

Máster Universitario en  
**Ingeniería Computacional y Sistemas Inteligentes**

Konputazio Zientziak eta Adimen Artifiziala Saila –  
Departamento de Ciencias de la Computación e Inteligencia Artificial

Tesis de Máster

**Técnicas no-lineales para el análisis  
no invasivo de comportamientos animales.**

**César Maldonado Córdova**

Tutor(a/es)

**M. Karmele López de Ipiña**

Departamento de Ingeniería de Sistemas y Automática  
Escuela de Ingeniería de Gipuzkoa

**Xavier Leinekugel**

Neurocentre Magendie INSERM, Bordeaux



K  
I  
S  
A  
  
I  
C  
S  
I



KZAA  
/CCIA

Septiembre 2016

## Resumen

El presente estudio realiza análisis y evaluación de modelos no-invasivos mediante técnicas no-lineales que permitan la clasificación de comportamientos animales. Los datos se obtienen de sensores de presión piezoeléctricos, los mismos que son adecuados para la medición de curvas o pulsaciones de presión altamente dinámicas y cuasi-estáticas. La información extraída serviría para implementaciones futuras de la Ingeniería Biomédica, para el tratamiento de enfermedades.

Se comienza con el pre-procesamiento de los datos en base a la frecuencia-temporal, que permite obtener parámetros lineales y no-lineales mediante la Densidad del Espectro de Potencia (PSD). Dichas características permite el aprendizaje del modelo clasificación realizada en la herramienta WEKA [23]. Las técnicas no-lineales como: Entropía Espectral de Shannon (EES), que permite entender la información y su validez, la Dimensión Fractal (DF) en base a los algoritmos de Higuchi [14] y Katz [16], mejoran el modelado del algoritmo para clasificar. Los parámetros se los obtiene por medio del software Matlab R2015a [12].

Para la experimentación sobre el aprendizaje del modelo de clasificación de comportamientos animales, se lo realiza mediante los algoritmos basados en: Análisis del Discriminante Lineal de Fisher (FLDA), Redes Neuronales Artificiales (Convolutional Neural Networks, CNN y Perceptron Multicapa, MLP) y árboles de decisiones (J48). Finalmente se analiza los algoritmos de clasificación por medio de sus tasas de acierto y la correlación entre variables (características). El análisis se centró en el aprendizaje del conocimiento para el clasificador automático basado en Redes Neuronales Artificiales, que con ayuda de los demás algoritmos permiten contrastar que técnica más representativa de la Minería de Datos (Data Mining), aporta mejores resultados.

## Índice de Contenido

1. Introducción.....	1
2.1. Objetivos del trabajo .....	1
2.2. Presentación del Problema (Marco de Estudio) .....	1
3. Motivación .....	6
4. Estado del arte .....	7
4.1. Estudio de animales de laboratorio. ....	7
4.2. Sensores Piezoeléctricos .....	8
4.3. Análisis no-invasivo. ....	9
4.4. Minería de Datos (Data Mining). ....	10
4.5. Técnicas no-lineales .....	13
5. Materiales .....	16
5.1. Consideraciones Éticas. ....	16
5.2. Animales para el Experimento. ....	16
5.3. Sistema de Evaluación del Comportamiento. ....	16
5.4. Adquisición de las señales de Presión Piezoeléctricos.....	19
5.5. Introducción a Matlab. ....	21
5.6. Pre-procesamiento de la señal Piezo .....	22
5.7. Herramienta WEKA. ....	23
6. Métodos.....	24
6.1. Extracción de parámetros. ....	24
6.1.1. Densidad Espectral de Potencia (PSD).....	24
6.1.2. Entropía Espectral de Shannon (EES) .....	24
6.1.3. Dimensión Fractal (DF).....	26
6.2. Clasificador Automático.....	28
7. Resultados.....	32
7.1. Análisis Discriminante Lineal de Fisher.....	32
7.2. Perceptron Multicapa (MLP).....	33
7.3. Redes Neuronales Convolucionales (CNN).....	34
7.4. Árboles de Decisiones (J48) .....	35
8. Discusión .....	36

9. Conclusiones.....	38
10. Líneas futuras.....	39
11. Referencias .....	40
12. Anexos.....	46
12.1. Cálculo de la Entropía Espectral de Shannon. ....	46
12.2. Cálculo de la Dimensión Fractal de Higuchi.....	48
12.3. Cálculo de la Dimensión Fractal de Katz (DFK).....	49

### Índice de Figuras

Figura 1: Diagrama del ratón moviéndose en XY durante un periodo de tiempo, su señal piezo y la MMPE para las coordenadas XY y su señal piezo.....	18
Figura 2 : Diagrama del ratón moviéndose en XY durante un periodo de tiempo, su señal piezo y la MMPE para las coordenadas XY y su señal piezo.....	19
Figura 3: Señal piezo del estado REM, en un periodo de tiempo. ....	20
Figura 4: Señal piezo del estado NREM ( <i>Slow-wave sleep</i> (SWS)) en un periodo de tiempo.....	20

### Índice de Tablas

Tabla 1: Porcentajes de acierto en base a las características mediante FDLA .....	33
Tabla 2: Porcentajes de acierto en base a las características mediante MLP .....	34
Tabla 3: Porcentajes de acierto en base a las características mediante CNN .....	34
Tabla 4: Porcentajes de acierto en base a las características mediante J48 .....	35

## **1. INTRODUCCIÓN**

### **2.1. OBJETIVOS DEL TRABAJO**

El análisis no-invasivo de comportamientos animales mediante sensores de presión piezoeléctricos, permite obtener valiosa información de aplicación en la Ingeniería Biomédica. El presente estudio está orientado a la investigación y la detección temprana de cambios en los comportamientos animales. En ese sentido se abordará los siguientes objetivos:

- Crear recursos de bio-señales en colaboración con los centros estudios autorizados y que son parte de la investigación. Siempre bajo criterios éticos, se utilizarán algunos recursos ya disponibles.
- Investigar nuevas metodologías de diseño de muestras biomédicas y normalización orientadas a la robustez de la evaluación médica.
- Utilizar señales no-invasivas.
- Investigar metodologías y parámetros bio-inspirados, centrándose tanto en parámetros lineales como no-lineales.

### **2.2. PRESENTACIÓN DEL PROBLEMA (MARCO DE ESTUDIO)**

El estudio del comportamiento con animales, se viene realizado desde hace muchos años con el fin de identificar características que permitan entender y comprender el funcionamiento biológico de especies que tienen enormes similitudes en cuanto a la constitución del *ácido desoxirribonucleico* (ADN) [1]. Dichos estudios centran su atención en los estados del sueño, para monitorear de forma no-invasiva (uso de sensores piezoeléctricos) el comportamiento en base a señales durante el sueño, en base a movimientos ocular rápido, denominada REM y ondas cortas NREM (Slow-wave sleep (SWS)). Dichas investigaciones aportan con información e insumos para la aplicación en la

Ingeniería Biomédica. El monitoreo de las señales de frecuencia–temporal arrojan grandes cantidades de información que son almacenadas en bases de datos y posteriormente con métodos y técnicas estadísticas y de las ciencias computacionales, se procesan y extraen información valida (entendible) que permita la toma de decisiones.

El desarrollo de la ciencia, muchas veces se fortalece por estudios a especies animales cuyos resultados permitirán ser utilizados en aplicaciones de la Ingeniería Biomédica [2]. Se obtienen patrones de la información de los animales observados aplicando modelos no-invasivos y técnicas no-lineales. Para el análisis de grandes volúmenes de información se lo hace mediante Minería de Datos, que es una aplicación de la estadística y de las ciencias computacionales, para extraer información entendible. Dichos resultados mediante su análisis médico permita la detección de estados tempranos e incluso preclínicos o prodrómicos, que faciliten la planificación para el futuro o incluso en el diagnóstico o tratamiento de enfermedades frecuentes en los seres humanos y animales.

El estudio del comportamiento a ratones es un experimento necesario en el desarrollo de la ciencia, de la academia, y de la industria. Es un proceso para explorar una multitud de modelos de transgénicos, y agentes farmacológicos potencialmente útiles. Si bien el estudio del comportamiento de animales en condiciones controladas puede ser usado para evaluar aspectos específicos de estado de ánimo o la cognición, tales como la ansiedad o de la memoria, el análisis del comportamiento espontáneo en condiciones sin supervisión también puede ser muy informativo sobre toda de factores psicológicos, motrices y componentes cognitivos. Por ejemplo, estudios previos basados en el análisis espontáneo de la rata o la actividad del ratón en un entorno de campo abierto, han permitido identificar la existencia de relación entre el comportamiento específico de diferentes cepas de ratón con medicamentos farmacológicos [3-4].

La identificación de comportamientos de ratones, se realiza con mayor frecuencia a partir de grabaciones de video [5-6]. Sobre la base de pixel por pixel mediante su análisis y procedimientos de cálculos sofisticados, tales como el reconocimiento de patrones o coincidencia de plantilla. Un número de comportamientos tales como: el descanso, la locomoción o aseo personal puede ser identificado automáticamente por algoritmos disponibles comercialmente, aunque con menos rendimiento por los operadores humanos entrenados. Una limitación de la señal de vídeo es que es a veces es difícil discriminar entre

conductas que implican movimiento "estacionario", tal como el aseo, el oler, comer o beber. Al respecto, los sensores de presión piezoeléctricos son muy sensibles y están disponibles para ayudar a detectar los movimientos leves de los animales [7; 8]. Un aspecto interesante de dicha señal relacionada con el movimiento, es que refleja la actividad de los músculos del animal. Dependiendo de la coordinación de estas miríadas de músculos, su forma mecánica puede resumir o se anulan entre sí. Por lo tanto, esta es una señal muy rica, pero también muy compleja que ha demostrado proporcionar un complemento útil a la señal de vídeo, para la identificación de una serie de comportamientos [7-10].

Sin embargo, la caracterización del comportamiento es un asunto importante y a su vez muy complejo que requiere de potentes herramientas de precisión y sensibilidad para clasificar el comportamiento, pues puede llegar a ser una tarea muy tediosa y difícil. En consecuencia, se ha informado que los operadores humanos entrenados para clasificar cada segundo de una sesión de grabación tenían grandes problemas, ocupaban el 44% de los segundos en atribuir a un comportamiento específico [10]. Esta información ha llevado el interés para utilizar parámetros no-lineales como la Permutación Entropía (PE), una medida física utilizada para caracterizar la dinámica no-lineal de bioseñales, un enfoque muy adecuado para el estudio de la complejidad del sistema biológico [11]. De hecho el PE directamente representa la información temporal contenida en la serie derivada de las señales caóticas, por lo que la serie más regular, predecible y menos compleja es menor su entropía. Este es, además, un método rápido, robusto y rentable que puede ser realizado automáticamente sin supervisión humana. Ya se ha utilizado con cierto éxito para caracterizar el comportamiento locomotor de ratas [12] o para detectar alteraciones de la conducta de los bancos de peces expuestos a contaminación [13].

Los sensores de presión piezoeléctricos permiten de forma no-invasiva observar y monitorear el comportamiento animal en sus diferentes actividades, en base a la fisiología del sueño. Las lecturas de las señales piezo emite información como señales eléctricas, para luego en base a técnicas estadísticas: pre-procesarlas, agruparlas o reducirlos con el fin de obtener patrones que permitan comprenderla, esto se lo realiza mediante la técnica de Minería de Datos [14], y posteriormente crear modelos, es decir representaciones abstractas de la realidad, mediante el Descubrimiento del Conocimiento (KDD) [15], que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a los patrones encontrados.

El KDD realiza extracción de información útil a partir de grandes cantidades de volumen de datos, por medio de la inducción, detección de dependencias de reglas, problemas de clasificación, clustering, reconocimiento de patrones y modelado predictivo. Los patrones, son expresiones que describen un subconjunto de los datos generales, además involucra un proceso iterativo e interactivo de búsqueda de modelos, patrones o parámetros que al ser descubiertos son válidos, novedosos para el sistema y potencialmente útiles.

Actualmente, existen muchos modelos que permiten extraer parámetros lineales y no-lineales de las señales piezo, las mismas que son obtenidas de los sensores de presión piezoeléctricos. La Densidad Espectral de Potencia (PSD) [16] que utiliza las Transformadas Rápida de Fourier (TTF) permite extraer parámetros en base a la frecuencia y el tiempo. Para luego realizar el aprendizaje automático de modelos de clasificación del comportamiento animal por medio de los estados REM y NREM.

Finalmente, mediante algoritmos de Redes Neuronales Artificiales (RNA), Análisis Discriminante Lineal de Fisher (FLDA) y Árboles de Decisión (J48) se modela el aprendizaje enfocados en las técnicas no-lineales referente a la información cruzada mutua, que se la obtiene por medio de la Entropía Espectral de Shannon (EES) y Dimensión Fractal (DF), que son la base fundamental para el modelamiento del clasificador automático con los diferentes algoritmos disponibles en la herramienta WEKA [20].

La DF, que es el estudio temporal o espacial de fenómenos continuos que muestran correlaciones en un amplio rango de escalas, permite medir el grado de fragmentación o irregularidad sobre múltiples series del borde o frontera de la estructura fractal bajo análisis, existen varios algoritmos que permiten calcular la DF de una serie temporal, los que se implementan en el presente trabajo se basan en los algoritmos de Higuchi (DFH) [17] y Katz (DFK) [18], que permiten el mejoramiento del modelo de clasificación del comportamiento animal.

Los materiales utilizados en el presente estudio se los explica a detalle en la sección 4, estableciendo la información extraída de observaciones de animales, mediante sensores de presión piezoeléctricos para el análisis de la presión. El tratamiento de datos por medio de la fase “pre-procesamiento” parte de la Minería de Datos, para el filtrado, eliminación de ruido

que se encuentre presente, y la obtención del conjunto de datos para el entrenamiento (train) y su evaluación por medio del conjunto destinada para test.

En la sección 5, se describen los *Métodos*, basados en técnicas no-lineales para la extracción de parámetros lineales y no-lineales, basándose en la PSD. Se clasifica las observaciones para el entrenamiento y test, fuente primaria para el descubrimiento de conocimiento en modelos de clasificación del comportamiento animal. Así mismo, se indican los modelos matemáticos que permiten calcular la Entropía Espectral de Shannon, que específica la validez e incertidumbre de la información; la Dimensión Fractal, que se lo realiza mediante la implementación de los algoritmos de Higuchi (DFH) y Katz (DFK). Dichos cálculos se ejecuta en el software MATLAB R2015a [19], usando recursos disponibles, que es parte inherente como objetivo del presente trabajo.

Con los datos previamente procesados, se realiza los experimentos en la herramienta WEKA [20]. La experimentación consiste en aprendizajes con algoritmos de clasificación de manera automática, basándose en las técnicas de la minería de datos que provienen de la inteligencia artificial y de la estadística. Dichas técnicas no son más que algoritmos que se aplican sobre un conjunto de datos para obtener resultados. Las técnicas son Redes Neuronales Artificiales, Análisis Discriminante Lineal de Fisher y Árboles de Decisión, poniéndose mayor énfasis en las primeras, el mismo que simula el comportamiento neuronal con enfoque a los algoritmos de Convolutional Neural Networks (CNN) y Perceptron Multicapa (MLP), que se los presenta en la sección 6.

En la sección 7, se realiza la discusión de los resultados obtenidos, contrastándolos con la teoría, que permite entender y realizar el debido análisis de las señales emitidas por los sensores de presión piezoeléctricos. Para definir el alcance del presente trabajo, y definir que técnica específica de la minería de datos para el modelamiento del clasificador automático, en base a los parámetros lineales y no-lineales de la señal piezo.

### **3. MOTIVACIÓN**

Encontrar patrones en grandes volúmenes de conjuntos de datos, es una importante área de investigación del Descubrimiento del Conocimiento (KDD) y de la Minería de Datos. Actualmente la exploración de datos, es parte del análisis estadístico y las ciencias computacionales, para procesar, explorar y analizar información con el objeto de obtener patrones y en base a ellos realizar diferentes investigaciones técnico-científicas que faciliten la toma de decisiones.

Durante muchos años se ha introducido en el desarrollo de la ciencia el estudio del comportamiento animal, por medio de técnicas no-invasivas con el enfoque de permitir el progreso científico, y que garantice el tratamiento futuro de enfermedades que aquejan al ser humano y animales. Pues, se conoce que todos los seres vivos poseen enormes similitudes en cuanto a la constitución del ácido desoxirribonucleico (ADN) [1], lo que garantiza que el estudio basado en animales pueda ser aplicado en los humanos, por medio de la Ingeniería Biomédica.

Los procedimientos o técnicas no-invasivos son aquellas que no involucran instrumentos que rompen la piel o que penetran físicamente en el cuerpo del animal observado, por lo que se considera importante su utilización para descubrir anomalías y comportamientos que permitan encontrar indicios para estudios sobre tratamientos en los seres humanos, y para ello es fundamental la minería de datos. La Minería de Datos se considera como un estudio y análisis de datos masivos, para extraer conclusiones e información relevante de ellos.

El presente trabajo se realiza el estudio del comportamiento animal, mediante sensores de presión piezoeléctricos. La información obtenida será un valioso insumo de aplicación en la Ingeniería Biomédica. Hay mucho trabajo realizado en la actualidad y también mucho trabajo por realizar, con lo cual se abre un gran abanico a futuras investigaciones para el diseño y construcción de productos y/o tecnologías para el beneficio de la salud como: equipos médicos, dispositivos de diagnóstico y de terapia. Es decir, es un campo infinito de alternativas de soluciones, basadas en el estudio y análisis del comportamiento animal, con el uso de las técnicas de la Minería de Datos.

## **4. ESTADO DEL ARTE**

### **4.1. ESTUDIO DE ANIMALES DE LABORATORIO.**

En el desarrollo de las ciencias médicas se utiliza preferiblemente los “animales de laboratorio”, que son considerados piezas fundamentales en las ciencias biomédicas. Permiten investigar y comprender las causas, diagnóstico y tratamiento de enfermedades que afectan al ser humano y a los animales. Son un elemento esencial en el desarrollo de avances para la prevención y tratamiento de enfermedades transmisibles y no transmisibles. En la actualidad los avances de la investigación en cáncer, cardiología, trasplantes de órganos, síndrome de inmunodeficiencia adquirida, enfermedad de Alzheimer, se deben a las contribuciones de los estudios realizados en animales de laboratorio [21-22].

Los animales de laboratorio son considerados como reactivos biológicos (RB), por lo que su pureza debe ser vigilada, controlada y contrastada, al igual que cualquier reactivo, sin descuidar su posible contaminación biótica. Por lo cual se requiere la producción de animales estandarizados o definidos, con características genéticas y sanitarias definidas, criados en ambientes controlados que respeten los requerimientos de cada especie, con el correcto cumplimiento de los principios éticos y de bienestar animal [23].

El Animal de laboratorio, es considerado como cualquier especie animal que se mantiene bajo condiciones determinadas y se utiliza con fines netamente científicos [24]. El estado sanitario de los animales de laboratorio está determinado por un complejo multifactorial en el que interactúan: además de la biología del animal, y el perfil genético, las condiciones ambientales del alojamiento, así como las prácticas y manejo al que son sometidos estos animales y sus insumos [24-25].

El ambiente en el cual se alojan los animales debe ser el apropiado para cada especie, permitiendo el despliegue de las conductas específicas y disminuir al mínimo conductas inducidas por estrés [26-27]. El ambiente de los animales de laboratorio está constituido por el microambiente y el macroambiente. El encierro secundario (macro ambiente) está constituido por la habitación: tamaño, iluminación, temperatura, ventilación y humedad relativa, ausencia de ruido y polvo, entre otros [24-28].

El microambiente en cambio constituye el encierro primario del animal, determinado por el habitáculo o jaula y todo lo que en él se incorpora, como: lecho; agua; alimento, número de animales. Sus componentes deben satisfacer las necesidades fisiológicas de conducta y las interacciones sociales entre individuos de la misma especie, así como el establecimiento de jerarquías dentro del encierro [24-28].

El ratón es considerado por muchos investigadores como un modelo animal casi perfecto porque además de su corto tiempo generacional, alta performance reproductiva y fácil mantenimiento, son los animales más sofisticados que pueden ser utilizados para conocer reacciones frente a una agresión, una intoxicación o una infección experimental, reacciones o trastornos inmunológicos, oncología, teratología y embriología [29-30].

Los estudios con animales de laboratorio deben estar normados y ligados al uso humanitario de los mismos, basándose en principios éticos para la experimentación y estudios científicos con animales. Estos aspectos se deben fundamentar en las tres "Rs": Reducir, Remplazar y Refinar, implicando la utilización racional e inteligente con el fin de minimizar el uso de animales y las causas de dolor y de diestres (estrés negativo) en los individuos de la misma especie a ser investigados. [31-32]

## **4.2.  SENSORES PIEZOELÉCTRICOS**

El funcionamiento de sensores se basa en el efecto piezoeléctrico descubierto por Jacques y Pierre Curie (1880) en los cristales de cuarzo. Este fenómeno se produce en determinados sistemas cristalinos, y consiste en la aparición de una cierta carga eléctrica cuando el cristal es sometido a algún tipo de deformación. Existen además otras sustancias naturales como la turmalina o la sal de Rochelle que poseen un similar comportamiento que el cuarzo [33].

Sin embargo, las señales que son capaces de producir las sustancias naturales son débiles lo que dificulta su lectura en los diferentes sistemas de instrumentación. Teniendo en cuenta la importancia del fenómeno y su aplicabilidad, se desarrollaron cristales que presenta las mismas propiedades que las materiales naturales, pero con mejoras en la potencia de las señales producidas, esto se logra por medio del titanato de bario o el titanato-circonato de plomo, aunque estos no aporta mucha estabilidad como los naturales, pero mejoran su nivel

de la señal de salida, para mejorar significativamente los parámetros de cualquier sistema de instrumentación.

Los materiales sintéticos producen el efecto piezoeléctrico gracias a la anisotropía de su estructura y a la distribución de sus cargas eléctricas, que se suele conseguir después de un tratamiento y un proceso de polarización con campos eléctricos muy intensos, lo que se consigue mediante una reordenación de los grupos de unidades cristalográficas, y el reforzamiento de los comportamientos anisótropos para obtener el efecto piezoeléctrico. Este efecto, dan lugar a los sensores, los mismos que son cristales de material piezoeléctrico con dos electrodos capaces de producir una fuerza cuando se actúe eléctricamente sobre el. La medida de fuerza se refiere a aceleraciones y vibraciones y los de medida de presión, este último son lo que se utilizan para el presente estudio, es decir, la utilización de sensores para medir la presión (sensores de presión piezoeléctricos).

#### **4.3. ANÁLISIS NO-INVASIVO.**

Los modelos no-invasivos se caracterizan por utilizar herramientas de medida que se encuentren en contacto con el individuo (animal) observado, pero de forma superficial o bien a cierta distancia. Enfocándose en la observación de variables como: presión arterial, frecuencia cardiaca, frecuencia respiratoria y temperatura. La presión es leída mediante sensores piezoeléctricos, que permiten obtener la información durante el comportamiento del animal de laboratorio.

Todos los animales de sangre caliente como parte fisiológica realizan estados de reposo, en la que su actividad se modifica o varía, así como también la sensibilidad. Además se considera patrón característico de los mamíferos y aves [1-2], es decir que permite el estudio y análisis de diferentes manifestaciones en el ciclo sueño-vigilia, que son ritmos biológicos que ocurren durante toda la existencia, pero de forma periódica, de ahí que muchos investigadores han realizado estudios de dicha actividad con el propósito de obtener información para ser aplicada para un futuro cercano relacionados con el avance de las ciencias médicas.

El sueño es un estado biológico presente en animales y personas. En la fisiología del sueño existe variación de los movimientos corporales y se activa una etapa denominada reposo, que es diferente en cada especie animal. Investigadores como Rechtschaffen y Kales (1968) establecieron criterios para el estudio de dichos cambios, por medio de la actividad eléctrica cerebral que permite el registro de señales fundamentales (EEG, EMG y EOG), es decir, electroencefalograma (EEG), electromiograma (EMG) y electrooculografía (EOG) [34], donde se identifican diferentes fases de sueño y los episodios de vigilia.

Existen dos tipos de sueño: el basado en movimientos oculares rápidos, denominado sueño REM y el sueño de ondas lentas, también conocido como No-REM (NREM). En el sueño REM existe movimientos rápidos de los ojos y músculos faciales, así como el aumento de pulso, respiración y presión arterial, que permite obtener información de la actividad mediante sensores piezoeléctricos para la medición de la presión, aceleración, tensión o fuerza, transformando las lecturas en señales eléctricas para monitorear la actividad en el comportamiento de los animales en estudio.

Actualmente, para realizar estudios del comportamiento durante el ciclo sueño-vigilia, se lo hace por medio de señales piezo, las mismas que son lecturas del sensor piezoeléctrico como modelo *no-invasivo*, es decir, sin alternar el normal funcionamiento del organismo. Como estas señales eléctricas no son estacionarias, es necesario la utilización de técnicas no-lineales que permita el análisis del comportamiento animal. Y en base a estos datos realizar la predicción y posterior clasificación.

#### **4.4. MINERÍA DE DATOS (DATA MINING).**

Al realizar las observaciones de los animales grabados y monitoreados se obtiene extensa información, que muchas veces pueden incluir ruido blanco [35], de ahí que debe ser pre-procesada, por lo que es necesario la utilización de la estadística y las ciencias de la computación, en especial del uso de la *Minería de Datos*. El propósito general de la minería de datos es procesar la información de la gran cantidad de datos almacenados o que se puedan generar, y desarrollar procedimientos para manejar los datos con el propósito de tomar en lo posterior decisiones. Una de las primeras tareas en el proceso de la minería de

datos consiste en resumir la información para comprender bien o de mejor manera su contenido.

La minería de datos permite extraer información de grandes volúmenes de datos y transformarla con el propósito de comprenderla, mediante la determinación de patrones que ayuden en la toma de decisiones. Dicha técnica estadística cumple con ciertos pasos específicos: 1) *Selección de datos*, que consiste en buscar la función objetivo, atributos apropiados que permitan interpretar la información disponible en los datos, 2) *Transformación de datos*, que consisten en operaciones para organizar los datos en la forma deseada, convirtiendo un tipo de datos en otro, definiendo nuevos atributos, reduciendo la dimensionalidad de los datos, removiendo ruidos, normalizando, decidir estrategias para manejar datos perdidos; 3) *Minería de datos*, los datos transformados son analizados o minados, utilizando una o más técnicas para extraer patrones de interés; 4) *Interpretación de resultados y validación*, para comprender el significado del conocimiento extraído y su rango de validez, debidamente probada, utilizando métodos de validación establecidos y probándolo con datos diferentes a los utilizados para crear el modelo; y 5) *Incorporación del conocimiento descubierto*, presentación de los resultados del modelo para poder comprobar o resolver conflictos con creencia o resultados anteriores, y aplicar el nuevo modelo [36].

El uso de la minería de datos en el presente trabajo, permite predecir tendencias o comportamientos por medio de los datos disponibles y previamente procesados. La predicción incorpora tareas de clasificación, regresión y análisis de series temporales. Para obtener la descripción y/o predicción para el modelo de clasificación mediante grupos u objetos en términos de sus parámetros o variables, enfocándose al aprendizaje supervisado [37], donde la información de salida es una clasificación discreta, es decir, teniendo un objeto y sus atributos de entrada, el resultado de la clasificación es una de las posibles clases exclusivas del problema. La tarea de clasificación tiende a descubrir alguna clase de relación entre los variables y las clases de salida. El conocimiento descubierto puede ser utilizado para predecir la clase de un nuevo objeto desconocido.

Las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas no son más que algoritmos que se aplican sobre un conjunto de datos para obtener resultados. Las técnicas más representativas son [38]:

- a) *Redes neuronales.* Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida.
  
- b) *Regresión lineal.* Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de dos variables.
  
- c) *Arboles de decisión.* Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial y el análisis predictivo, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.
  
- d) *Modelos estadísticos.* Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.
  
- e) *Agrupamiento o Clustering.* Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se trataría de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes.
  
- f) *Reglas de asociación.-* Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. Según el objetivo del análisis de los datos, los algoritmos utilizados se clasifican en supervisados y no supervisados. Algoritmos supervisados (o predictivos): predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.

#### 4.5. TÉCNICAS NO-LINEALES

Aunque las señales emitidas por los sensores de presión piezoeléctricos no son estacionarias, puede ser modelado con técnicas lineales, como modelos interesantes por representar soluciones confiables y sobre todo por su bajo coste computacional. El Análisis Discriminante, es una de las técnicas de análisis multivariante más conocidas, cuyo objetivo es encontrar la combinación lineal de las variables independientes que mejor permitan diferenciar a las clases, por medio de una función discriminante, la cual podrá ser utilizada para clasificar nuevos casos.

Existen distintas técnicas de análisis discriminante, tales como Análisis Discriminante Lineal (Linear Discriminant Analysis, LDA), Análisis Discriminante Cuadrático (Quadratic Discriminant Analysis, QDA) entre otras. Permiten el estudio de patrones que pueden ser enfocándose en los modelos ocultos de markov (HMM) y en base al LDA, para el reconocimiento de patrones y aprendizaje de máquinas con el fin de encontrar una combinación lineal de rasgos que caracterizan o separan dos o más clases de los estados observados. La combinación resultante puede ser utilizada como un clasificador lineal, o, más comúnmente, para la reducción de dimensiones antes de la posterior clasificación.

El propósito del LDA, es aprovechar la información contenida en las variables independientes para crear una función por medio de la combinación lineal, capaz de diferenciar lo más posible a las diferentes clases. El LDA proporciona una respuesta rápida con recursos computacionales limitados y en algunos estudios realizados es usado para clasificar los patrones encontrados en las señales cerebrales en dos clases, no obstante es posible extender el método para diferenciar entre cualquier número de clases [49].

Sin embargo, como se indicó anteriormente las señales piezo, no se caracterizan por ser estables e incluso podrían ser no lineales [47], por lo que analizar estas señales con modelos no-lineales puede brindar información de interés. A todo esto se suma que un sistema de monitoreo basado en análisis no-invasivo con sensores de presión piezoeléctricos, se compone de múltiples variables, por lo que antes de utilizar un esquema de red neuronal se necesita un procesamiento previo de la señal, para extraer sus

características más relevantes [49] mediante técnicas estadísticas e ingeniería computacional.

El análisis no-lineal se basa en la existencia de sistemas caóticos caracterizados por un comportamiento no-periódico y por su sensibilidad a las condiciones iniciales. Muchos sistemas biológicos exhiben un comportamiento no-lineal, por lo cual resulta difícil predecir su comportamiento frente a un estímulo dado. Los sistemas fisiológicos tienen un comportamiento sumamente irregular e impredecible del individuo, de ahí que en la actualidad los métodos de análisis no-lineal se han utilizado para estudiar multitud de señales biomédicas y de esta manera comprender el funcionamiento de los procesos fisiológicos de los órganos o sistemas que los generan [39].

Las técnicas no-lineales para el procesamiento de las señales eléctricas son mediante la información mutua cruzada, que es una función estadística que proporciona una cuantificación de los componentes lineales y no-lineales del acoplamiento de variables, y se basa por lo general en la entropía de Shannon, la misma que mide la aleatoriedad de la fuente de información.

Las señales biomédicas se obtiene de los procesos que tienen lugar en el dominio del tiempo. Los parámetros necesarios para el estudio del comportamiento de los animales en base a la actividad de sueño REM y sueño NREM, son extraídos mediante la función de Densidad Espectral de Potencia (Power Spectral Density, PSD). Existen varios métodos que permiten la obtención del PSD de una señal, los más conocidos son: (1) Métodos no Paramétricos basados en la Transformada de Fourier Discreta (DTF), (2) Métodos Paramétricos basados en Modelos ARMA y variantes, y (3) Métodos de tiempo-frecuencia que asumen la no estacionariedad de las señales con el tiempo [40].

1. *Métodos no paramétricos*: Están basados en la DFT, y se utiliza de forma intensiva la aproximación directa, que se obtiene como la magnitud al cuadrado de la DFT obtenida mediante la Transformada de Fourier Rápida (Fast Fourier transform, FFT) y aplicada directamente sobre los datos previamente procesados por medio de función ventana hamming. Los métodos paramétricos son computacionalmente eficientes y el PSD obtenido es directamente proporcional a la potencia de las componentes espectrales del proceso.

2. *Métodos paramétricos:* Se conoce que la auto-correlación responde a una determinada forma paramétrica o modelo. Los modelos Autorregresivos (AR) es un método paramétrico, que calcula la PSD estimando primero los coeficientes de un hipotético sistema lineal que genera la señal junto a los Modelos Autorregresivos de Media Móvil (Auto-Regressive Moving Average Models, ARMA) y los modelos de Medias Móviles (MA).
3. *Métodos de tiempo-frecuencia:* Tanto los métodos paramétricos como los no paramétricos hacen la suposición de que la señal es estacionaria. En caso contrario, se debe implementar un esquema de segmentación que asegure una ventana de datos libre de transitorios. Los métodos de tiempo-frecuencia, asumen la no estacionariedad de las señales con el tiempo.

## **5. MATERIALES**

En el presente estudio se utilizaron varios materiales que permitieron la obtención de la información por medio de los sensores de presión piezoeléctricos, el pre-procesamiento (filtrado y eliminación de ruido) y obtención de las características o parámetros lineales y no-lineales en base a las señales. Estos procesos se lo desarrollo mediante software especializado para minería de datos, como es el entorno de desarrollo MATLAB.

### **5.1. CONSIDERACIONES ÉTICAS.**

Todos los procedimientos experimentales se realizaron de conformidad con la legislación Francesa y de la Unión Europea (UE), en materia de protección de los animales utilizados con fines experimentales y científicos, 86/609/EEC y 2010/63/EU. Los experimentos han sido aprobados por el comité de ética CEEA50 (autorización legal 50120167).

### **5.2. ANIMALES PARA EL EXPERIMENTO.**

Los ratones (animales de laboratorio) fueron alojados en una instalación para animales, libres de gérmenes patógenos específicos (SPF) y se mantuvieron en observación durante un ciclo de 12 horas entre luz y oscuridad, con habitual acceso a comida y agua. Todos los experimentos se han realizado durante el período de luz.

### **5.3. SISTEMA DE EVALUACIÓN DEL COMPORTAMIENTO.**

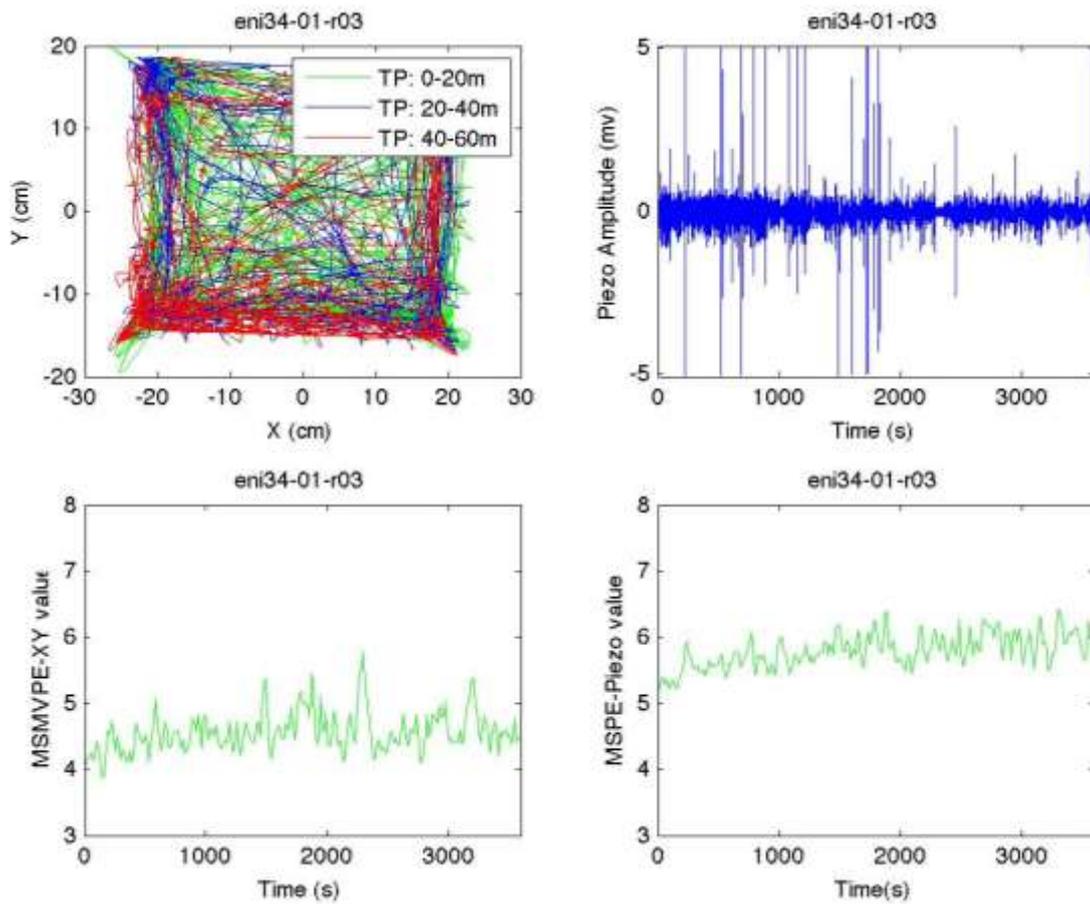
Debido a que el número de animales estudiados es todavía demasiado pequeño se trata de un *estudio piloto*. Los ratones se introducen individualmente para ser grabando durante sesiones en la plataforma de grabación (Phenact 2, Addenfi, Paris, Francia) [41], en un entorno con poco iluminación en campo abierto (45x35 cm), rodeados por paredes de 50 cm de altura, equipadas con monitoreo por video. En este sistema, la placa de piso está

descansando en sensores de presión piezoeléctricos, que proporciona la señal analógica continua generada por los cambios sutiles en la presión baja de la placa debido al movimiento de los ratones. Este ambiente era novedoso para el animal. La señal de video se adquirió a una frecuencia de muestreo de 20 frames sobre segundo, con una cámara web colocada a 1,5 m por encima de la plataforma.

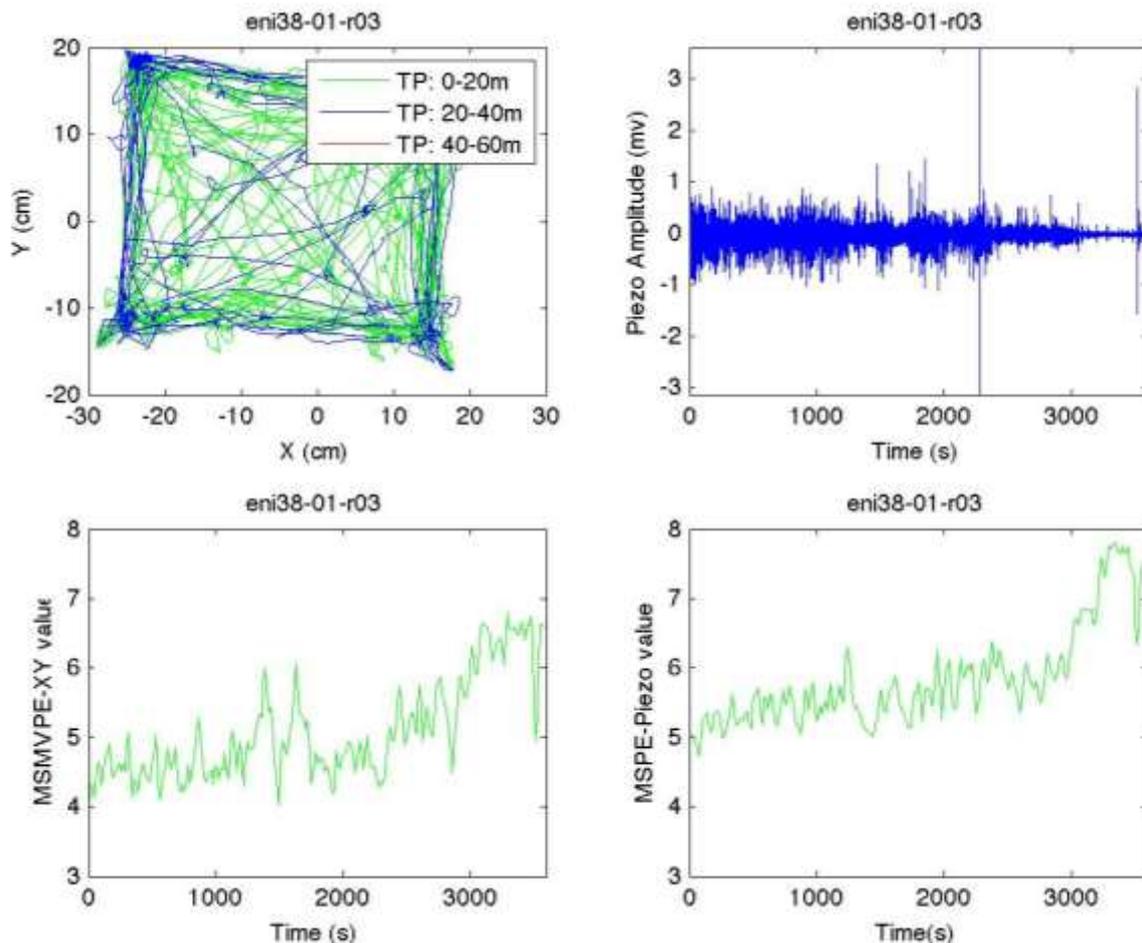
La señal de piezo se registró continuamente a una frecuencia de muestreo de 20 kHz. Las señales se adquirieron de forma sincrónica utilizando software Spike2 (CED) y almacenados en un computador personal (PC) para su análisis con el software EthovisionXT (Noldus) y con secuencias de comandos en MATLAB (MathWorks). El software Ethovision, proporciona la detección automatizada de los animales y el seguimiento (coordenadas x-y).

A continuación, se presentan dos figuras de 2 ratones moviéndose X-Y durante 2 periodos de tiempo, su señal piezo y una entropía calculada la Multiscale Multivariate Permutation Entropy (MMPE) para las coordenadas XY extraídas con imágenes. Tanto la Figura 1 y la Figura 2, se visualiza cuatro gráficas. En la parte superior izquierda se representa el movimiento del ratón en las coordenadas XY, a su derecha el espectro de la señal piezo en relación al tiempo. En la parte inferior dos figuras, en la izquierda se muestra el valor de la entropía con MMPE de las coordenadas XY, y la de la derecha el cálculo de la entropía de la señal piezo, las dos con relación al tiempo.

**Figura 1:** Diagrama del ratón moviéndose en XY durante un periodo de tiempo, su señal piezo y la MMPE para las coordenadas XY y su señal piezo.



**Figura 2** : Diagrama del ratón moviéndose en XY durante un periodo de tiempo, su señal piezo y la MMPE para las coordenadas XY y su señal piezo.



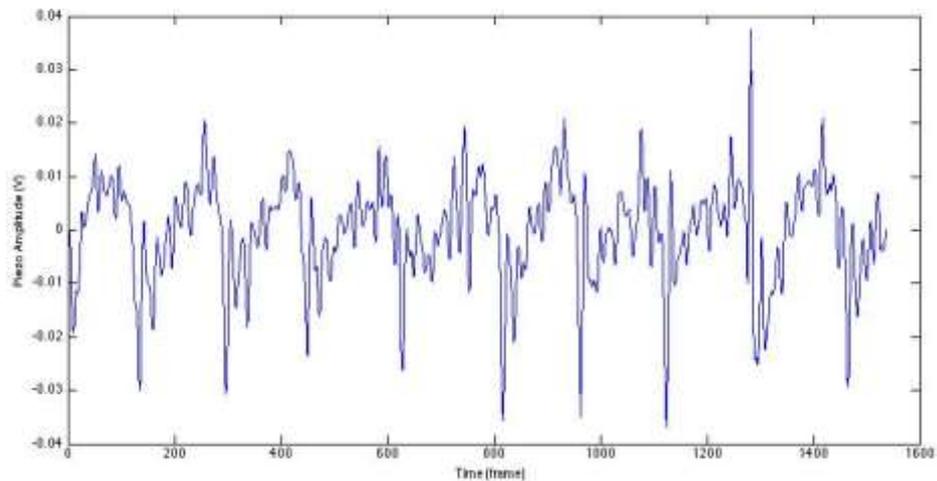
#### 5.4. ADQUISICIÓN DE LAS SEÑALES DE PRESIÓN PIEZOELÉCTRICOS.

La señal se muestra en instantes de tiempo de acuerdo a una frecuencia de muestreo determinada para poder capturar las características más relevantes de dicha señal. Dicha señal piezo se obtiene por medio de los sensores de presión piezoeléctricos, conectados en la plataforma donde se observa a los ratones. Permiten grabar estados de sueño tanto en REM y NREM (*Slow-wave sleep* (SWS)), además cuando se encuentra en actividades específicas de comportamiento. Sin embargo las señales biológicas varían de manera continua, así como su representación en señales eléctricas. Por lo que no pueden ser

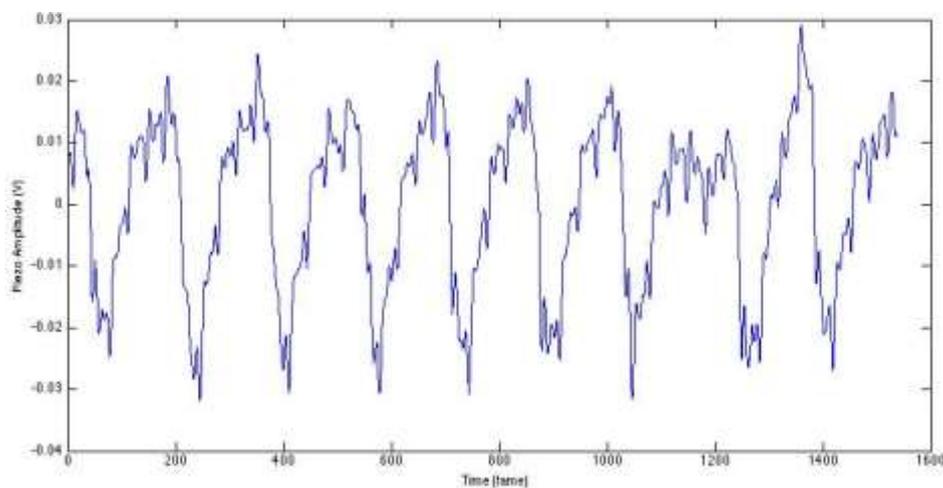
almacenadas mediante archivos o sistema de base de datos, así como tampoco realizar tratamiento alguno con la información que contienen.

Para solucionar esto se toman series de muestras, es decir valores de la señal en un instante de tiempo representados en forma de series de números, y ésta debe ser representativa de la señal, y además que perder la información que contiene. Para en lo posterior a dicha información almacenarla en registros o base datos por medio del software MATLAB [43]. Registrando observaciones de cada animal en sus diferentes estados, especificado o detallados en la sección 4.3.

**Figura 3:** Señal piezo del estado REM, en un periodo de tiempo.



**Figura 4:** Señal piezo del estado NREM (*Slow-wave sleep (SWS)*) en un periodo de tiempo



## 5.5. INTRODUCCIÓN A MATLAB.

MATLAB [43] (abreviatura de MATrix LABoratory, "laboratorio de matrices") es un software matemático que ofrece un entorno de desarrollo con un lenguaje de programación propio denominado lenguaje M. Entre sus capacidades básicas están: la manipulación de tablas de datos, la representación de datos y funciones, la implementación de algoritmos, y la comunicación con programas en otros lenguajes y dispositivos hardware. Además se puede ampliar sus posibilidades con las cajas de herramientas, denominadas "toolboxes" que son actividades específicas para una gran variedad de aplicaciones, y que abordan para casi todas las necesidades y requerimientos de las investigaciones.

Actualmente, es considerado como el software más usado tanto en universidades y centros de investigación y desarrollo. En los últimos años ha aumentado el número de prestaciones, como la de programar directamente procesadores digitales de señal. MATLAB puede utilizarse para realizar cálculos sencillos de manera directa, como si se tratase de una calculadora, o bien programar tareas complejas para realizar otros cálculos de manera más automatizada. Esta herramienta tecnológica dispone también de un amplio abanico de programas de apoyo especializado, que extienden significativamente el número de funciones incorporadas en el programa principal, los mismos que cubren casi todas las áreas principales en el mundo de la ingeniería y la simulación.

En el presente estudio se utiliza MATLAB para: el almacenamiento de datos registrados en el estudio de los ratones por medio de los sensores de presión piezoeléctricos, detallado en la sección 4.3; pre-procesamiento de los datos que corresponde al filtrado y eliminación de ruidos, mediante la Densidad Espectral de Potencia (PSD), con el uso de la Transformada Rápida de Fourier (TTF); extracción de parámetros lineales y no-lineales de la señal piezo con PSD; y finalmente para el cálculo de la Entropía de Espectral de Shannon (EES) y la Dimensión Fractal (DF), basados tanto de algoritmo de Higuchi como el implementado por Katz, insumos que son fundamentales para el aprendizaje del modelo de clasificación automático.

## 5.6. PRE-PROCESAMIENTO DE LA SEÑAL PIEZO

El pre-procesamiento de los datos es el primer paso de varios para la toma de decisiones, y uno de los más decisivos en el resultado. Es parte de la Minería de Datos para detectar y eliminar ruidos en las señales, y además transformarlos o reducirlos para hacerlos más manejables. En el pre-procesamiento en general se requiere extraer de las señales los segmentos (filtrado) que contengan información útil para caracterizarla. La segmentación consiste en dividir la señal en segmentos que posean cierta estacionariedad, es decir señales (valores) que no varíen con el tiempo o que si lo hacen sea de forma lenta y suave [44].

Para el pre-procesamiento de la señal es necesario realizar el tratamiento de los mismos, dicho paso consiste en seleccionar las frecuencias deseadas de la señal y rechazar las no deseadas. El punto de corte del filtro es un valor de frecuencia determinado por las características de frecuencia de la señal que se estudia. Para lo cual se puede utilizar varios tipos de filtro:

- a. *Filtro pasa-bajas*: deja pasar las bajas frecuencias y lo que esté por encima de un determinado valor no pasa.
- b. *Filtro pasa-altas*: al contrario que el anterior, este filtro deja pasar las altas frecuencias pero no las bajas.
- c. *Filtro pasa-banda*: combinación de pasa-altas y pasa-bajas.
- d. *Filtro Notch o supresor de banda*: se programa para impedir el paso específico de una señal que no interesa.

Todas las señales piezo se filtraron a 128 Hz de orden del filtro, de respuesta de impulso finito en MATLAB, con una frecuencia de corte inferior de 0,5 Hz, y una frecuencia de corte superior a 11 Hz. El límite inferior permite la carga de corriente continua para ser eliminada, mientras que el punto de corte superior se utilizó para reducir la interferencia de las frecuencias más altas fuera del rango en el que los movimientos de interés tienen la mayor parte de su energía. Si existen señales por de frecuencia por encima de ese rango podría ser considerados ruido y no se las utiliza.

Luego del filtrado de las señales se procede con la eliminación de señal no deseada, ruidos y/o señal considerada artefactos, con el propósito de dejarla limpia, para quedarse con datos que posea información útil. Se basa en la frecuencia y longitudes temporales, esto se logra mediante la Transformada Discreta de Fourier (DFT), que permite el tratamiento de señales obtenidas por los sensores de presión piezoeléctricos, sin embargo en la actualidad por costo computacional, rapidez y eficiencia del algoritmo se utiliza la Transformada Rápida de Fourier (FFT). La FFT permite el tratamiento digital de señales y filtrado digital, para la reducción de ruido en señales como el ruido blanco, análisis de frecuencia de cualquier señal discreta.

El análisis Fourier, se basa en frecuencias determinadas, que caracterizan la señal que se desea analizar en el dominio del tiempo y que transforma la señal en su representación de dominio de frecuencia discreta. Como último paso se realiza la selección de los datos para el conjunto de entrenamiento (train) y de test. Los mismos que permiten el aprendizaje automático para el clasificador, mediante el modelado de algoritmos disponibles en la herramienta WEKA, o que son implementados para su experimentación como el caso del Convolutional Neural Networks (CNN) [45]

## **5.7. HERRAMIENTA WEKA.**

WEKA, acrónimo de Waikato Environment for Knowledge Analysis, es un conjunto de librerías JAVA para la extracción de conocimiento desde bases de datos. Está constituido por una serie de paquetes de código abierto con diferentes técnicas de pre-procesado, clasificación, agrupamiento, asociación y visualización. Se trata de un software desarrollado en la Universidad de Waikato (Nueva Zelanda) bajo licencia GNU-GPL lo cual ha impulsado que sea una de las suites más utilizadas en la actualidad para la Minería de Datos. Se trata de una herramienta de gran potencia, por lo que se la utiliza en el presente trabajo investigativo como: aprendizaje automático y generación de conocimiento para modelar el algoritmo de clasificación, en base a los que corren bajo en entorno de WEKA. La aplicación disponible de la herramienta que se utilizó para los el aprendizaje del modelo es EXPLORER, la misma que permite entre otras tareas llevar a cabo la ejecución de los algoritmos de clasificación y regresión, implementados sobre los ficheros de entrada.

## **6. MÉTODOS.**

### **6.1. EXTRACCIÓN DE PARÁMETROS.**

#### **6.1.1. Densidad Espectral de Potencia (PSD).**

Para el cálculo de la Densidad Espectral de Potencia (PSD) se utiliza la transformada de Fourier para intuir que el grueso de la potencia estará centrado, y de esta manera conocer cómo se encuentra distribuida la potencia en función de la frecuencia. El análisis espectral constituye un método de análisis poderoso que pone en evidencia periodicidades ocultas en una serie temporal. La PSD es una función real positiva de una variable de frecuencia asociada con un proceso estocástico o una función determinística del tiempo que tiene unidades de potencia por hercio o Hertz (Hz).

Para obtener la PSD se utilizó una función ventana hamming, con la cual se permite determinar las características de potencia de la señal previamente procesadas. Los valores que se obtiene no son infinitos, pues están divididos por el periodo de la señal. La PSD muestra como está dispersa la potencia en función de la frecuencia, por lo que se obtienen nueve parámetros. Dichos parámetros o características permiten el aprendizaje de los algoritmos de clasificación, además del cálculo de funciones necesarias para la aplicación de las técnicas no-lineales como son: la Entropía Espectral de Shannon (EES) y la Dimensión Fractal (DF) basados en los algoritmos Higuchi y Katz.

#### **6.1.2. Entropía Espectral de Shannon (EES)**

La Entropía de Shannon (H) se la define como una medida de incertidumbre media de la cantidad de información enviada en un mensaje, esto implica que cuanto mayor sea el valor de H, mayor será la incertidumbre. Es decir, los sucesos que son inciertos contienen mucha información, por lo que la entropía asociada a ellos será elevada. Por tanto, puede considerarse que la entropía de Shannon es una medida de la cantidad de información asociada a un determinado suceso. Dada una variable aleatoria discreta Z, cuya distribución

de probabilidad es  $P = (p_1, p_2, \dots, p_N)$ , la entropía de Shannon  $H(X)$  para dicha variable viene dada por la siguiente expresión [52]:

$$H(X) = - \sum_{x_i \in \Theta} p(x_i) \log p(x_i) = - E[\log p(x_i)]$$

Desde el punto de vista del procesamiento de señal, se realizaron variaciones de la entropía, por lo cual se la denominada Entropía Espectral de Shannon (EES, Shannon Spectral Entropy), en la que la irregularidad de una señal se estima en términos de la uniformidad del espectro de potencia normalizado. De esta manera, un espectro de potencia plano con un contenido espectral uniforme obtendría una EES elevada. Por otro lado, un espectro de potencia estrecho formado por unas pocas componentes espectrales tendría una EES baja. Así pues, la EES puede considerarse como un cuantificador del desorden de una señal, cuyo significado original implica incertidumbre en la información en términos de desorden, discrepancia y diversidad. El algoritmo utilizado en el presente Trabajo de Final de Master para calcular la EES se detalla a continuación:

1. Se obtiene la PSD de la señal previamente pre-procesada, a partir de la Transformada de Fourier, y de la función de auto-correlación. Representan la distribución de la potencia de la señal en función de la frecuencia.
2. Se normaliza con respecto a toda la potencia en la banda de frecuencias comprendida, entre 0,5 Hz y 11 Hz, permitiendo obtener una función de densidad de probabilidad.
3. La EES se obtiene como la entropía de Shannon de la función de densidad de probabilidad en el intervalo de frecuencias considerado.

$$EES = \sum_{f=r_1}^{r_2} P_x(f) * \ln \left[ \frac{1}{P_x(f)} \right]$$

Donde  $r_1$  y  $r_2$  son las frecuencias discretas correspondientes a las frecuencias de corte.

La EES permite cuantificar la distribución de potencia del espectro de una señal. Un valor de EES elevado indica que el espectro de la señal es uniforme y tiene una distribución en frecuencia bastante amplia, mientras que un valor bajo se corresponde con un espectro donde la potencia se encuentra condensada en un rango de frecuencias menor. Además, si se compara varias señales entre sí, un valor menor de Entropía Espectral sugiere que esa señal es más regular y predecible.

### **6.1.3. Dimensión Fractal (DF)**

La Dimensión Fractal (FD) es una de las características más populares presente en la naturaleza. Un fractal es un objeto en el cual todas sus partes tienen alguna relación con el todo, los fractales son objetos cuya dimensión es no entera o fraccionaria, que representa de forma cuantitativa como una medida relativa del número de partes que forma el patrón. En otros términos, la DF permite cuantificar la complejidad y auto-similitud de las señales piezo [46-47].

La dimensión fractal, es una de las medidas más utilizadas para caracterizar la complejidad de un sistema [50-51] y es importante realizar su cálculo y utilizarlo en el aprendizaje del modelado, por lo que es imprescindible elegir un algoritmo para su medición. Existen varios algoritmos de este tipo, en presente estudio se centra en las alternativas que son especialmente adecuados para el análisis de series de tiempo y que no necesita modelos anteriores del sistema. Dos de dichos algoritmos son Higuchi [16] y Katz [17], por el nombre de sus autores. Dichos algoritmos se son utilizados por su disponibilidad y fácil implementación en el presente trabajo [48].

#### **6.1.3.1. Dimensión Fractal de Higuchi (DFH)**

El algoritmo de Higuchi para calcular la Dimensión Fractal de una serie temporal requiere que las señales a analizar sean estacionarias. Sin embargo, la Dimensión Fractal Higuchi (DFH) permite analizar intervalos cortos de señal, donde es probable que la condición de estacionariedad se presente. Higuchi [16] ha propuesto un algoritmo para la medición de la

dimensión fractal de las secuencias de tiempo discretos directamente de las series de tiempo  $x(1), x(2), \dots, x(n)$ , donde  $N$  es el tamaño de la señal. No es necesario entrar en detalles, pero básicamente el algoritmo calcula la longitud  $L_m(k)$  (véase la Ecuación 1) para cada valor de  $m$  y  $k$  que cubre toda la serie [51].

$$L_m(k) = \frac{\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |x(m+ik) - x(m+(i-1)k)| (n-1)}{\lfloor \frac{N-m}{k} \rfloor k} \quad (1)$$

Después de eso, una suma de todas las longitudes  $L_m(k)$  para cada  $k$  se determina con la Ecuación 2.

$$L(k) = \sum_{m=1}^k L_m(k) \quad (2)$$

Y, por último, la pendiente de la curva de  $\ln(L(k)) / \ln(1/k)$  se estima utilizando mínimos cuadrados lineal mejor ajuste, y el resultado es la dimensión fractal Higuchi (HFD).

### 6.1.3.2. Dimensión Fractal de Katz (DFK)

El término Dimensión Fractal (DF) fue introducido por Mandelbrot para el estudio temporal o espacial de fenómenos continuos que muestran correlaciones en un amplio rango de escalas. En términos estrictos, la DF mide el grado de fragmentación o irregularidad sobre múltiples escalas del borde o frontera de la estructura fractal bajo análisis.

El cálculo de la dimensión fractal de Katz (DFK) es ligeramente más lento que otros algoritmos, pero se deriva directamente de la forma de onda, eliminando la etapa de pre-procesamiento de crear una secuencia binaria. Katz [17] ha propuesto una fórmula normalizada de la dimensión fractal (ver la ecuación 3).

$$FD = \frac{\log(n)}{\log(n) + \log\left(\frac{d}{L}\right)} \quad (3)$$

En este caso L es la longitud y d es la extensión de la curva, y se normalizan utilizando el promedio paso a = L / n usando las Ecuaciones 4 y 5:

$$L = \sum_{i=1}^n l_{i,i+1} \quad (4)$$

$$d = \max\{l_{i,j}\} \quad (5)$$

Donde l es la distancia euclidiana entre dos puntos como en la Ecuación 6.

$$l_{ij} = \sqrt{(y_i - y_j)^2 + (x_i - x_j)^2} \quad (6)$$

## 6.2. CLASIFICADOR AUTOMÁTICO

### 6.2.1. Análisis Discriminante Lineal Fisher (FLDA)

Uno de los experimentos para el modelo de la clasificación automática de comportamientos animales, basado en los estados REM y NREM se los realiza en base del Análisis Discriminante Lineal de Fisher, que permite de una manera rápida y eficiente la clasificación entre clases. Estos valores permiten comprobar el desenvolvimiento del modelo automático basado en las Redes Neuronales Artificiales. El *análisis discriminante*, es una técnica de clasificación para asignar nuevas observaciones a grupos ya conocidos, es decir, corresponde al aprendizaje supervisado para el reconocimiento de patrones encontrando una combinación lineal entre dos o más clases conocidas.

Análisis discriminante de Fisher, permite diferenciar los grupos para el proceso de clasificación. Son combinaciones lineales de las variables, interviniendo cada una con un peso diferente que indica las que más discriminan. La función discriminante de Fisher  $D$  se obtiene como función lineal de  $k$  variables explicativas  $X$ , es decir:

$$D = u_1X_1 + u_2X_2 + \dots + u_kX_k$$

El análisis discriminante de Fisher se lo hace correr en la herramienta WEKA, en base al modelo utilizado todo el conjunto de datos para entrenamiento (use training set). Se lo analiza en base a la tasa de acierto: Correcta Clasificación de la Instancia (CCI), la Kappa Estadística (KE) y la Incorrecta Clasificación de la Instancia (ICI) con un valor de umbral  $1,0 \times 10^{-6}$ . Acotar que los CCI, KE e ICI serán para los tres tipos de algoritmos o técnicas representativas en la Minería de Datos.

### **6.2.2. Perceptrón Multicapa (MLP)**

El Perceptrón Multicapa es una red neuronal artificial (RNA) formada por múltiples capas, esto le permite resolver problemas que no son linealmente separables. La arquitectura de perceptrón multicapa se caracteriza porque tiene sus neuronas agrupadas en capas de diferentes niveles. Cada una de las capas está formada por un conjunto de neuronas y se distinguen tres tipos de capas diferentes: la capa de entrada, las capas ocultas y la capa de salida.

Las neuronas de la capa de entrada no actúan como neuronas propiamente dichas, sino que se encargan únicamente de recibir las señales o patrones del exterior y propagar dichas señales a todas las neuronas de la siguiente capa. La última capa actúa como salida de la red, proporcionando al exterior la respuesta de la red para cada uno de los patrones de entrada. Las neuronas de las capas ocultas realizan un procesamiento no-lineal de los patrones recibidos.

La configuración para determinar el modelo de clasificación automático mediante el algoritmo MLP, se base en una red oculta de 100 (hiddenLayers) y con un tiempo de aprendizaje de 1,0 (learningRate).

### **6.2.3. Red Neuronal Convolutacional (CNN)**

Una Red Neuronal Convolutacional es un tipo de red neuronal artificial donde las neuronas corresponden a campos receptivos, este tipo de red es una variación de un perceptrón multicapa, pero su funcionamiento las hace mucho más efectivas en la clasificación de imágenes y señales de voz. Sin embargo, es adaptable en el procesamiento de datos en arreglos multidimensionales. La principal ventaja es que dispone menos parámetros a entrenar que una red multicapa con conexión total del mismo número de capas ocultas, por lo que su entrenamiento es más rápido. Además, presenta invarianza a la traslación de los patrones a identificar.

La red neuronal profunda de convolución permite la delegación en tarea de seleccionar las características importantes a las capas iniciales de convolución de la red. En éstas, la importancia de las características se refleja en los pesos del kernel (núcleo) de cada mapa de características. Los parámetros se definieron en base al balance entre la cantidad de información y los requerimientos de memoria y tiempo de procesamiento. Para la presente de investigación se utiliza experimentos para el entrenamiento basados en la topología de las redes neuronales, que se basaron en pruebas con:

- ✓ 2 capas, tanto para 100,100 neuronas y de 50, 50.
- ✓ 2 capas de 50 o 100 mapas de características (feature maps) y con varios valores para patch-size y también para la propiedad max-pooling.
- ✓ Variaciones en la iteración: 100, 50, 10 y 1000 para elegir el mejor modelo.

La configuración establecida para realizar los experimentos con CNN es con dos capas de mapas de características de 100, además al ser los valores de entrada vectores multidimensionales con “patch-size” de 2x2 y con una 1x1 de operación “max-pooling”.

#### **6.2.4. Árboles de Decisión**

La Clasificación, mediante un árbol de decisión produce datos con una estructura binaria arborescente, lo cual facilita la interpretación e identificación de las variables significativas. Un modelo de Árbol de decisión contiene normas para predecir la variable de destino. En la clasificación se parte de un conjunto de registros de observaciones, en el que cada registro consta de varios campos o atributos. Los atributos son numéricos o categóricos. Uno de los atributos, denominado campo de etiqueta de clase, indica la clase a la que pertenece la observación. El objetivo de la clasificación basada en árboles de decisión, es crear un modelo de la etiqueta de clase basándose en los demás atributos. Tras crear un modelo, se puede utilizar para determinar la etiqueta de clase de los registros no clasificados.

En comparación con otros métodos, los árboles de decisión se pueden crear de una forma relativamente rápida. Además los modelos de árbol de decisión son simples y fáciles de comprender. Un árbol de decisión es un discriminador de clases que realiza la partición de modo recursivo al conjunto de entrenamiento hasta que cada partición consta completamente o predominantemente de observaciones pertenecientes a una sola clase. Cada nodo sin hojas del árbol contiene un punto de división que es una prueba de uno o varios atributos y determina cómo se realiza la partición de los datos.

## **7. RESULTADOS**

Con el fin de obtener las características necesarias para el modelamiento del clasificador automático en los diferentes algoritmos, que permita realizar un análisis en base a los resultados arrojados de los parámetros lineales y no-lineales obtenidos por medio de PSD. Utilizando las técnicas no-lineales que permiten mejorar el comportamiento de las señales no estacionarias, basadas en el cálculo de Entropía Espectral de Shannon y de la Dimensión Fractal, por medio de la implantación de algoritmos disponibles para su cálculo (Higuchi y Katz).

Con estas características se procede a realizar el modelo con el conjunto de entrenamiento, con 3500 observaciones (50% REM y 50% NREM). De esta manera se realiza el aprendizaje del modelo de la clasificación automática de comportamientos animales, observados en base a sus registros de dos estados (REM y NREM). Para este análisis se dispone de la correcta clasificación del algoritmo, de los que no han sido clasificados adecuadamente y del estadístico Kappa, que permite el análisis de la correlación entre las variables o características.

Se utiliza para ellos las técnicas específicas de la minería de datos, es decir, en base a los algoritmos de Análisis Discriminante Lineal, Redes Neuronales Artificiales (CNN y MLP) y Árboles de Decisión, que permiten en base a sus resultados y con las combinaciones entre parámetros no-lineales como la Entropía Espectral de Shannon (EES) y la Dimensión Fractal en base a Higuchi y Katz, con el propósito de mejorar el aprendizaje de los modelos de clasificación.

### **7.1. ANÁLISIS DISCRIMINANTE LINEAL DE FISHER.**

Los resultados obtenidos con el algoritmo de clasificación basado en el Análisis Discriminante Lineal de Fisher (FLDA), en base a experimentos con las características obtenidas por medio de la Densidad Espectral de Potencia (PSD), después del debido filtrado y pre-procesamiento de la información. Luego se realiza pruebas con la combinación de cada característica obtenida, Entropía Espectral de Shannon (EES) y con las

dimensiones fractales calculadas tanto para el algoritmo de Higuchi (DFH) y de Katz (DFK), para finalmente realizar modelados con todas las características, y de esta manera tomar decisiones sobre el comportamiento (rendimiento) del clasificador.

**Tabla 1:** Porcentajes de acierto en base a las características mediante FDLA

<b>CARACTERÍSTICAS</b>	<b>% CCI</b>	<b>KE</b>	<b>% ICI</b>
<b><i>PSD</i></b>	79,54	0,59	20,46
<b><i>PSD + EES</i></b>	79,80	0,60	20,20
<b><i>PSD + DFH</i></b>	79,74	0,59	20,26
<b><i>PSD + DFK</i></b>	79,80	0,60	20,20
<b><i>PSD + EES + DFH</i></b>	79,83	0,60	20,17
<b><i>PSD + EES + DFK</i></b>	79,83	0,60	20,17
<b><i>PSD + DFH + DFK</i></b>	79,80	0,60	20,20
<b><i>PSD + EES + DFH + DFK</i></b>	80,11	0,60	19,89

## **7.2. PERCEPTRON MULTICAPA (MLP)**

Este es uno de los tipos de redes neuronales más comunes. Se basa en otra red más simple llamada perceptrón simple solo que el número de capas ocultas puede ser mayor o igual que una. Las neuronas de la capa oculta usan como regla de propagación la suma ponderada de las entradas con los pesos sinápticos y sobre esa suma ponderada se aplica una función de transferencia de tipo sigmoide, que es acotada en respuesta.

Los experimentos con el algoritmo de redes neuronales MLP, se lo realizó en base a 100 capas ocultas y con un tiempo de aprendizaje de 1,0. Esto permitió lograr mejores resultados de varios experimentos realizados. En la Tabla 2, se presenta los resultados obtenidos con los datos de entrenamiento, con función a cada característica obtenida en el presente estudio.

**Tabla 2:** Porcentajes de acierto en base a las características mediante MLP

<b>CARACTERÍSTICAS</b>	<b>% CCI</b>	<b>KE</b>	<b>% ICI</b>
<b><i>PSD</i></b>	89,06	0,78	10,94
<b><i>PSD + EES</i></b>	88,94	0,78	11,06
<b><i>PSD + DFH</i></b>	88,29	0,77	11,71
<b><i>PSD + DFK</i></b>	89,94	0,80	10,06
<b><i>PSD + EES + DFH</i></b>	88,63	0,77	11,37
<b><i>PSD + EES + DFK</i></b>	88,89	0,78	11,11
<b><i>PSD + DFH + DFK</i></b>	88,71	0,77	11,29
<b><i>PSD + EES + DFH + DFK</i></b>	89,14	0,78	10,86

### **7.3. REDES NEURONALES CONVOLUCIONALES (CNN)**

La configuración del algoritmo CNN está basado en dos capas de 100 cada una (feature maps) y con un valor de 2x2 para patch-size y la propiedad de max-pooling de 1x1. Se obtiene los resultados presentados en la tabla 3. Los mismo que han sido seleccionados en base a pruebas con diferentes mapas de características y números de capas, así como variando el número de interacciones con el algoritmo.

**Tabla 3:** Porcentajes de acierto en base a las características mediante CNN

<b>CARACTERÍSTICAS</b>	<b>% CCI</b>	<b>KE</b>	<b>% ICI</b>
<b><i>PSD</i></b>	83,51	0,67	16,49
<b><i>PSD + EES</i></b>	84,46	0,69	15,54
<b><i>PSD + DFH</i></b>	83,94	0,68	16,06
<b><i>PSD + DFK</i></b>	84,14	0,68	15,85
<b><i>PSD + EES + DFH</i></b>	84,80	0,70	15,20
<b><i>PSD + EES + DFK</i></b>	84,14	0,68	15,85
<b><i>PSD + DFH + DFK</i></b>	83,40	0,67	16,60
<b><i>PSD + EES + DFH + DFK</i></b>	83,34	0,67	16,66

#### 7.4. ÁRBOLES DE DECISIONES (J48)

El algoritmo J48 de WEKA, es una implementación del algoritmo C4.5, uno de los algoritmos de minería de datos más utilizado. Se trata de un refinamiento del modelo generado con OneR. Supone una mejora moderada en las prestaciones, y puede conseguir una probabilidad de acierto ligeramente superior al del anterior clasificador.

El parámetro más importante que se debe tener en cuenta es el factor de confianza para la poda "confidence level", que influye en el tamaño y capacidad de predicción del árbol construido. Para cada operación de poda, define la probabilidad de error que se permite a la hipótesis de que el empeoramiento debido a esta operación es significativo. A probabilidad menor, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. Para el presente estudio se construye un árbol de decisión con un valor del factor de confianza para la poda del 25%.

**Tabla 4:** Porcentajes de acierto en base a las características mediante J48

<b>CARACTERÍSTICAS</b>	<b>% CCI</b>	<b>KE</b>	<b>% ICI</b>
<b><i>PSD</i></b>	84,43	0,67	15,57
<b><i>PSD + EES</i></b>	84,94	0,70	15,06
<b><i>PSD + DFH</i></b>	84,51	0,69	15,49
<b><i>PSD + DFK</i></b>	85,37	0,71	14,63
<b><i>PSD + EES + DFH</i></b>	85,63	0,71	14,37
<b><i>PSD + EES + DFK</i></b>	85,20	0,70	14,80
<b><i>PSD + DFH + DFK</i></b>	85,51	0,69	15,49
<b><i>PSD + EES + DFH + DFK</i></b>	84,71	0,69	15,29

Los resultados expuestos son en base a experimentos realizados en la herramienta WEKA que sirve para analizar el aprendizaje de los diferentes algoritmos en el aprendizaje de un modelo de clasificación, para comportamientos animales basados en los estados REM y NREM (ratones de laboratorio), basados en métodos no-invasivos (sensores de presión piezoeléctricos) y técnicas no-lineales como características obtenidas por medio de la Entropía de Espectral de Shannon y la Dimensión Fractal.

## 8. DISCUSIÓN

En este estudio se ha enfocado al análisis de rendimiento de las técnicas de clasificación basadas en reglas que implementa el software WEKA, sobre los datos obtenidos de comportamientos animales. Se ha utilizado 3500 observaciones para el conjunto de entrenamiento (50 % de cada clase) y en base a ellos se realiza experimentos con los algoritmos de clasificación para el aprendizaje del modelo en base a: Análisis Discriminante Lineal, Redes Neuronales y Árboles de decisiones, por lo que consiste en clasificación de estados de sueño tanto en REM y NREM (*Slow-wave sleep* (SWS)).

En los resultados del Análisis Discriminante Lineal de Fisher se observa que con la utilización de los parámetros en base a la técnica no-lineal (EES, DFH y DFK) y las características obtenidas con la Densidad Espectral de Potencia (PSD) mejora en 1% respecto únicamente a los parámetros del conjunto de PSD. Aunque respecto a las mismas pruebas con los algoritmos en base a redes neuronales mejora con Perceptron Multicapa (MLP) en un 9% y con el Redes Neuronal Convulsión (CNN) hasta un 3%. Estos porcentajes se diferencian únicamente al utilizar el tipo de clasificador basado en la inteligencia artificial.

Con el algoritmo MLP se observa la mejor clasificación con las características del conjunto PSD se obtiene 89 % en aciertos, combinando con la Dimensión Fractal de Katz (DFK), se produce un mejor rendimiento con 89,94 %. Con la utilización de los características PSD más el valor de EES, con las dimensiones fractales DFH y DFK mejora. Es importante indicar que con el algoritmo MLP se obtiene un rendimiento muy estable con las diferentes características de la señal del comportamiento de los animales, notándose una mejora considerable (9 %) con respecto al discriminante lineal.

El algoritmo CNN presta una mejora clasificación con respecto al FDLA, pero no así en relación al MLP. Sin embargo la Entropía Espectral de Shannon (EES) mejora en un 1% respecto al conjunto de características extraídas con PSD, si a esto se le combina con la Dimensión Fractal de Higuchi (DFH) se obtiene el mejor porcentaje de clasificación (84,80 %) notándose que tanto EES y DFH permiten elevar el rendimiento del algoritmo.

En cambio, el Árbol de Decisión J48 se observa que presenta una ligera mejora con respecto a CNN, sin embargo está por debajo de MLP. Al igual que en los otros clasificadores ayuda a mejorar el comportamiento la característica de la Entropía Espectral de Shannon (EES) cuasi en un 0,5%. Mientras que con relación a la Dimensión Fractal, el algoritmo de Katz presenta mejores resultados con un 1% de mejoría. Con la combinación de todos los parámetros obtenidos para el aprendizaje supervisado hay mejoría con respecto a solo el conjunto de características extraídas con PSD.

Con el contraste de las cuatro tablas se puede indicar que la técnica para clasificación que mejor resultados brinda es en base a las Redes Neuronales Artificiales, específicamente con el Multicapa Perceptron. Sin embargo es importante acotar que tanto las Redes Neuronales, como los Árboles de Decisión presentan mejor rendimiento que el Análisis Discriminante Lineal, por lo que las técnicas no-lineales presentan mejoras para el tratamiento de señales obtenidas de los sensores de presión piezoeléctricos, como lo indica [47; 49].

Al ser un estudio piloto, pues los resultados expuestos y analizados representan al monitoreo y estudio de dos ratones de laboratorio en sus periodos REM y NREM, por lo que actualmente no se puede apreciar mejoras significativas en las tasas de acierto por medio la utilización de los parámetros no-lineales como: la Entropía de Espectral de Shannon y Dimensión Fractal. Sin embargo, estos resultados sin ninguna duda mejorarían al estudiar más animales e incrementar parámetros no-lineales Multivariate Multiscale Permutation Entropy.

Con todo ello, se puede manifestar que la presente investigación continuara con el estudio incorporando más animales, siempre conforme a las normas éticas establecidas por el país de Francia y la Unión Europea. Además con el análisis de parámetros no-lineales como Permutación de la Entropía, tanto Multivariante y Multiescala, dicho trabajo futuro presentará mejorías notables y contribuirá de mejor manera a la Ingeniería Biomédica, objetivo del presente investigación.

## 9. CONCLUSIONES

En esta sección se indicara las conclusiones que se llega después de haber concluido el presente trabajo académico sobre: Técnicas no-lineales para el análisis no invasivo de comportamientos animales.

El principal objetivo del presente trabajo es el análisis de las características en el comportamiento de los animales observados, orientado a la investigación y la detección temprana de cambios en los comportamientos animales, siendo estas características de fundamental importancia en la aplicación futura de la Ingeniería Biomédica. Características relativas del comportamiento animal en base a los sensores de presión piezoeléctricos, el dominio espectral y la dimensión fractal fueron analizados como características no-lineales.

Se ha añadido funciones como los algoritmos de Higuchi y Katz, para el proceso de aprendizaje del modelo de clasificación que permiten apreciar ciertas mejoras con dominio espectral y a su vez con las dimensiones fractales. Aunque no hay una mejoría significativa entre los conjuntos de características evaluados, se puede proyectar que con más datos de animales (ratones) observados mejorara significativamente estos registros, pues hay que considerar que es una estudio piloto, por el bajo número de animales grabados y monitoreados.

Las técnicas de Minería de Datos respecto a Redes Neuronales Artificiales aportan mejores rendimientos en el aprendizaje para el modelo de clasificación automático, como se puede observar en las tablas ilustradas en la sección resultados, respecto a las otras técnicas como Análisis Discriminante Lineal y Árboles de Decisiones, que el presente estudio muestra como contraste para el análisis de señales piezo obtenidas de los sensores de presión piezoeléctricos.

En el aspecto académico a título personal hay una significativa mejoría respecto a los conocimientos, pues se conocía aspectos conceptuales y práctica con datos realizados en base a la Minería de Datos, pero nada comparado a un estudio real con enfoque al avance de la ciencias médicas, donde se involucró capacitación e investigación de métodos y técnicas que se utilizan en la actualidad para resolver estudios avanzados como este.

Sin embargo, dicho trabajo mejoraría sustancialmente al incorporar nuevas cepas de ratones para su estudio, y en base a la implementación de nuevos parámetros no-lineales, los mismos que son capaces de detectar cambios sutiles en la señal. Los resultados se podrán mejorar en un futuro con una base de datos más amplia de animales, pues se trata de un estudio piloto y los actuales resultados son esperanzadores.

## **10. LÍNEAS FUTURAS.**

El presente trabajo es trata de un estudio piloto, porque se observa a dos animales de laboratorio (ratones) en sus estados REM y NREM (*Slow-wave sleep* (SWS)). Sin embargo, los resultados se podrán mejorar en un futuro con una base de datos más amplia de animales y con nuevos parámetros no-lineales como la Multivariate Multiscale Permutation Entropy, que permitirá sin ninguna duda mejores rendimientos en los aciertos del clasificador automático, a las presentadas en la sección resultados. Por lo que aún queda mucho trabajo por realizar respecto al presente trabajo de estudio.

## 11. REFERENCIAS

- [1] HHMI. 2001. "The Genes We Share with Yeast, Flies, Worms, & Mice: New Clues to Human Health & Disease" ("Los Genes que compartimos con la levadura, las moscas, los gusanos y los ratones: nuevas pistas para la salud humana y las enfermedades"). Disponible en online: <http://www.actionbioscience.org/esp/genomica/hhmi.html>
  
- [2] G.M. Rubin et al., "Comparative genomics of the eukaryotes," *Science* 287 (March 24, 2000): 2210-11.
  
- [3] N. Maubourguet, A. Lesne, J. P. Changeux, U. Maskos, and P. Faure, "Behavioral sequence analysis reveals a novel role for beta2\* nicotinic receptors in exploration," *PLoS. Comput. Biol.*, vol. 4, no. 11, p. e1000229, Nov.2008.
  
- [4] U. Maskos, B. E. Molles, S. Pons, M. Besson, B. P. Guiard, J. P. Guilloux, A. Evrard, P. Cazala, A. Cormier, M. Mameli-Engvall, N. Dufour, I. Cloez-Tayarani, A. P. Bemelmans, J. Mallet, A. M. Gardier, V. David, P. Faure, S. Granon, and J. P. Changeux, "Nicotine reinforcement and cognition restored by targeted expression of nicotinic receptors," *Nature*, vol. 436, no. 7047, pp. 103-107, July2005.
  
- [5] R. K. Schwarting, R. Goldenberg, H. Steiner, J. Fornaguera, and J. P. Huston, "A video image analyzing system for open-field behavior in the rat focusing on behavioral asymmetries," *J. Neurosci. Methods*, vol. 49, no. 3, pp. 199-210, Sept.1993.
  
- [6] A. J. Spink, R. A. Tegelenbosch, M. O. Buma, and L. P. Noldus, "The EthoVision video tracking system--a tool for behavioral phenotyping of transgenic mice," *Physiol Behav.*, vol. 73, no. 5, pp. 731-744, Aug.2001.
  
- [7] A. A. Megens, J. Voeten, J. Rombouts, T. F. Meert, and C. J. Niemegeers, "Behavioral activity of rats measured by a new method based on the piezo-electric principle," *Psychopharmacology (Berl)*, vol. 93, no. 3, pp. 382-388, 1987.

- [8] G. M. Mang, J. Nicod, Y. Emmenegger, K. D. Donohue, B. F. O'Hara, and P. Franken, "Evaluation of a piezoelectric system as an alternative to electroencephalogram/ electromyogram recordings in mouse sleep studies," *Sleep*, vol. 37, no. 8, pp. 1383-1392, 2014.
- [9] A. E. Flores, J. E. Flores, H. Deshpande, J. A. Picazo, X. S. Xie, P. Franken, H. C. Heller, D. A. Grahn, and B. F. O'Hara, "Pattern recognition of sleep in rodents using piezoelectric signals generated by gross body movements," *IEEE Trans. Biomed. Eng*, vol. 54, no. 2, pp. 225-233, Feb.2007.
- [10] J. Brodtkin, D. Frank, R. Grippo, M. Hausfater, M. Gulinello, N. Achterholt, and C. Gutzen, "Validation and implementation of a novel high-throughput behavioral phenotyping instrument for mice," *J. Neurosci. Methods*, vol. 224, pp. 48-57, Mar.2014.
- [11] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," *Phys. Rev. Lett.*, vol. 88, no. 17, p. 174102, Apr.2002.
- [12] M. P. Paulus, M. A. Geyer, L. H. Gold, and A. J. Mandell, "Application of entropy measures derived from the ergodic theory of dynamical systems to rat locomotor behavior," *Proc. Natl. Acad. Sci. U. S. A*, vol. 87, no. 2, pp. 723-727, Jan.1990.
- [13] H. Eguiraun, K. Lpez-de-Ipia, and I. Martinez, "Application of Entropy and Fractal Dimension Analyses to the Pattern Recognition of Contaminated Fish Responses in Aquaculture," *Entropy*, vol. 16, no. 11, pp. 6133-6151, 2014.
- [14] Disponible online: [https://es.wikipedia.org/wiki/Minera\\_de\\_datos](https://es.wikipedia.org/wiki/Minera_de_datos) (fecha de consulta 10 de septiembre 2016).
- [15] Revista de la Facultad de Ingeniera Industrial Vol. (7) 2: pp. 83-86 (2004) UNMSM ISSN: 1560-9146 (impreso) / ISSN: 1810-9993 (electrnico).
- [16] BRAVO. Milton. Profesor de UPSE: Materia de Comunicaciones. Densidad Espectral de Energa y Potencia. (2011).

- [17] Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D*, vol 31, nº 2, 1988, pp 277 - 83.
- [18] Katz MJ. Fractals and the analysis of waveforms. *Computer in Biology and Medicine*, vol 18, nº 3, 1988, pp 145 - 56.
- [19] Matlab. Disponible en: [http://www.mathworks.com/products/product\\_listing/index.html](http://www.mathworks.com/products/product_listing/index.html). (Fecha de consulta 11 de julio de 2016).
- [20] WEKA. Disponible online: <http://www.cs.waikato.ac.nz/ml/weka/>. (Fecha de consultado 15 de Agosto de 2016).
- [21] Zuñiga J, Tur M, Milocco S, Piñeiro R. *Ciencia y tecnología en protección y experimentación animal*. México: McGraw-Hill Interamericana, 2001; p. 682.
- [22] Balch C, Bleyer A, Krakoff I, et al. The vital role of animal research in advancing cancer diagnosis and treatment. *Cancer Bulletin* 1990; 42: 266-269.
- [23] *Guía para el cuidado y uso de animales de laboratorio*. Institute of laboratory animals National Research Council. México: Academia Nacional de Medicina, 2002, p. 148.
- [24] *Institutional Animal Care and Use Committee Guidebook*. 2<sup>nd</sup> ed. Office of Laboratory Animal Welfare. National Institutes of Health Department of Health and Human Services. Bethesda, MD. 2002; p. 210
- [25] Festing M, Vaughn S. The choice of animal model and reduction. *ATLA Alternatives to Laboratory Animals* 2004; 32(supplement 2) 59-64.
- [26] Balcombe J. Laboratory environments and rodents behavioural needs, a review. *Laboratory Animals* 2006; 40 (3): 217-235.
- [27] Gent N. A study in refining husbandry techniques for the in house breeding of rats and mice. *Animal Technology and Welfare* 2006; 5(1): 3-8.

- [28] Olfert ED, Cross BM and Mc William AA. Guide to the care and use of experimental animals, Vol 1. 2<sup>nd</sup> ed. Ontario: Canadian Council on Animal Care, 1998; p. 211.
- [29] Atchley W, Fitch W. Genes trees and the origins of inbred strain of mice. Science 1991; 254: 554-558.
- [30] Benavides F, Guénet J. Manual de genética de roedores de laboratorio: principios básicos y aplicaciones. Alcalá: Sociedad Española para las Ciencias del Animal de Laboratorio (SECAL)- Universidad de Alcalá, 2003; 312.
- [31] Russell W, Burch R. The principles of humane experimental technique. London: Methuen, 1992 (reprint is available from UFAW).
- [32] Vanda-Cantón B. La experimentación biomédica en animales en los códigos bioéticos. Lab Acta 2003;15: 69-73.
- [33] Monje C. Daniel. Conceptos electrónicos en la medida de aceleración y vibración. 2009, Escuela Politécnica de Sevilla.
- [34] Rechtschaffen A, Kales A, editors. Los Angeles: Brain Information Service/Brain Research Institute, University of California; 1968. A manual of standardized terminology, techniques and scoring system of sleep stages in human subjects.
- [35] Guamizo L, Cristian. Análisis de reducción de ruido en señales EEG orientado al reconocimiento de patrones; 2008. Revista Tecnológicas No 21.
- [36] Dapozo, Gladys t al., Técnicas de preprocesamiento para mejorar la calidad de los datos en un estudio de caracterización de ingresantes universitarios. Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura Universidad Nacional del Nordeste. 9 de Julio N° 1449. CP 3400.
- [37] Disponible online: [https://es.wikipedia.org/wiki/Aprendizaje\\_supervisado](https://es.wikipedia.org/wiki/Aprendizaje_supervisado). (Fecha de consulta 20 de septiembre de 2016).

- [38] Disponible online [https://www.ecured.cu/Minería\\_de\\_Datos](https://www.ecured.cu/Minería_de_Datos). (Fecha de consulta, 20 de septiembre de 2016).
- [39] Gómez P. Carlos. Análisis no lineal de registros Magnetoencefalográficos para la ayuda en el Diagnóstico de la enfermedad de Alzheimer (2009). Tesis Doctoral de la Universidad Valladolid, Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática.
- [40] Juan f. Guerrero Martínez. Estimación Espectral (2010 – 2011). Curso de formación para la Ingeniería Biomédica.
- [41] Disponible online: <http://www.addenfi.com/sitenew/index.php/produits/phenact/57-plateforme> (Fecha de consulta, 24 de septiembre 2016).
- [42] Steven R. Gooch. A method for non-invasive, Automated behavior classification In mice, using piezoelectric Pressure sensors. (2014).
- [43] Disponible online: <https://www.mathworks.com/products/matlab/>. (Fecha de consulta 15 de septiembre de 2016).
- [44] Guamizo, Cristian. Segmentación automática de señales no estacionarias utilizando la Transformada wavelet estacionaria (2007). Scientia et Technica Año XIII, No 34, Mayo de 2007. Universidad Tecnológica de Pereira. ISSN 0122-1701.
- [45] Disponible online: <http://weka.wikispaces.com/Unofficial+packages+for+WEKA+3.7> (Fecha de consulta, 24 julio de 2016).
- [46] Accardo A, Affinito M, Carrozzi M, Fouquet F. Use of the fractal dimension for the analysis of electroencephalographic time series. Biological Cybernetics, vol 77, nº 5, 1997, pp 339 - 50.

- [47] Eguiraun, H.; Lopez-de-Ipina, K.; Martinez, I. Application of entropy and fractal dimension analyses to the pattern recognition of contaminated fish responses in aquaculture. *Entropy* 2014, 16, 6133–6151, DOI: 10.3390/e16116133.
- [48] Disponible online: [https://www.mathworks.com/matlabcentral/fileexchange/50290-higuchi-and-katz-fractal-dimension-measures/content/Fractal\\_dimension\\_measures/Katz\\_FD.m](https://www.mathworks.com/matlabcentral/fileexchange/50290-higuchi-and-katz-fractal-dimension-measures/content/Fractal_dimension_measures/Katz_FD.m)
- [49] TSOI, A. C.; SO, D. S.; SERGEJEW, A.: Classification of electroencephalogram using artificial neural networks. *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 6*, J. D. COWAN, G. TESAURO, AND J. ALSPECTOR, ED. SAN FRANCISCO, CA: MORGAN KAUFMANN, PP. 1151- 1158, 1994.
- [50] Mandelbrot BB. How long is the coast of Britain? Statistical selfsimilarity and fractional dimension. *Science*, vol 156, 1967, pp 636 - 8.
- [51] Ezeiza, A., López de Ipiña, K., Hernández, C. et al. *Cogn Comput* (2013) 5: 545. doi:10.1007/s12559-012-9165-0
- [52] Eguiraun, H.; López-de-Ipiña, K.; Martinez, I. Shannon Entropy in a European Seabass (*Dicentrarchus labrax*) System during the Initial Recovery Period after a Short-Term Exposure to Methylmercury. *Entropy* 2016, 18, 209.

## 12. ANEXOS

### 12.1. CÁLCULO DE LA ENTROPÍA ESPECTRAL DE SHANNON.

Código utilizado para el cálculo de la Entropía Espectral de Shannon, implementado en el software Matlab. Previamente se obtiene los parámetros en base a la Densidad Espectral de Potencia (PSD). Código donde se extrae los parámetros lineales y no lineales pre-procesados.

```
function [results, results1] = Parametros

[datosTrain, datosTest] = preprocesoPiezo;

data = datosTrain{1};           % Se Obtiene los REM
[T x_dim] = size(data{1}');
nEx = length(data);
obs = T*nEx;
results = zeros(obs*2, x_dim); % Para REW y SWS
j = 1;

for i = 1:nEx,
    aux = data{i}';
    for k = 1:T,
        results(j,1:9) = aux(k,:);
        j = j + 1;
    end
end

data = datosTrain{2};           % Se Obtiene los SWS
for i = 1:nEx,
    aux = data{i}';
    for k = 1:T,
        results(j,1:9) = aux(k,:);
        j = j + 1;
    end
end

% Para obtener el conjunto de test

datos = datosTest{1};           % Obtengo los REM
[T1 x_dim1] = size(datos{1}');
nEx1 = length(datos);
obs1 = T1*nEx1;
results1 = zeros(obs1*2, x_dim1); % Para REW y SWS
j1 = 1;
```

```

for i = 1:nEx1,
    aux = datos{i}';
    for k = 1:T1,
        results1(j1,1:9) = aux(k,:);
        j1 = j1 + 1;
    end
end

datos = datosTest{2}; % Obtengo los SWS
for i = 1:nEx1,
    aux = datos{i}';
    for k = 1:T1,
        results1(j1,1:9) = aux(k,:);
        j1 = j1 + 1;
    end
end

```

Obtenidas los parámetros o características que permitirán el aprendizaje para el modelo del calificador automático, se procede a calcular la Entropía Espectral de Shannon.

```

function [H, H1] = Entropy

    close all;
    clear all;

    results, results1] = Parametros;      % Densidad Espectral de Potencia
    normalize = normc(results');         % Se normaliza las características
    resultados = normalize';

    [T x_dim] = size(results);

    H = zeros(T,1);

    for i = 1:T,
        e = wentropy(resultados(i,:), 'shannon');
        H(i) = e;
    end

    % Para el conjunto de Test

    normalizeTest = normc(results1');
    resultadosTest = normalizeTest';

    [T1 x_dim1] = size(results1);
    H1 = zeros(T1,1);

    for i = 1:T1,
        e = wentropy(resultadosTest(i,:), 'shannon');
        H1(i) = e;
    end

```

## 12.2. CÁLCULO DE LA DIMENSIÓN FRACTAL DE HIGUCHI.

```
function [HFD] = Higuchi_FD(serie, Kmax)
%{
Script for computing the Higuchi Fractal Dimension (HDF) of a signal.

INPUT:
  Serie: is the temporal series that one wants to analyze by HDF.
  It must be a row vector.
  Kmax: maximum number of sub-series composed from the original. To
  determine its values, we have followed the recommendation of Doyle et
  al at "Discriminating between elderly and young using a fractal
  dimension analysis of centre of pressure".

OUTPUT:
  HFD: the HFD of the temporal series.

PROJECT: Research Master in signal theory and bioengineering - University
of Valladolid

DATE: 02/03/2014

AUTHOR: Jesús Monge Álvarez
%}
%% Checking the input parameters:

control = ~isempty(serie);
assert(control, 'The user must introduce a series (first input).');
control = ~isempty(Kmax);
assert(control, 'The user must introduce the Kmax parameter (second
input).');

%% Processing:
% Composing of the sub-series:
N = length(serie);
X = NaN(Kmax, Kmax, N);
for k = 1:Kmax
    for m = 1:k
        limit = floor((N-m)/k);
        j = 1;
        for i = m:k:(m + (limit*k))
            X(k,m,j) = serie(i);
            j = j + 1;
        end
    end
end

% Computing the length of each sub-serie:
L = NaN(1, Kmax);
for k = 1:Kmax
    L_m = zeros(1, k);
    for m = 1:k
        R = (N - 1) / (floor((N - m) / k) * k);
```

```

        aux = squeeze(X(k,m,logical(~isnan(X(k,m,:))))); %We get the sub-
serie without the NaNs.
        for i = 1:(length(aux) - 1)
            L_m(m) = L_m(m) + abs(aux(i+1) - aux(i));
        end
        L_m(m) = (L_m(m) * R)/k;
    end
    L(k) = sum(L_m)/k;
end

% Finally, we compute the HFD:
x = 1./(1:Kmax);
aux = polyfit(log(x),log(L),1);
HFD = aux(1); %We only want the slope, not the independent term.

```

Mediante el algoritmo de Higuchi, se procede al cálculo de la DFH respecto a las características del presente trabajo.

```

[results] = Parametros;
[T x_dim] = size(results);
DFH = zeros(T,1);

for i = 1:T,
    dfh = Higuchi_FD(results(i,:),2); % sub-series de 2
    DFH(i) = dfh;
end

```

### 12.3. CÁLCULO DE LA DIMENSIÓN FRACTAL DE KATZ (DFK)

```

function [KFD] = Katz_FD(serie)
%{
Script for computing the Katz Fractal Dimension (KFD).

INPUT:
    Serie: is the temporal series that one wants to analyze by KFD.
    It must be a row vector.
OUTPUT:
    KFD: the KFD of the temporal series.

TIP: the KFD of a straight line must be exactly 1. Otherwise, the
implementation of the algorithm is wrong.

PROJECT: Research Master in signal theory and bioengineering - University
of Valladolid

DATE: 08/07/2014

```

```

AUTHOR: Jesús Monge Álvarez
%}
%% Checking the input parameters:

control = ~isempty(serie);
assert(control, 'The user must introduce a series (first input).');

%% Processing:
% Computing 'L':
% 'L' is the total length of the curve, that is to say, the sum of the
% distance between successive points. The distance between two points of the
% waveform is defined as the Ecludian distance: dist(s1,s2) = sqrt[(x1-
x2)^2 + (y1 - y2)^2];
% For this case: (x1 - x2) = 1 for all samples.
L = 0;
N = length(serie);
n = N - 1; %'n' is the number of steps in the waveform.
for i = 1:(N - 1)
    aux = sqrt(1 + ((serie(i) - serie(i+1))^2));
    L = L + aux;
    clear('aux');
end

% Computing 'd':
% 'd' is the planar extent or diameter of the waveform. It is estimated as
% the distance between the first point of the sequence and the point of the
% sequence that provides the farthest distance: d = max(dist(1,i));
i=2...N.
dist = NaN(1,N-1); %Predifinition variable for computatinoal efficiency.
for i = 2:N
    dist(i) = sqrt(((1 - i)^2) + ((serie(1) - serie(i))^2));
end
d = max(dist);

% Computing of KFD:
% The KFD is computed as follows: KFD = log10(n) / [log10(n) + log10(d/L)];
KFD = log10(n)/(log10(n) + log10(d/L));

```

Mediante el algoritmo de Higuchi, se procede al cálculo de la DFH respecto a las características del presente trabajo.

```

[results] = Parametros;
[T x_dim] = size(results);
DFK = zeros(T,1);

for i = 1:T,
    dfk = Katz_FD(results(i,:));
    DFK(i) = dfk;
end

```