



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

EMPLEO DE REDES NEURONALES ARTIFICIALES Y MEDIDAS COMPLEJAS PARA  
CLASIFICAR UNA CONDICIÓN CLÍNICA DEL CEREBRO HUMANO

T E S I S  
QUE PARA OPTAR POR EL GRADO DE  
MAESTRO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

PRESENTA:  
JORGE RICARDO GERSENOWIES ROSAS

Directora de tesis:  
DRA. SUEMI RODRÍGUEZ ROMO  
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN, UNAM

CIUDAD UNIVERSITARIA, CD. MX. ENERO, 2018



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



## Agradecimientos

---

A mis padres, ya que gracias a su apoyo incondicional y su guía constante me han dado el ejemplo y me han aportado la fuerza necesaria para que el presente trabajo fuera posible.

A la Dra. Suemi Rodríguez Romo ya que, gracias a ella pude conocer el mundo del cómputo científico y gracias a sus enseñanzas, apoyo y confianza fue posible llevar a buen término el presente trabajo.

Al programa de Posgrado en ciencia e Ingeniería de la Computación.

Al CONACYT por la beca otorgada que hizo posible la realización del presente trabajo sin distracciones.

A la DGAPA-UNAM por el apoyo mediante el PAPIIT 101817.

A mis compañeros, que actualmente considero mis amigos, con quienes integré un gran equipo de trabajo, con el cual espero seguir colaborando a futuro.

## Contenido

Resumen .....	3
Abstract.....	4
Capítulo 1 Introducción, motivación y objetivos.....	5
1.1 Introducción .....	5
1.2 Motivación .....	6
1.3 Hipótesis.....	6
1.4 Objetivo General .....	7
1.5 Objetivos particulares .....	7
Capítulo 2. Antecedentes .....	8
2.1 Cerebro como sistema complejo .....	8
2.2 El EEG y su estudio .....	10
2.3 Caos en el cerebro .....	13
2.4 Fractales y Multifractales (el formalismo multifractal de Kolmogorov).....	14
2.5 Regularidad Hölder Puntual.....	16
2.6 Dimensión de Hausdorff y Espectro Multifractal .....	17
2.6.1 $\varepsilon$ -cubrimiento.....	17
2.6.2 Medida de Hausdorff $\delta$ -dimensional. ....	17
2.7 Espectro Multifractal. ....	18
2.7.1 El Algoritmo del Análisis Multifractal de Fluctuaciones sin tendencia(MFDFA)..	18
2.8 Teoría de la información .....	21
2.8.1 Entropía de Shannon .....	21
2.8.2 Entropía de Tsallis .....	23

2.9 Redes neuronales artificiales como clasificadores .....	25
2.9.1 Elementos de una Red Neuronal .....	27
2.9.2 Backpropagation .....	28
2.9.3 Ajuste de los pesos en la capa de salida .....	31
2.9.4 Ajuste de pesos en las capas ocultas .....	32
2.9.5 Derivaciones .....	33
Capítulo 3. Obtención de medidas complejas .....	34
3.1 MFDFA .....	34
3.2 Entropía de Tsallis .....	39
3.3 Solución propuesta .....	41
3.4 La base de datos .....	41
3.5 Tabla de entrenamiento .....	43
Capítulo 4. Entrenamiento de la red neuronal .....	44
4.1 Entrenamiento de la red neuronal .....	44
Capítulo 5. Resultados .....	51
5.1 Diseño de Software .....	51
5.2 Resultados .....	53
5.3 Discusión .....	59
5.4 Conclusiones .....	60
Bibliografía .....	62
Índice de figuras .....	65
Índice de Tablas .....	66

## Resumen

En el presente trabajo de tesis se expone la forma en que se realizó la investigación que tuvo como finalidad elaborar un clasificador que, como núcleo tiene un conjunto de redes neuronales, dichas redes neuronales fueron alimentadas en su fase de entrenamiento con medidas complejas tomadas de electroencefalogramas mediante el enfoque del análisis multifractal de fluctuaciones sin tendencia (MFDFA por sus siglas en inglés) y la entropía según la teoría de la información teniendo en cuenta el modelo presentado por Constantino Tsallis.

Para realizar los cálculos de entropía y multifractal, se empleó en común el mismo momento de fluctuaciones  $q$ , teniendo los valores desde -10 a 10, con revalorizaciones cada 0.2. Dicho momento representa el comportamiento de las fluctuaciones dependiendo de su valor, siendo las fluctuaciones más rápidas correspondientes a -10 y las fluctuaciones más lentas correspondientes a 10.

Se encontró que, las cuatro medidas que empleamos, las cuales son el momento de fluctuaciones  $q$ , el exponente de Hurst, el exponente de masa y la entropía según Tsallis, son suficientes (aun cuando no se pudo demostrar que sean necesarias) para poder realizar la clasificación de presencia o ausencia de una condición clínica en particular, en este caso el alcoholismo, la utilización de las medidas antes mencionadas constituye parte de la contribución del presente trabajo.

El tipo de redes neuronales empleadas en el presente trabajo, corresponde a las redes neuronales multicapa y entrenadas mediante el algoritmo de BackPropagation (propagación desde atrás). Para realizar el entrenamiento de las redes neuronales se emplearon dos aplicaciones en específico, una de pago (Neuro Solutions, en versión de demostración) y una de código abierto (Multiple BackPropagation (MBP)), y se comparó en términos de la optimización del error, entre ambas plataformas. Para lograr una mejor optimización, en el caso del MBP se dividieron para entrenar las redes neuronales los momentos  $q$  en positivos y negativos, luego se tomó cada momento. El error es minimizado

al usar una red neuronal por momento de fluctuación  $q$ , que también forma parte de la contribución del trabajo.

## Abstract

In this thesis, we introduce a method for decision-making over a particular system. The core of the method relies on a set of neural networks which are trained with a set of four variables taken from the set of complex measures extracted from a database of electroencephalograms using multifractal detrended fluctuation analysis (MFDFA) and information entropy by Tsallis.

For obvious reasons, the same fluctuation moment  $q$ 's value was use in both; the MFDFA and the Tsallis entropy. We find that a set of four variables is sufficient (no necessary though) condition to perform the classification of the clinical conditions; a) alcoholic and b) normal patients. We call these universality classes

The neural networks used in this paper are multi-layer neural networks and has been trained with the back-propagation algorithm. To train the neural networks use two applications; one commercial (Neuro solutions in demo version) and other open source (Multiple back-propagation). We compare them in terms of error optimization, in both applications. We refine the accuracy of our approach by using a neural network per fluctuation moment.

# Capítulo 1 Introducción, motivación y objetivos.

## 1.1 Introducción

El estudio del comportamiento humano es una disciplina en la cual se fundamenta el campo de la inteligencia artificial, muchos modelos de inteligencia artificial se han establecido a partir de modelos de lógica relacional y de otras ramas como la Psicología, la Filosofía y las Neurociencias (Hilera & Martinez, 2000).

Si bien estas ciencias arrojan numerosos resultados, la rama de las neurociencias se ha especializado en el estudio directo del cerebro, siendo su enfoque central el cerebro como fuente del pensamiento e inteligencia, ocupando su respuesta electrofisiológica como una de las principales fuentes de información para su estudio (Hilera & Martinez, 2000).

La respuesta electrofisiológica cerebral medida a través del electroencefalograma; es un reflejo de la actividad eléctrica llevada a cabo por el cerebro, es decir, se supone que es una ventana para ver como realiza sus tareas, como interpreta, almacena y accede a la información, muchos estudios en neurociencias tienen como objetivo el establecer la correlación que puede existir entre la actividad eléctrica cerebral, medida en el electroencefalograma, y alguna función cognitiva, por ello el estudio de esta respuesta nos lleva a tener una mejor comprensión de los fenómenos que se presentan en el cerebro y nos puede permitir la detección de estados anómalos presentes en el mismo (Conn, 2003).

En el desarrollo del presente proyecto se construye un clasificador, que a partir de la relación que encontramos entre medidas multifractales y de la teoría de la información obtenidas de electroencefalogramas, pueda decidir si una condición anómala está o no está presente en el cerebro en análisis, permitiendo discriminar entre pacientes alcohólicos de los no alcohólicos.

Cabe destacar que, si bien el MF DFA y la entropía han sido utilizados en el estudio de electroencefalogramas, el momento de fluctuaciones  $q$ , el exponente de Hurst, el

exponente de masa y la entropía según Tsallis, no han sido utilizados juntos antes para relacionarlos con una condición clínica particular.

## 1.2 Motivación

Los sistemas de salud en la actualidad no se dan abasto en la atención a pacientes y muchas veces las enfermedades son diagnosticadas de manera tardía, cuando ya nada se puede hacer. En este sentido el apoyo mediante sistemas computacionales que permitan la toma de decisiones preliminar puede ser de gran ayuda, ya que pueden acortar el tiempo entre la toma de datos y la canalización del paciente al especialista en cuestión, además, el proporcionar un sistema de inteligencia artificial que pueda ser utilizado en sitio puede apoyar a que la atención del paciente y la utilización de recursos se vuelven más eficientes.

Aunque en el presente trabajo se ocupa como problema de salud el alcoholismo, la técnica presentada puede ser más general y puede ser la puerta para en un futuro poder clasificar más enfermedades mentales y ayudar con su detección temprana. También se extrapola la existencia de clases de universalidad para el comportamiento cerebral en estas condiciones, con lo que tenemos una nueva medida que puede determinar la condición del cerebro humano y con ello podemos conocer más de cerca cómo se comporta.

## 1.3 Hipótesis

- Existirán al menos dos clases de universalidad con las cuales se puedan clasificar el estado sano (no alcohólico) y un estado clínicamente enfermo (alcohólico) del cerebro humano a partir del estudio de electroencefalogramas (EEG) y empleando Redes Neuronales Artificiales como clasificadores.
- Existirá al menos una red neuronal que sea capaz de clasificar una condición particular del cerebro humano, teniendo como entrada medidas complejas del electroencefalograma.
- Existirá una solución estimada (ansatz) que, dado el momento de fluctuaciones  $q$ , el exponente de Hurst, el exponente de masa y la entropía según Tsallis, como

variables independientes (entradas) ajuste una variable dependiente (salida), en este caso la condición alcohólico y no alcohólico.

#### 1.4 Objetivo General

Diseñar e implementar un clasificador, el cual tendrá como núcleo una red neuronal artificial, que pueda clasificar en base a el momento de fluctuaciones  $q$ , el exponente de Hurst, el exponente de masa y la entropía según Tsallis, a pacientes alcohólicos y no alcohólicos.

#### 1.5 Objetivos particulares

Se obtendrán las medidas complejas a partir de electroencefalogramas, aplicando análisis multifractal de fluctuaciones sin tendencia (MFDFA por sus siglas en inglés), y se obtendrá la cantidad de entropía según Tsallis.

Se obtendrá una red neuronal que clasifique paciente alcohólicos y no alcohólicos, teniendo como entradas las medidas complejas mencionadas en el punto anterior, y como salidas 1 para alcohólicos y 0 para no alcohólicos.

Se programará una interfase donde se pueda introducir las medidas complejas y obtener la clasificación en pantalla.

## Capítulo 2. Antecedentes

En el presente capítulo se abordan los fundamentos teóricos necesarios para la elaboración del trabajo de tesis, primeramente, hablamos del cerebro, dando un repaso breve a su estudio como sistema complejo, posteriormente se aborda los temas de fractales y multifractales para luego terminar con teoría de la información y entropía.

### 2.1 Cerebro como sistema complejo

Los sistemas complejos se caracterizan por ser sistemas que no se pueden reducir a versiones más simplificadas; frecuentemente son resultado de un fenómeno colectivo con muchos grados de libertad en interacción o con fuerte interdependencia, presentando una dinámica no lineal y suelen describirse bajo el formalismo de procesos estocásticos. Es también común que el fenómeno bajo análisis sea de carácter general con una variedad de manifestaciones específicas.

Dado que el cerebro puede ser considerado una máquina biológica, construida por miles unidades de procesamiento llamadas neuronas y por las interconexiones que hay entre ellas, aparentemente al azar, el hecho de que la interacción entre neuronas se pueda hacer en su localidad, dentro las mini columnas las cuales se organizan en módulos que redundan la función denominados Columnas Corticales, entre conjuntos de columnas corticales o entre diferentes cortezas. Hace que el número de grados de libertad presente sea muy elevado, cabe destacar que una sola neurona no tiene un comportamiento representativo del cerebro y que se requiere de redes neuronales biológicas completas para poder estudiar su comportamiento, en este sentido el cerebro es un sistema complejo (Lamb, 1999).

Sin embargo, al igual que ocurre con el resto de las formaciones orgánicas como lo es el aparato circulatorio, cuyo crecimiento es del tipo fractal, se puede encontrar que entre individuos de una misma especie apenas cambia la distribución de sus regiones operativas, es decir así como encontramos la vena femoral corriendo paralela al fémur en la parte

interior de la pierna, en el cerebro encontramos que las cortezas encargadas de procesar las diferentes señales provenientes de los diferentes órganos sensores están ubicadas en aproximadamente la misma región del cerebro de cada individuo (Pennycuick, 1992).

La existencia de patrones no solo se encuentra en la parte morfológica, sino que, cuando analizamos las señales eléctricas cerebrales obtenidas a través de electroencefalografía, encontramos que existen patrones de onda montadas dentro de la actividad del electroencefalograma durante y para la realización de determinadas actividades, las cuales son conocidas como potenciales evocados (Conn, 2003).

Por medio de los electroencefalogramas podemos obtener información acerca de la condición del cerebro. Ya en la bibliografía nos encontramos con métodos "clásicos" de análisis dinámico no lineal aplicados a los electroencefalogramas, como pueden ser la dimensión de correlación, los exponentes de Lyapunov, la entropía, el análisis de fluctuaciones sin tendencia, la mayoría utilizando un solo parámetro, aunque por si mismos no pueden describir completamente el comportamiento del electroencefalograma (Aijun He, 2007).

En este caso el formulismo multifractal revela más información "oculta" de EEG utilizando el espectro de singularidad para caracterizar su dinámica no lineal, ya que nos permite examinar más de cerca las pequeñas fluctuaciones, pero al mismo tiempo ver las fluctuaciones más grandes (Aijun He, 2007).

Dentro de la literatura podemos encontrar trabajos, como en (Aijun He, 2007), (Likens & Amazeen, 2014) y (MA, NING, WANG, & BIAN, 2006) donde se explora la capacidad del análisis multifractal para diferenciar entre distintos tipos de señales de EEG.

En el caso de la medida de entropía se ha encontrado que tienen una amplia difusión en el estudio del comportamiento del cerebro, sobre todo en la medición de estados de anestesia profunda, sueño y coma, en los cuales se consideran reflejo de la cantidad de actividad cerebral existente, ya que ayuda a medir la complejidad del estado del cerebro, correspondiendo los estados más complejos a los de mayor actividad cerebral y los menos complejos a los de menor actividad, incluso, existen equipos médicos que umbralizan el

estado de conciencia con el de inconciencia en el caso de la anestesia profunda usando a la entropía (Zhenhu, y otros, 2015).

Por ello la entropía y el análisis multifractal podrían rastrear los cambios en el patrón de EEG durante diferentes estados. De lo anterior se puede extraer lo siguiente:

- a) El cerebro se puede considerar como una red compleja, con un comportamiento complejo factible de ser medido.
- b) El cerebro genera una respuesta eléctrica medida a través del electroencefalograma, que contiene señales auto-similares en el tiempo.

## 2.2 El EEG y su estudio

En un electroencefalograma tenemos la señal eléctrica de voltaje que podemos medir cuando se minimiza la resistencia de la piel a casi cero, las señales obtenidas son reflejo de la actividad cerebral, aunque también se toman lecturas de otros procesos electrofisiológicos como los movimientos oculares y de la mandíbula, sin embargo, gran parte de la señal está relacionada con la actividad cerebral (Collura, 1993).

El electroencefalograma es una técnica electrofisiológica para examinar la actividad eléctrica cerebral (Collura, 1993), de manera regular su análisis se lleva a cabo utilizando la forma de onda en un lapso de tiempo y analizando el espectro en el dominio de la frecuencia, dividiendo el espectro de frecuencias en diferentes bandas de la siguiente manera (Fergus, Hignett, Hussain, Al-Jumeily, & Abdel-Aziz, 2015):

Banda	Rango de Frecuencia	Voltaje
<b>Alfa</b>	8 a 13 Hz	20-60 $\mu$ V
<b>Beta</b>	>13 Hz (generalmente 18-25 Hz)	5-10 $\mu$ V
<b>Theta</b>	4 a 7,5 Hz	20-60 $\mu$ V
<b>Delta</b>	< 3,5 Hz	20-60 $\mu$ V

*Tabla 1 Bandas y frecuencias en electroencefalogramas*

El electroencefalograma se registra del cuero cabelludo de la cabeza del paciente con una serie de electrodos de plata clorurada colocados según la nomenclatura estándar de la posición de los electrodos, mostrada en la figura 1 (American Clinical Neurophysiology Society, 2006):

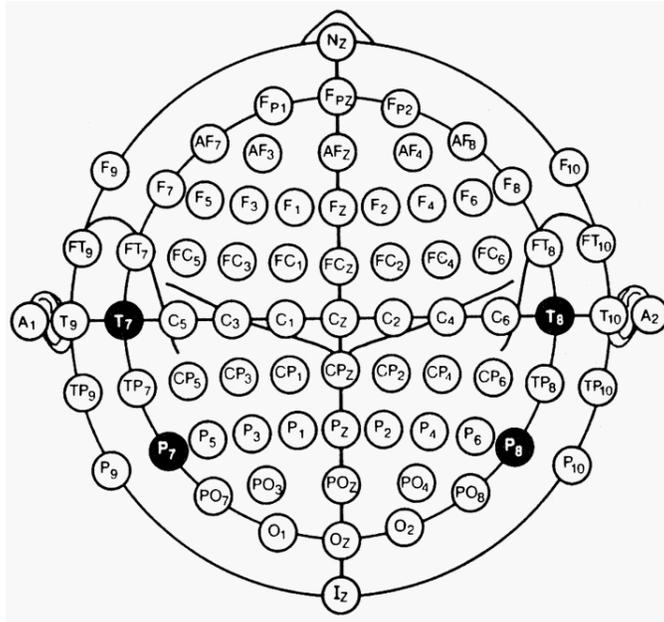


Figura 1 Nomenclatura y posición de electrodos para el sistema 10-20 Internacional de la American Clinical Neurophysiology Society, donde A1 y A2 o Nz son tomados como referencia

Lo que se obtiene es un conjunto de series de tiempos, las cuales resultan de las muestras de las señales y que son representadas de manera gráfica de la siguiente manera:

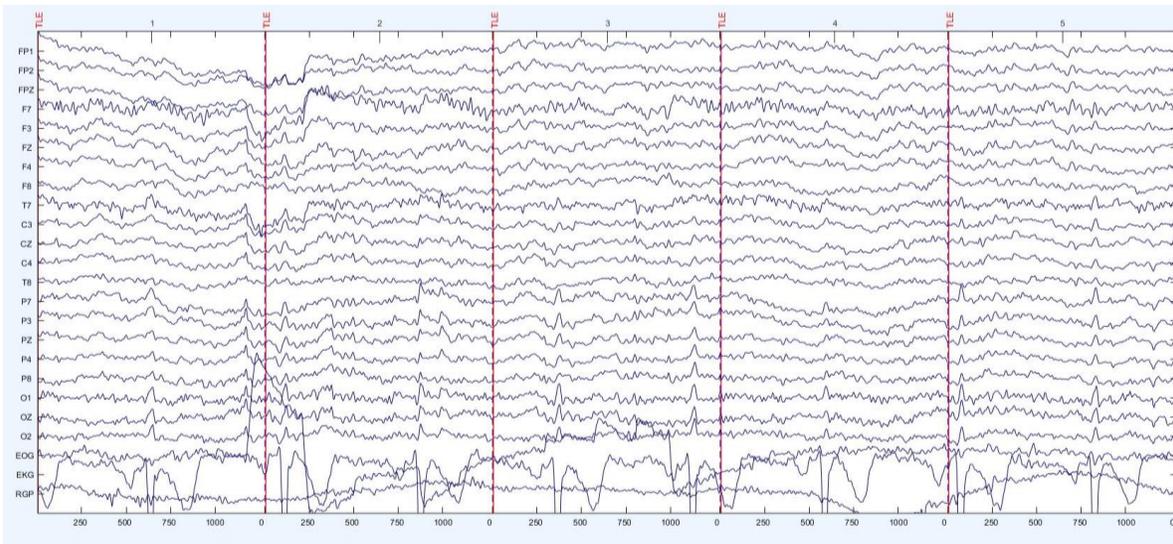


Figura 2. Ejemplo de electroencefalograma

Como se puede observar en la figura 2 el electroencefalograma está constituido por las señales de cada uno de los electrodos que se colocaron en la cabeza del paciente, y que están etiquetados de acuerdo al estándar internacional conocido como sistema 10-20

internacional (American Clinical Neurophysiology Society, 2006). Colocados en un plano donde el eje horizontal es el tiempo, dividido en segmentos de muestras, sobre el eje vertical se colocan las señales de voltaje de la parte frontal (arriba) a la occipital (abajo), siendo las últimas señales fuentes de ruido, EOG o electrooculograma que mide la actividad eléctrica debida al movimiento de los ojos, el EKG o electrocardiograma, que mide la actividad eléctrica proveniente del corazón y el RGP que es la respuesta galvánica de la piel tomada en la mano, sobre todo para medir sudoración como reflejo de comodidad, durante el muestreo del electroencefalograma.

El EOG se considera ruido y se ocupa para la aplicación de filtros sobre las demás señales, el EKG se emplea como referencia para evitar que los electrodos se conecten sobre un vaso circulatorio, que también se puede considerar ruido en este caso, en general se utilizan con el fin de eliminar la mayor cantidad de fuentes (Faure & Korn, 2001).

Por otro lado, algo que nos enseña el estudio del cerebro mediante los electroencefalogramas y otras técnicas, es la evidencia de la estratificación a través de centros de procesamiento en el cerebro, cuya corteza puede ser dividida en múltiples cortezas comunicadas entre si y a órganos sensoriales mediante diferentes vías sensoriales (Conn, 2003).

Se muestra a en la figura 3 un mapa de las diferentes cortezas en el cerebro con las áreas de Brodmann marcadas, cabe destacar que el área de Brodmann 9 está asociada a los procesos mentales superiores tales como el juicio, la voluntad y el razonamiento (Conn, 2003), razón por la cual ocuparemos el canal Fz (frontal central) para nuestro estudio.

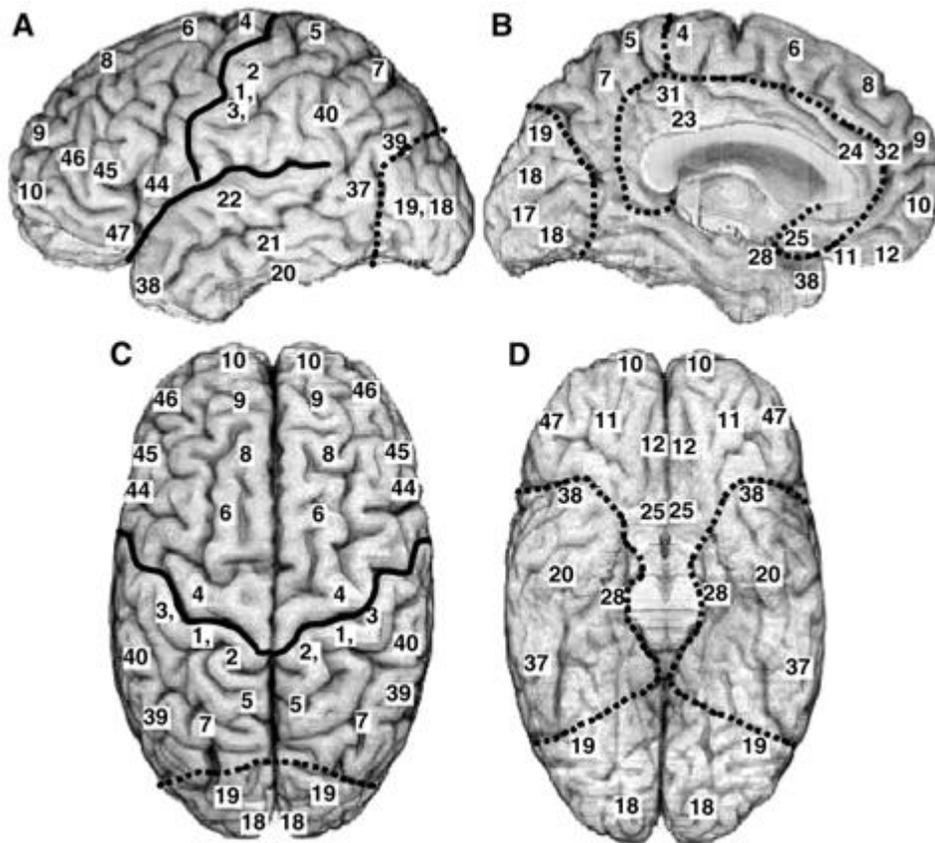


Figura 3 Vistas del cerebro Lateral (A), mesial (B), superior (C), e inferior (D), representando puntos de demarcación importantes y áreas de Brodmann (números) (Conn, 2003).

### 2.3 Caos en el cerebro

Se ha mencionado que el cerebro es visto como un sistema complejo debido a la alta cantidad de grados de libertad que presenta y a su irreductibilidad, sin embargo, nos encontramos ante el problema de estudiarlo como sistema caótico (Faure & Korn, 2001).

Para poder modelar un sistema en términos de la teoría de caos, es necesario establecer la sensibilidad que existe de la respuesta a los parámetros de entrada, y en el caso del cerebro, podemos observar que, si bien existe una determinación fuerte de un estado ante entradas, es decir, es un sistema sensible a las condiciones iniciales, no podemos obtener una misma respuesta en casos estrictamente controlados del mismo sistema, ya que, el cerebro es un sistema que aprende, por ende nunca dará una respuesta plenamente igual a estímulos iguales, se tiene que recordar que cuando se enfrenta a un

estímulo nuevo, se tiende a prestarle atención y a ser consciente de él, sin embargo, conforme se va haciendo familiar el estímulo va quedando fuera de nuestro foco de atención, es decir cada vez somos menos conscientes de su presencia, hasta que el estímulo queda relegado a la parte inconsciente de la atención, condición llamada habituación, esto también tiene consecuencia en la respuesta, ya que una vez aprendido que el estímulo, se habrá formado una nueva red cerebral, por lo que la respuesta eléctrica global habrá cambiado (Schrödinger, 1983).

De cierta manera se puede decir que esto aleja el sistema del caos determinista al agregar más fuentes de aleatoriedad y menor dependencia al estado inicial, conduce a presentar al cerebro mediante otra herramienta dentro del paradigma de la dinámica de sistemas complejos, por lo que en el presente trabajo se opta por un acercamiento con un enfoque multifractal y de la teoría de la información.

#### 2.4 Fractales y Multifractales (el formalismo multifractal de Kolmogorov)

En los sistemas complejos se presentan ciertos principios organizativos, los que aparecen en fenómenos muy dispares entre sí, como los pertenecientes a la física, la ecología, la biología o la economía, estos principios se pueden agrupar en dos grandes categorías: la invariancia escalar y la universalidad (Rosenblatta & Figliolab, 2007).

El primero de estos principios considera que los sistemas complejos suelen tener una organización jerárquica cuya manifestación se modela a través de una ley de potencia de alguna variable relevante del problema. El exponente de esta ley de potencia resulta una magnitud que caracteriza al sistema (Rosenblatta & Figliolab, 2007).

La universalidad se entiende como la tendencia que tienen sistemas de naturaleza diferente a tener el mismo comportamiento, es decir a seguir la misma ley de potencias, estos se agrupan en clases que son conocidas como “clases de universalidad”, entonces si la misma ley de potencia modela a diferentes sistemas con diferentes elementos y de naturaleza diferente, estos sistemas ocuparán la misma clase de universalidad. Que

distintos sistemas se agrupan sugiere la existencia de mecanismos similares responsables del comportamiento de invariancia de escala (Rosenblatta & Figliolab, 2007).

Estos conceptos pueden rastrearse en los trabajos de Kolmogorov de los principios de los años cuarenta del siglo XX, cuando se encontraba estudiando el fenómeno de la turbulencia. Con una notable intuición, Kolmogorov propone que en el caso de la turbulencia totalmente desarrollada con vórtices de diferentes escalas que interactúan entre sí la energía de la disipación se transmite a través de los vórtices más grandes a los más pequeños a través de una ley de potencia cuya base es la dimensión característica del vórtice y el exponente de dicha función define características propias del sistema. En estas ideas se halla el origen de lo que actualmente se conoce como formalismo o análisis multifractal (Rosenblatta & Figliolab, 2007).

Los principios organizativos de los sistemas complejos pueden ser cuantificados a través de este formalismo, pues proporciona un mecanismo de invariancia escalar para el análisis y la generación de señales complejas. Esencialmente se basa en el cálculo de dos conjuntos de números o funciones asociadas a la señal: los exponentes Hölder, que cuantifican la regularidad; y el espectro multifractal, que cuantifica la multifractalidad (Rosenblatta & Figliolab, 2007).

El análisis multifractal describe la distribución de los diferentes exponentes Hölder “ $\alpha$ ” de una función mediante el cálculo del espectro multifractal, que asocia a cada “ $\alpha$ ” la dimensión de Hausdorff del conjunto de puntos donde la función tiene regularidad puntual de exponente Hölder “ $\alpha$ ”. Recordemos que la regularidad puntual de una función localmente acotada está caracterizada por el exponente Hölder  $\alpha$  de cada punto, en un intervalo  $0 \leq \alpha \leq +\infty$ , que es un parámetro que capta la rugosidad o suavidad de una función en cada punto (Rosenblatta & Figliolab, 2007).

A mediados de la década de los noventa del siglo XX Peng (Peng, Havlin, Stanley, & Goldberger, 1995 ), desarrolló el Análisis de fluctuaciones sin tendencia (DFA por sus siglas en inglés) que resultó ser una técnica de amplio uso para la detección de propiedades de escala con comportamiento fractal de series temporales no estacionarias. El DFA tuvo gran

éxito en la aplicación a diferentes campos tales como secuencias de DNA, señales asociadas a variables económicas o a la física del estado sólido. Años más tarde, Kantelhart y Stanley extendieron este algoritmo para series temporales con comportamiento multifractal, desarrollando el Análisis multifractal de fluctuaciones sin tendencia (MFDFA por su nombre en inglés). Estos dos métodos están bien probados, en principio, para el caso de medidas de Borel no negativas, singulares y de soporte acotado. Esto incluye el caso de aquellas cuyo soporte es un conjunto de Cantor generalizado. En estos casos, se puede definir una función de partición cuya naturaleza depende del método, y a partir de ella, construir el correspondiente formalismo multifractal que permite analizar la regularidad Hölder de funciones que integran dichas medidas (Peng, Havlin, Stanley, & Goldberger, 1995 ).

## 2.5 Regularidad Hölder Puntual

Sean  $\alpha \geq 0$  y  $x_0 \in R^d$ . Una función acotada  $f : R^d \rightarrow R$  es  $C^\alpha(x_0)$  si existe una constante  $C > 0$  y un polinomio  $P_{x_0}$  de grado menor  $\alpha$  tal que, en un entorno de  $x_0$

$$|f(x) - P_{x_0}| \leq C|x - x_0|^\alpha \quad (1)$$

En el caso  $\alpha=0$  la condición (1) es  $|f(x)| \leq C$  y convenimos que al polinomio nulo le asignamos grado  $-\infty$ .

Luego el exponente Hölder de  $f$  en  $x_0$  es:

$$H_f(x_0) = \text{sub}\{\alpha: f \in C^\alpha(x_0)\} \quad (2)$$

El exponente Hölder describe la suavidad o la rugosidad del gráfico de una función en cada punto. Si  $f : R \rightarrow R$  tiene exponente Hölder  $H_f(x_0) = \alpha$ , con  $0 \leq \alpha < 1$ , de la definición se puede deducir que  $f$  no es derivable en  $x_0$ ; entonces, en este caso, el exponente capta la rugosidad que tiene el gráfico de la función en esta singularidad. Por ejemplo, si  $f(x) = |x|^{1/n}$ , con  $n$  un número natural, se puede ver que el exponente Hölder en  $x_0 = 0$  es  $\frac{1}{n}$ . Es decir que, cuánto más pequeño es el exponente, más rugosa es una función en ese punto y cuánto más cerca de 1 está el exponente resulta menos rugosa.

## 2.6 Dimensión de Hausdorff y Espectro Multifractal

Los fenómenos naturales suelen presentar características multifractales, es decir, las funciones que los modelan están caracterizadas por diferentes exponentes Hölder, que además pueden fluctuar punto a punto. La cuestión es entonces caracterizar al conjunto de puntos  $x_0$  que tienen el mismo exponente Hölder puntual  $H_f(x_0) = \alpha$ . La dimensión de Hausdorff de este conjunto es la medición que mejor se ajusta para este propósito ya que muchas funciones multifractales suelen tener el mismo exponente Hölder puntual sobre conjuntos densos y la dimensión de Hausdorff puede ser diferente en cada conjunto denso, lo cual permite distinguirlos. Otras nociones de dimensión o medida asignan el mismo valor a los conjuntos densos, lo que hace que no sean apropiadas para describir a estos conjuntos. El siguiente paso es definir la dimensión de Hausdorff, que es la definición matemática de la dimensión fractal, y el espectro multifractal.

### 2.6.1 $\varepsilon$ -cubrimiento.

Sea  $A \subset \mathbb{R}^d$  y  $n > 0$ . Decimos que  $\{A_i\}_{i \in \mathbb{N}}$  es un  $\varepsilon$ -cubrimiento de  $A$  si  $\{A_i\}_{i \in \mathbb{N}}$  es un cubrimiento de  $A$  por conjuntos acotados  $A_i$  con diámetro  $|A_i| \leq n$ .

### 2.6.2 Medida de Hausdorff $\delta$ -dimensional.

Cubrimos al conjunto  $A$  con todos los  $\varepsilon$ -cubrimientos posibles. Entonces, para cada  $\delta \in [0, d]$ , se calcula

$$M_\varepsilon^\delta = \inf_{\varepsilon\text{-cubrimientos}} \left\{ \sum_i |A_i|^\delta \right\} \quad (3)$$

Luego se define la medida de Hausdorff  $\delta$ -dimensional de  $A$  como

$$m_\delta(A) = \lim_{\varepsilon \rightarrow 0} M_\varepsilon^\delta \quad (4)$$

Dimensión de Hausdorff.

Sea  $A \subset \mathbb{R}^d$  luego existe  $0 \leq \delta_0 \leq d$  tal que

$$\forall \delta < \delta_0, m_\delta(A) = +\infty \quad (5)$$

$$\forall \delta > \delta_0, m_\delta(A) = 0 \quad (6)$$

Entonces la dimensión de Hausdorff del conjunto A será el valor de  $\delta_0$ .

## 2.7 Espectro Multifractal.

El espectro multifractal  $d(\alpha)$  es la función que asocia a cada  $\alpha$  la dimensión de Hausdorff del conjunto de puntos donde la función tiene regularidad puntual de exponente Hölder  $\alpha$ .

Es decir, si  $0 \leq \alpha \leq +\infty$  y  $f : R^d \rightarrow R$ ,

$$d(\alpha) = \dim\{x_0 \in R^d : H_f(x_0) = \alpha\} \quad (7)$$

En funciones que modelan fenómenos naturales, donde el exponente Hölder fluctúa punto a punto, no es posible obtenerlo en forma directa. Entonces se hace una estimación indirecta a través de una fórmula llamada formalismo multifractal (Jaffard, 1997); que computa valores globales de la función. Esta fórmula tiene variantes que dependen del espacio de funciones al que pertenece la función a estudiar. En este caso supondremos que las funciones tienen características auto-similares. Una función  $f$  tiene características auto-similares si existen subconjuntos disjuntos donde el gráfico de la función restringido a cada uno de estos subconjuntos es una transformación afín de  $f$ .

### 2.7.1 El Algoritmo del Análisis Multifractal de Fluctuaciones sin tendencia(MFDFA).

El objetivo del MFDFA es encontrar el espectro de singularidades en series de tiempo tanto estacionarias como no estacionarias, este algoritmo se puede dividir en 5 partes (Ryszard, 2008):

Parte 1:

En el primer paso definimos la secuencia de los desplazamientos (perfiles) como sigue:

$$Y(j) = \sum_{k=1}^j (x_k - \langle x \rangle), j = 1, \dots, N \quad (8)$$

Donde  $x_k$  es el elemento k de la serie de tiempos y el valor medio  $\langle x \rangle$  está dado por:

$$\langle x \rangle = \frac{1}{N} \sum_{k=1}^N x_k \quad (9)$$

Parte 2:

Dividimos nuestra serie de tiempos en  $N_s = INT\left(\frac{N}{S}\right)$  segmentos de tamaño  $S$  sin solapamiento, dado que los elementos en la serie de tiempo pueden sobrar, de manera análoga se hará la división ahora iniciando en el lado opuesto de la serie, de tal manera que al final tendremos  $2N_s$  segmentos.

Parte 3:

El objetivo de esta parte es quitar la tendencia de los perfiles  $Y(j), j = 1, \dots, N$  en cada segmento de longitud dada  $s$ , separadamente, el ajuste de mínimos cuadrados se hace mediante la determinación de funciones  $x^2$  (varianzas) por cada segmento  $v = 1, \dots, N_s$ .

$$F^2(u, s) = \frac{1}{s} \sum_{j=1}^s [Y((v-1)s + j) - w_v^n(j)]^2 \quad (10)$$

Mientras que para los segmentos  $v = N_s + 1, \dots, 2N_s$ , las varianzas correspondientes son definidas como:

$$F^2(u, s) = \frac{1}{s} \sum_{j=1}^s [Y((N-v-N_s)s + j) - w_v^n(j)]^2 \quad (11)$$

Donde  $w_v^n(j)$  es el polinomio de ajuste en el segmento  $v$  de orden  $n$  fijos para el resto del procedimiento.

Cabe destacar que con lo anterior se duplica la cantidad de datos en la serie de tiempos

Parte 4:

En esta parte distinguimos esencialmente dos casos:

Caso 1:  $q \neq 0$

Para este caso definimos la función de fluctuación como sigue:

$$F_q(s) = \left[ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(v, s)]^{q/2} \right]^{1/q} \quad (12)$$

Estamos interesados en cómo estas funciones de fluctuación q-dependientes dependen de la escala de tiempo s para diferentes valores de q-herencia, las partes 2 a 4 se pueden repetir para varias escalas de tiempo s.

Caso 2:  $q \rightarrow 0$

Para este caso definimos la función de fluctuación como sigue:

$$F_{q \rightarrow 0}(s) \rightarrow F_0(s) = \exp \left[ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln F^2(v, s) \right] \sim s^{h(q=0)} \quad (13)$$

Parte 5:

En esta parte determinamos el comportamiento a escala de las funciones de fluctuación por el análisis log-log de las trazas de  $F_q(s)$  vs s para diferentes valores de q llamada:

$$F_q(s) \sim s^{h(q)} \quad (14)$$

Donde  $h(q)$  es en exponente de Hurst generalizado. Para  $q=2$  la ecuación (12) es equivalente a la empleada en el análisis de fluctuaciones in tendencia. En la referencia (Kantelhardt, S.A., E., S., & Bunde, 2012), se muestra el hecho de que  $H(q)$  está relacionada con el exponente del formalismo multifractal  $\tau(q)$ , esto es:

$$\tau(q) = qH(q) - 1 \quad (15)$$

Por lo tanto, las dimensiones generalizadas se pueden obtener:

$$D(q) = \frac{\tau(q)}{q-1} = \frac{qH(q) - 1}{q-1} \quad (16)$$

## 2.8 Teoría de la información

Durante la primera mitad del siglo pasado y debido a la acelerada difusión y especialización que experimentaron los medios de comunicación se desarrolló un modelo del proceso de la comunicación, llamado Teoría de la información o teoría matemática de la comunicación. En la definición del concepto de comunicación de la teoría de la información quedan incluidos todos los procedimientos mediante los cuales una mente puede influir a otra, es decir se consideran todas las formas en las cuales se pueden transmitir las ideas, es decir desde la palabra escrita, hasta las comunicaciones por medios electrónicos (Shannon & Weaver, 1961).

Las ideas son en esencia información, por lo que la transmisión de esta información se puede generalizar a un más, de tal manera que, todo aquello que nos permita comprender y analizar un fenómeno de manera organizada será considerado que entrega o transmite una determinada información y por el contrario todo aquello que se aleje del orden será considerado lo opuesto a la información, en algunos casos se le nombra entropía o desorden (Jaynes, 1983).

### 2.8.1 Entropía de Shannon

La Teoría de la información es uno de los pocos campos científicos suficientemente afortunados de tener un principio identificable en 1948 mediante el papel de Claude Shannon. A menudo escuchamos que Claude Shannon es llamado el padre de la era Digital.

Tal vez el más eminente de los resultados de Shannon fue el concepto que cada canal de comunicación tenía un límite de velocidad, medido en dígitos binarios por segundo: éste es el famoso límite de Shannon, ejemplificado por la fórmula famosa y familiar para la capacidad de un canal de ruido blanco Gaussiano (Shannon & Weaver, 1961).

La mala noticia es que es matemáticamente imposible obtener comunicación libre de errores por encima del límite. No importa que tan sofisticado sea el sistema de corrección de error utilizado, o que tanto se pueda comprimir los datos, no se puede hacer que el canal vaya más rápido que el límite sin perder información (Shannon & Weaver, 1961).

La buena noticia es que, por debajo del límite de Shannon, es posible transmitir información con error cero. Shannon demostró matemáticamente que había formas de codificación de información que le permiten llegar al límite sin errores: independientemente de la cantidad de ruido o estática, o debilidad de la señal. En esencia, por debajo del límite, se podría hacer la probabilidad de error tan bajos como se requiera.

El Teorema de codificación de canal ruidoso es lo que dio lugar a todo el campo de códigos de corrección de errores y teoría de la codificación de canal: el concepto de la introducción de redundancia en la representación digital para proteger contra la corrupción.

Ahora se necesita una medida de la complejidad de la información transferida. Lo que realmente queremos aquí, es un promedio ponderado. Si observamos el símbolo  $x_i$ , llegaremos a estar recibiendo información de  $\log_b \frac{1}{p(x_i)}$  de la observación del símbolo  $x_i$  en particular. En un largo plazo (digamos N) de las observaciones, vamos a ver (aproximadamente) N  $x_i$  ocurrencias de símbolo  $x_i$  (en el sentido de los frecuentistas, que es lo que significa decir que la probabilidad de ver  $x_i$  es  $p(x_i)$ ). Así, en la observación de N (independiente), obtendremos información total:

$$I(x) = \sum_{i=1}^n p(x_i) \log_b \left( \frac{1}{p(x_i)} \right) \quad (17)$$

Que también puede ser expresada como:

$$S(x) = I/Np(x_i) = \sum_{i=1}^n p(x_i) \log_b \frac{1}{p(x_i)} \quad (18)$$

Donde  $S(x)$  es la entropía de Shannon,  $x_i$  es un símbolo o elemento en el mensaje y  $p(x_i)$  es la probabilidad del elemento  $x_i$ .

La expresión:

$$\log_b \frac{1}{p(x_i)} \quad (19)$$

También se conoce como "de incertidumbre" o "de sorpresa", para menores probabilidades  $p(x_i)$ , por ejemplo, para  $p(x_i) \rightarrow 0$ , mayor es la incertidumbre para  $u_i \rightarrow \infty$ , para el resultado  $x_i$ .

La ecuación (16) es conocida como entropía de Shannon (Shannon & Weaver, 1961).

### 2.8.2 Entropía de Tsallis

Considérese un sistema con  $n$  estados posibles, cada uno con probabilidad  $p_i$ , ( $1 \leq i \leq n$ ) tales que  $\sum_{i=1}^n p_i = 1$ . La Entropía de Boltzman–Gibbs–Shannon (BGS) es el fundamento de la mecánica estadística clásica y se define para  $p = (p_1, \dots, p_n)$  mediante

$$S(p) = k_B \sum_{i=1}^n p_i \log \left( \frac{1}{p_i} \right) \quad (20)$$

donde  $k_B$  es la constante de Boltzman, normalmente se escoge un sistema de unidades tal que  $k_B=1$ , y  $\log$  es el logaritmo natural (Tsallis, 1988).

Si bien la propuesta de Boltzmann–Gibbs ha mostrado efectividad (Borja, Cartagena, Loaiza, Molina, & Puerta, 2010), su uso se restringe a los sistemas conocidos como aditivos o extensivos; tales sistemas son aquellos que cumplen la propiedad de aditividad  $S(A + B) = S(A)+S(B)$ , donde  $A$  y  $B$  son dos sistemas independientes en el sentido probabilístico. Para sistemas no extensivos es necesario generalizar la Entropía de forma tal que se cubran los casos subaditivos o superaditivos. En este sentido, la estadística no extensiva de Tsallis es considerada como un nuevo paradigma en mecánica estadística.

La Entropía  $S_q$ , definida por Tsallis, para cada real  $q$ , con  $q \neq 1$ , está dada por

$$S_q(p) = k \frac{1 - \sum_{i=1}^n (p_i)^q}{q - 1} \quad (21)$$

y toma el valor  $S(p)$  si  $q = 1$ .

La entropía  $S_q$  de Tsallis tiene las siguientes propiedades:

1.  $S_q$  es extensión de  $S$  en el siguiente sentido: si  $q \rightarrow 1$  se tiene  $S_q \rightarrow S$ .
2.  $S_q(P) = 0$  solo en el caso de que alguna  $p_i = 1$ .
3. El máximo para  $S_q$ , bajo la única restricción  $\sum_i p_i = 1$ , se obtiene para  $p_i = \frac{1}{n}$ , para

$$1 \leq i \leq n.$$

4. Propiedad de no extensividad. Supóngase que se tienen dos sistemas independientes  $A$  y  $B$ , en sentido probabilístico, esto es  $P_{ij}^{A+B} = P_i^A + P_j^B$ . Entonces la Entropía de Tsallis se comporta como sigue.

$$\frac{S_q(A+B)}{k} = \frac{S_q(A)}{k} + \frac{S_q(B)}{k} + (1-q) \frac{S_q(A)S_q(B)}{k^2} \quad (22)$$

Del último numeral se concluye que: para  $q > 1$  se tienen sistemas superextensivos, para  $q = 1$  (el caso de Boltzmann–Gibbs–Shannon) se tienen sistemas extensivos y para  $q < 1$  sistemas subextensivos.

## 2.9 Redes neuronales artificiales como clasificadores

Las redes neuronales artificiales son un modelo simplificado del cerebro humano, el cual se basa en la habilidad de conseguir conocimiento a través de la experiencia, siendo un nuevo modelo de tratamiento de la información cuya unidad básica de procesamiento está basada en la unidad fundamental del sistema nervioso humano, la neurona. La neurona se interconecta con otras para generar una unidad de cómputo en conjunto muy poderosa (Hilera & Martinez, 2000).

### Historia de las Redes Neuronales

En 1943 Warren S. McCulloch y Walter Pitts. Fueron los primeros teóricos que concibieron los fundamentos de las redes neuronales, al publicar una teoría acerca de la forma de trabajar de las neuronas (Un Cálculo Lógico de la Inminente Idea de la Actividad Nerviosa - Boletín de Matemática Biofísica 5: 115-133). Propusieron el primer modelo de neurona artificial (McCulloch & Pitts, 1943).

Más tarde en 1949 - Donald O. Hebb. Fue el primero en explicar los procesos del aprendizaje desde un punto de vista psicológico, desarrollando una regla de como el aprendizaje ocurría. Este es el fundamento bajo el cual se encuentran la mayoría de las funciones de aprendizaje que pueden aplicarse en una red neuronal. Su idea fue que el aprendizaje ocurría cuando se activaban ciertos cambios en una neurona (Hebb, 1949).

Frank Rosenblatt, en 1957 comenzó el desarrollo del Perceptrón que es la red neuronal más antigua; utilizándose para aplicaciones como reconocedor de patrones. Este modelo era capaz, después de haber aprendido una serie de patrones, de reconocer otros patrones similares, diferentes a los que se le han presentado anteriormente. Sin embargo, tenía una serie la limitación de ser incapaz de clasificar clases no separables linealmente (Rosenblatt, 1957).

Ya para 1960 - Bernard Widrow y Marcial Hoff. Desarrollaron el modelo Adaline (ADAPTative LINear Elements). Esta fue la primera red neuronal aplicada a un problema real (filtros adaptativos para eliminar ecos en las líneas telefónicas) que se ha utilizado comercialmente durante varias décadas. Y en 1961 Karl Steinbuch desarrolla la Lernmatrix (memoria asociativa que es capaz de recuperar una imagen o conocimiento, aun cuando el medio este viciado). Para 1969 surgieron críticas que frenaron, hasta 1982, el crecimiento que estaban experimentando las investigaciones sobre redes neuronales. Minsky y Papert, del Instituto Tecnológico de Massachussets (MIT), publicaron un libro Perceptrons, en el cual probaron (matemáticamente) que el Perceptrón no era capaz de resolver problemas relativamente fáciles, tales como el aprendizaje de una función no-lineal. A pesar del libro, algunos investigadores continuaron su trabajo. Tal fue el caso de James Anderson, que desarrolló un modelo lineal, llamado Lineal Associator (Widrow, 1960).

En 1974 Paul Werbos desarrolló la idea básica del algoritmo de aprendizaje de propagación hacia atrás (backpropagation); cuyo significado quedó definitivamente aclarado en 1985, pero antes en 1982 John J. Hopfield provoca el renacimiento de las redes neuronales con su libro: "Computación neuronal de decisiones en problemas de optimización." Durante 1986 Rumelhart y Hinton redescubren el algoritmo de aprendizaje de propagación hacia atrás (backpropagation). A partir de 1986, el panorama fue alentador con respecto a las investigaciones y el desarrollo de las redes neuronales (Werbos, 1974), (Hilera & Martinez, 2000) y (Valencia, Yáñez, & Sánchez, 2006).

Durante la primera parte del siglo XXI, los sistemas basados en algoritmos de aprendizaje se han propagado en nuestra vida diaria, desde los buscadores de internet y asistentes personales como Cirí, hasta en los campos más delicados como en las predicciones económicas, es un campo que está actualmente en desarrollo sin lugar a dudas.

### 2.9.1 Elementos de una Red Neuronal

Las redes neuronales están integradas por unidades de cómputo se conocen como neuronas o neurodos, estos pueden ser de tres tipos, de entrada, los cuales reciben la información o datos del exterior de la red neuronal, unidades ocultas, las cuales interactúan en las capas internas de la red neuronal y finalmente están las unidades de salida, las cuales son las encargadas de entregar la respuesta de la red neuronal, ubicadas en la capa de salida, en la siguiente figura se representa una red neuronal con la disposición de estos elementos (Valencia, Yáñez, & Sánchez, 2006).

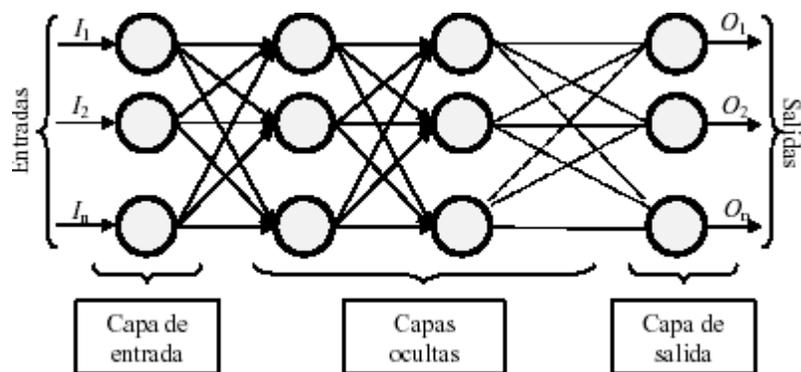


Figura 4 : Esquema de una red Neuronal (Valencia, Yáñez, & Sánchez, 2006)

Las flechas que unen a las diferentes neuronas entre si son denominadas sinapsis.

La neurona artificial pretende imitar las características más importantes de las neuronas biológicas. Cada neurona  $i$ -ésima está caracterizada en cualquier instante por un valor numérico denominado valor o estado de activación  $a_i(t)$ , asociado a cada unidad, existe una función de salida,  $f_i$ , se transforma el estado actual de activación en una señal de salida. Dicha señal es enviada a través de los canales de comunicación unidireccionales a otras unidades de la red; en estos canales la señal se modifica de acuerdo con la sinapsis (el peso  $w_{ij}$ ) asociada a cada uno de ellos según determinada regla. Las señales moduladas que han llegado a la unidad  $j$ -ésima se combinan entre ellas, generando así la entrada total  $Net_j$ .

$$Net_j = \sum_i y_i w_{ij} \quad (23)$$

Una función de activación,  $F$ , determina el nuevo estado de activación  $a_i(t + 1)$  de la neurona, teniendo en cuenta la entrada total calculada y el anterior estado de activación  $a_i(t)$ .

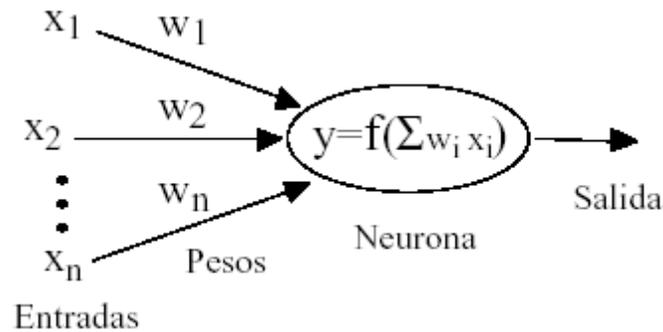


Figura 5 Neurona de una red neuronal artificial (Valencia, Yáñez, & Sánchez, 2006).

### 2.9.2 Backpropagation

Según (Colmenares, 2000) el algoritmo backpropagation, también conocido como retropropagación del error o método para calcular el gradiente del error, es desarrollado como un mecanismo para el aprendizaje de redes neuronales multicapa.

Es un método de entrenamiento general para redes multicapa, requiriendo diferenciabilidad en la capa de salida. Es decir, la función debe ser diferenciable (sigmoide). En la siguiente figura podemos observar una neurona básica en backpropagation, con sus capas de entrada, oculta y de salida. La de salida activada con una función sigmoide que describe valores de  $z$  entre 0 y 1.

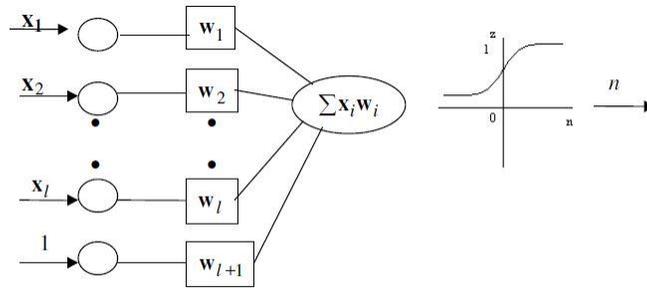


Figura 6 Neurona básica en backpropagation

Donde:

$$z = \frac{1}{1 + e^{-n}} \quad (24)$$

Si  $n \gg 0$  entonces  $z \rightarrow 1$

Si  $n \ll 0$  entonces  $z \rightarrow 0$

Con las siguientes características:

1. Las neuronas transforman una entrada no- restringida en una señal limitada z
2. La función sigmoideal restringe el rango de z entre 0 y 1.
3. Es diferenciable. Hay que tener presente que la no linealidad es una característica resaltante de este algoritmo.
4. Las entradas a la unidad de procesamiento son ejemplos de entrada o salidas de la capa previa

En una red cuya arquitectura incluye capas ocultas con varias neuronas (nodos) ocultas, podemos observar las siguientes características de diseño.

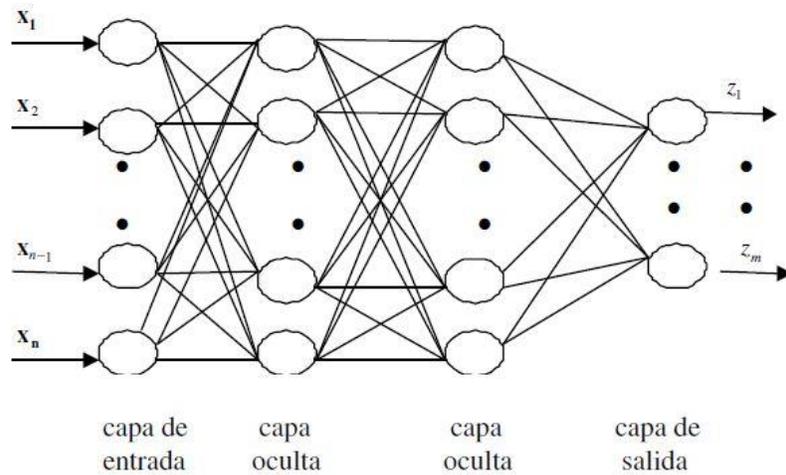


Figura 7 Red neuronal multicapa

Está conectada hacia adelante.

Está compuesta con una capa de entrada varias ocultas y una de salida.

La capa de entrada sirve únicamente como distribuidor.

Una neurona está asociada con un conjunto de pesos que conecta a sus entradas.

Los pesos en la capa uno, terminan en las neuronas de la capa uno.

En backpropagation, el método general de entrenamiento se resume en los siguientes cuatro pasos:

- 1.- Selecciona aleatoriamente y sin repetición un vector de entrada desde el conjunto de entrenamiento
- 2.- Aplica esta entrada a la red y calcula la salida.
- 3.- Calcular el error entre la salida calculada y la salida deseada.
- 4.- Ajustar los pesos para que el error cometido entre la salida calculada y la salida deseada disminuya.

5.- Se repiten los pasos para todas las entradas del conjunto de entrenamiento, hasta que el error global sea aceptablemente bajo.

### 2.9.3 Ajuste de los pesos en la capa de salida

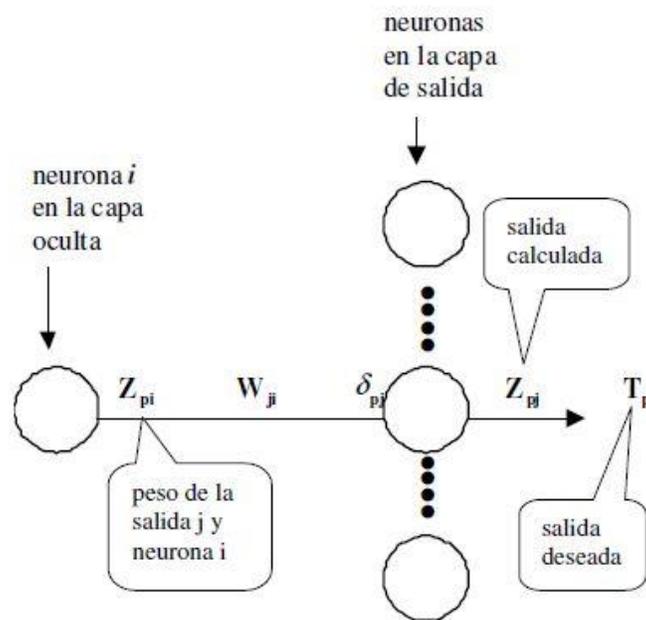


Figura 8 Ajuste de pesos en capa de salida

Según (Colmenares, 2000) el valor de salida deseado  $T_{pj}$  está disponible en la capa de salida, entonces los pesos  $w_{ji}$  son ajustados utilizando algún algoritmo tal como la regla delta y las constantes de aprendizaje.

Para un patrón de entrada  $p$  determinado, si  $w_{ji}$  es el peso que se corresponde con la salida  $j$  y la neurona oculta  $i$ , entonces el valor de corrección entre la salida deseada  $T_{pj}$  y la calculada  $z_{pi}$  para ese patrón  $p$  expresada por  $\delta_{pj}$  y la salida correspondiente a la neurona  $i$  dada por  $z_{pi}$ , podría corregir el valor del peso  $w_{ji}$ , es decir:

$$\Delta_p w_{ji} = \eta \delta_{pj} z_{pi} \quad (25)$$

Donde:

$$\delta_{pj} = z_{pj}(1 - z_{pj})(T_{pj} - z_{pj}) \quad (26)$$

#### 2.9.4 Ajuste de pesos en las capas ocultas

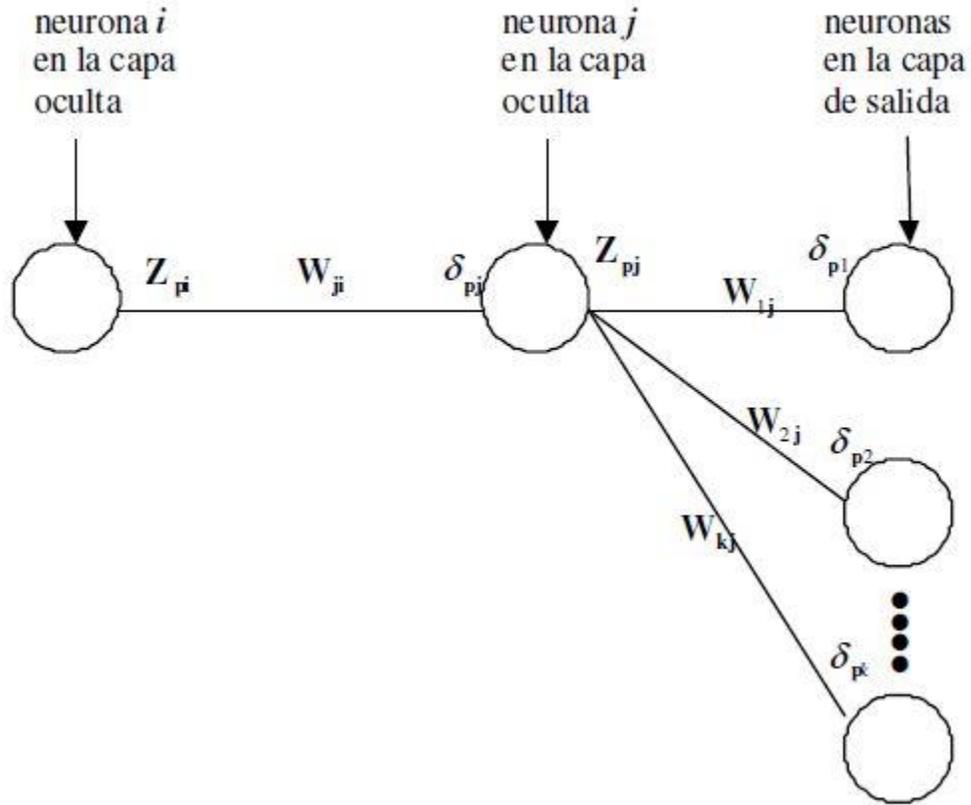


Figura 9 Ajuste de pesos en las capas ocultas

Según (Colmenares, 2000) cuando la actualización de pesos es en las capas ocultas se calcula  $\delta$  propagando hacia atrás desde la capa de salida.

$$\delta_{pj} = z_{pj}(1 - z_{pj}) \left( \sum_k \delta_{pk} w_{kj} \right) \quad (27)$$

Y luego

$$\Delta_p w_{ji} = \eta \delta_{pj} z_{pi} \quad (28)$$

Repetiendo para cada capa oculta en la red.

### 2.9.5 Derivaciones

Podemos derivar, según (Colmenares, 2000), de la siguiente manera:

- 1.- Backpropagation es un método de descenso por el gradiente.
- 2.- Minimiza el cuadrado de las diferencias entre los valores calculados y deseados de las salidas sumados sobre todas las unidades de salida y todos los pares, entradas y salidas, del conjunto de entrenamiento.
- 3.- Para un conjunto de entrenamiento  $p$  (entradas y salidas) y  $n$  unidades de salidas

$$E_p = \frac{1}{2} \sum_{j=1}^n (T_{pj} - z_{pj})^2 \quad (29)$$

- 4.- El error total sobre todo el conjunto sería:

$$E = \sum_p E_p \quad (30)$$

## Capítulo 3. Obtención de medidas complejas

En el presente capítulo se aborda la obtención de las variables de entrada de la red neuronal que se implementará en el siguiente capítulo.

Primeramente, hablaré del análisis multifractal de fluctuaciones sin tendencia (MFDFA por sus siglas en inglés), en el cual obtendremos dos de las principales variables que emplearemos para entrenar a la red neuronal. Posteriormente obtendremos la tercera medida, mediante la utilización del concepto de entropía según Tsallis.

Para poder obtener las medidas complejas de los electroencefalogramas, se siguió el siguiente procedimiento:

1. Se tomo únicamente la serie de tiempos del canal Fz de cada uno de los electroencefalogramas en la base de datos
2. A la serie de tiempos se le aplicó el análisis multifractal de fluctuaciones sin tendencia, del cual solo se tomó el exponente de Hurts de orden  $q$  y el exponente de masa de orden  $q$ .
3. Se calculo la entropía según Tsallis sobre la serie de tiempos, considerando el espectro de Fourier normalizado como distribución de probabilidad.
4. Posteriormente se formó la tabla se entrenamiento con el exponente de Hurts de orden  $q$ , el exponente de masa de orden  $q$ , la entropía según Tsallis y  $q$ .

### 3.1 MFDFA

A través del algoritmo presentado en la sección 2.7.1 del capítulo anterior, vamos a obtener una serie de medidas, las cuales están relacionadas al comportamiento multifractal de la serie de tiempos obtenida a través del electroencefalograma.

La primera medida a obtener es el exponente de Hurst generalizado, que dentro de la geometría fractal es detonado por  $H$  o  $H_q$  en honor a Harold Edwin Hurst y Ludwig Otto Hölder por Benoît Mandelbrot (Mandelbrot & Wallis, 1968), y está relacionado con la dimensión fractal,  $D$ , y es una medida de la aleatoriedad “leve” o “alta” de una serie de datos

(Mandelbrot & Hudson, 2006). La segunda medida a tener en cuenta es el exponente de masa, que es la transformada de Legendre de el exponente de Hurst  $H_q$ .

También nos va a interesar el momento de fluctuaciones  $q$ , el cual vamos a utilizar tanto en el MF DFA como en la obtención de entropía de Tsallis.

A continuación, se resume el algoritmo del MF DFA:

Tenemos una serie de tiempos  $x(t), t = 1, 2, \dots, N$ .

Construye una su suma acumulativa  $u(t) = \sum_{i=1}^t x(i), t = 1, 2, \dots, N$

$u(t)$  es particionada en  $N_s$  segmentos disjuntos del mismo tamaño  $s$ , donde  $N_s = N/s$ .

Cada segmento se puede denotar por  $u_v(i) = u(l + i)$  para  $1 \leq i \leq s$  donde  $l = (v - 1)s$

En cada segmento  $u_v$  determinamos la tendencia local  $\tilde{u}_v$  mediante un ajuste polinómico.

Obtenemos la secuencia residual  $\epsilon_v = u_v(i) - \tilde{u}_v(i), 1 \leq i \leq s$

La fluctuación sin tendencia general es:

Para  $q \neq 0$

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [F(v, s)]^q \right\}^{1/q}$$

Para  $q = 0$

$$F_0(q) = \exp \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} \ln[F(v, s)] \right\}$$

Variando los valores de  $s$  podemos determinar la dependencia de ley de potencias de la función de fluctuación sin tendencia  $F_q(s)$  en la escala  $s$ .

$$F_q(s) \sim s^{H_q}$$

Donde el exponente  $H_q$  es el exponente de Hurst generalizado.

Es exponente de masa  $\tau_q$  se obtiene para cada  $q$ :

$$\tau_q = qH_q - 1$$

De los datos obtenidos al aplicar el MF DFA nos interesan tres de las variables resultantes que son  $q$ ,  $H_q$  y  $\tau_q$ .

En las figuras 10 y 11 podemos observar el comportamiento del exponente de Hurts contra  $q$  para un paciente alcohólico y uno no alcohólico respectivamente.

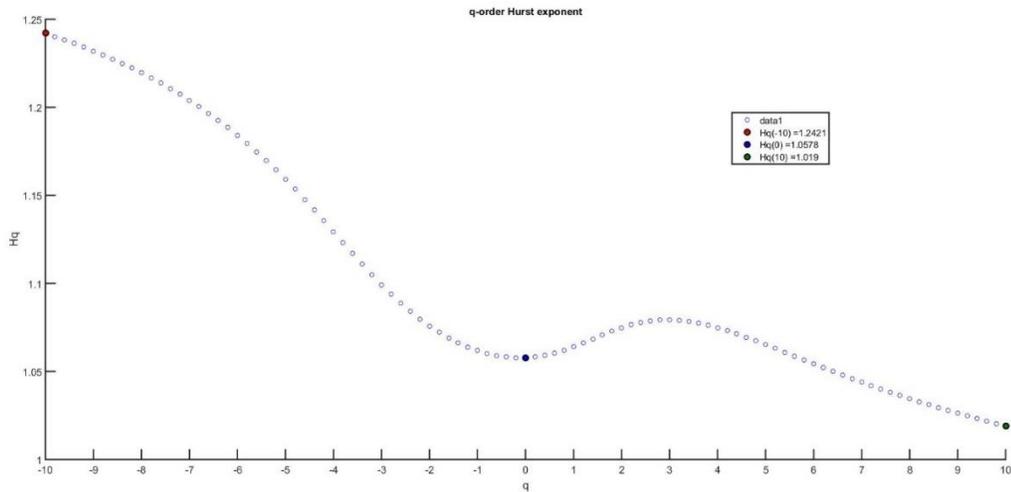


Figura 10 Exponente de Hurst de orden  $q$  no Alcohólico

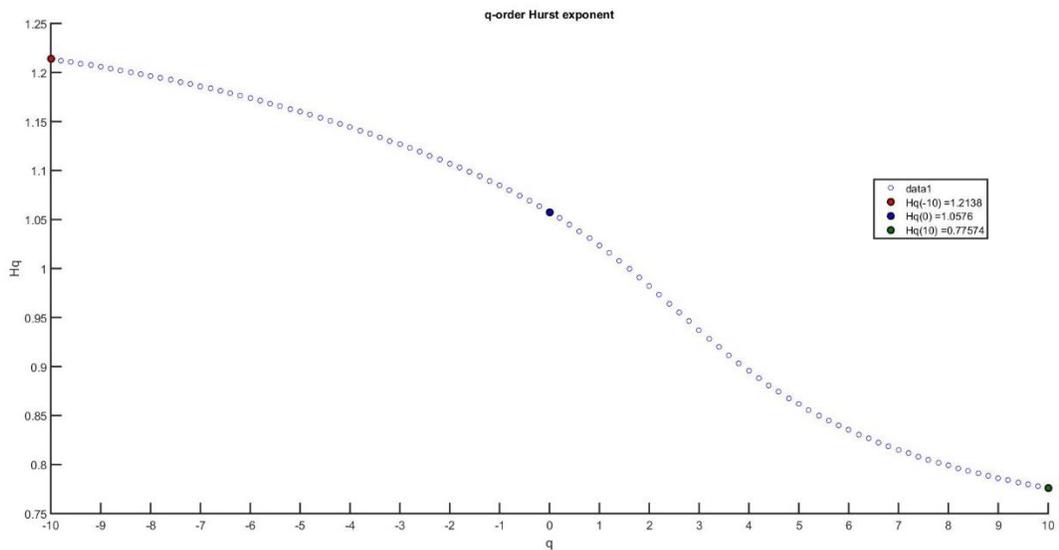


Figura 11 Exponente de Hurst de orden  $q$  Alcohólico

Como podemos observar en las figuras 19 y 11 el comportamiento del exponente de Hurst es diferente en ambos casos, ahora, si obtenemos un ajuste polinomial para cada una de las curvas  $H_q(q)$  mostradas en las figuras 10 y 11, podemos tener una función que relaciona ambas variables y que nos permite ver mejor las diferencias entre ambos comportamientos, en las figuras 12 y 13 podemos observar dicho ajuste polinomial trazado en rojo.

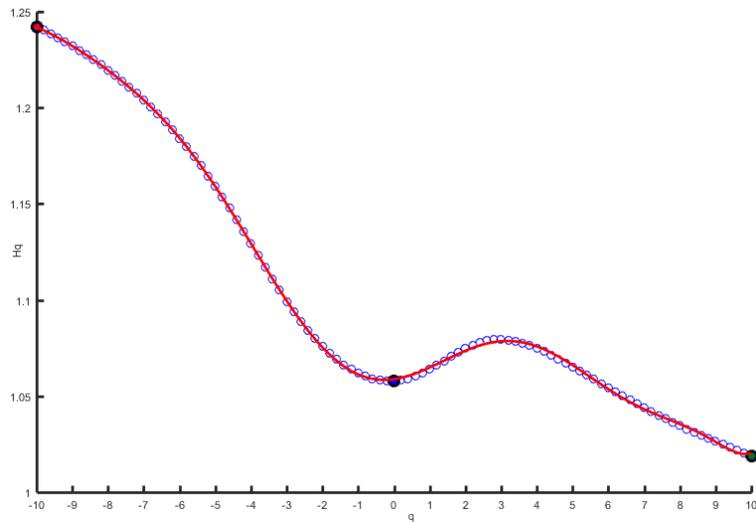


Figura 12 Exponente de Hurst de orden  $q$  no Alcohólico con su ajuste polinomial

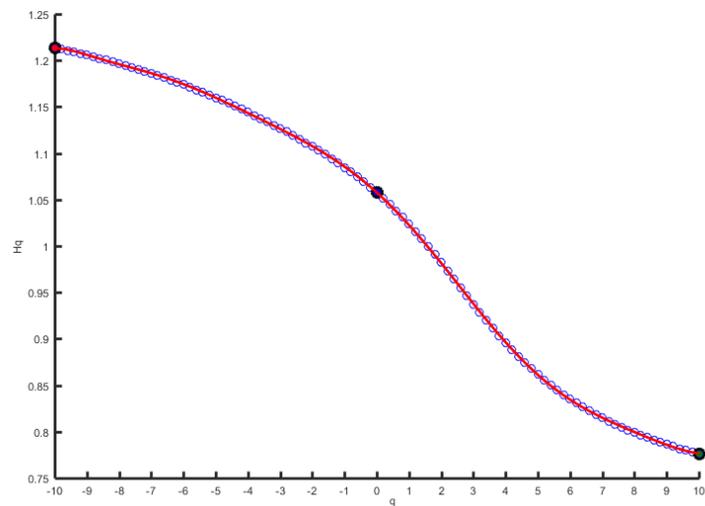


Figura 13 Exponente de Hurst de orden  $q$  Alcohólico con su ajuste polinomial

Podemos expresar los polinomios obtenidos del ajuste polinomial de la siguiente manera:

Para el paciente no alcohólico:

$$H_q = 6.8 \times 10^{-11} q^{10} + 5.9 \times 10^{-10} q^9 - 2 \times 10^{-8} q^8 - 1.7 \times 10^{-7} q^7 + 2.4 \times 10^{-6} q^6 \\ + 1.8 \times 10^{-5} q^5 - 0.00014 q^4 - 0.00082 q^3 + 0.0044 q^2 + 0.0026 q \\ + 1.1$$

Para el paciente alcohólico:

$$H_q = -3.8 \times 10^{-11} q^{10} + 3.3 \times 10^{-10} q^9 + 1.2 \times 10^{-8} q^8 - 7.9 \times 10^{-8} q^7 - 1.6 \times 10^{-6} q^6 \\ + 6 \times 10^{-6} q^5 + 0.0001 q^4 - 4.3 \times 10^{-5} q^3 - 0.0037 q^2 - 0.031 q \\ + 1.1$$

Donde podemos apreciar que son diferentes funciones.

Luego podemos graficar el comportamiento del exponente de Hurst en todos los pacientes, tanto los alcohólicos como lo no alcohólicos, para corroborar que existe gráficamente, alguna diferencia en el comportamiento de esta variable. Que es lo que se realiza en las figuras 14 y 15, en donde vemos concentrados todos los exponentes de Hurst para todos los pacientes, y en donde podemos apreciar que en efecto existe una diferencia en dicho exponente, lo que revela también una diferencia en la dimensión multifractal en las series de tiempos de los eegs, para cada tipo de paciente.

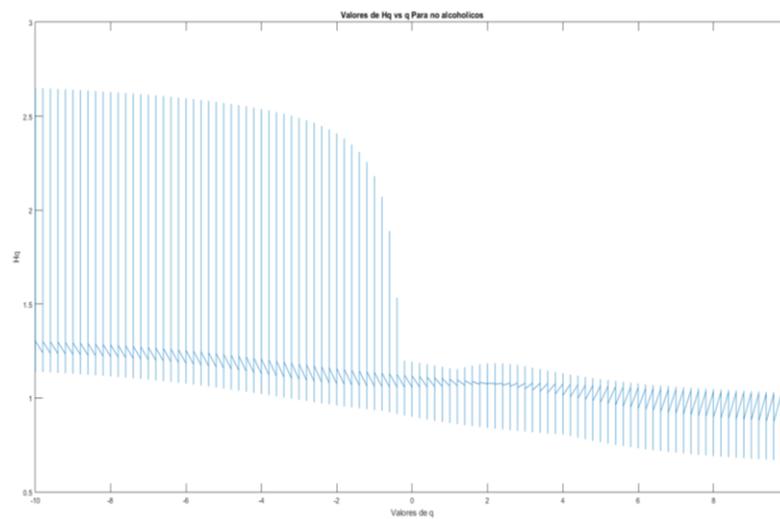


Figura 14 Exponente de Hurst  $H_q$  vs  $q$  para todos los pacientes no Alcohólicos

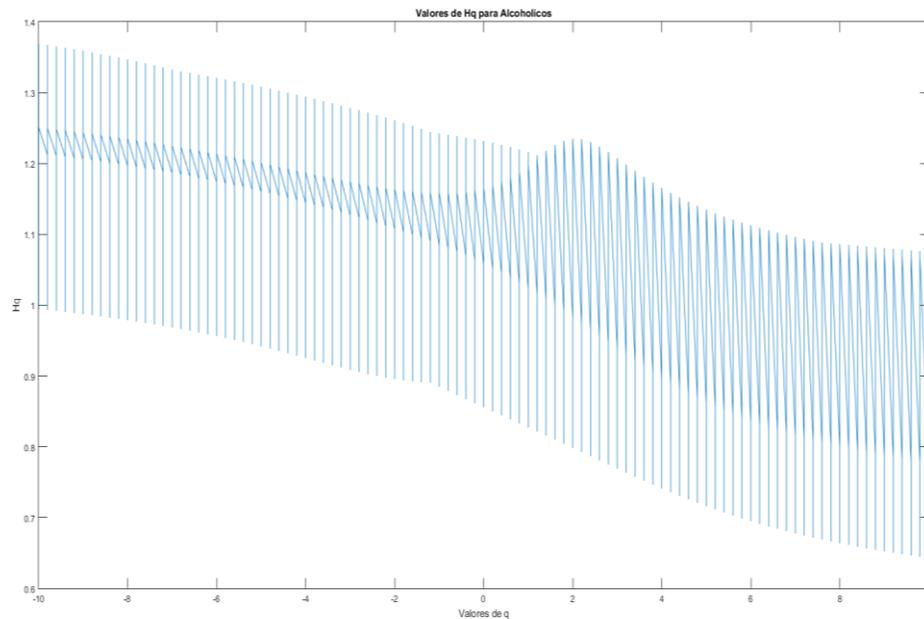


Figura 15 Exponente de Hurst Hq vs q para todos los pacientes Alcohólicos

### 3.2 Entropía de Tsallis.

Una vez que obtuvimos las características multifractales de las series de tiempos, se precedió a calcular la entropía de Tsallis, para lo cual se consideró el siguiente algoritmo:

Aplica FFT a la serie de tiempos X(t)

$p_i$  será la parte real normalizada de la FFT de la serie de tiempos x(t)

Si  $q \neq 1$

$$S_q(p) = k \frac{1 - \sum_{i=1}^n (p_i)^q}{q - 1}$$

Si  $q = 1$

$$S(p) = k_B \sum_{i=1}^n p_i \log \left( \frac{1}{p_i} \right)$$

En este caso, si graficamos las entropías calculadas para cada q en podemos observar que tienen un comportamiento muy similar, tanto en los pacientes alcohólicos como en los pacientes ni alcohólicos.

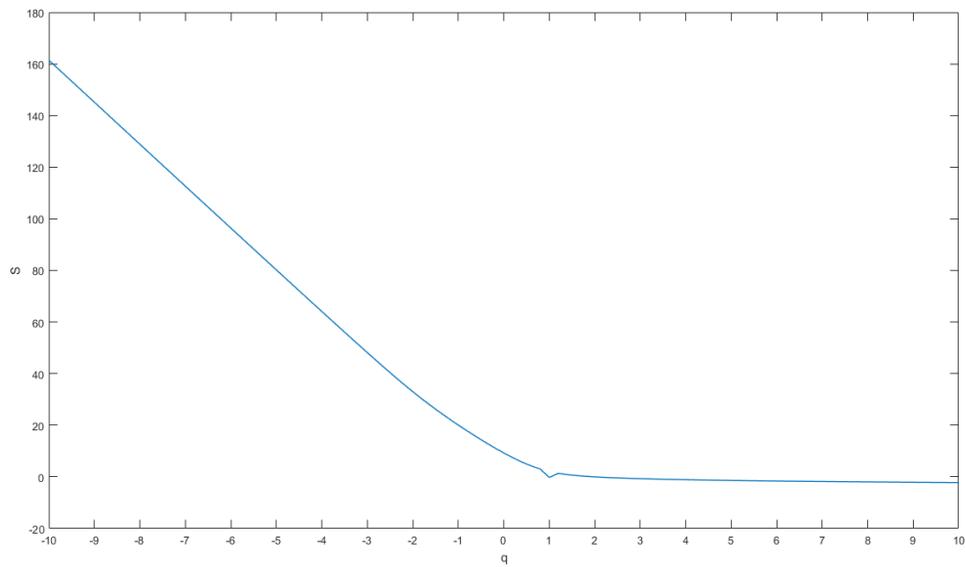


Figura 16 Entropía Vs q para un paciente no Alcohólico

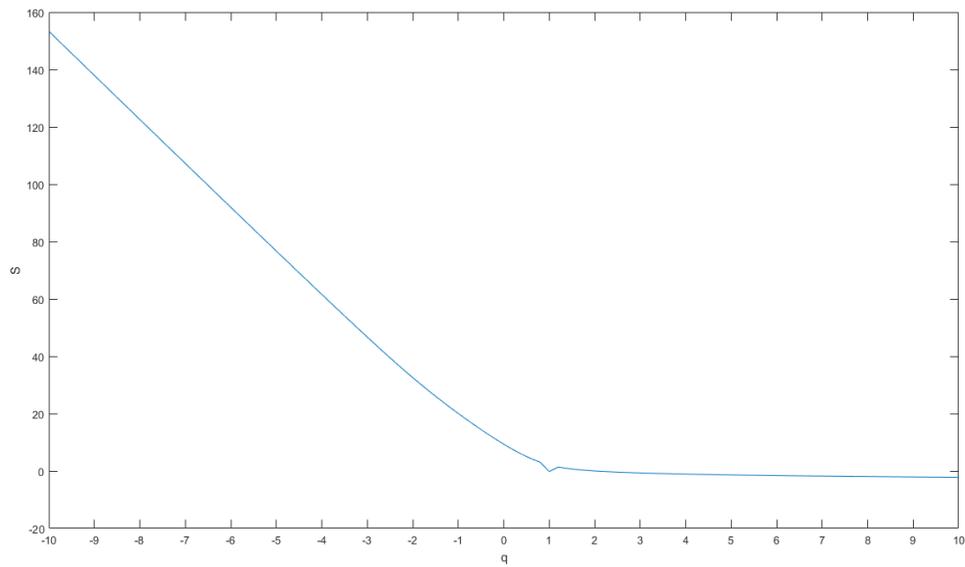


Figura 17 Entropía Vs q para un paciente Alcohólico

En el caso de las figuras 16 y 17, es lo que se está mostrando, donde podemos ver que la mayor diferencia es que para las fluctuaciones más pequeñas, la entropía mayor en el caso de la persona alcohólica y para las fluctuaciones más grandes la entropía se hace asintótica en cero para ambos casos.

### 3.3 Solución propuesta

Se propone que existe una relación tal que dada las variables obtenidas en las secciones 3.1 y 3.2 nos permita clasificar entre los pacientes alcohólicos y los pacientes no alcohólicos, esto es, que dado un conjunto  $\Omega$  de entradas se obtenga una de dos posibles soluciones, 1 para el caso de alcohólico y 0 para el caso de no alcohólico.



Figura 18 Propuesta de solución

En la figura 18 se muestra la propuesta de solución donde  $\Omega$  se define cómo:

$$\Omega = \{S_q, H_q, \tau_q, q\}$$

Que es el conjunto de entradas.

Se resalta que en este caso el conjunto  $\Omega$ , y la relación que existe entre  $\Omega$  y la salida, constituyen la aportación central del presente trabajo, ya que se considera que el conjunto de entrada contiene la cantidad de información suficiente para determinar si existe o no una determinada condición en el cerebro humano.

En este caso la relación será una red neuronal artificial multicapa, para lo cual se propone utilizar un método de entrenamiento supervisado, como lo es backpropagation, que será quien actúe como clasificador. En el capítulo siguiente se aborda el entrenamiento de la red neuronal.

### 3.4 La base de datos

Se ha hablado anteriormente de que se realiza el MF DFA y se obtiene la entropía según Tsallis, esto se aplica sobre un conjunto de datos tomados de una base de datos pública la cual se describe a continuación.

Se escogió para realizar el presente trabajo una base de datos pública de electroencefalogramas, tomada de Psychophysics, siendo el propietario original Henri Begleiter, del laboratorio de neurodinámica, de la State University of New York, Health Center, ubicada en Brooklyn, New York, siendo el donador Lester Ingber, POB 06440 Sears Tower (Center for Machine Learning and Intelligent Systems, 1999), conteniendo los datos de 122 voluntarios que realizaron tareas de reconocimiento visual (en este caso son pacientes alcohólicos y no alcohólicos), registrando electroencefalogramas de 64 canales a una frecuencia de muestreo de 256Hz.

Los registros de la base de datos se obtuvieron exponiendo a cada uno de los sujetos aleatoriamente a cualquiera de las dos situaciones siguientes:

- a) fue expuesto a un solo estímulo (S1)
- b) fue expuesto a dos estímulos (S1 y S2)

Como estímulos se eligieron fotografías de objetos elegidos de la base de datos de imágenes de 1980 de Snodgrass y Vanderwart (Snodgrass & Vanderwart, 1980). Cuando ambos estímulos se muestran, se presentaron una de dos condiciones, la primera donde la pareja S1 y S2 era idéntica y la segunda donde S1 era diferente de S2. Cada uno de los 122 sujetos completaron 120 muestras donde se mostraron diferentes estímulos.

Las posiciones de los electrodos se encuentran preestablecidos de acuerdo a la Standard Electrode Position Nomenclature, American Electroencephalographic Association (American Clinical Neurophysiology Society, 2006). Se tiene en promedio 1,447,776 muestras por paciente, dando como resultado 176,628,672 datos de series de tiempos de electroencefalogramas.

Cabe destacar que, para fines del presente trabajo, solo se empleó esta base de datos, y que a futuro se pretende hacer una prueba más extensa con un mayor volumen de casos clínicos, ya que, no fue posible validar el clasificador fuera de la base de datos.

### 3.5 Tabla de entrenamiento

Una vez que se obtuvieron todas las medias, se procedió a juntarlas en una tabla de entrenamiento, la cual contiene todas las medias obtenidas, un ejemplo de esta tabla de entrenamiento se puede ver la tabla 2, en la cual se presenta una muestra de los vectores de entrenamiento empleados para entrenar la red neuronal.

Hq	Tq	s	q	Alcohólico / no alcohólico
<b>1.06350749</b>	-1.2127015	81114.3158	-0.2	1
<b>1.01349318</b>	-1.20269864	80616.4469	-0.2	1
<b>1.10286318</b>	-1.22057264	62593.3618	-0.2	1
<b>0.95033356</b>	-1.19006671	73444.6853	-0.2	1
<b>0.86154734</b>	-1.17230947	114815.849	-0.2	1
<b>0.95443411</b>	-1.19088682	98561.7513	-0.2	1
<b>0.98790672</b>	-1.19758134	98543.6653	-0.2	1
<b>1.09961818</b>	-1.21992364	77828.0678	-0.2	1
<b>1.03273161</b>	-1.20654632	92452.0964	-0.2	1

Tabla 2 Muestra de tabla de entrenamiento

Se presenta de tal manera que los primeros 4 valores corresponden a las cuatro variables de  $\Omega$  y la quinta columna corresponde la clasificación previa en la base de datos, correspondiente a 0 para no alcohólico y a 1 para alcohólico, teniendo como resultado un total de 11,111 vectores de entrenamiento, de los cuales se emplean un 70% para entrenar y el resto para probar la red neuronal.

## Capítulo 4. Entrenamiento de la red neuronal

En el presente capítulo se aborda el entrenamiento de la red neuronal, Los diferentes frameworks que se emplearon para ello y los problemas que se solucionaron para la implementación final del clasificador.

### 4.1 Entrenamiento de la red neuronal

Para entrenar la red neuronal artificial se emplearon varios frameworks, el primero de ellos fue el NeuroSolutions, que una herramienta de pago, la cual nos permite manejar grandes cantidades de datos y emplear computo distribuido para el entrenamiento de redes neuronales, en un principio se emplea para corroborar la existencia de la solución propuesta en la sección 3.3 del capítulo anterior.

En el encontramos que, si se emplea una red neuronal del tipo backpropagation de 3 capas ocultas y de 100 neuronas por capa, esta puede clasificar con un error de menor al 3%, a ambos grupos de pacientes, como se muestra en la tabla 3.

Red Neuronal	Resultado	Error
0.98275887	1	0.01724113
0.98275887	1	0.01724113
0.98275887	1	0.01724113
0.98275887	1	0.01724113
0.98275887	1	0.01724113
0.98275887	1	0.01724113
0.98275887	1	0.01724113
0.02190846	0	0.02190846
0.02190846	0	0.02190846
0.02190846	0	0.02190846
0.02190846	0	0.02190846
0.98503859	1	0.01496141
0.98503859	1	0.01496141
0.98503859	1	0.01496141
0.98503859	1	0.01496141

<b>0.98503859</b>	1	0.01496141
<b>0.98503859</b>	1	0.01496141
	Error promedio	0.01753472
	Porcentaje de Éxito	98.2465284

*Tabla 3 . Resultados de res neuronal entrenada en neurosolutions*

Sin embargo, al ser un producto de pago, el framework no permitió emplear la red neuronal fuera de él, ya que se empleó la versión de prueba del mismo, por lo que se procedió a buscar otra solución al problema.

Como segunda aproximación se empleó el software Multiple back propagation, que tiene las siguientes características que es una aplicación libre (bajo licencia GPL V3), con código abierto, para entrenar redes neuronales con los algoritmos Back-Propagation y Multiple Back-Propagation.

Fue diseñado por Noel de Jesus Mendonça Lopes, del Instituto Politécnico da Guarda, de 1999 a 2016, Se emplea en el presente trabajo sobre todo porque es un software libre y tiene el código abierto, lo que facilita el poder realizar (alguna) modificaciones en la aplicación en caso de que fuera necesario.

El software en cuestión emplea los algoritmos es entrenamiento propuestos y discutidos en (Lopes & Ribeiro, Hybrid learning multi neural architecture, 2001), (Lopes & Bernardete, 2003), (Lopes & Bernardete, 2009), (Lopes & Bernardete, 2010) y (Lopes & Bernardete, 2011), obteniendo con la arquitectura presentada en la figura 19, una red neuronal operativa con un error promedio de 0.186, demasiado alto para nuestros fines.

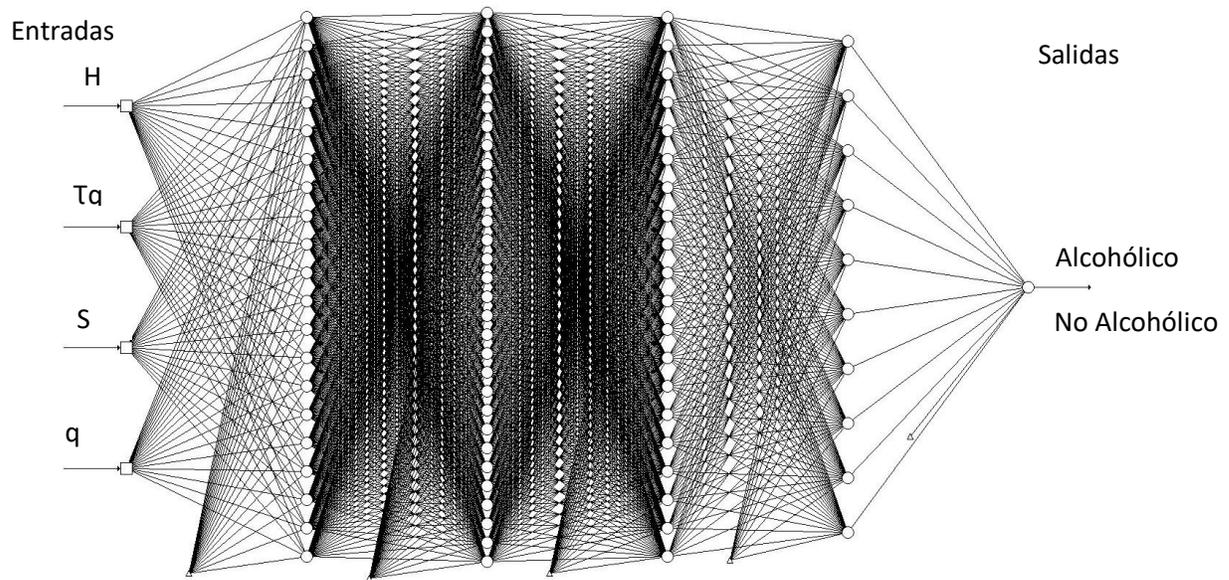


Figura 19 Topología de la red neuronal con 4 capas ocultas (20, 30, 20 y 10 neuronas en cada capa oculta)

El empleo de la arquitectura presentada es debido a limitaciones del propio software.

En este caso, obtenemos otra aproximación a la solución propuesta que si bien con el problema de tener el error muy alto, cercano al 18%. Por lo que se planteó fraccionar la tabla de entrenamiento de la red neuronal, primero por valores positivos y negativos de  $q$  y luego por cada valor de  $q$ , con lo cual el error presente en las redes resultantes se redujo, las redes neuronales resultantes tienen la misma arquitectura presentada en la figura 19.

En la tabla 4 se concentran los errores obtenidos en las 101 redes neuronales resultantes de separar en cada valor  $q$  la tabla de entrenamiento, con las cuales se logró reducir el error a 0.02028 (2.028%) en promedio, con una desviación estándar de 0.01896 y una varianza de 0.000359439 que es una cantidad bastante más cercana a 0.01753 (1.753%), que se obtuvo con el NeuroSolutions.

Q	error	q	error	q	error	q	error
10	0.03262	5	0.00999	0	0.04766	-5	0.009988
9.8	0.0099986	4.8	0.00999	-0.2	0.00991	-5.2	0.00999
9.6	0.00999	4.6	0.00999	-0.4	0.00991	-5.4	0.03381
9.4	0.009974	4.4	0.009954	-0.6	0.00998	-5.6	0.03379
9.2	0.00997	4.2	0.00999	-0.8	0.00998	-5.8	0.009976
9	0.009998	4	0.00998	-1	0.00998	-6	0.009992
8.8	0.009987	3.8	0.00992	-1.2	0.00989	-6.2	0.00999
8.6	0.009987	3.6	0.00996	-1.4	0.00997	-6.4	0.07033
8.4	0.009999	3.4	0.00994	-1.6	0.00999	-6.6	0.04788
8.2	0.009986	3.2	0.00965	-1.8	0.00999	-6.8	0.03344
8	0.00996	3	0.00965	-2	0.00998	-7	0.04767
7.8	0.00999	2.8	0.00956	-2.2	0.00999	-7.2	0.047911
7.6	0.00998	2.6	0.00988	-2.4	0.01	-7.4	0.00999
7.4	0.00999	2.4	0.00989	-2.6	0.00999	-7.6	0.01
7.2	0.009997	2.2	0.00993	-2.8	0.0378	-7.8	0.01
7	0.00998	2	0.00997	-3	0.0476	-8	0.061444
6.8	0.01246	1.8	0.00994	-3.2	0.0477	-8.2	0.04768
6.6	0.009999	1.6	0.00995	-3.4	0.04767	-8.4	0.009986
6.4	0.009999	1.4	0.00999	-3.6	0.0676	-8.6	0.0476
6.2	0.00994	1.2	0.00993	-3.8	0.05876	-8.8	0.01716
6	0.00996	1	0.00992	-4	0.06761	-9	0.04767
5.8	0.00998	0.8	0.00998	-4.2	0.05862	-9.2	0.01
5.6	0.00999	0.6	0.00998	-4.4	0.06742	-9.4	0.04776
5.4	0.00998	0.4	0.00998	-4.6	0.00998	-9.6	0.009993
5.2	0.00999	0.2	0.00995	-4.8	0.0826	-9.8	0.05158
						-10	0.009996

Tabla 4. Errores en las redes neuronales creadas para cada valor de q

En la figura 20 podemos observar el histograma con la representación de los errores obtenidos en cada una de las redes neuronales, donde nos podemos observar en cuales valores de q es error es más grande.

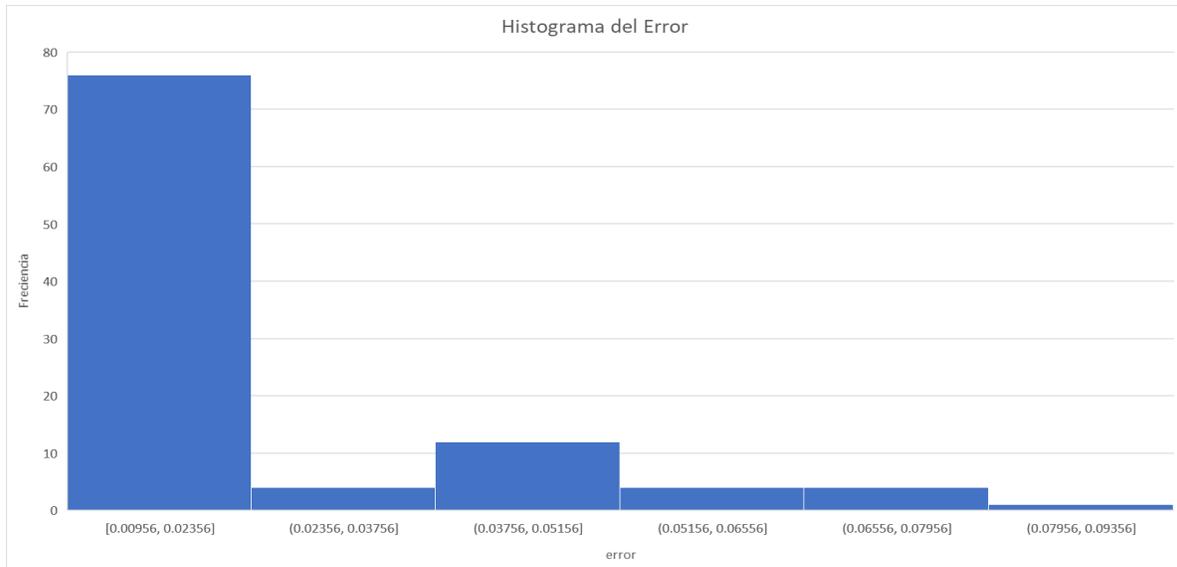


Figura 20 Histograma del error

En el caso de las de las redes neuronales entrenadas con MBP se encontró que el resultado emitido por las diferentes redes no tiene una tendencia como en el caso de la red entrenada con NeuroSolutions, si no que presenta un comportamiento más aleatorio, tanto en el caso de los ceros (no alcohólico) como el caso de los unos (alcohólico), en las tablas 5 a 8 se presenta una muestra de los valores arrojados como resultado de la red neuronal y se comparan con el valor esperado, teniendo en cuenta los casos donde se separó qs positivas y negativas, para cada valor de q y para todas las q.

Para q positivas	Para q negativas	Con todas las q	q=0	q=1	q=2						
0	0.000001	1	0.782043	1	0.99968	1	0.979956	1	0.992983		
0	0.893901	1	0.902774	1	0.738974	1	0.999984	1	0.98959	1	0.976843
0	0.000001	0	0.244414	0	0.313353	1	0.997686	1	0.984775	1	0.985057
1	0.847863	0	0.247446	1	0.968647	1	0.999928	1	0.966987	1	0.983287
0	0.000001	1	0.864142	1	0.930687	1	0.999124	1	0.990205	1	0.995568
0	0.000001	1	0.900743	1	0.366356	1	0.998384	1	0.995385	1	0.984467
0	0.000001	0	0.001833	0	0.999996	1	0.999744	1	0.994745	1	0.99957
1	0.911373	0	0.129	0	0.058389	1	0.999871	0	0.001566	0	0.000016
0	0.000128	0	0.239914	1	0.999996	1	0.998935	0	0.124302	0	0.024636
0	0.000001	1	0.879771	0	0.122102	1	0.995515	0	0.000105	0	0.000589
1	0.860599	0	0.237798	1	0.793065	0	0.001619	0	0.000004	0	0.015576
0	0.000001	1	0.903744	1	0.247078	0	0.000024	0	0.004402	0	0.011613
1	0.850028	1	0.783631	0	0.238075	0	0.00007	0	0.00031	0	0.057055
1	0.855112	0	0.003275	0	0.031899	0	0.001126	0	0.014037	0	0.006459

Tabla 5. Valor deseado contra calculado para redes entrenadas con MBP para q positivas, negativas, todas ,0, 1 y 2.

q=3	q=4	q=5	q=6	q=7	q=8						
1	0.999553	1	0.999997	1	0.991216	1	0.992798	1	0.995946	1	0.957396
1	0.998181	1	0.99667	1	0.998418	1	0.999561	1	0.999666	1	0.988996
1	0.982741	1	0.898448	1	0.999978	1	0.966796	1	0.998002	1	0.992411
1	0.991978	1	0.999715	1	0.999619	1	0.985383	1	0.99978	1	0.985381
1	0.999977	1	0.987395	1	0.999674	1	0.998053	1	0.924764	1	0.99705
1	0.949604	1	0.978807	1	0.997213	1	0.998492	1	0.999993	1	0.98327
1	0.995847	1	0.994747	1	0.999995	1	0.924704	1	0.989775	1	0.995583
0	0.000102	0	0.001428	0	0.000006	0	0.000009	0	0.000043	0	0.124551
0	0.005269	0	0.056199	0	0.001785	0	0.007574	0	0.075515	0	0.009338
0	0.01729	0	0.006379	0	0.000304	0	0.000002	0	0.004876	0	0.002596
0	0.000077	0	0.022906	0	0.004417	0	0.03779	0	0.008452	0	0.00583
0	0.046342	0	0.011984	0	0.000818	0	0.000841	0	0.081398	0	0.000004
0	0.065065	0	0.000002	0	0.008667	0	0.000002	0	0.070573	0	0.000034
0	0.001391	0	0.010394	0	0.148494	0	0.015777	0	0.000001	0	0.000373

Tabla 6. Valor deseado contra calculado para redes entrenadas con MBP para q 3,4,5,6,7 y 8

q=9	q=10	q=-1	q=-2	q=-3	q=-4						
1	0.984339	1	0.999999	1	0.999753	1	0.991761	1	0.998289	1	0.998002
1	0.994344	1	0.994022	1	0.993976	1	0.99913	1	0.999865	1	0.998002
1	0.999703	1	0.999777	1	0.980802	1	0.974508	1	0.999997	1	0.998898
1	0.998548	1	0.999786	1	0.996775	1	0.995881	1	0.999432	1	0.926966
1	0.997212	1	0.999999	1	0.999557	1	0.997846	1	0.999996	1	0.998369
1	0.9804	1	0.999301	1	0.968937	1	0.993761	1	0.99982	1	0.999589
1	0.974421	1	0.998287	1	0.985036	1	0.987246	1	0.998659	1	0.999995
0	0.116741	0	0.000174	0	0.003522	0	0.000021	0	0.000178	0	0.001612
0	0.00181	0	0.000001	0	0.025272	0	0.024817	0	0.001585	0	0.005733
0	0.000009	0	0.005065	0	0.000001	0	0.008863	0	0.000472	0	0.002976
0	0.005797	0	0.001845	0	0.101865	0	0.01329	0	0.002582	0	0.00913
0	0.000509	0	0.003352	0	0.000009	0	0.130431	0	0.000039	0	0.000698
0	0.015115	0	0.005766	0	0.016605	0	0.000001	0	0.000305	0	0.000001
0	0.023089	0	0.001136	0	0.031197	0	0.00469	0	0.000003	0	0.000752

Tabla 7. Valor deseado contra calculado para redes entrenadas con MBP para q 9,10,-1,-2,-3 y -4



## Capítulo 5. Resultados

### 5.1 Diseño de Software

Se diseñó, para el caso de la toma de decisión, tres programas, el primero contiene como núcleo de la toma de decisión una única red neuronal, la cual considera los datos totales en el entrenamiento y que tiene un error de 0.18, el programa fue hecho mediante Visual studio 2013 y se empleó Windows Forms para generar la interface de usuario, el programa tarda en hacer la toma de decisión una vez que tiene los datos ingresados una razón de 13 milisegundos.

El segundo programa considera tres redes neuronales, una para valores de  $q$  negativos, otra para valores de  $q$  positivos y la última para valores de  $q$  cero, el programa tarda en hacer la toma de decisión una vez que tiene los datos ingresados una razón de 14 milisegundos.

El tercer programa considera una red neuronal para cada valor de  $q$ , teniendo, así como núcleo del sistema de toma de decisiones las 101 redes neuronales antes mencionadas, el programa tarda en hacer la toma de decisión una vez que tiene los datos ingresados una razón de 15 milisegundos, en este caso, como parte de la contribución del presente trabajo, podemos tomar como referencia el método presentado para poder reducir significativamente el error en la clasificación a la hora de emplear redes neuronales, y también se puede reducir los tiempos que toma entrenar a una red neuronal en particular.

En la figura 21 podemos apreciar la interface gráfica diseñada, cabe destacar que en cuestión grafica los tres programas tienen el mismo aspecto, lo que cambia es el cómo

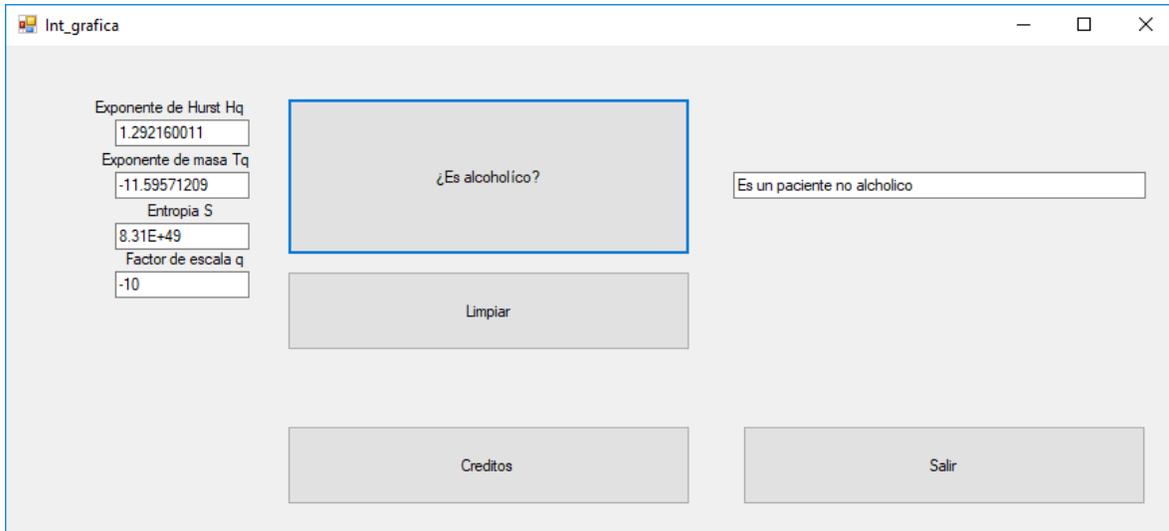


Figura 21 Interface gráfica.

realizan las clasificaciones.

Se le agregó un medidor de tiempo de ejecución, midiendo número de ciclos de reloj que toma el grupo de instrucciones y convirtiéndolo en convierte a milisegundos, por lo que el tiempo resultante dependerá de la velocidad del procesador donde se ejecute el programa.

## 5.2 Resultados

Mediante la aplicación del MFDFA y la entropía de Tsallis empleando Matlab se puede observar las diferencias que existen entre el comportamiento de un cerebro alcohólico y un cerebro no alcohólico, lo que llevó a proponer el conjunto de entradas  $\Omega$ , que se usaron para entrenar a las redes neuronales

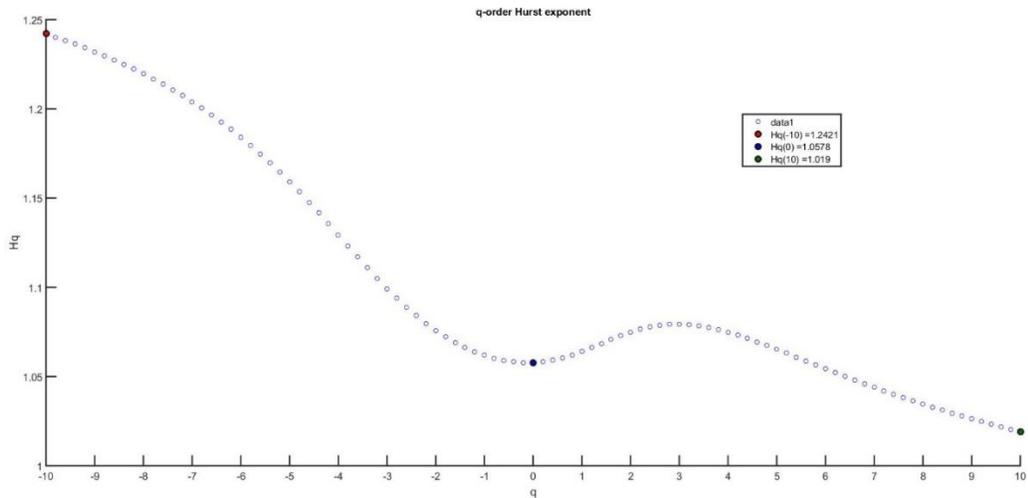


Figura 22 Exponente de Hurst vs factor de escala  $q$  para pacientes no alcohólicos

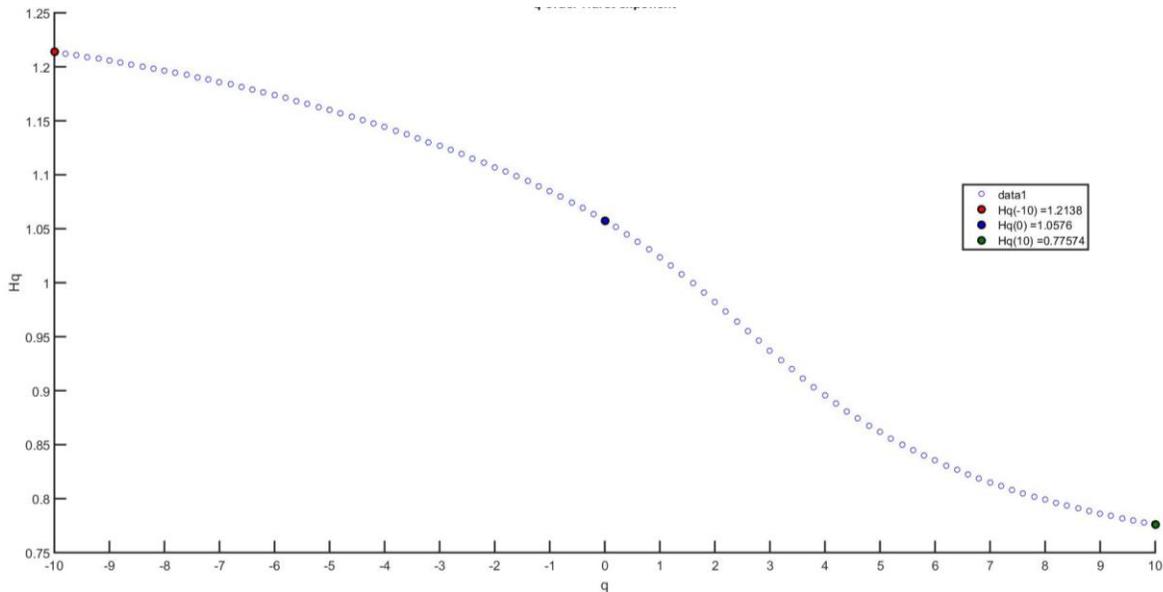


Figura 23 Exponente de Hurst vs factor de escala  $q$  para pacientes alcohólicos.

Recordando lo observado en el capítulo 3, podemos observar en las figuras 22 y 23, que existe una clara diferencia en el comportamiento del exponente de Hurst, para distintos valores de  $q$ , de un paciente alcohólico a uno no alcohólico, lo que nos marca la posibilidad de poder encontrar una clasificación para cada tipo de paciente, por otro lado, si consideramos el comportamiento de todos los pacientes en la base de datos podemos observar aún más diferencias.

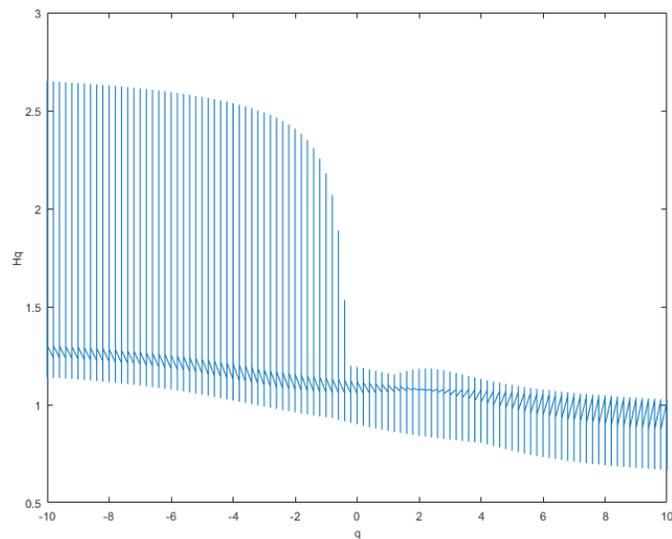


Figura 24 Exponente de Hurst vs factor de escala  $q$  para todos los pacientes no alcohólicos en la base de datos.

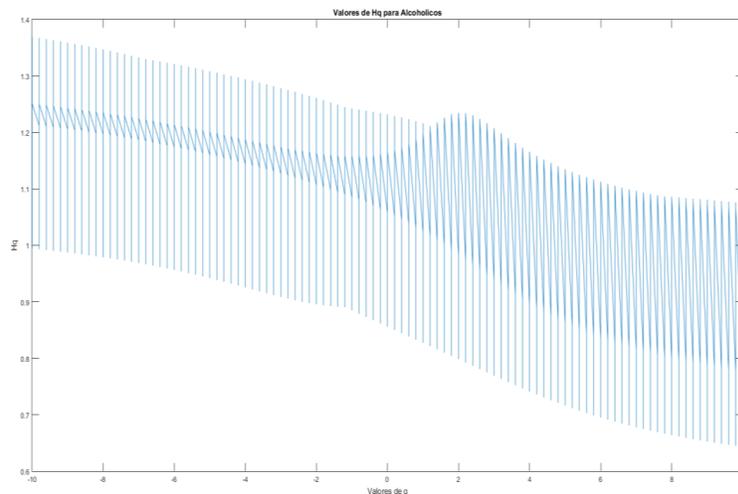


Figura 25 Exponente de Hurst vs factor de escala  $q$  para todos los pacientes alcohólicos en la base de datos

El patrón mostrado al graficar todos los exponentes de Hurst de los electroencefalogramas de los pacientes contenidos en la base de datos es claramente diferenciable entre ambos tipos de pacientes cómo se muestra en las figuras 24 y 25. Es más, si se obtiene el valor promedio de todos los exponentes de Hurts podemos observar aún más clara la diferencia existente entre el comportamiento de dicho exponente en cada tipo de paciente.

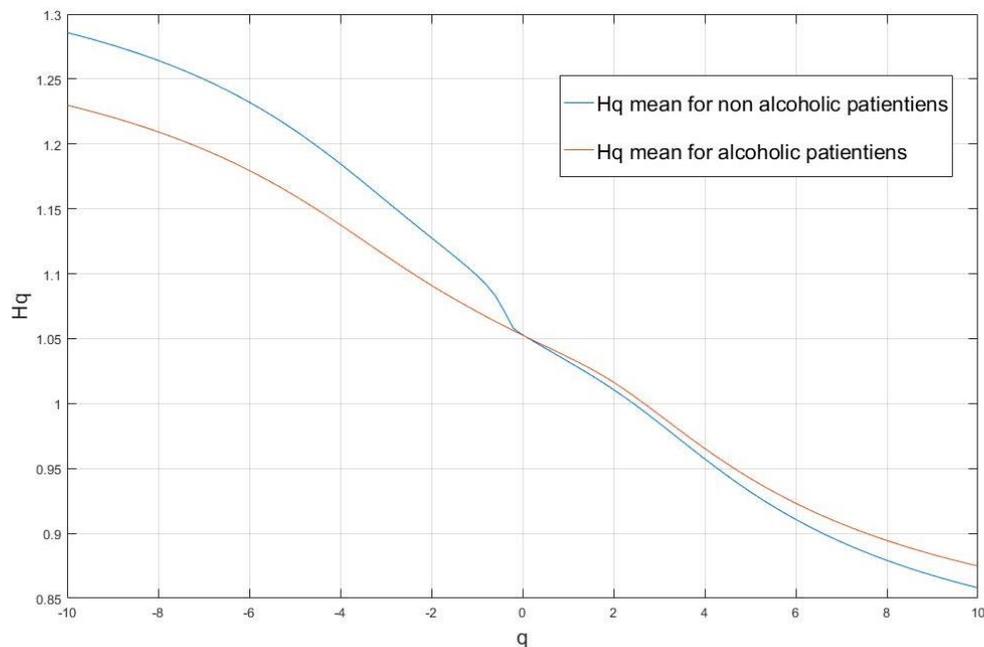


Figura 26 Exponente de Hurts promedio contra q, en rojo para pacientes alcohólicos y en azul para pacientes no alcohólicos

Como se puede observar en la figura 26, el problema aparece como linealmente separable, en los casos extremos, con q negativa y con q positiva, mientras más se acerca a cero el valor de q se observa menos clara esa separación lineal. Debemos recordar que cuanto más pequeña es q más pequeña es la fluctuación que estamos analizando, es decir que mientras más pequeña es q la frecuencia de la señal del EEG es mayor.

Podemos observar, en las figuras 27 y 28, el comportamiento del exponente de masa  $T_q$  donde se muestra nuevamente linealmente separable sobre todo en las fluctuaciones pequeñas.

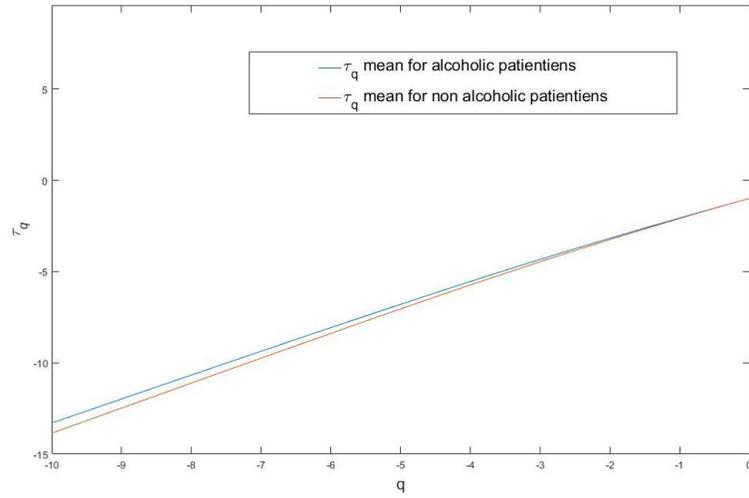


Figura 27 Exponente de masa contra  $q$ , para valores de  $q$  negativos

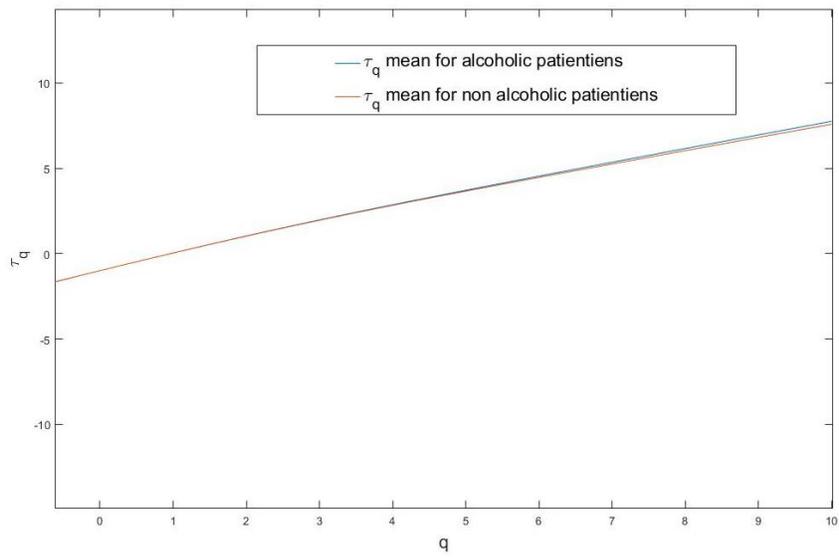


Figura 28 Exponente de masa contra  $q$ , para valores de  $q$  positivos

Con referencia a la entropía según Tsallis se muestra en las figuras 29 y 30, que con todos los pacientes se conserva la tendencia mostrada en el capítulo 3, donde vimos que la entropía tiene un comportamiento similar en ambos casos, pero con un valor superior para el caso de pacientes no alcohólicos sobre en el caso de las fluctuaciones pequeñas.

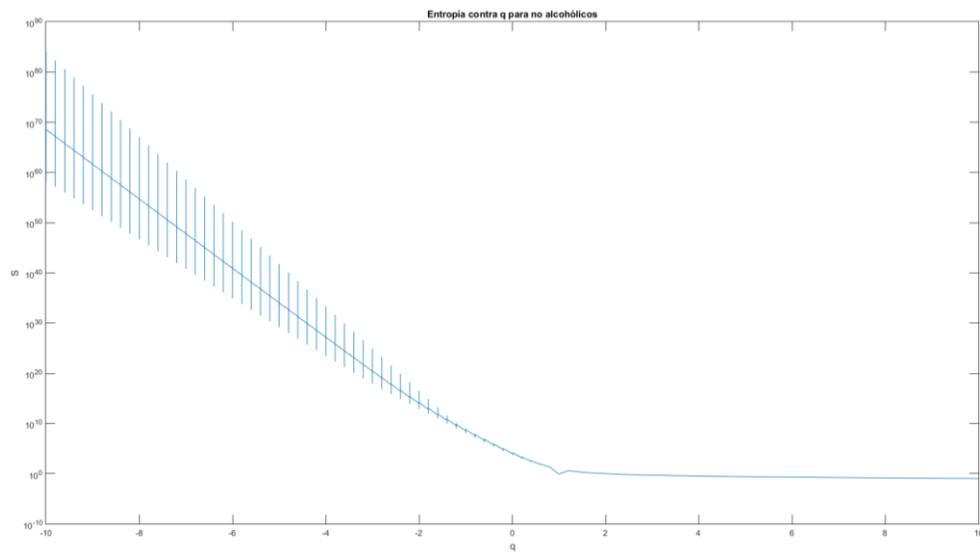


Figura 29 Entropía de Tsallis vs factor de escala  $q$  para todos los pacientes no alcohólicos en la base de datos.

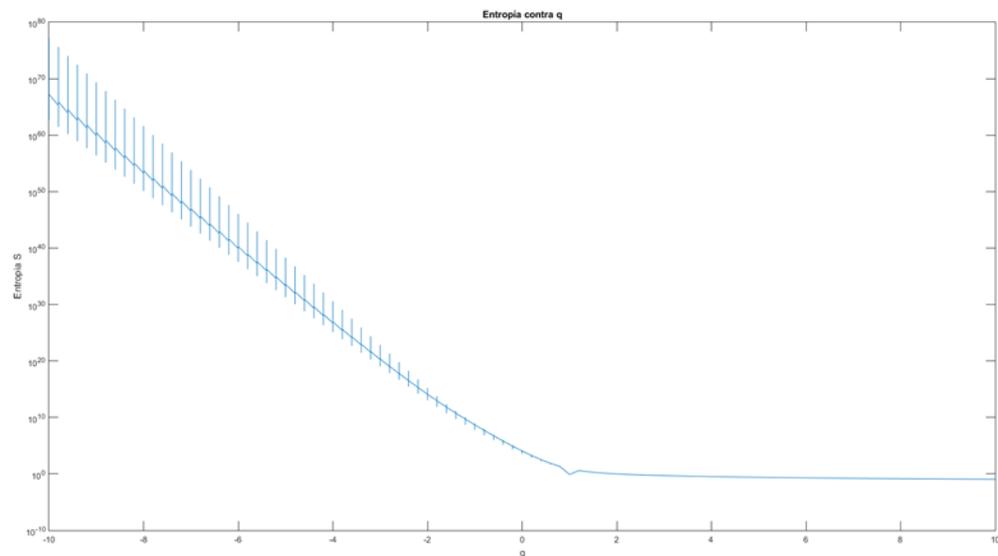


Figura 30 Entropía de Tsallis vs factor de escala  $q$  para todos los pacientes alcohólicos en la base de datos.

Para poder observar más de cerca se toma en cuenta el valor promedio de la entropía de Tsallis para todos los pacientes y podemos observar en la figura 31 que nuevamente en las fluctuaciones pequeñas se encuentra una clara diferencia, entre los valores de entropía presentados por los pacientes alcohólicos y los pacientes no alcohólicos.

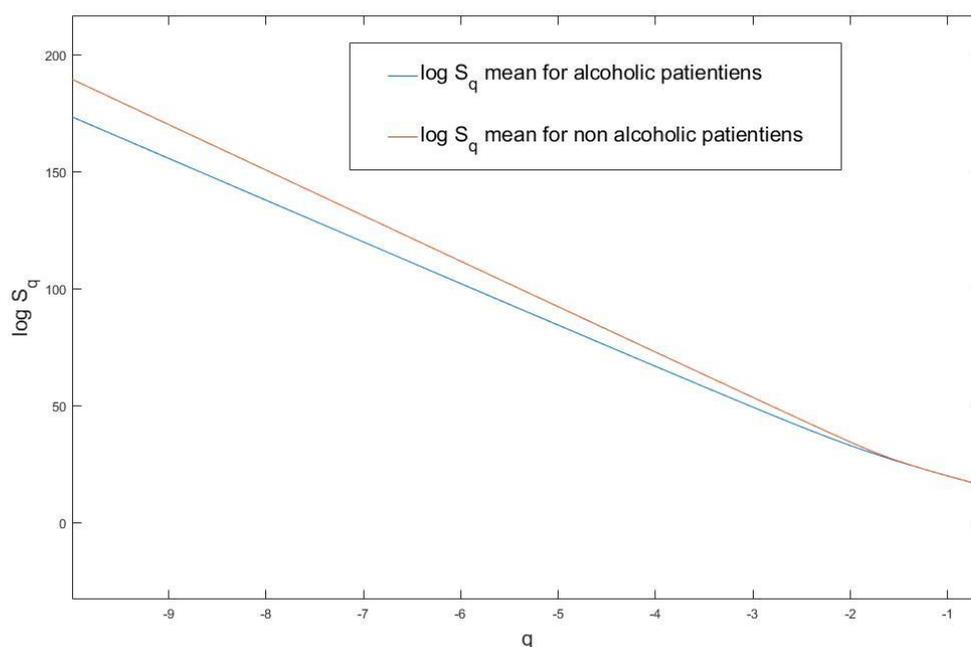


Figura 31 Entropía de Tsallis promedio vs factor de escala  $q$

Se encontró que, mediante el análisis multifractal, el exponente de Hurts para el caso de pacientes alcohólicos tiende a conservar su valor promedio tanto en fluctuaciones pequeñas como en fluctuaciones grandes, en cambio, el exponente de Hurts tiende a disminuir mucho su valor medio en el caso de los pacientes no alcohólicos.

Por otra parte, se encontró que mediante la obtención de la entropía no extensiva de Tsallis, se tiene en general un comportamiento similar en los dos tipos de pacientes, aunque es mayor la entropía en pacientes no alcohólicos que en pacientes alcohólicos.

Se puede encontrar una red neuronal mediante neuro-solutions que clasificó con un éxito del 98.24% a los eeg en la base de datos.

Se encontró otra red neuronal mediante el esquema de multiple back propagation que pudo clasificar los eeg en la base de datos con un 82% de éxito.

Por otro lado, al partir el problema en múltiples redes neuronales, se pudo reducir el error a 0.02028, por lo que al final se obtuvo una precisión del 97.972% al diferenciar pacientes alcohólicos y no alcohólicos, presentes en la base de datos.

### 5.3 Discusión

En este caso se puede corroborar que las medidas complejas empleadas en el estudio son suficientes para general clasificador capaz de determinar con una eficiencia mayor al 90% si un paciente tiene una determinada condición clínica, por lo que se puede asegurar que, al menos para el caso particular de estudio, fue posible determinar mediante estas medidas complejas y el empleo de un clasificador, la condición de un individuo.

Por lo que se puede considerar extender a otras condiciones, siempre y cuando afecten de manera directa a la respuesta eléctrica cerebral, ya sea generando colisiones o bien como en este caso, reducciones en la complejidad el sistema mediante la eliminación de una cierta cantidad aleatoria de elementos.

El hecho de que exista una función tal que, dadas las medidas complejas de electroencefalograma, sea capaz de determinar si existe o no existe un estado alterado en el cerebro de un paciente, demuestra que existen al menos dos clasificaciones del comportamiento dinámico complejo de los cerebros humanos.

## 5.4 Conclusiones

En el presente trabajo se corroboró el procedimiento de análisis que se requiere para obtener las medidas complejas necesarias, más no se sabe si son las mínimas necesarias, para poder clasificar entre un electroencefalograma de un cerebro alcohólico y un cerebro no alcohólico, se destaca que no existe comprobación de unicidad en la solución, por lo que se puede, a futuro, probar con algún otro conjunto de medidas u algún otro enfoque.

Se enumeran las contribuciones hechas a continuación:

- a) Primero el empleo de el momento de fluctuaciones  $q$ , el exponente de Hurst, el exponente de masa y la entropía según Tsallis, como conjunto de entrada:

$$\Omega = \{S_q, H_q, \tau_q, q\}$$

- b) Se comprobó que se puede clasificar el comportamiento electrofisiológico del cerebro humano, mediante la relación de  $\Omega$  y la condición, aunque se requiere de más datos para corroborar si existen solo dos o más clases, que correspondan a los diferentes daños provocados por la gran variedad de enfermedades mentales existentes.
- c) Se comprobó que mediante la separación del problema en unidades más pequeñas se puede optimizar el error de una manera más eficiente y se alcanzó una certidumbre cercana a la de un entorno de cobro con un entorno de código abierto.
- d) Por otro lado, se probaron diversas formas de obtener redes neuronales como clasificadores, y se determinó la existencia de un clasificador eficaz, el cual tiene la capacidad de clasificar entre dos estados del cerebro humano, en este caso las redes neuronales obtenidas son todas ellas soluciones posibles a nuestro problema.
- e) Se observó que mediante el entrenamiento con el framework de código propietario se tendía a tener dos posibles valores para uno, sin embargo, cuando se realizó el entrenamiento con el framework de código libre se observó que esta tendencia no se repite, por lo que se puede considerar como parte del error existente en la red

neuronal este comportamiento o bien a la existencia de más de una clase de alcohólico, por lo que se requieren de más datos para poder corroborar esa tendencia.

Como trabajo a futuro se plantea realizar una comprobación fuera de la base de datos empleada en el experimento, por lo que aún no es posible generalizar a todos los individuos y se considera que los resultados aplican solo dentro de los individuos en la base de datos.

También se plantea la posibilidad de obtener mediante otros modelos diferentes a redes neuronales, soluciones posibles, de tal manera que podamos encontrar la solución posible más óptima para representar la relación entre  $\Omega$  y una o varias condiciones en el cerebro humano.

## Bibliografía

- A.F.Ihlen, E. (2012). Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology*.
- Aijun He, X. Y. (2007). Multifractal Analysis of Epilepsy in Electroencephalogram. *IEEE/ICME International Conference on Complex Medical Engineering*.
- American Clinical Neurophysiology Society. (2006). Guideline 5: Guidelines for Standard Electrode Position Nomenclature. *American Clinical Neurophysiology Society*.
- Borja, T. R., Cartagena, M. C., Loaiza, O. G., Molina, V. G., & Puerta, Y. M. (2010). Sobre el parámetro de no extensividad para algunos sistemas super-aditivos. *Ingeniería y Ciencia*, 143–158.
- Center for Machine Learning and Intelligent Systems. (13 de 10 de 1999). *UC Irvine Machine Learning Repository*. Obtenido de EEG Database Data Set: <http://archive.ics.uci.edu/ml/datasets/EEG+Database>
- Collura, T. F. (1993). History and evolution of Electroencephalographic Instruments and Techniques. *Journal of clinical neurophysiology*, 476- 508.
- Colmenares, G. (2000). *Análisis Multivariante Y sus Aplicaciones; Parte 5 Inteligencia Artificial*. Mérida, Venezuela: IIES.
- Conn, P. M. (2003). *Neuroscience in Medicine*. New Jersey: Humana Press Totowa.
- Faure, P., & Korn, H. (2001). Is there chaos in the brain? 1. Concepts of nonlinear dynamics and methods of investigation. *C.R. Acad Paris; Sciences de la vie / Life Sciences*, 773-793.
- Fergus, P., Hignett, D., Hussain, A., Al-Jumeily, D., & Abdel-Aziz, K. (2015). Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques. *Hindawi Publishing Corporation BioMed Research International*.

- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. London: John Wiley & Sons.
- Hilera, J. R., & Martinez, V. J. (2000). *Redes Neuronales Artificiales*. Madrid: Alfaomega-Rama.
- Jaffard, S. (1997). Multifractal formalism for functions; Parts I and II. *SIAM Journal of Mathematical Analysis*, 944 -998.
- Jaynes, E. T. (1983). *Papers on probability, statistics and statistical physics*. Reidel, Dordrecht, Holland: R. D. Rosenkrantz.
- Kantelhardt, J., S.A., Z., E., K. B., S., H., & Bunde, A. S. (2012). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*.
- Lamb, S. M. (1999). *Pathways of the brain in the neurocognitive basis of language*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Likens, A. D., & Amazeen, P. G. (2014). Neural signatures of team coordination are revealed by multifractal analysis. *SOCIAL NEUROSCIENCE*, 219–234.
- Lopes, N., & Bernardete, R. (2003). An efficient gradient-based learning algorithm applied to neural networks with selective actuation neurons. *Neural, Parallel and Scientific Computations*, 253-272.
- Lopes, N., & Bernardete, R. (2009). GPU implementation of the multiple back-propagation algorithm. In *10th International Conference on Intelligent Data Engineering and Automated Learning* (págs. 449-456). Springer.
- Lopes, N., & Bernardete, R. (2010). Stochastic GPU-based multithread implementation of multiple back-propagation. *Second International Conference on Agents and Artificial Intelligence* (págs. 271-276). ICAART 2010.
- Lopes, N., & Bernardete, R. (2011). An evaluation of multiple feed-forward networks on GPUs. *International Journal of Neural Systems*, 31-47.

- Lopes, N., & Ribeiro, B. (2001). Hybrid learning multi neural architecture. *INNS-IEEE International Joint Conference on Neural Networks*, (págs. 2788-2793).
- MA, Q., NING, X., WANG, J., & BIAN, C. (2006). A new measure to characterize multifractality of sleep electroencephalogram. *Chinese Science Bulletin*, 3059—3064.
- Mandelbrot, B. B., & Hudson, R. L. (2006). *The (mis)Behavior of Markets*. New York: Basic Books.
- Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water resources research*, 909-918.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 115–133 .
- Peng, C., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995 ). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *CHAOS, American Institute of Physics*.
- Pennycuick, C. J. (1992). *Newton rules Biology A physical approach to biological problems*. Miami: Oxford University Press.
- Rosenblatt, F. (1957). The perceptron, a perceiving and recognizing automaton Project Para. *Cornell Aeronautical Laboratory*.
- Rosenblatta, M., & Figliolab, A. (2007). Estudio del espectro multifractal de una señal no estacionaria con “Multifractal detrended fluctuation analysis (MFDFA)”. *Mecánica Computacional*, 2162-2179.
- Ryszard, K. (2008). Algorithm of the Multifractal Detrended Fluctuation Analysis (MF-DFA) and some remarks concerning our case. *Warsaw*.
- Schrödinger, E. (1983). *Mente y Materia ¿Qué procesos biológicos están directamente relacionados con la conciencia?* España: Tusquets.

- Shannon, C., & Weaver, W. (1961). *The Mathematical Theory of Communication*. University of Illinois Press.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*.
- Valencia, R. M., Yáñez, M. C., & Sánchez, F. L. (2006). *Algoritmo Backpropagation para Redes Neuronales: conceptos y aplicaciones*. México Distrito Federal: IPN CIC.
- Werbos, P. J. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. *Doctoral Dissertation, Applied Mathematics, Harvard University, MA*.
- Widrow, B. &. (1960). Adaptive switching circuits. *STANFORD UNIV CA STANFORD ELECTRONICS LABS*.
- Zhenhu, L., Yinghua, W., Xue, S., Duan, L., Logan, J. V., Jamie, W. S., . . . Xiaoli, L. (2015). EEG ENTROPY MEASURES IN ANESTHESIA. *Frontiers in computational neuroscience*.

## Índice de figuras

<i>Figura 1 Nomenclatura y posición de electrodos para el sistema 10-20 Internacional de la American Clinical Neurophysiology Society, donde A1 y A2 o Nz son tomados como referencia</i>	11
<i>Figura 2. Ejemplo de electroencefalograma</i>	11
<i>Figura 3 Vistas del cerebro Lateral (A), mesial (B), superior (C), e inferior (D), representando puntos de demarcación importantes y áreas de Brodmann (números) (Conn, 2003).</i>	13
<i>Figura 4 : Esquema de una red Neuronal (Valencia, Yáñez, &amp; Sánchez, 2006)</i>	27
<i>Figura 5 Neurona de una red neuronal artificial (Valencia, Yáñez, &amp; Sánchez, 2006).</i>	28
<i>Figura 6 Neurona básica en backpropagation</i>	29
<i>Figura 7 Red neuronal multicapa</i>	30
<i>Figura 8 Ajuste de pesos en capa de salida</i>	31
<i>Figura 9 Ajuste de pesos en las capas ocultas</i>	32

Figura 10 Exponente de Hurst de orden $q$ no Alcohólico	36
Figura 11 Exponente de Hurst de orden $q$ Alcohólico	36
Figura 12 Exponente de Hurst de orden $q$ no Alcohólico con su ajuste polinomial	37
Figura 13 Exponente de Hurst de orden $q$ Alcohólico con su ajuste polinomial	37
Figura 14 Exponente de Hurst $H_q$ vs $q$ para todos los pacientes no Alcohólicos	38
Figura 15 Exponente de Hurst $H_q$ vs $q$ para todos los pacientes Alcohólicos	39
Figura 16 Entropía Vs $q$ para un paciente no Alcohólico	40
Figura 17 Entropía Vs $q$ para un paciente Alcohólico	40
Figura 18 Propuesta de solución	41
Figura 19 Topología de la red neuronal con 4 capas ocultas (20, 30, 20 y 10 neuronas en cada capa oculta)	46
Figura 20 Histograma del error	48
Figura 21 Interface gráfica.	52
Figura 22 Exponente de Hurst vs factor de escala $q$ para pacientes no alcohólicos	53
Figura 23 Exponente de Hurst vs factor de escala $q$ para pacientes alcohólicos.	53
Figura 24 Exponente de Hurst vs factor de escala $q$ para todos los pacientes no alcohólicos en la base de datos.	54
Figura 25 Exponente de Hurst vs factor de escala $q$ para todos los pacientes alcohólicos en la base de datos	54
Figura 26 Exponente de Hurts promedio contra $q$ , en rojo para pacientes alcohólicos y en azul para pacientes no alcohólicos	55
Figura 27 Exponente de masa contra $q$ , para valores de $q$ negativos	56
Figura 28 Exponente de masa contra $q$ , para valores de $q$ positivos	56
Figura 29 Entropía de Tsallis vs factor de escala $q$ para todos los pacientes no alcohólicos en la base de datos.	57
Figura 30 Entropía de Tsallis vs factor de escala $q$ para todos los pacientes alcohólicos en la base de datos.	57
Figura 31 Entropía de Tsallis promedio vs factor de escala $q$	58

## Índice de Tablas

Tabla 1 Bandas y frecuencias en electroencefalogramas.....	10
Tabla 2 Muestra de tabla de entrenamiento.....	43
Tabla 3 . Resultados de res neuronal entrenada en neurosolutions .....	45
Tabla 4. Errores en las redes neuronales creadas para cada valor de $q$ .....	47
Tabla 5. Valor deseado contra calculado para redes entrenadas con MBP para $q$ positivas, negativas, todas ,0, 1 y 2.....	48
Tabla 6. Valor deseado contra calculado para redes entrenadas con MBP para $q$ 3,4,5,6,7 y 8 .....	49

*Tabla 7. Valor deseado contra calculado para redes entrenadas con MBP para  $q$  9,10,-1, -2,-3 y -4..... 49*

*Tabla 8. Valor deseado contra calculado para redes entrenadas con MBP para  $q$  -5, -6,-7, -8, -9 y -10. .... 50*