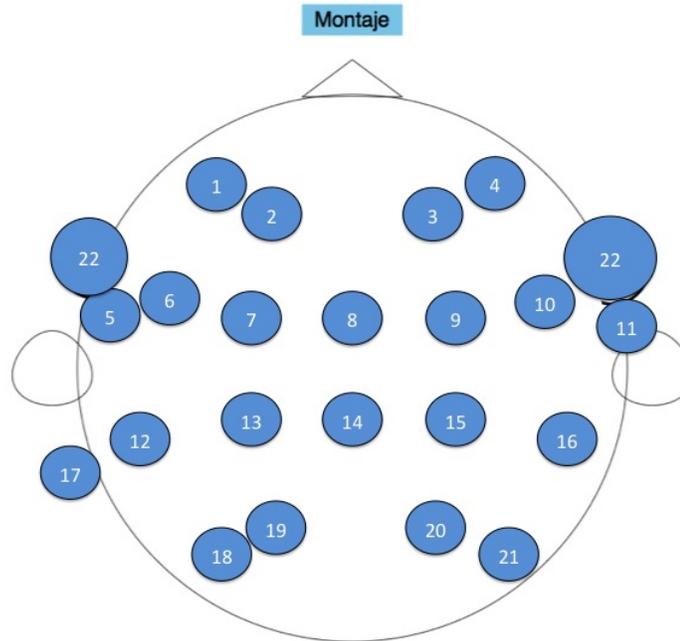


Figura 4.28: Distribución de residuales de la red bayesiana.

La Figura 4.29 muestra la semaforización de los valores de correlación, siendo los rojos los más altos. En este caso los resultados muestran las diferencias de conectividad entre mujeres y hombre. A mayor cantidad de rojos, mayor conectividad.

Estimaciones de coeficientes de correlación (Mujeres)

	C3_P3	C4_P4	F3_C3	F4_C4	F7_T7	F8_T8	FP1_F3	FP1_F7	FP2_F4	FP2_F8	FT9_FT10	FT10_T8	P3_O1	P4_O2	P7_O1	P7_T7	P8_O2	T7_FT9	T8_P8
C3_P3	inf		0.2076855				0.124506	0.0464298					0.2723189						
C4_P4		inf	0.5601945						0.2029259	0.0948077				0.3653207					
F3_C3	0.2076855		inf				0.6443406	0.2401309					0.0817941						
F4_C4		0.5601945		inf					0.2191893	0.1048092				0.2598606					
F7_T7					inf				0.2623377		0.0701069		0.0988179	0.0694819	0.3393424		0.1313555		
F8_T8						inf				0.3609729	-0.000112	0.0885281		0.094203			0.1876541	0.2763186	
FP1_F3	0.124506		0.6443406				inf	0.1164368					0.0522219						
FP1_F7	0.0464298		0.2401309		0.2623377		0.1164368	inf					0.0571906	0.0341942	0.1265252				
FP2_F4		0.2029259		0.2191893					inf	0.3478636				0.0778743					
FP2_F8		0.0948077		0.1048092		0.3609729			0.3478636	inf				0.0788991		0.0909934		0.1267091	
FT9_FT10					0.0701069	-0.000112					inf	0.0114347	0.0802989	0.0066308	0.0804299	0.2037671	0.0161389	0.1545287	0.0396231
FT10_T8						0.0885281					0.0114347	inf		0.1278017			0.240182	0.4507409	
P3_O1	0.2723189		0.0817941		0.0988179		0.0522219	0.0571906			0.0802989		inf	0.5185728	0.2129923		0.1600257		
P4_O2		0.3653207		0.2598606		0.094203			0.0778743	0.0788991	0.0066308	0.1278017		inf			0.4279554	0.2273703	
P7_O1					0.0694819			0.0341942			0.0804299		0.5185728	inf	0.1897101		0.1434798		
P7_T7					0.3393424			0.1265252			0.2037671		0.2129923	0.1897101	inf		0.6860937		
P8_O2						0.1876541				0.0909934	0.0161389	0.240182		0.4279554		inf		0.4710394	
T7_FT9					0.1313555						0.1545287		0.1600257	0.1434798	0.6860937		inf		
T8_P8						0.2763186				0.1267091	0.0396231	0.4507409		0.2273703		0.4710394		inf	

Estimaciones de coeficientes de correlación (Hombres)

	C3_P3	C4_P4	F3_C3	F4_C4	F7_T7	F8_T8	FP1_F3	FP1_F7	FP2_F4	FP2_F8	FT9_FT10	FT10_T8	P3_O1	P4_O2	P7_O1	P7_T7	P8_O2	T7_FT9	T8_P8
C3_P3	inf		0.0696754				0.0795977	0.0755025					0.1556525						
C4_P4		inf	0.3849062						0.0943544	0.0248336				0.3107725					
F3_C3	0.0696754		inf				0.6899615	0.2336807					0.0541308						
F4_C4		0.3849062		inf					0.3055632	0.1169195				0.1280207					
F7_T7					inf				0.2229491		0.064318		0.0354706		0.1059922	0.1936947		0.1291126	
F8_T8						inf				0.4410138	0.0554135	0.1920111		0.0355682			0.0942283	0.1375858	
FP1_F3	0.0795977		0.6899615				inf	0.0763494					0.0391195						
FP1_F7	0.0755025		0.2336807		0.2229491		0.0763494	inf					0.0233163		0.0442276	0.0833514			
FP2_F4		0.0943544		0.3055632					inf	0.4186748				0.0435293					
FP2_F8		0.0248336		0.1169195		0.4410138			0.4186748	inf				0.0298028			0.0374797	0.0677829	
FT9_FT10					0.064318	0.0554135					inf	0.108178	0.0153174	0.0019916	0.0200554	0.1319192	0.0517979	0.1115981	0.0994655
FT10_T8						0.1920111					0.108178	inf		0.0867852			0.1913132	0.4687923	
P3_O1	0.1556525		0.0541308		0.0354706		0.0391195	0.0233163				0.0153174	inf		0.3198454	0.102503	0.0513159		
P4_O2		0.3107725		0.1280207		0.0355682			0.0435293	0.0298028	0.0019916	0.0867852		inf			0.552818	0.2006422	
P7_O1					0.1059922			0.0442276			0.0200554		0.3198454		inf	0.2856235	0.0632651		
P7_T7					0.1936947			0.0833514			0.1319192		0.102503		0.2856235	inf	0.5629515		
P8_O2						0.0942283			0.0374797	0.0517979	0.1913132		0.552818		inf			0.3782508	
T7_FT9					0.1291126						0.1115981		0.0513159		0.0632651	0.5629515		inf	
T8_P8						0.1375858				0.0677829	0.0994655	0.4687923		0.2006422		0.3782508		inf	

Figura 4.29: Estimaciones de coeficientes de correlación

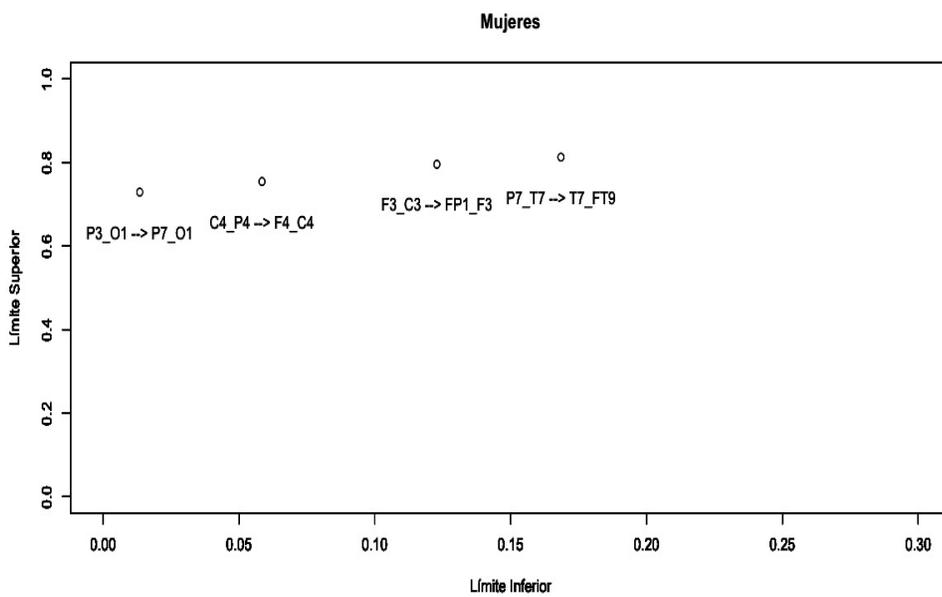


Figura 4.30: Limites Mujeres

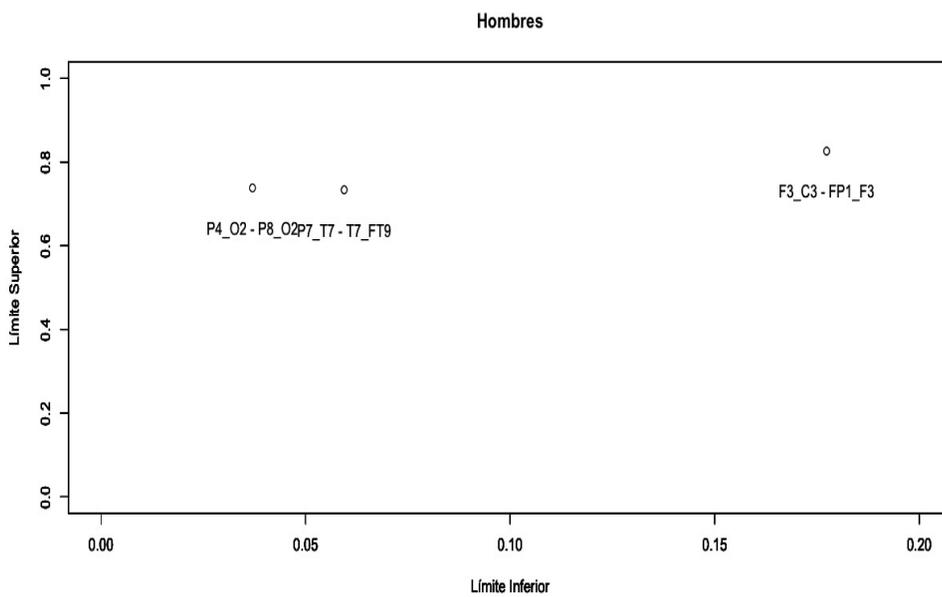


Figura 4.31: Limites Hombres

Capítulo 5

Análisis de resultados

5.1. Análisis Longitudinal de Conglomerados para Conectividad Funcional Cerebral, a partir de EEG

En este trabajo hay una propuesta de agrupamiento de patrones de cambio de las CFC a lo largo del tiempo. Los resultados obtenidos son evidencia de un patrón de cambio en el tiempo, diferente para cada grupo (Figuras 4.9-4.10) representadas por graficas animadas. Este análisis fue realizado en señales EEG de personas diagnosticadas con Epilepsia, y se pretende realizar a futuro en personas sanas para comparar con los resultados obtenidos en este trabajo.

5.2. Análisis Longitudinal y Topográfico de la Conectividad Funcional Cerebral. Detección de patrones espaciales y Temporales de Conectividad Funcional Cerebral

Es notable que por sexo las mujeres tienen mayor conectividad en los canales P4 O2, T8 P8, P8 O2, FT10 T8, entre otros. Esto corresponde al hemisferio derecho, centro de las facultades visual y espaciales, sensaciones, sentimientos y otras similares. Mientras que los hombres tiene mayor conectividad en el hemisferio izquierdo, que corresponde al centro de las facultades intelectuales, como la logica, el habla, entre otras.

5.3. Análisis de correlaciones basado en modelos gráficos gaussianos.

Hay evidencia notable de dimorfismo sexual. En hombres, los valores más grandes se encuentran en las zonas Frontal, y Temporal. Las zonas con menor norma en Hombres son: la Occipital, Central y Parietal. En mujeres, con menor diferencia entre los valores máximo

y mínimo, en comparación con hombres. Los mayores valores de la norma se encuentra, en mujeres, en las zonas Frontal, Temporal, Parietal y Occipital.

Los resultados son que la conectividad de la RB en Mujeres es mayor que en Hombres de acuerdo a la obtención de las matrices de correlación semaforizadas.

5.4. Comparación con el Estado del Arte.

Se presenta la comparación del estado del arte con lo presentado en este trabajo, es importante mencionar que dichos trabajos no son estrictamente afines al objetivo de esta investigación. La tabla 5.1 y 5.2 muestra la comparativa realizada con respecto a la literatura reportada.

Autor	Métodos	Resultados	Comparación
Schaltenbrand, et. al.(1993)	Usaron reglas de inferencia para clasificar las etapas del sueño mediante EEG.	80 % de precisión	En este trabajo se usaron métodos cuantitativos.
Bugli y Lambert (2007)	Usa PCA e ICA para detectar patrones en la señal electroencefalográfica. Todo programado en Matlab. No busca patrones de cambio en el tiempo.	ICA es muy prometedor para analizar las señales biomédicas multidimensionales y mucho más eficiente que el PCA	Uso de análisis longitudinal jerárquico para detectar patrones de cambio en el tiempo. Programación de algoritmos en entorno de programación libre para hacer más directa la transferencia de la tecnología al usuario final.
Petr Berka, et. al. (2009)	Uso del análisis de Componentes Principales de la Matriz de Covarianza para la segmentación de los estados en pacientes comatosos	99 % de exactitud	Se usa el análisis de Componentes Principales de la Matriz de Covarianza, tanto de la señal de cada canal como del arreglo multicanal de todo el montaje electroencefalográfico. Se introduce el logito de la variación explicada por el primer componente principal para detectar patrones de cambio en el tiempo.

Tabla 5.1: Comparativa del Estado del Arte de algoritmos.

Autor	Métodos	Resultados	Comparación
Subasi y Gursoy (2010)	La finalidad de este trabajo es predecir eventos convulsivos. Para abordar el problema de señal no estacionaria se utiliza la transformada de Wavelet además de PCA.	LDA 100% PCA 98.75%	Para abordar la señal no estacionaria se usa la Short time Fast Fourier, la reducción de la dimensionalidad se obtiene mediante PCA y el logito de la variación explicada por el primer componente principal.
Kavita Mahajan (2011)	Usan transformada wavelet y análisis de componentes principales junto con redes neuronales para estimar la distribución de los coeficientes wavelets y predecir normal y anormal en los datos de pacientes epilépticos	PCA 93,63%. ICA 96,75%	Se utiliza la Short time Fast Fourier Transform para contemplar el carácter no estacionario de la señal.
Yoselyn Ortega (2018)	Se utiliza la Short time Fast Fourier Transform para contemplar el carácter no estacionario de la señal. Se usa el análisis de Componentes Principales de la Matriz de Covarianza, tanto de la señal de cada canal como del arreglo multicanal de todo el montaje electroencefalografico. Se introduce el logito de la variación explicada por el primer componente principal para detectar patrones de cambio en el tiempo.	Es notable que por sexo las mujeres tienen mayor conectividad en los canales P4 O2,T8P8,P8 O2, FT10 T8, entre otros. Mientras que los hombres tiene mayor conectividad en el hemisferio izquierdo, que corresponde al centro de las facultades intelectuales, como la lógica, el habla, entre otras	

Tabla 5.2: Comparativa del Estado del Arte de algoritmos.

Capítulo 6

Conclusiones y Trabajo a futuro

6.1. Conclusiones

En este trabajo hay una propuesta de aplicaciones informáticas de carácter libre para diferentes análisis cuantitativos de electroencefalografía. Los resultados obtenidos son representaciones gráficas interpretables del fenómeno neurológico, que es un sistema de señales eléctricas cerebrales. Las premisas fundamentales para estas elaboraciones son que las señales biofísicas son no estacionarias y por lo tanto no susceptibles de análisis bajo los supuestos tradicionales de los modelos de análisis de señales. La otra premisa es que la electroencefalografía multicanal, siendo generada por un mismo individuo, consta de canales de señales con estructura de correlación compleja, que es difícil de analizar por el carácter no estacionario mencionado anteriormente. Un aspecto que es destacable es el énfasis en la detección de patrones de cambio en el tiempo, lo que permitiría describir los procesos de maduración normales y anormales en individuos infantiles. Por último, también se destaca la programación en entorno libre para potenciar su interacción con otros sistemas y dispositivos.

6.2. Trabajo a futuro

- Como trabajo a futuro se menciona la necesidad de validación por especialistas en salud mental, dado que todas las pruebas se realizaron con datos de individuos en estado basal, siendo de interés su evaluación cuando los individuos realizan actividades físicas o cognitivas.
- Otro aspecto a considerar es descubrir los límites de estabilidad de las herramientas, analizando en efecto de ruidos y artefactos, el costo computacional de los algoritmos, y su puesta en marcha en un servidor web de aplicaciones.

Referencias

- Alba, J., y Cid, J. (2006). Reconocimiento de patrones. *Carlos III, Madrid, Universidad Carlos III, Madrid*. Retrived from <http://www.gts.tsc.uvigo.es/pi/Reconocimiento.pdf>.
- Anderson, C. J. (2005). Applied longitudinal data analysis: Modeling change and event occurrence.
- Berka, P. (2009). *Data mining and medical knowledge management: Cases and applications: Cases and applications*. IGI Global.
- Bishop, C. M. (2006). Pattern recognition and machine learning, 2006.
- Bruce, E. N. (2001). *Biomedical signal processing and signal modeling*. Wiley New York:.
- Bugli, C., y Lambert, P. (2007). Comparison between principal component analysis and independent component analysis in electroencephalograms modelling. *Biometrical Journal*, 49(2), 312–327.
- Carpenter, G. A., y Grossberg, S. (1990, abril 3). System for self-organization of stable category recognition codes for analog input patterns. (US Patent 4,914,708)
- De la Garza García, J., Morales Serrano, B., y González Cavazos, B. (2013). *Análisis estadístico multivariante*. McGraw Hill.
- Flexer, A. (2000). Data mining and electroencephalography. *Statistical Methods in Medical Research*, 9(4), 395–413.
- Freeman, W., y Quiroga, R. Q. (2012). *Imaging brain function with eeg: advanced temporal and spatial analysis of electroencephalographic signals*. Springer Science & Business Media.
- Gotman, J., Kobayashi, E., Bagshaw, A. P., Bénar, C.-G., y Dubeau, F. (2006). Combining eeg and fmri: a multimodal tool for epilepsy research. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 23(6), 906–920.
- Kubat, M., Pfurtscheller, G., y Flotzinger, D. (1994). Ai-based approach to automatic sleep classification. *Biological Cybernetics*, 70(5), 443–448.
- Lampe, J., y Voss, H. (2013). Large-scale tikhonov regularization of total least squares. *Journal of Computational and Applied Mathematics*, 238, 95–108.
- Lee, H., y Choi, S. (2003). Pca+ hmm+ svm for eeg pattern classification. , 1, 541–544.
- Li, R., Chen, K., Fleisher, A. S., Reiman, E. M., Yao, L., y Wu, X. (2011). Large-scale

- directional connections among multi resting-state neural networks in human brain: a functional mri and bayesian network modeling study. *Neuroimage*, 56(3), 1035–1042.
- Liu, X., Wang, G., Gao, J., y Gao, Q. (2017). A quantitative analysis for eeg signals based on modified permutation-entropy. *IRBM*.
- Mahajan, K., Vargantwar, M., y Rajput, S. M. (2011). Classification of eeg using pca, ica and neural network. *International Journal of Engineering and Advanced Technology*, 1(1), 80–83.
- María Luisa Córdoba Tlaxcalteco, Y. N. O. G., y del Rey Torres Rodríguez, A. (2017). Red bayesiana para detectar patrones de distribución espacial de la conectividad funcional cerebral con electroencefalografía. *Ingeniantes*, 1(2), 53-58.
- Nomenclature, S. E. P. (1991). American electroencephalographic society guidelines for. *Journal of clinical Neurophysiology*, 8(2), 200–202.
- Roberts, S., y Tarassenko, L. (1992). New method of automated sleep quantification. *Medical and Biological Engineering and Computing*, 30(5), 509–517.
- Schaltenbrand, N., Lengelle, R., y Macher, J.-P. (1993). Neural network model: application to automatic analysis of human sleep. *Computers and Biomedical Research*, 26(2), 157–171.
- Shi, L.-C., Duan, R.-N., y Lu, B.-L. (2013). A robust principal component analysis algorithm for eeg-based vigilance estimation. En *Engineering in medicine and biology society (embc), 2013 35th annual international conference of the ieee* (pp. 6623–6626).
- Smith, L. I. (2002). A tutorial on principal components analysis.
- Sprott, D. A. (2008). *Statistical inference in science*. Springer Science & Business Media.
- Subasi, A., y Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. *Expert Systems with Applications*, 37(12), 8659–8666.
- Yoselyn Nohemi Ortega Gijón, A. d. R. T. R., y Tlaxcalteco, M. L. C. (2017). Análisis discriminante longitudinal de la conectividad funcional cerebral usando componentes principales y regresión logística. *CIM Orizaba*, 2(5), 704-709.

Anexos

6.3. Extracción de datos .edf

```
1 tryCatch({
2   library(edf)
3   df<-read.edf('C:/Users/maria/Desktop/pacientesEpilepsia/paciente1/Toma1/chb01_01.edf')
```

La señal *chb01_01.edf* corresponde al paciente 01 y toma 01.

6.3.1. Gráfica de las señales

El archivo contiene las 22 señales de los canales llamados: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, y T8-P8. A continuación se muestra el código para generar las gráficas de los 22 canales para el paciente *chb01_01.edf*.

```
1 graf1<-plot(1:921600,df$signal$FP1_F7$data[1:921600],type="l",xlab="",ylab="FP1_F7")
2 graf2<-plot(1:921600,df$signal$F7_T7$data[1:921600],type="l",xlab="",ylab="F7_T7")
3 graf3<-plot(1:921600,df$signal$T7_P7$data[1:921600],type="l",xlab="",ylab="T7_P7")
4 graf4<-plot(1:921600,df$signal$P7_O1$data[1:921600],type="l",xlab="",ylab="P7_O1")
5 graf5<-plot(1:921600,df$signal$FP1_F3$data[1:921600],type="l",xlab="",ylab="FP1_F3")
6 graf6<-plot(1:921600,df$signal$F3_C3$data[1:921600],type="l",xlab="",ylab="F3_C3")
7 graf7<-plot(1:921600,df$signal$C3_P3$data[1:921600],type="l",xlab="",ylab="C3_P3")
8 graf8<-plot(1:921600,df$signal$P3_O1$data[1:921600],type="l",xlab="",ylab="P3_O1")
9 graf9<-plot(1:921600,df$signal$FP2_F4$data[1:921600],type="l",xlab="",ylab="FP2_F4")
10 graf10<-plot(1:921600,df$signal$F4_C4$data[1:921600],type="l",xlab="",ylab="F4_C4")
11 graf11<-plot(1:921600,df$signal$C4_P4$data[1:921600],type="l",xlab="",ylab="C4_P4")
12 graf12<-plot(1:921600,df$signal$P4_O2$data[1:921600],type="l",xlab="",ylab="P4_O2")
13 graf13<-plot(1:921600,df$signal$FP2_F8$data[1:921600],type="l",xlab="",ylab="FP2_F8")
14 graf14<-plot(1:921600,df$signal$F8_T8$data[1:921600],type="l",xlab="",ylab="F8_T8")
15 graf15<-plot(1:921600,df$signal$T8_P8$data[1:921600],type="l",xlab="",ylab="T8_P8")
16 graf16<-plot(1:921600,df$signal$P8_O2$data[1:921600],type="l",xlab="",ylab="P8_O2")
17 graf17<-plot(1:921600,df$signal$FZ_CZ$data[1:921600],type="l",xlab="",ylab="FZ_CZ")
18 graf18<-plot(1:921600,df$signal$CZ_PZ$data[1:921600],type="l",xlab="",ylab="CZ_PZ")
19 graf19<-plot(1:921600,df$signal$P7_T7$data[1:921600],type="l",xlab="",ylab="P7_T7")
20 graf20<-plot(1:921600,df$signal$T7_FT9$data[1:921600],type="l",xlab="",ylab="T7_FT9")
21 graf21<-plot(1:921600,df$signal$FT9_FT10$data[1:921600],type="l",xlab="",ylab="FT9_FT10")
22 graf22<-plot(1:921600,df$signal$FT10_T8$data[1:921600],type="l",xlab="",ylab="FT10_T8")
23 }, error = function(e) {
24   message(e)
25   return(NULL)
26 }
```

6.3.2. Ejemplo para el canal FP1-F7

```
1 tiempo<-df$signal$FP1\_F7$t
2 data1<-df$signal$FP1\_F7$data
```

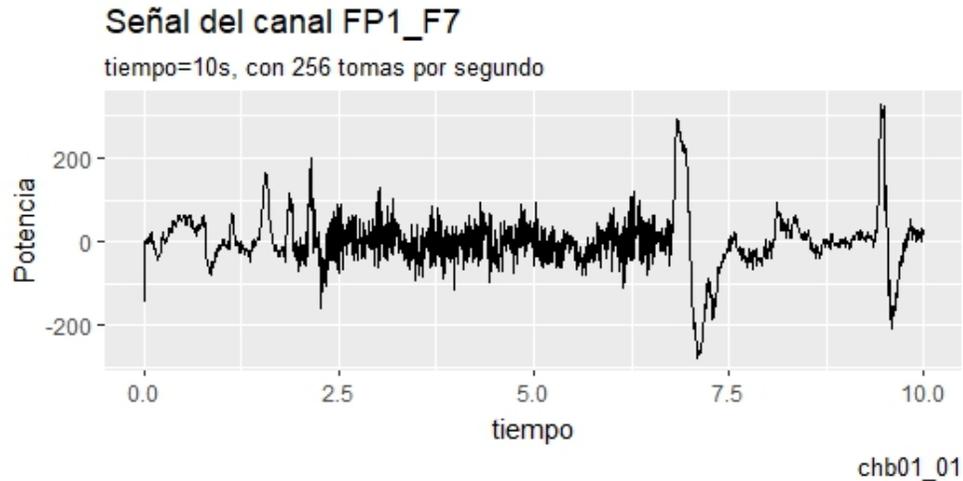


Figura 6.1: Gráfica del canal FP1-F7

6.4. Extracción de Bandas

En MATLAB: La extracción de bandas, en todos los casos se obtiene a través de la función "waveletFunction", con la ondícula Daubechies 8.

```

1 waveletFunction = 'db8';
2 [C,L] = wavedec(dataFP1_F7_data.x,8,waveletFunction);
3 cD1 = detcoef(C,L,1); %NOISY
4 cD2 = detcoef(C,L,2); %NOISY
5 cD3 = detcoef(C,L,3); %NOISY
6 cD4 = detcoef(C,L,4); %NOISY
7 cD5 = detcoef(C,L,5); %GAMA
8 cD6 = detcoef(C,L,6); %BETA
9 cD7 = detcoef(C,L,7); %ALPHA
10 cD8 = detcoef(C,L,8); %THETA
11 cA8 = appcoef(C,L,waveletFunction,8); %DELTA
12 %%%Calculation the Details Vectors
13 D1 = wrcoef('d',C,L,waveletFunction,1); %NOISY
14 D2 = wrcoef('d',C,L,waveletFunction,2); %NOISY
15 D3 = wrcoef('d',C,L,waveletFunction,3); %NOISY
16 D4 = wrcoef('d',C,L,waveletFunction,4); %NOISY
17 D5 = wrcoef('d',C,L,waveletFunction,5); %GAMMA
18 D6 = wrcoef('d',C,L,waveletFunction,6); %BETA
19 D7 = wrcoef('d',C,L,waveletFunction,7); %ALPHA
20 D8 = wrcoef('d',C,L,waveletFunction,8); %THETA
21 A8 = wrcoef('a',C,L,waveletFunction,8); %DELTA
22 GammaFP1_F7 = D5;
23 ThetaFP1_F7 = D8;
24 AlphaFP1_F7 = D7;
25 BetaFP1_F7 = D6;
26 DeltaFP1_F7 = A8;
27 figure;
28 plot(1:1:length(dataFP1_F7_data.x), dataFP1_F7_data.x); title('FP1-F7');
29 figure;
30 plot(1:1:length(GammaFP1_F7),GammaFP1_F7); title('GAMMA FP1-F7');
31 figure;

```

```

32 plot(1:1:length(BetaFP1_F7), BetaFP1_F7); title('BETA FP1-F7');
33 figure;
34 plot(1:1:length(AlphaFP1_F7), AlphaFP1_F7); title('ALPHA FP1-F7');
35 figure;
36 plot(1:1:length(ThetaFP1_F7), ThetaFP1_F7); title('THETA FP1-F7');
37 figure;
38 plot(1:1:length(DeltaFP1_F7), DeltaFP1_F7); title('DELTA FP1-F7');
39 figure; subplot(5,1,1);
40 plot(1:1:length(GammaFP1_F7), GammaFP1_F7); title('GAMMA FP1-F7');
41 subplot(5,1,2);
42 plot(1:1:length(BetaFP1_F7), BetaFP1_F7); title('BETA FP1-F7');
43 subplot(5,1,3);
44 plot(1:1:length(AlphaFP1_F7), AlphaFP1_F7); title('ALPHA FP1-F7');
45 subplot(5,1,4);
46 plot(1:1:length(ThetaFP1_F7), ThetaFP1_F7); title('THETA FP1-F7');
47 subplot(5,1,5);
48 plot(1:1:length(DeltaFP1_F7), DeltaFP1_F7); title('DELTA FP1-F7');

```

Al extraer las bandas, se obtienen cinco archivos en formato .mat: DeltaFP1_F7, ThetaFP1_F7, AlphaFP1_F7, BetaFP1_F7 y ThetaFP1_F7 (Figura.6.2).

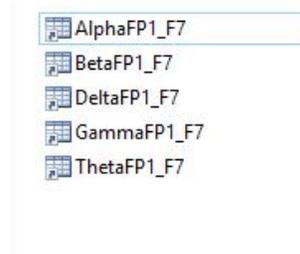


Figura 6.2: Bandas del canal FP1-F7

6.4.1. Gráficas de las bandas

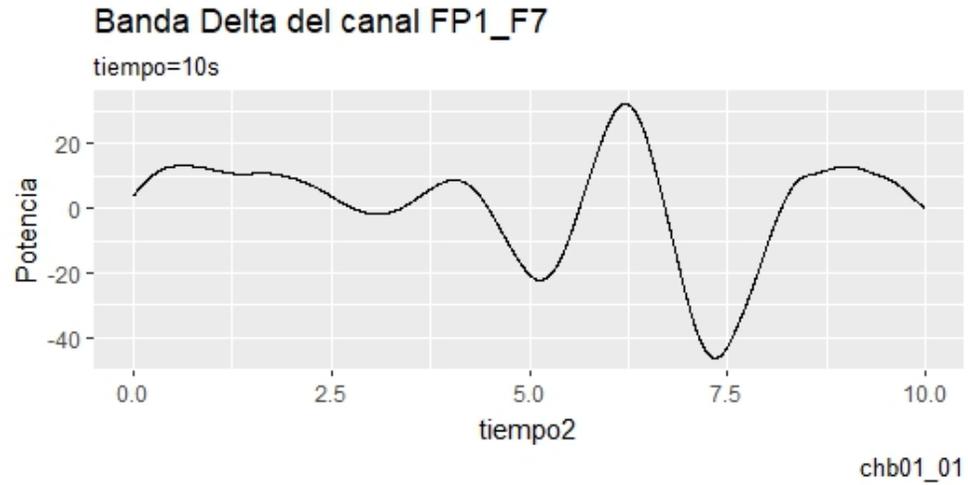


Figura 6.3: Banda Delta del canal FP1-F7

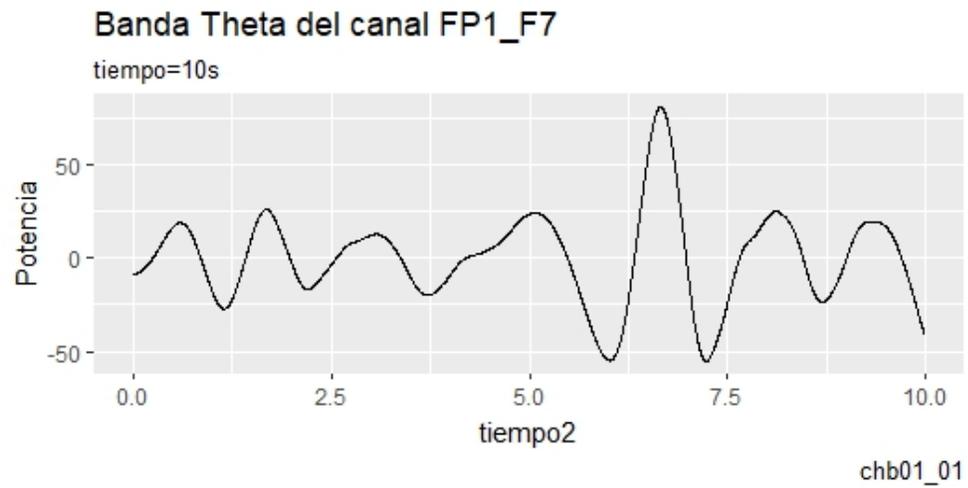


Figura 6.4: Banda Theta del canal FP1-F7

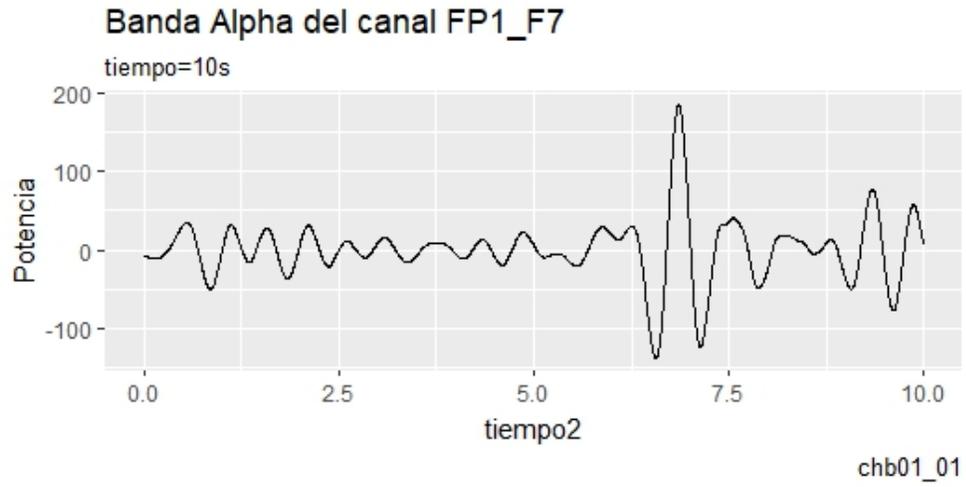


Figura 6.5: Banda Alpha del canal FP1-F7

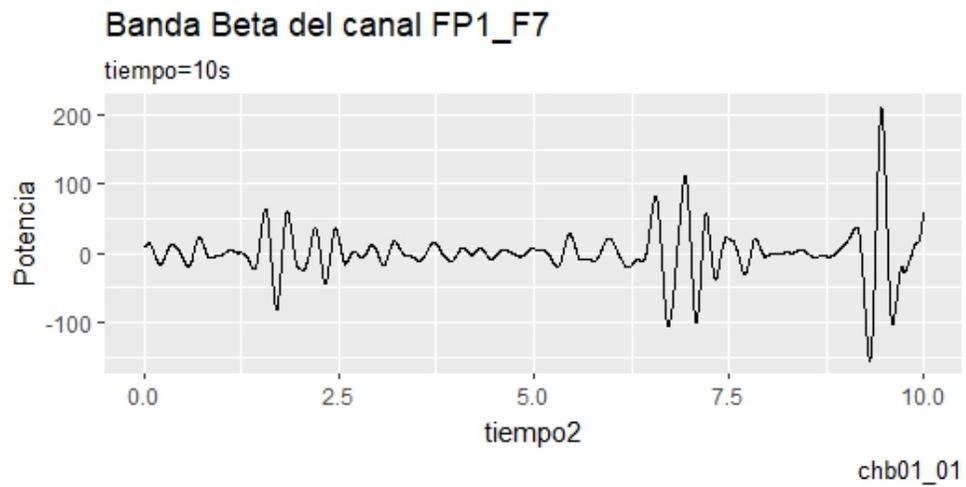


Figura 6.6: Banda Beta del canal FP1-F7

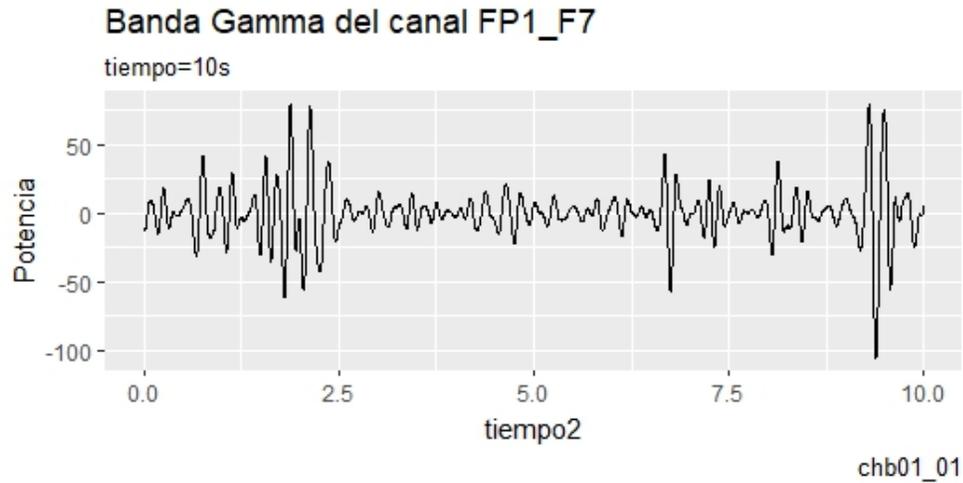


Figura 6.7: Banda Gamma del canal FP1-F7

6.5. Modelos gráficos

A partir de este momento se trabaja con la banda Alpha de cada uno de los pacientes, las 22 tomas y los 22 canales mencionados anteriormente.

6.5.1. Red de Bayes

Como primer paso, utilizamos la librería *DiagrammeR* para poder hacer la red de Bayes:

```
1 library (DiagrammeR)
```

Definimos las conexiones entre los canales para crear la red de Bayes:

```
1 grViz("digraph boxes_and_circles {
2   node [shape = circle
3     fontname= Helvetica
4     penwidth= 2.0]
5   F7FP1; T7F7; P7T7; O1P7; F3FP1; C3F3; P3C3;
6   O1P3; F4FP2; C4F4; P4C4; O2P4; F8FP2; T8F8;
7   P8T8; O2P8; T8FT10; FT10FT9; FT9T7;
8   edge [arrowhead = diamond]
9   O1P3->P3C3
10  C3F3->F3FP1
11  O1P3->O1P7
12  O1P7->P7T7
13  P7T7->T7F7
14  T7F7->F7FP1
15  F3FP1->F7FP1
16  P3C3->C3F3
17  O2P8->P8T8
18  O2P4->O2P8
19  O2P4->P4C4
20  P4C4->C4F4
21  C4F4->F4FP2
```

```

22| F4FP2->F8FP2
23| T8F8->F8FP2
24| P8T8->T8F8
25| P7T7->FT9T7
26| T7F7->FT9T7
27| P8T8->T8FT10
28| T8F8->T8FT10
29| FT9T7->FT10FT9
30| T8FT10->FT10FT9
31| edge[arrowhead = diamond]
32| graph[nodsep = 0.7]
33| }”)

```

Se crea una matriz para almacenar los valores propios por paciente y por canal.

```
1 matrizValorespropios<-matrix(0,19,22)
```

Se define la función para aplicar el PCA:

```

1 obtenerPCA<-function(datosE){
2   n<-length(datosE)
3   promedio<-rep(mean(datosE),1000)
4   datosM<-((datosE-promedio)%*%t(datosE-promedio))
5   datosM<-(1/(n-1))*datosM
6   eigenvalues<-eigen(datosM)$values
7   eigenVectors<-eigen(datosM)$vectors
8   results <- list()
9   results$first <- eigenvalues[1]
10  results$second <-eigenVectors[,1]
11  return(results)
12  }
13 valoresPropios<-rep(0,22)
14 df<-read.edf('chb01_01.edf')

```

Para cada uno de los pacientes, se lee el archivo .mat que contiene los datos del paciente, número de toma y la banda correspondiente:

paciente1/toma1/alphas1_1/AlphaFP1_F7.mat

Los resultados obtenidos se guardan en un archivo “matrizValorespropios.csv” que contiene todos los datos de la matriz “matrizValorespropios”.

```

1 #1
2 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFP1_
3 F7.mat')
4 datos<-readMat(paste(direc1, '/AlphaFP1_F7.mat', sep=""))
5 datos<-datos$Alpha[1:1000]
6 resultados<-obtenerPCA(datos)
7 valoresPropios[1]<- resultados$first
8 #2
9 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaF7_
10 T7.mat')
11 datos<-readMat(paste(direc1, '/AlphaF7_T7.mat', sep=""))
12 resultados<-obtenerPCA(datos$Alpha[1:1000])
13 valoresPropios[2]<- resultados$first
14 #3
15 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaT7_
16 P7.mat')
17 datos<-readMat(paste(direc1, '/AlphaT7_P7.mat', sep=""))
18 resultados<-obtenerPCA(datos$Alpha[1:1000])
19 valoresPropios[3]<- resultados$first

```

```

17 #4
18 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaP7_
    O1.mat')
19 datos<-readMat(paste(direc1, '/AlphaP7_O1.mat', sep=""))
20 resultados<-obtenerPCA(datos$Alpha[1:1000])
21 valoresPropios[4]<- resultados$first
22 #5
23 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFP1_
    F3.mat')
24 datos<-readMat(paste(direc1, '/AlphaFP1_F3.mat', sep=""))
25 resultados<-obtenerPCA(datos$Alpha[1:1000])
26 valoresPropios[5]<- resultados$first
27 #6
28 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaF3_
    C3.mat')
29 datos<-readMat(paste(direc1, '/AlphaF3_C3.mat', sep=""))
30 resultados<-obtenerPCA(datos$Alpha[1:1000])
31 valoresPropios[6]<- resultados$first
32 #7
33 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaC3_
    P3.mat')
34 datos<-readMat(paste(direc1, '/AlphaC3_P3.mat', sep=""))
35 resultados<-obtenerPCA(datos$Alpha[1:1000])
36 valoresPropios[7]<- resultados$first
37 #8
38 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaP3_
    O1.mat')
39 datos<-readMat(paste(direc1, '/AlphaP3_O1.mat', sep=""))
40 resultados<-obtenerPCA(datos$Alpha[1:1000])
41 valoresPropios[8]<- resultados$first
42 #9
43 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFZ_
    CZ.mat')
44 datos<-readMat(paste(direc1, '/AlphaFZ_CZ.mat', sep=""))
45 resultados<-obtenerPCA(datos$Alpha[1:1000])
46 valoresPropios[9]<- resultados$first
47 #10
48 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaCZ_
    PZ.mat')
49 datos<-readMat(paste(direc1, '/AlphaCZ_PZ.mat', sep=""))
50 resultados<-obtenerPCA(datos$Alpha[1:1000])
51 valoresPropios[10]<- resultados$first
52 #11
53 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFP2_
    F4.mat')
54 datos<-readMat(paste(direc1, '/AlphaFP2_F4.mat', sep=""))
55 resultados<-obtenerPCA(datos$Alpha[1:1000])
56 valoresPropios[11]<- resultados$first
57 #12
58 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaF4_
    C4.mat')
59 datos<-readMat(paste(direc1, '/AlphaF4_C4.mat', sep=""))
60 resultados<-obtenerPCA(datos$Alpha[1:1000])
61 valoresPropios[12]<- resultados$first
62 #13
63 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaC4_
    P4.mat')
64 datos<-readMat(paste(direc1, '/AlphaC4_P4.mat', sep=""))
65 resultados<-obtenerPCA(datos$Alpha[1:1000])
66 valoresPropios[13]<- resultados$first
67 #14
68 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaP4_
    O2.mat')
69 datos<-readMat(paste(direc1, '/AlphaP4_O2.mat', sep=""))

```

```

70 resultados<-obtenerPCA(datos$Alpha[1:1000])
71 valoresPropios[14]<- resultados$first
72 #15
73 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFP2_
    F8.mat')
74 datos<-readMat(paste(direc1, '/AlphaFP2_F8.mat', sep=""))
75 resultados<-obtenerPCA(datos$Alpha[1:1000])
76 valoresPropios[15]<- resultados$first
77 #16
78 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaF8_
    T8.mat')
79 datos<-readMat(paste(direc1, '/AlphaF8-T8.mat', sep=""))
80 resultados<-obtenerPCA(datos$Alpha[1:1000])
81 valoresPropios[16]<- resultados$first
82 #17
83 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaT8_
    P8.mat')
84 datos<-readMat(paste(direc1, '/AlphaT8_P8.mat', sep=""))
85 resultados<-obtenerPCA(datos$Alpha[1:1000])
86 valoresPropios[17]<- resultados$first
87 #18
88 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaP8_
    O2.mat')
89 datos<-readMat(paste(direc1, '/AlphaP8-O2.mat', sep=""))
90 resultados<-obtenerPCA(datos$Alpha[1:1000])
91 valoresPropios[18]<- resultados$first
92 #19
93 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaP7_
    T7.mat')
94 datos<-readMat(paste(direc1, '/AlphaP7_T7.mat', sep=""))
95 resultados<-obtenerPCA(datos$Alpha[1:1000])
96 valoresPropios[19]<- resultados$first
97 #20
98 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaT7_
    FT9.mat')
99 datos<-readMat(paste(direc1, '/AlphaT7_FT9.mat', sep=""))
100 resultados<-obtenerPCA(datos$Alpha[1:1000])
101 valoresPropios[20]<- resultados$first
102 #21
103 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFT9_
    FT10.mat')
104 datos<-readMat(paste(direc1, '/AlphaFT9_FT10.mat', sep=""))
105 resultados<-obtenerPCA(datos$Alpha[1:1000])
106 valoresPropios[21]<- resultados$first
107 #22
108 #datos<-readMat('/Users/maria/Desktop/pacientesEpilepsia/paciente1/toma1/alphas1_1/AlphaFT10_
    T8.mat')
109 datos<-readMat(paste(direc1, '/AlphaFT10_T8.mat', sep=""))
110 resultados<-obtenerPCA(datos$Alpha[1:1000])
111 valoresPropios[22]<- resultados$first
112
113 matrizValorespropios[17,]<-valoresPropios
114
115 nombrecsv<-paste('datosElectro', paciente, '-', toma, '.csv', sep="")
116 write.csv(matrizValorespropios, nombrecsv)

```

Ahora, ya podemos acceder al archivo “matrizValorespropios.csv” con la intrucción:

```

1 valoresPrincipalesBayes <- read.csv(file="matrizValorespropios.csv", header=
    TRUE, sep=",")

```

Se aplica la transformación “logito” a los datos: donde n es el número de paciente.

```

1 valoresPrincipalesBayes [n, 2:23] <- log10( (valoresPrincipalesBayes [1, 2:23] / sum(
   valoresPrincipalesBayes [1, 2:23])) / (1 - (valoresPrincipalesBayes [1, 2:23] / sum(
   valoresPrincipalesBayes [1, 2:23]))))

```

Se crea una tabla que contiene los datos de la transformación “logito”, junto con sus dos características: sexo y edad. El dataframe se guarda en un archivo “tabla1.csv”:

```

1 sexo<-c(0,1,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0)
2 edad<-c(11,11,14,22,7,1.5,14.5,3.5,10,3,12,2,3,9,7,6,13,9,6)
3
4 colnames (tabla1) <-c("Id", "FP1_F7", "F7_T7", "T7_P7", "P7_O1", "FP1_F3", "F3_C3",
5 "C3_P3", "P3_O1", "FZ_CZ", "CZ_PZ", "FP2_F4", "F4_C4", "C4_P4",
6 "P4_O2", "FP2_F8", "F8_T8", "T8_P8", "P8_O2", "P7_T7", "T7_FT9",
7 "FT9_FT10", "FT10_T8", "edad", "sexo")
8
9 tabla1<-data.frame (valoresPrincipalesBayes$V1, valoresPrincipalesBayes$V2,
10 valoresPrincipalesBayes$V3, valoresPrincipalesBayes$V4,
11 valoresPrincipalesBayes$V5, valoresPrincipalesBayes$V6,
12 valoresPrincipalesBayes$V7, valoresPrincipalesBayes$V8,
13 valoresPrincipalesBayes$V11, valoresPrincipalesBayes$V12,
14 valoresPrincipalesBayes$V13, valoresPrincipalesBayes$V14,
15 valoresPrincipalesBayes$V15, valoresPrincipalesBayes$V16,
16 valoresPrincipalesBayes$V17, valoresPrincipalesBayes$V18,
17 valoresPrincipalesBayes$V20, valoresPrincipalesBayes$V21,
18 valoresPrincipalesBayes$V22)
19
20 write.csv (tabla1, 'tabla1.csv')

```

Se aplica la librería “bnlearn” para la red de Bayes:

```

1 library (bnlearn)
2 #primer caso
3 ug = empty.graph (names (tabla1))
4 ug<-set.arc (ug, from="P3_O1", to="C3_P3")
5 ug<-set.arc (ug, from="P3_O1", to="P7_O1")
6 ug<-set.arc (ug, from="P7_O1", to="T7_P7")
7 ug<-set.arc (ug, from="T7_P7", to="F7_T7")
8 ug<-set.arc (ug, from="F7_T7", to="FP1_F7")
9 ug<-set.arc (ug, from="FP1_F3", to="FP1_F7")
10 ug<-set.arc (ug, from="C3_P3", to="F3_C3")
11 ug<-set.arc (ug, from="F3_C3", to="FP1_F3")
12 ug<-set.arc (ug, from="P8_O2", to="T8_P8")
13 ug<-set.arc (ug, from="P4_O2", to="P8_O2")
14 ug<-set.arc (ug, from="P4_O2", to="C4_P4")
15 ug<-set.arc (ug, from="C4_P4", to="F4_C4")
16 ug<-set.arc (ug, from="F4_C4", to="FP2_F4")
17 ug<-set.arc (ug, from="FP2_F4", to="FP2_F8")
18 ug<-set.arc (ug, from="F8_T8", to="FP2_F8")
19 ug<-set.arc (ug, from="T8_P8", to="F8_T8")
20 ug<-set.arc (ug, from="T7_P7", to="T7_FT9")
21 ug<-set.arc (ug, from="F7_T7", to="T7_FT9")
22 ug<-set.arc (ug, from="F8_T8", to="FT10_T8")
23 ug<-set.arc (ug, from="T8_P8", to="FT10_T8")
24 ug<-set.arc (ug, from="T7_FT9", to="FT9_FT10")
25 ug<-set.arc (ug, from="FT10_T8", to="FT9_FT10")

```

Para el grupo de pacientes Hombres y el grupo de pacientes Mujeres, se aplica la red de Bayes y se guardan los resultados en tablas correspondientes al sexo correspondiente.

```

1 fitted = bn.fit (ug, data = tabla2)
2 #MUJERES
3 valoresPrincipalesBayes3<-valoresPrincipalesBayes [c(1,3,5:7,9,11:19), 2:23]

```

```

4 tabla3<-data.frame(valoresPrincipalesBayes3$V1, valoresPrincipalesBayes3$V2,
5   valoresPrincipalesBayes3$V3, valoresPrincipalesBayes3$V4,
6   valoresPrincipalesBayes3$V5, valoresPrincipalesBayes3$V6,
7   valoresPrincipalesBayes3$V7, valoresPrincipalesBayes3$V8,
8   valoresPrincipalesBayes3$V11, valoresPrincipalesBayes3$V12,
9   valoresPrincipalesBayes3$V13, valoresPrincipalesBayes3$V14,
10  valoresPrincipalesBayes3$V15, valoresPrincipalesBayes3$V16,
11  valoresPrincipalesBayes3$V17, valoresPrincipalesBayes3$V18,
12  valoresPrincipalesBayes3$V20, valoresPrincipalesBayes3$V21
13  , valoresPrincipalesBayes3$V22)
14 colnames(tabla3) <- c("FP1_F7", "F7_T7", "T7_P7", "P7_O1", "FP1_F3", "F3_C3",
15  "C3_P3", "P3_O1", "FP2_F4", "F4_C4", "C4_P4",
16  "P4_O2", "FP2_F8", "F8_T8", "T8_P8", "P8_O2",
17  "T7_FT9", "FT9_FT10", "FT10_T8")
18 fitted = bn.fit(ug, data = tabla3)
19
20 #HOMBRES
21 valoresPrincipalesBayes4<-valoresPrincipalesBayes2[c(2,4,8,10),2:23]
22 tabla4<-data.frame(valoresPrincipalesBayes4$V1, valoresPrincipalesBayes4$V2,
23   valoresPrincipalesBayes4$V3, valoresPrincipalesBayes4$V4,
24   valoresPrincipalesBayes4$V5, valoresPrincipalesBayes4$V6,
25   valoresPrincipalesBayes4$V7, valoresPrincipalesBayes4$V8,
26   valoresPrincipalesBayes4$V11, valoresPrincipalesBayes4$V12,
27   valoresPrincipalesBayes4$V13, valoresPrincipalesBayes4$V14,
28   valoresPrincipalesBayes4$V15, valoresPrincipalesBayes4$V16,
29   valoresPrincipalesBayes4$V17, valoresPrincipalesBayes4$V18,
30   valoresPrincipalesBayes4$V20, valoresPrincipalesBayes4$V21,
31   valoresPrincipalesBayes4$V22)
32 colnames(tabla4) <- c("FP1_F7", "F7_T7", "T7_P7", "P7_O1", "FP1_F3", "F3_C3",
33  "C3_P3", "P3_O1", "FP2_F4", "F4_C4", "C4_P4",
34  "P4_O2", "FP2_F8", "F8_T8", "T8_P8", "P8_O2",
35  "T7_FT9", "FT9_FT10", "FT10_T8")
36 fitted2 = bn.fit(ug, data = tabla4)
37
38 write.csv(tabla3, 'tablaMujeres.csv')
39 write.csv(tabla4, 'tablaHombres.csv')

```

Los parámetros que nos interesan analizar están almacenados en los vectores “fitted” y “fitted2”. Se calcula la norma y la diferencia entre cada uno de los vectores correspondientes al grupo de Hombres y Mujeres. Es importante notar los vectores tienen diferente número de parámetros, de acuerdo con el lugar correspondiente en la red de Bayes.

```

1 arrayP8O2<- c(t(fitted$P8_O2$coefficients), t(fitted2$P8_O2$coefficients))
2 NormaP8O2M<-sqrt(fitted$P8_O2$coefficients[1]^2+fitted$P8_O2$coefficients[2]^2)
3 NormaP8O2H<-sqrt(fitted2$P8_O2$coefficients[1]^2+fitted2$P8_O2$coefficients[2]^2)
4 P8_O2<-sqrt((fitted$P8_O2$coefficients[1]-fitted2$P8_O2$coefficients[1])^2 +
5   (fitted$P8_O2$coefficients[2]-fitted2$P8_O2$coefficients[2])^2)
6
7 arrayP4O2<- c((fitted$P4_O2$coefficients[1]),(fitted2$P4_O2$coefficients[1]))
8 NormaP4O2M<-sqrt(fitted$P4_O2$coefficients[1]^2)
9 NormaP4O2H<-sqrt(fitted2$P4_O2$coefficients[1]^2)
10 P4_O2<- sqrt((fitted$P4_O2$coefficients[1]-fitted2$P4_O2$coefficients[1])^2)
11
12 arrayC4P4<- c((fitted$C4_P4$coefficients[1]),(fitted2$C4_P4$coefficients[1]))
13 NormaC4P4M<-sqrt(fitted$C4_P4$coefficients[1]^2+fitted$C4_P4$coefficients[2]^2)
14 NormaC4P4H<-sqrt(fitted2$C4_P4$coefficients[1]^2+fitted2$C4_P4$coefficients[2]^2)
15 C4_P4<- sqrt((fitted$C4_P4$coefficients[1]-fitted2$C4_P4$coefficients[1])^2+
16   (fitted$C4_P4$coefficients[2]-fitted2$C4_P4$coefficients[2])^2)
17
18 arrayF4C4<- c((fitted$F4_C4$coefficients[1]),(fitted2$F4_C4$coefficients[1]))
19 NormaF4C4M<-sqrt(fitted$F4_C4$coefficients[1]^2+fitted$F4_C4$coefficients[2]^2)
20 NormaF4C4H<-sqrt(fitted2$F4_C4$coefficients[1]^2+fitted2$F4_C4$coefficients[2]^2)

```

```

21 | F4_C4<- sqrt((fitted$F4_C4$coefficients[1]-fitted2$F4_C4$coefficients[1])^2+
22 |           (fitted$F4_C4$coefficients[2]-fitted2$F4_C4$coefficients[2])^2)
23 |
24 | arrayFP2F4<- c((fitted$FP2_F4$coefficients[1]),(fitted2$FP2_F4$coefficients[1]))
25 | NormaFP2F4M<-sqrt(fitted$FP2_F4$coefficients[1]^2+fitted$FP2_F4$coefficients[2]^2)
26 | NormaFP2F4H<-sqrt(fitted2$FP2_F4$coefficients[1]^2+fitted2$FP2_F4$coefficients[2]^2)
27 | FP2_F4<- sqrt((fitted$FP2_F4$coefficients[1]-fitted2$FP2_F4$coefficients[1])^2+
28 |           (fitted$FP2_F4$coefficients[2]-fitted2$FP2_F4$coefficients[2])^2)
29 |
30 | arrayFP2F8<- c((fitted$FP2_F8$coefficients[1]),(fitted2$FP2_F8$coefficients[1]))
31 | NormaFP2F8M<-sqrt(fitted$FP2_F8$coefficients[1]^2+fitted$FP2_F8$coefficients[2]^2
32 |           +fitted$FP2_F8$coefficients[3]^2)
33 | NormaFP2F8H<-sqrt(fitted2$FP2_F8$coefficients[1]^2+fitted2$FP2_F8$coefficients[2]^2
34 |           +fitted2$FP2_F8$coefficients[3]^2)
35 | FP2_F8<- sqrt((fitted$FP2_F8$coefficients[1]-fitted2$FP2_F8$coefficients[1])^2+
36 |           (fitted$FP2_F8$coefficients[2]-fitted2$FP2_F8$coefficients[2])^2+
37 |           (fitted$FP2_F8$coefficients[3]-fitted2$FP2_F8$coefficients[3])^2)
38 |
39 | arrayF8T8<- c((fitted$F8_T8$coefficients[1]),(fitted2$F8_T8$coefficients[1]))
40 | NormaF8T8M<-sqrt(fitted$F8_T8$coefficients[1]^2+fitted$F8_T8$coefficients[2]^2)
41 | NormaF8T8H<-sqrt(fitted2$F8_T8$coefficients[1]^2+fitted2$F8_T8$coefficients[2]^2)
42 | F8_T8<- sqrt((fitted$F8_T8$coefficients[1]-fitted2$F8_T8$coefficients[1])^2+
43 |           (fitted$F8_T8$coefficients[2]-fitted2$F8_T8$coefficients[2])^2)
44 |
45 | arrayT8P8<- c((fitted$T8_P8$coefficients[1]),(fitted2$T8_P8$coefficients[1]))
46 | NormaT8P8M<-sqrt(fitted$T8_P8$coefficients[1]^2+fitted$T8_P8$coefficients[2]^2)
47 | NormaT8P8H<-sqrt(fitted2$T8_P8$coefficients[1]^2+fitted2$T8_P8$coefficients[2]^2)
48 | T8_P8<- sqrt((fitted$T8_P8$coefficients[1]-fitted2$T8_P8$coefficients[1])^2+
49 |           (fitted$T8_P8$coefficients[2]-fitted2$T8_P8$coefficients[2])^2)
50 |
51 | arrayFP1F7<- c((fitted$FP1_F7$coefficients[1]),(fitted2$FP1_F7$coefficients[1]))
52 | NormaFP1F7M<-sqrt(fitted$FP1_F7$coefficients[1]^2+fitted$FP1_F7$coefficients[2]^2
53 |           +fitted$FP1_F7$coefficients[3]^2)
54 | NormaFP1F7H<-sqrt(fitted2$FP1_F7$coefficients[1]^2+fitted2$FP1_F7$coefficients[2]^2
55 |           +fitted2$FP1_F7$coefficients[3]^2)
56 | FP1_F7<- sqrt((fitted$FP1_F7$coefficients[1]-fitted2$FP1_F7$coefficients[1])^2+
57 |           (fitted$FP1_F7$coefficients[2]-fitted2$FP1_F7$coefficients[2])^2+
58 |           (fitted$FP1_F7$coefficients[3]-fitted2$FP1_F7$coefficients[3])^2)
59 |
60 | arrayF7T7<- c((fitted$F7_T7$coefficients[1]),(fitted2$F7_T7$coefficients[1]))
61 | NormaF7T7M<-sqrt(fitted$F7_T7$coefficients[1]^2+fitted$F7_T7$coefficients[2]^2)
62 | NormaF7T7H<-sqrt(fitted2$F7_T7$coefficients[1]^2+fitted2$F7_T7$coefficients[2]^2)
63 | F7_T7<- sqrt((fitted$F7_T7$coefficients[1]-fitted2$F7_T7$coefficients[1])^2+
64 |           (fitted$F7_T7$coefficients[2]-fitted2$F7_T7$coefficients[2])^2)
65 |
66 | arrayT7P7<- c((fitted$T7_P7$coefficients[1]),(fitted2$T7_P7$coefficients[1]))
67 | NormaT7P7M<-sqrt(fitted$T7_P7$coefficients[1]^2+fitted$T7_P7$coefficients[2]^2)
68 | NormaT7P7H<-sqrt(fitted2$T7_P7$coefficients[1]^2+fitted2$T7_P7$coefficients[2]^2)
69 | T7_P7<- sqrt((fitted$T7_P7$coefficients[1]-fitted2$T7_P7$coefficients[1])^2+
70 |           (fitted$T7_P7$coefficients[2]-fitted2$T7_P7$coefficients[2])^2)
71 |
72 | arrayP7O1<- c((fitted$P7_O1$coefficients[1]),(fitted2$P7_O1$coefficients[1]))
73 | NormaP7O1M<-sqrt(fitted$P7_O1$coefficients[1]^2+fitted$P7_O1$coefficients[2]^2)
74 | NormaP7O1H<-sqrt(fitted2$P7_O1$coefficients[1]^2+fitted2$P7_O1$coefficients[2]^2)
75 | P7_O1<- sqrt((fitted$P7_O1$coefficients[1]-fitted2$P7_O1$coefficients[1])^2+
76 |           (fitted$P7_O1$coefficients[2]-fitted2$P7_O1$coefficients[2])^2)
77 |
78 | arrayP3O1<- c((fitted$P3_O1$coefficients[1]),(fitted2$P3_O1$coefficients[1]))
79 | NormaP3O1M<-sqrt(fitted$P3_O1$coefficients[1]^2)
80 | NormaP3O1H<-sqrt(fitted2$P3_O1$coefficients[1]^2)
81 | P3_O1<- sqrt((fitted$P3_O1$coefficients[1]-fitted2$P3_O1$coefficients[1])^2)
82 |
83 | arrayC3P3<- c((fitted$C3_P3$coefficients[1]),(fitted2$C3_P3$coefficients[1]))
84 | NormaC3P3M<-sqrt(fitted$C3_P3$coefficients[1]^2+fitted$C3_P3$coefficients[2]^2)

```

```

85 NormaC3P3H<-sqrt( fitted2$C3_P3$coefficients [1]^2+ fitted2$C3_P3$coefficients [2]^2)
86 C3_P3<- sqrt(( fitted$C3_P3$coefficients [1] - fitted2$C3_P3$coefficients [1]) ^2+
87 ( fitted$C3_P3$coefficients [2] - fitted2$C3_P3$coefficients [2]) ^2)
88
89 arrayF3C3<- c(( fitted$F3_C3$coefficients [1]), (fitted2$F3_C3$coefficients [1]))
90 NormaF3C3M<-sqrt( fitted$F3_C3$coefficients [1]^2+ fitted$F3_C3$coefficients [2]^2)
91 NormaF3C3H<-sqrt( fitted2$F3_C3$coefficients [1]^2+ fitted2$F3_C3$coefficients [2]^2)
92 F3_C3<- sqrt(( fitted$F3_C3$coefficients [1] - fitted2$F3_C3$coefficients [1]) ^2+
93 ( fitted$F3_C3$coefficients [2] - fitted2$F3_C3$coefficients [2]) ^2)
94
95 arrayFP1F3<- c(( fitted$FP1_F3$coefficients [1]), (fitted2$FP1_F3$coefficients [1]))
96 NormaFP1F3M<-sqrt( fitted$FP1_F3$coefficients [1]^2+ fitted$FP1_F3$coefficients [2]^2)
97 NormaFP1F3H<-sqrt( fitted2$FP1_F3$coefficients [1]^2+ fitted2$FP1_F3$coefficients [2]^2)
98 FP1_F3<- sqrt(( fitted$FP1_F3$coefficients [1] - fitted2$FP1_F3$coefficients [1]) ^2+
99 ( fitted$FP1_F3$coefficients [2] - fitted2$FP1_F3$coefficients [2]) ^2)
100
101 arrayT7FT9<- c(( fitted$T7_FT9$coefficients [1]), (fitted2$T7_FT9$coefficients [1]))
102 NormaT7FT9M<-sqrt( fitted$T7_FT9$coefficients [1]^2+ fitted$T7_FT9$coefficients [2]^2
103 +fitted$T7_FT9$coefficients [3]^2)
104 NormaT7FT9H<-sqrt( fitted2$T7_FT9$coefficients [1]^2+ fitted2$T7_FT9$coefficients [2]^2
105 +fitted2$T7_FT9$coefficients [3]^2)
106 T7_FT9<- sqrt(( fitted$T7_FT9$coefficients [1] - fitted2$T7_FT9$coefficients [1]) ^2+
107 ( fitted$T7_FT9$coefficients [2] - fitted2$T7_FT9$coefficients [2]) ^2+
108 ( fitted$T7_FT9$coefficients [3] - fitted2$T7_FT9$coefficients [3]) ^2)
109
110 arrayFT10T8<- c(( fitted$FT10_T8$coefficients [1]), (fitted2$FT10_T8$coefficients [1]))
111 NormaFT10T8M<-sqrt( fitted$FT10_T8$coefficients [1]^2+ fitted$FT10_T8$coefficients [2]^2
112 +fitted$FT10_T8$coefficients [3]^2)
113 NormaFT10T8H<-sqrt( fitted2$FT10_T8$coefficients [1]^2+ fitted2$FT10_T8$coefficients [2]^2
114 +fitted2$FT10_T8$coefficients [3]^2)
115 FT10_T8<- sqrt(( fitted$FT10_T8$coefficients [1] - fitted2$FT10_T8$coefficients [1]) ^2+
116 ( fitted$FT10_T8$coefficients [2] - fitted2$FT10_T8$coefficients [2]) ^2+
117 ( fitted$FT10_T8$coefficients [3] - fitted2$FT10_T8$coefficients [3]) ^2)
118
119 arrayFT9FT10<- c(( fitted$FT9_FT10$coefficients [1]), (fitted2$FT9_FT10$coefficients [1]))
120 NormaFT9FT10M<-sqrt( fitted$FT9_FT10$coefficients [1]^2+ fitted$FT9_FT10$coefficients [2]^2
121 +fitted$FT9_FT10$coefficients [3]^2)
122 NormaFT9FT10H<-sqrt( fitted2$FT9_FT10$coefficients [1]^2+ fitted2$FT9_FT10$coefficients [2]^2
123 +fitted2$FT9_FT10$coefficients [3]^2)
124 FT9_FT10<- sqrt(( fitted$FT9_FT10$coefficients [1] - fitted2$FT9_FT10$coefficients [1]) ^2+
125 ( fitted$FT9_FT10$coefficients [2] - fitted2$FT9_FT10$coefficients [2]) ^2+
126 ( fitted$FT9_FT10$coefficients [3] - fitted2$FT9_FT10$coefficients [3]) ^2)
127
128
129 diferenciasX<-c(P8_O2,P4_O2,C4_P4,F4_C4,FP2_F4,FP2_F8,F8_T8,T8_P8,
130 FP1_F7,F7_T7,T7_P7,P7_O1,P3_O1,C3_P3,F3_C3,FP1_F3,
131 T7_FT9,FT10_T8,FT9_FT10)
132 nombresNodos<-c("P8_O2","P4_O2","C4_P4","F4_C4","FP2_F4","FP2_F8",
133 "F8_T8","T8_P8","FP1_F7","F7_T7","T7_P7","P7_O1","P3_O1",
134 "C3_P3","F3_C3","FP1_F3","T7_FT9","FT10_T8","FT9_FT10")
135 diferencias<-data.frame(diferenciasX , nombresNodos)

```