



UNIVERSIDAD DE VALLADOLID



Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad de Valladolid

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA DE TECNOLOGÍAS DE
TELECOMUNICACIÓN

PROCESADO DE SEÑALES DE ACTIGRAFÍA: EXTRACCIÓN,
SELECCIÓN DE CARACTERÍSTICAS Y DISEÑO DE UN
SISTEMA EXPERTO PARA EL DIAGNÓSTICO DEL TRASTORNO
POR EL DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD

AUTOR: CHRISTIAN FORTICH JASSAN
TUTOR: CARLOS ALBEROLA LÓPEZ

14 de julio de 2016

TÍTULO: PROCESADO DE SEÑALES DE ACTIGRAFÍA: EXTRACCIÓN, SELECCIÓN DE CARACTERÍSTICAS Y DISEÑO DE UN SISTEMA EXPERTO PARA EL DIAGNÓSTICO DEL TRASTORNO POR EL DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD

AUTOR: CHRISTIAN FORTICH JASSAN

TUTOR: CARLOS ALBEROLA LÓPEZ

DEPARTAMENTO: TSCIT

Miembros del Tribunal

PRESIDENTE: CARLOS ALBEROLA LÓPEZ

SECRETARIO: FEDERICO SIMMROSS WATTENBERG

VOCAL: MARCOS MARTÍN FERNÁNDEZ

SUPLENTE 1: SANTIAGO AJA FERNÁNDEZ

SUPLENTE 2: RODRIGO DE LUIS GARCÍA

CALIFICACIÓN:

Resumen del proyecto

En el presente trabajo se desarrolla un sistema objetivo, no intrusivo y de bajo coste para la detección del Trastorno por el Déficit de Atención e Hiperactividad (TDAH) basado en actimetría. Mediante la actimetría se capturan una serie de patrones de actividad del individuo y, a partir de un conjunto de muestras de pacientes sanos y patológicos, se identifican las características más discriminantes, se entrenan los clasificadores y se cuantifican sus prestaciones.

Palabras clave

TDAH, actimetría, actigrafía, diagnóstico, características, trastorno, déficit, atención, hiperactividad

Keywords

ADHD, assesstment, actimetry, actigraphy, features, attention, deficit, hyperactivity, disorder

Abstract

The present work develops an objective, non-intrusive and inexpensive method for detection of Attention-Deficit/Hyperactivity Disorder (ADHD) based on actimetry. By means of actimetry, a number of patterns of the patient activity are captured and based on a set of samples from cases and controls, the most discriminant features will be identified, classifiers will be trained and their performances will be measured.

AGRADECIMIENTOS

En primer lugar, quisiera agradecer a mis padres y mi hermano, por la paciencia, la comprensión, el apoyo y todos esos: *“Hijo, ¿te falta mucho para terminar?”* o *“Descansa algo, que llevas mucho tiempo trabajando”*. Sin ellos, habría sido simplemente imposible haber llegado hasta donde me encuentro ahora y les debo todo lo que soy, lo he alcanzado y nunca podré expresar enteramente mi agradecimiento por esto. De corazón, *mis viejos*, ¡muchísimas gracias!

En segundo lugar, me gustaría agradecer a mi tutor, Carlos Alberola López, por toda la dedicación y la inestimable ayuda prestada a lo largo del desarrollo de este proyecto: revisiones, correcciones, sugerencias y explicaciones. Sin él, no habría sido lo mismo. En tercer lugar, agradecer a todos los compañeros del Laboratorio de Procesado de Imagen; valoro mucho la disposición de todos los miembros para ayudarme en temas tanto académicos como en cualquier otro aspecto sobre el que me surgieran dudas.

Al margen de lo académico, quisiera agradecer mis compañeros de clase, aquellos que siguieron mi trabajo en mayor o menor medida y con los que compartí clases, ratos libres, cañas y salidas; por el interés de saber a qué me dedicaba y el ánimo del que me revistieron para avanzar, muchas gracias.

Este trabajo es solo el último peldaño de la primera planta del conocimiento, en el que aspiro seguir subiendo cada vez más. Estoy ansioso por saber qué me deparará el próximo reto.

ÍNDICE GENERAL

1. Introducción	1
1.1. La patología	1
1.2. Uso de actimetría	2
1.3. Objetivos	3
1.4. Fases	3
1.5. Medios necesarios empleados para el desarrollo	4
2. Materiales y métodos	5
2.1. Materiales	5
2.2. Métodos	5
2.2.1. Adecuación de la información	6
2.2.2. Preprocesamiento de la información	6
2.2.3. Extracción de características	7
2.2.3.1. Parámetros estadísticos de la señal	7
2.2.3.2. Características no lineales de la señal	7
2.2.3.3. Parámetros del ritmo circadiano	11
2.2.3.4. Parámetros del sueño	11
2.2.3.5. Características espectrales de la señal	12
2.2.4. Desestacionalización	12
2.2.5. Selección de características y clasificación	13
2.3. Procedimiento de selección de características y de entrenamiento y test de los clasificadores	14
3. Resultados y discusión	16
3.1. Clasificadores de más de dos dimensiones	16
3.2. Desempeño del clasificador	17
3.2.1. Desempeño en términos de precisión, sensibilidad y especificidad	17
3.2.2. Desempeño del clasificador en términos de la función de distribución de la precisión	18
3.3. Mejores clasificadores bidimensionales	19
3.4. Utilidad del proceso de desestacionalización	20
3.5. Otras características reseñables	21
3.6. Uso de características espectrales	21
4. Conclusiones y líneas futuras	23
4.1. Desestacionalización	23
4.2. Características discriminantes	23
4.3. Clasificadores	24
4.4. Características espectrales	24
4.5. Seguimiento	25

A. Tablas de resultados	26
A.1. Desempeño de los clasificadores	26
A.1.1. Desempeño del clasificador kNN	26
A.1.2. Desempeño del clasificador SVM con kernel lineal	27
A.1.3. Desempeño del clasificador SVM con kernel cuadrático	27
A.2. Parejas de características que ofrecen mejor desempeño al clasificador	27
A.2.1. Clasificador kNN	28
A.2.2. Clasificador SVM lineal	28
A.2.3. Clasificador SVM cuadrático	29
A.3. Otras características reseñables	29
A.3.1. Dinámica simbólica	30
A.3.2. Características del ritmo circadiano	30
A.3.3. Estadísticos	31

ÍNDICE DE TABLAS

3.1. Breve comparativa de las prestaciones de los clasificadores 3D y 4D vs clasificador 2D	17
3.2. Desempeño del clasificador SVM con kernel RBF	18
3.3. Mejores 5 parejas de características para el clasificador SVM RBF	20
A.1. Desempeño del clasificador kNN	26
A.2. Desempeño del clasificador SVM con kernel lineal	27
A.3. Desempeño del clasificador SVM con kernel cuadrático	27
A.4. Mejores 5 parejas de características para el clasificador kNN	28
A.5. Mejores 5 parejas de características para el clasificador SVM lineal	28
A.6. Mejores 5 parejas de características para el clasificador SVM cuadrático	29
A.7. Mejores 5 parejas de características de las que al menos una es dinámica simbólica	30
A.8. Mejores 5 parejas de características de las que al menos una es de ritmo circadiano	30
A.9. Mejores 5 parejas de características de las que al menos una es un estadístico	31

Capítulo 1

INTRODUCCIÓN

En este capítulo se hace una conceptualización del problema. Se describe en qué consiste el trastorno, los métodos tradicionales de diagnóstico y las dificultades que plantea el mismo. Además, plantea una prometedora técnica que ofrece sobreponer esas dificultades, manteniendo las ventajas de los métodos tradicionales.

1.1 LA PATOLOGÍA

El trastorno por déficit de atención e hiperactividad (TDAH) es un trastorno disruptivo del comportamiento que se caracteriza principalmente por la presencia de un patrón persistente de inatención y/o hiperactividad e impulsividad [1]. Para que se considere que la patología esté presente, los síntomas de hiperactividad, impulsividad o falta de atención deben manifestarse desde antes de los siete años de edad, aunque no es estrictamente necesario, debido a que muchos individuos son diagnosticados después de mostrar los síntomas durante varios años después de esta edad. La sintomatología se ha de presentar por lo menos en la casa y el colegio (en el caso de los niños) o el trabajo (en el caso de los adultos) y debe afectar significativamente la operatividad del paciente, bien sea de forma social, académica o laboral. También debe tenerse en cuenta que tales síntomas no deben poder explicarse solo debido a una fase del trastorno pervasivo del comportamiento, la esquizofrenia, o algún otro desorden psicótico.

El déficit de atención se puede manifestar de muchas maneras: en el ámbito académico, ocupacional, o en situaciones sociales. Los individuos afectados por el trastorno tienen problemas para mantener la concentración o el orden en sus efectos personales, instrumentos de trabajo y demás, así como un descuido y una falta de atención para con sus interlocutores y una marcada desobediencia de las reglas en juegos y actividades. Tampoco suelen ser capaces de seguir las instrucciones dadas para llevar a cabo un trabajo o una tarea, y suelen cambiar de actividad con regularidad, saltando entre tareas incompletas o bien abandonando éstas debido a un estímulo irrelevante al que una persona sana no respondería. También es habitual que olviden cosas rutinarias, como llevar merienda u olvidar una cita concertada.

La hiperactividad se manifiesta por una fuerte intranquilidad, retorcimientos en el asiento

y la incapacidad de mantenerse sentado cuando se espera que lo esté. Los pacientes patológicos presentan excesiva actividad en situaciones inapropiadas, tienen dificultad para realizar actividades que requieren que esté quieto y están, literalmente, *en marcha* o como *impulsados por un motor*. Conceptualmente, depende en gran medida de la edad del paciente y su nivel de desarrollo, de manera que el diagnóstico de la hiperactividad debe hacerse con mucho cuidado en este sentido.

La impulsividad se manifiesta bajo la forma de impaciencia, dificultad en demorar las respuestas o responder preguntas no formuladas en su totalidad. Está caracterizada por interrupciones frecuentes, conversaciones fuera de lugar, incapacidad de esperar el turno o tocar cosas que no deben, comportamientos que pueden causar dificultades en el ámbito laboral, personal o académico.

En lo sucesivo, la impulsividad y la hiperactividad se considerarán, a efectos prácticos, agrupadas bajo el concepto de hiperactividad.

1.2 USO DE ACTIMETRÍA

La actimetría se define como el proceso de cuantificación del movimiento en los tres ejes cartesianos, facilitando la obtención de mediciones objetivas del mismo a lo largo de periodos de tiempo extendidos [2]. Esta técnica lleva dos décadas siendo una de las herramientas valorativas más usadas en investigación y medicina del sueño, lo cual se puede constatar con el aumento de las publicaciones basadas en actimetría en detrimento de otros métodos clásicos -como la polisomnografía (PSG)- pasando de una publicación de actimetría por cada diez de PSG a una publicación de actimetría por cada cuatro de PSG en menos de veinte años [3]. Como ejemplos del uso de la actimetría se pueden mencionar los trabajos hechos en el Laboratorio de Procesado de Imagen¹ de la Universidad de Valladolid, ámbito en el cual se desarrolla el presente trabajo. Entre las contribuciones hechas en este ámbito se pueden mencionar [4, 5, 6, 7]. Esto demuestra que la actimetría tiene un potencial para el estudio de ciertas patologías o condiciones médicas, asociadas sobre todo a los niveles de actividad.

No obstante, las contribuciones citadas anteriormente hacen poco hincapié en aspectos como la homogeneidad y el tamaño de la muestra, el análisis espectral, la desestacionalización de los datos o la dimensionalidad del clasificador, y que en el presente trabajo se considerarán con más nivel de detalle.

El uso de la actimetría se plantea como una alternativa económica y no intrusiva para la detección del TDAH, puesto que no representa ninguna intromisión a la actividad diaria del paciente. El test clínico de diagnóstico del TDAH involucra entrevistas periódicas durante meses con diversos profesionales de la salud, depende de la subjetividad de las respuestas que los tutores del paciente proporcionen acerca de su comportamiento, y de las respuestas que ellos mismos proporcionen acerca de su actividad diaria [8].

La actimetría, en cambio, aborda el diagnóstico desde un punto de vista alternativo, encargándose del registro de la actividad del paciente para ofrecer un caudal de información que describa

¹<http://ipi.tel.uva.es>

de forma objetiva su comportamiento de forma totalmente desatendida. También presenta la ventaja de que tras coleccionar los registros actigráficos del paciente, el diagnóstico puede dictaminarse con mayor rapidez, tardando solo el tiempo en que el sistema computacional procesa la información. Esto es de especial relevancia en el ámbito de la detección precoz, pues es de capital importancia que el tratamiento de un paciente patológico empiece cuanto antes, a fin de mejorar su calidad de vida y de las personas de su alrededor.

1.3 OBJETIVOS

En este trabajo se pretende desarrollar un sistema experto de detección del TDAH en un paciente en función de su registro actigráfico, obtenido a lo largo de un día. El sistema de detección debe ser rápido, fiable y no intrusivo para el paciente a la vez que económico, pues aunque la prevalencia del trastorno hace que sea el desorden neuroconductual más frecuente en la población infantil [8], el hecho de que se tenga poco conocimiento sobre los aspectos epidemiológicos que rodean el trastorno y su etiología, lo convierten en un problema de salud pública [9].

En la consecución de este objetivo, los estudios presentados en [5] sirven como punto de partida para el presente trabajo, dando como válida la hipótesis de que el resultado del diagnóstico es dependiente de la heterogeneidad en los datos, razón por la cual se estudiará una muestra homogénea compuesta por los registros actigráficos de los pacientes cuyas edades están se encuentren entre 6-8 años. En consecuencia, y en aras de la completitud, en este trabajo se persigue (a) diseñar un sistema experto de detección del TDAH a partir de una muestra homogénea de pacientes cuyas edades estén entre 6 y 8 años inclusive, empleando parámetros derivados del análisis no lineal de las señales de actimetría como características (b) analizar si la desestacionalización de las características mejora sus prestaciones (c) estudiar nuevas características derivadas del análisis espectral de las señales (d) discutir el número óptimo de dimensión sobre la cual construir los clasificadores.

1.4 FASES

El objetivo planteado en la sección anterior describe el problema que se desea resolver. En la consecución de dicho objetivo se seguirá la siguiente línea de eventos:

- La información actigráfica de la que se dispone se estructurará de la manera más conveniente para realizar un procesamiento eficiente, una vez se haya elegido la solución computacional de procesamiento.
- Se analizará la información: eliminación de datos redundantes, erróneos y purga de valores de saturación, *outliers* (valores extremos), *NaNs* (resultados de operaciones inválidas sobre los datos) y demás valores que no aporten información al procesamiento.
- Se fragmentarán los registros de información y se extraerán las características más relevantes de los mismos.

- Se construirán clasificadores basados en las características extraídas.
- Se estudiará el desempeño de los clasificadores (en términos de precisión, sensibilidad y especificidad).

1.5 MEDIOS NECESARIOS EMPLEADOS PARA EL DESARROLLO

Para la realización del presente trabajo, se hará uso de los siguientes medios, todos ellos provistos por el Laboratorio de Procesado de Imagen:

1. Registros actigráficos previamente capturados e identificados.
2. Ordenador de sobremesa con distribución Linux (procesador Intel Core i7-4790 @ 3.60 GHz *quad-core*, 16GB de memoria RAM, 6TB de almacenamiento en red sobre RAID).
3. Servidor en red para acceso remoto y cálculo.
4. Conexión a internet.
5. Software de cálculo numérico: MATLAB R2015a, con *toolbox* de cálculo estadístico.

MATERIALES Y MÉTODOS

En este capítulo se detallarán los materiales y los métodos necesarios para el procesamiento. Se detallará lo que se requiere para procesar y el procedimiento mediante el cual se va a efectuar el desarrollo en sí, planteando los pasos intermedios requeridos para la consecución del sistema. A este respecto, se hablará de:

- La manipulación preliminar de la información en bruto.
- La extracción de *features* (se detallarán los usados).
- La desestacionalización.
- El proceso de selección de *features*, clasificación y decisión.

2.1 MATERIALES

Se dispone de información actigráfica de 74 pacientes que padecen TDAH según el test DSM-IV [1] (quienes en lo sucesivo serán designados como casos) y 75 pacientes que, en principio, no se tiene constancia de que padezcan TDAH ni ningún trastorno mental, neurológico o de sueño (quienes en lo sucesivo serán designados como controles). Dicha información fue facilitada por pediatras y psicólogos del Complejo Asistencial Universitario y del Centro de Salud Jardinillos, ambos ubicados en la ciudad de Palencia, España. No obstante, y siendo consecuentes con las asunciones de partida, —detalladas en la sección 1.3— el hecho de considerar una muestra homogénea en términos de edad obliga a incluir en el procesamiento sólo a aquellos pacientes cuyas edades estén en el rango de 6 a 8 años—inclusive—, con lo cual la muestra se verá reducida hasta 39 casos y 42 controles.

2.2 MÉTODOS

La señal que procede del actímetro es procesada en varias etapas, según lo descrito en [5]. En concreto:

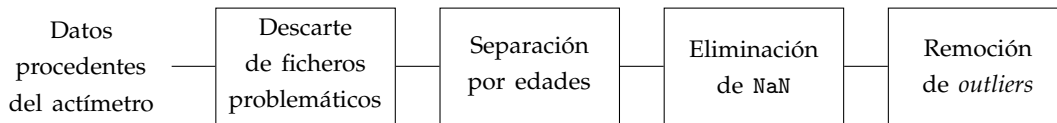


FIGURA 2.1: Etapas de la adecuación de los registros de información, previa al preprocesado

2.2.1 ADECUACIÓN DE LA INFORMACIÓN

La información procedente del actímetro como datos en bruto posee información redundante, descartable o no válida de acuerdo con los criterios elegidos para el procesamiento de la información, (como por ejemplo el criterio de edad antes mencionado), razón por la cual se hace necesaria una etapa de adecuación de las señales que permitan disponerlas para que el procesado sea más eficaz. Para ello:

- Las señales se purgan de valores que saturan el conversor A/D del dispositivo captador (o se descartan si en su mayoría son valores de saturación, pues ficheros con tales valores serán problemáticos en etapas posteriores).
- Se separan los casos y controles según el criterio de edades antes elegido.
- Se eliminan registros que contengan valores no válidos (NaN).
- Se eliminan *outliers* (valores extremos).
- Se eliminan registros que, aún no cumpliendo ninguna de las características anteriores, capturan actividad del paciente en un intervalo de duración inferior a 22 horas.

Todos estos pasos tienen como objetivo que las siguientes etapas del procesado reciban únicamente información útil, tratando de afectar lo menos posible al carácter discriminante que se presume poseen los datos. En la figura 2.1 se muestra la secuencia de pasos que se siguen desde que se reciben los datos hasta que están listos para la etapa de preprocesado.

2.2.2 PREPROCESAMIENTO DE LA INFORMACIÓN

La etapa de preprocesado de la información, tras el proceso de adecuación, se divide en tres partes: en una primera instancia se realiza la detección automática de los intervalos de actividad y sueño según el método presentado en [4]. La elección de este método respecto a otros métodos clásicos —como los presentados en [10, 11]— presenta ventajas asociadas a la inicialización. El primer algoritmo calcula automáticamente el inicio y el final del periodo de sueño mientras que los últimos dos requieren de dichos parámetros como entrada al mismo y que otros parámetros de entrada al mismo se calculen de manera empírica, en función del dispositivo de captura. Al detectar los intervalos de sueño, se puede fragmentar el registro completo de actividad en subregistros que corresponden a intervalos del día.

En la segunda etapa, y haciendo uso de los intervalos de actividad y descanso obtenidos en la etapa anterior, se fragmenta la información en diferentes tramos del día. Concretamente, se adquieren subregistros correspondientes al día completo, vigilia, vigilia solo matutina, vigilia

solo vespertina, intervalo completo de sueño, tramo inicial del sueño, tramo intermedio del sueño y tramo final del sueño. Con esta división, se efectúa un análisis más detallado de los niveles de actividad del paciente a lo largo del día, y se consiguen extraer características discriminantes que no se encontrarían únicamente en el registro completo, como se recoge en [4].

Posteriormente, se diezman todos los registros obtenidos en la etapa anterior. Esto es importante; permite mejorar la calidad de la información relacionada a los periodos de actividad, enmascarada por la gran cantidad de muestras de valor nulo que proceden de la señal original. Es inmediato advertir que la elección de la tasa de muestreo no es trivial; las tasas de diezmo elegidas deben permitir aislar actividades concretas (periodos cortos de tiempo, entre 1s y 1 min) y entre tareas (entre 5 y 30 min) manteniendo un compromiso entre cantidad de información disponible y reducción de muestras redundantes. Una vez decididas las tasas de muestreo, se diezman todos los fragmentos con cada una de dichas tasas.

2.2.3 EXTRACCIÓN DE CARACTERÍSTICAS

El proceso de extracción de las características (conocidas en el argot, y en adelante, como *features*) tiene lugar tras la fragmentación de los registros. La extracción se hace para cada uno de los fragmentos del día y para cada uno de los diezmos en aquellos registros cuya longitud se haya determinado adecuada. Se obtienen pues, características temporales de la señal tales como:

2.2.3.1 PARÁMETROS ESTADÍSTICOS DE LA SEÑAL

Parámetros estadísticos clásicos de las señales, que se pueden encontrar en cualquier referencia de estadística, como [12]: la media, la desviación típica, la mediana, el rango intercuartil, la asimetría estadística (*skewness*), la curtosis (medida de la forma) y el coeficiente de variación, definido como el cociente entre la desviación típica y la media.

2.2.3.2 CARACTERÍSTICAS NO LINEALES DE LA SEÑAL

Características no lineales de la señal, entre las que se pueden mencionar la medida de tendencia central, la dinámica simbólica, la complejidad binaria de Lempel Ziv, y la entropía aproximada:

- La medida de tendencia central [13] (CTM, por sus siglas en inglés) es una magnitud no lineal basada en la representación gráfica de las diferencias de primer orden de tres muestras consecutivas. Para una secuencia discreta $x[n]$ de longitud N , el eje de abscisas es $x[n+2] - x[n-1]$ y el de ordenadas es $x[n+1] - x[n]$. La CTM se define como el número de muestras que caen dentro de un disco de radio ρ centrado en el origen de coordenadas. De acuerdo con esto

$$\text{CTM} = \frac{1}{N-2} \sum_{k=1}^{N-2} \delta_k \quad (2.1)$$

siendo

$$\delta_k = \begin{cases} 1 & \sqrt{\frac{(x[k+2] - x[k+1])^2 + (x[k+1] - x[k])^2}{2}} \leq \rho \\ 0 & \text{resto} \end{cases} \quad (2.2)$$

El cálculo de la CTM solo requiere que se fije el parámetro $\rho \geq 0$ para definirla completamente. Siendo un radio, su elección establece un compromiso entre qué tan grande debe ser como para excluir el ruido inherente a las señales, y qué tan pequeño para captar señales de pequeña amplitud. De la definición arriba expresada se puede intuir que las señales regulares tendrán un valor alto de CTM, mientras que las señales que presenten una alta variabilidad tendrán menores valores de CTM.

- La dinámica simbólica [14, 15] (SD, por sus siglas en inglés) de una señal $x[n]$ es un parámetro que se basa en el cálculo de las frecuencias relativas de aparición de varias palabras en $c[n]$, una versión discretizada de la señal $x[n]$. El cálculo de la SD requiere como entrada al algoritmo un parámetro de longitud de alfabeto y otro de símbolos por palabra. En el presente trabajo se ha elegido un alfabeto de tres símbolos ($N_A = 3$, $\mathcal{A} = \{0, 1, 2\}$), los cuales representarán los niveles de actividad bajo, medio y alto, respectivamente. En cuanto la longitud del alfabeto, se emplearán palabras de dos símbolos de longitud, que representan la transición entre los 3 niveles antes mencionados. Fijando estos dos parámetros, la SD queda completamente definida.

El primer paso para el cálculo de la SD es construir la señal discretizada. Debido a que se usan 3 niveles de actividad, se definen dos umbrales q_1, q_2 que se corresponden con los percentiles 33 y 66 de $x[n]$ respectivamente. De esta manera, se puede expresar la señal discretizada $c[n]$ de la siguiente manera:

$$c[n] = \begin{cases} 0 & \text{si } x[n] < q_1 \\ 1 & \text{si } q_1 \leq x[n] < q_2 \\ 2 & \text{si } x[n] \geq q_2 \end{cases} \quad (2.3)$$

La caracterización de la cadena de símbolos se hace teniendo en cuenta la distribución de probabilidades de $\omega_\ell[k]$, que se puede extraer de las muestras de $c[n]$ como sigue:

$$\omega_\ell[k] = \begin{bmatrix} c[k(\ell - \tau) + \tau] \\ c[k(\ell - \tau) + \tau - 1] \\ \vdots \\ c[k(\ell - \tau) + \tau - (\ell - 1)] \end{bmatrix} \quad (2.4)$$

siendo τ el factor de solapamiento entre palabras, definido como $\tau = \ell - 1$. Con los parámetros dados para la SD ($\ell = 2$, $N_A = 3$), entonces la ecuación 2.4 particulariza a

$$\omega_2[k] = \begin{bmatrix} c[k] \\ c[k+1] \end{bmatrix} \quad (2.5)$$

Hechas estas particularizaciones, el alfabeto se compone de $3^2 = 9$ palabras, a saber:

$$\mathcal{W} \equiv \{00, 01, 02, 10, 11, 12, 20, 21, 22\}$$

Las decisiones que se tomen basadas en este parámetro estarán asociadas a la probabilidad de aparición de cada palabra, $p_i = P(\omega_2[k] = \omega_{2,i})$, con $\omega_{2,i} \in \mathcal{W}$, $i = 1, 2, \dots, 9$

La entropía de Shannon [16], definida como:

$$E_{\text{Sh}} = \sum_{i=1}^{(N_A)^l} p_i \log_2(p_i) \quad (2.6)$$

Donde p_i sigue siendo la probabilidad de la palabra i -ésima, es parámetro de agregación de las mismas. La entropía de Shannon indica la variabilidad de una señal: su valor es tanto más alto cuanto más alta es la variabilidad de la señal.

- La complejidad binaria de Lempel-Ziv [17] (acortado como LZC Bin) es una medida de la aleatoriedad de una secuencia, basada en la umbralización de la misma. Como mencionan Lempel y Ziv, una clasificación binaria es suficiente. De esta manera, solo se requiere fijar un único umbral γ . Definido el umbral, se genera la secuencia $s[n]$ de la siguiente manera

$$s[n] = \begin{cases} 0 & \text{si } x[n] \leq \gamma \\ 1 & \text{si } x[n] \geq \gamma \end{cases} \quad \forall n \in [0, N-1] \quad (2.7)$$

La secuencia resultante, P , será entonces una secuencia binaria. Para la inicialización del algoritmo, se definen las subsecuencias S , $Q \ni P$, su concatenación, SQ , la secuencia resultante de eliminar el último carácter, $SQ\pi$ y el conjunto de todas las posibles subsecuencias de $SQ\pi$ (o vocabulario), $v(SQ\pi)$. Además, se inicializa el contador de complejidad $c = 1$, $S = s[1]$, $Q = s[2]$ y $SQ\pi = s[1]$. Bajo estos supuestos, la LZC Bin se puede calcular mediante los siguientes pasos:

1. Para una iteración r , se verifica que la subsecuencia $Q = s[r+1]$ se encuentra en la subsecuencia $S = \{s[1], s[2] \dots s[r]\}$, es decir, si $Q \in v(SQ\pi)$.
2. Se actualiza la subsecuencia Q , que ahora tomará el valor $Q = \{s[r+1], s[r+2]\}$ y se repite el paso anterior hasta que se encuentre un índice i tal que la subsecuencia $Q = \{s[r+1], s[r+2] \dots s[r+i-1]\} / Q \notin v(SQ\pi)$, lo cual provoca que se aumente el contador de complejidad c en una unidad.
3. Las subsecuencias S , Q se actualizan a los valores $S = \{s[1], s[2] \dots s[r+i-1]\}$ y $Q = s[r+i+1]$

4. Al llegar a $r = N$, el contador c se normaliza por el valor $\frac{N}{\log_\alpha N}$ con el propósito de que el valor de complejidad sea independiente de la longitud de la secuencia temporal $x[n]$
- La entropía aproximada (ApEn) [18] es una medida logarítmica de la regularidad de una secuencia discreta, basándose en las probabilidades de aparición de determinados patrones.

Dada una secuencia discreta $x[n]$, $n = 1, 2, \dots, N$, la caracterización de la ApEn toma 2 parámetros: uno de tamaño de subsecuencia, m , uno de distancia máxima, r . El nombre y definición formal de estos parámetros proceden de la teoría del caos y en la literatura al respecto [19] se les conoce como *embedded dimension* o dimensión incrustada y distancia de filtrado, respectivamente. Fijada esta caracterización, el cálculo de la ApEn se lleva a cabo como sigue:

1. Partiendo de la secuencia discreta $x[n]$, se calculan las secuencias $u^{(m)}[1], \dots, u^{(m)}[N - 1 + m]$, que se definen como $u^{(m)}[i] = [x[i], x[i + 1], \dots, x[i + m - 1]]$, con i comprendido entre 1 y $N - 1 + m$. Esta definición permite generar secuencias de m muestras contiguas de la secuencia x , que comienzan en el elemento i -ésimo de la misma.
2. Definiendo la distancia entre dos secuencias $u^{(m)}[i]$, $u^{(m)}[j]$ como

$$d\left(u^{(m)}[i], u^{(m)}[j]\right) = \max_k \left(\left| u^{(m)}[i; k] - u^{(m)}[j; k] \right| \right) \quad (2.8)$$

donde $u^{(m)}[i; k]$ representa la componente k -ésima de la subsecuencia $u^{(m)}[i]$, y k es un valor comprendido entre 0 y $m - 1$, se calcula la probabilidad de que las dos subsecuencias estén separadas a una distancia a lo sumo r como la frecuencia relativa de los índices $j \in [1, N - 1 + m]$ que satisfagan que $d\left(u^{(m)}[i] - u^{(m)}[j]\right) \leq r$ normalizada por el número de los mismos, es decir, por un factor $\frac{1}{N - 1 + m}$. Dicha probabilidad se designa como $C_i^{(m)}(r)$.

3. Con la definición de $C_i^{(m)}(r)$ vista en el literal anterior, se construye la función $\Phi^{(m)}(r)$ como el promedio de la esperanza del logaritmo de dicha probabilidad. Entonces, la función Φ se expresa de la siguiente forma:

$$\Phi^{(m)}(r) = \frac{1}{N - 1 + m} \sum_{i=1}^{N-1+m} \log C_i^{(m)}(r) \quad (2.9)$$

Con todas estas definiciones, la ApEn se puede expresar como:

$$\text{ApEn}(m, r) = \lim_{N \rightarrow \infty} \left[\Phi^{(m)}(r) - \Phi^{(m+1)}(r) \right] \quad (2.10)$$

Basada en la definición asintótica que se recoge en [18] y que, como se indica, es de poca utilidad práctica. Lo que es más, debido a que no se dispone de un número infinito de muestras, la definición dada en 2.10 se hace depender del número de muestras de la secuencia discreta mediante la expresión

$$\text{ApEn}(m, r, N) = \Phi^{(m)}(r) - \Phi^{(m+1)}(r) \quad (2.11)$$

Donde la dependencia de ApEn con N está implícita en la función Φ .

2.2.3.3 PARÁMETROS DEL RITMO CIRCADIANO

Los parámetros del ritmo circadiano proceden de un ajuste de mínimos cuadrados de una secuencia discreta en el dominio del tiempo a una senoide [20]. El ajuste sinusoidal implica la estimación de tres parámetros: la amplitud de la senoide, la fase de la senoide (conocida como acrofase) y el mesor, definido como el valor medio de un ciclo. Estos parámetros del ritmo circadiano se calculan tomando como entradas una secuencia discreta, los instantes de tiempo correspondientes a dicha secuencia y la longitud del ciclo. En este trabajo se considera que la actividad es periódica con periodo de 24 horas, de manera que la frecuencia de la senoide se asume conocida.

2.2.3.4 PARÁMETROS DEL SUEÑO

Parámetros de puntuación del sueño (SSA, por sus siglas en inglés) extraídos mediante el método de Sadeh-Acebo [21, 22], que consiste en la estimación empírica de la probabilidad de que tenga lugar una etapa de sueño silencioso (PQS), la probabilidad de una etapa de sueño activo (PAS) o de una etapa de vigilia (PAW). La estimación empírica se hace a lo largo de una ventana de tiempo (que se corresponde con la etapa a la que hacen referencia las probabilidades) teniendo en cuenta las siguientes entradas:

- El número de minutos sin actividad en dicha ventana (nzw)
- El número de minutos con baja y alta actividad (ntl , nth)
- La desviación típica de la ventana sumada a los cinco minutos anteriores a la ventana considerada ($s5$)
- El nivel de actividad medio en el minuto medido y el anterior a éste ($m1$)
- El nivel más bajo de actividad medido durante la ventana formada por el minuto medido y los cuatro anteriores a éste ($lw4$)

El método de Sadeh-Acebo estima los parámetros mediante la resolución del sistema

$$\begin{pmatrix} \text{PQS} \\ \text{PAS} \\ \text{PAW} \end{pmatrix} = \underbrace{\begin{pmatrix} 15,94 & 3,223 & 2,138 & 1,1036 & 0,466 & 0,00292 & 0,0106 \\ 5,134 & 1,696 & 2,062 & 0,9568 & 0,0586 & 0,00556 & 0,0105 \\ -25,638 & 1,714 & 3,0168 & 4,064 & 0,1066 & 0,386 & 0,016 \end{pmatrix}}_M \begin{pmatrix} 1 \\ \text{nzw} \\ \text{ntl} \\ \text{nth} \\ \text{s5} \\ \text{ml} \\ \text{lw4} \end{pmatrix} \quad (2.12)$$

donde la matriz $M \in \mathcal{M}_{3 \times 6}$ está compuesta por observaciones empíricas.

Los elementos $M_{1,5}$ y $M_{3,5}$ se escalan por 10 en la matriz para mejorar las prestaciones de la estimación, puesto que al ser parámetros empíricos, dependen del actígrafo empleado y el autor de la referencia emplea un modelo distinto al usado para capturar las señales que se procesan en el presente trabajo.

2.2.3.5 CARACTERÍSTICAS ESPECTRALES DE LA SEÑAL

Como característica procedente del análisis espectral se emplea la dimensión fractal de Maragos-Sun, que es una técnica consistente en la medición de la fragmentación (o la presencia de irregularidades) en un entorno de la frontera de un fractal, visto a distintas escalas. Aplicado a una señal unidimensional fractal y variante en el tiempo, la dimensión fractal de Maragos-Sun [23] se calcula llevando a cabo los siguientes pasos:

1. Dilatar y contraer la secuencia fractal $x[n]$, a lo largo de los distintos valores de escala provistos.
2. Calcular el área que cubren dichas escalas, definida como la diferencia entre las dilataciones y las contracciones de la señal.
3. Hacer un ajuste polinómico de grado 1 (en el sentido de los mínimos cuadrados) de una nube de puntos calculada a partir de la señal fractal n -dimensional cuya dimensión fractal se quiere determinar. Las abscisas de la nube vienen dadas por el logaritmo del inverso de las distintas escalas, normalizadas por $\frac{2}{N}$, siendo N la longitud de la secuencia, mientras que la ordenada i -ésima se calcula como el logaritmo del área de la escala i -ésima, normalizada por su abscisa elevada al cuadrado.
4. La dimensión fractal de Maragos-Sun es, entonces, la pendiente del ajuste polinómico lineal.

Nótese que la dimensión fractal no tiene por qué ser un número entero.

2.2.4 DESESTACIONALIZACIÓN

En [24] se establece una relación entre el nivel de actividad de los pacientes y la fecha del año en que se mida. En un intento por eliminar la dependencia del diagnóstico con el momento

del año en que éste tenga lugar, se lleva a cabo un proceso de *de-trending* o desestacionalización que remueva efectos periódicos de actividad no asociados directamente con el trastorno; en concreto, se elimina de los *features* la influencia de la fecha del año en el nivel de actividad, asumiendo que la media y la desviación típica del nivel de actividad siguen, de forma fisiológica, un ritmo aproximadamente circanual. De esta manera, se lleva a cabo un ajuste lineal de dichos parámetros, obteniendo un valor medio y una desviación típica para cada mes del año del que se tenga al menos un caso y un control. A continuación, a los *features* de cada caso y de cada control que correspondan al mismo mes se les resta la media y se dividen por la desviación típica, generando efectivamente un conjunto de datos de igual tamaño que el conjunto de datos de entrada, pero que no contiene la influencia estacional antes mencionada.

Se define el vector $f_{r,m}$, con $1 \leq r \leq 12$ y $1 \leq m \leq K$ como el vector que recoge todas las muestras del *feature* m -ésimo, estimado a partir de los datos de casos y controles adquiridos durante el mes r -ésimo. De esta manera, hay 12 vectores de muestras para cada *feature*, para los cuales se ha estimado tanto la media muestral como la desviación típica muestral:

$$\hat{\eta}_{r,m} = \frac{1}{N_{r,m}} \mathbf{1}_{N_{r,m}}^T \mathbf{f}_{r,m} \quad (2.13)$$

$$\hat{\sigma}_{r,m} = \sqrt{\frac{1}{N_{r,m}} \mathbf{f}_{r,m}^T \mathbf{f}_{r,m} - \hat{\eta}_{r,m}^2} \quad (2.14)$$

siendo $N_{r,m}$ el número de componentes del vector $f_{r,m}$, las cuales son modeladas mediante las siguientes curvas cosenoidales:

$$\eta_{r,m} = A_{\eta_m} + B_{\eta_m} \cdot \cos\left(\frac{2\pi}{12}r + \Theta_{\eta_m}\right), \quad (2.15)$$

$$\sigma_{r,m} = A_{\sigma_m} + B_{\sigma_m} \cdot \cos\left(\frac{2\pi}{12}r + \Theta_{\sigma_m}\right), \quad (2.16)$$

donde A_{η_m} , B_{η_m} , Θ_{η_m} , A_{σ_m} , B_{σ_m} y Θ_{σ_m} son elegidas de manera que se ajusten a las secuencias $\hat{\eta}_{r,m}$ y $\hat{\sigma}_{r,m}$ ($1 \leq r \leq 12$), respectivamente, en el sentido de los mínimos cuadrados. Por último, las características desestacionalizadas, $\tilde{f}_{r,m}$, se obtienen como sigue:

$$\tilde{f}_{r,m} = \frac{f_{r,m} - \eta_{r,m}}{\sigma_{r,m}}. \quad (2.17)$$

donde las operaciones se deben interpretar como operaciones componente a componente.

2.2.5 SELECCIÓN DE CARACTERÍSTICAS Y CLASIFICACIÓN

El análisis estadístico de la separabilidad (el poder discriminante) de los *features* extraídos en los pasos anteriores se lleva a cabo mediante la prueba del *p-valor* de Wilcoxon (es una práctica habitual considerar que un *feature* es discriminante si su *p-valor* es inferior a 0.05, como se discute

en [12] y en otras obras clásicas de referencia del análisis estadístico). Esta prueba es equivalente a la prueba de separabilidad de medianas de las dos poblaciones que forman parte del estudio que ocupa el presente trabajo (casos y controles).

En cuanto a la clasificación, se han empleado los siguientes tipos de clasificadores¹:

- *k-nearest-neighbors* (kNN) con $k = 3$ vecinos.
- *Support vector machine*, empleando los *kernels* cuadrático (quad), *radial basis function* (rbf) y lineal (lin).

En este trabajo emplearemos los clasificadores como meras “cajas negras,” de forma que no se abordarán sus detalles de funcionamiento. En la literatura al respecto existe abundante información (véase, por ejemplo, [25])

Por último, se contrasta la distribución teórica del trastorno, esbozada por el análisis de casos y controles de que se dispone, con el registro actigráfico del paciente que se desea diagnosticar. Se debe emitir un diagnóstico binario: paciente sano o patológico.

2.3 PROCEDIMIENTO DE SELECCIÓN DE CARACTERÍSTICAS Y DE ENTRENAMIENTO Y TEST DE LOS CLASIFICADORES

Para cumplir los objetivos planteados en la sección 1.3 se lleva a cabo una serie de pasos de implementación, que se describen a continuación:

SELECCIÓN DE CARACTERÍSTICAS QUE SE INTRODUCIRÁN AL CLASIFICADOR

El conjunto de selección hará uso de todo el espacio de *features* calculados en el capítulo anterior. Como parte de este proceso de selección, la significatividad de cada uno de los *features* se evaluará a lo largo de veinte iteraciones, mediante la prueba del *p-valor*. En cada una de estas iteraciones (o *folds*), el *feature* en cuestión se extraerá de un subconjunto aleatorio de casos y controles. Como consecuencia del tamaño muestral reducido y la enorme cantidad de *features* será necesario hacer múltiples comparaciones, lo cual hará más probable la aparición de *features* que se evalúen como significativos, razón por la cual se emplea una técnica de corrección por comparaciones múltiples, que consiste en forzar a que los *features* que se evalúen como significativos tengan que serlo en todos y cada uno de los *folds*, lo cual reduce de manera efectiva el *p-valor*. Esto es equivalente a aplicar una condición muy restrictiva sobre el *p-valor* como la superposición de varias condiciones progresivamente más restrictivas.

ELECCIÓN DE LOS ÍNDICES DE FEATURES SIGNIFICATIVOS, ENTRENAMIENTO Y TESTEO DE LOS CLASIFICADORES

Tras determinar los mejores *features*, se procede a la construcción de los clasificadores bidimensionales. Para esto, se generan las $\binom{N}{2}$ parejas a partir de los N índices significativos

¹Además de los clasificadores que se mencionan, se usó un clasificador en árbol, pero sus prestaciones son sensiblemente inferiores a las de los demás clasificadores, de manera que dicho clasificador no se vuelve a mencionar en el resto del trabajo.

(con N en torno a 1200) calculados en el epígrafe anterior. A continuación, y siguiendo una estrategia de aleatorización similar a la de antes, para cada pareja de índices se generan veinte permutaciones aleatorias del conjunto de todos los casos y controles, las cuales se subdividirán en dos particiones: una partición se encargará de entrenar a cada clasificador, lo cual generará modelos de clasificación. Dichos modelos son empleados, en conjunción con la partición restante, para generar las predicciones de clasificación. De forma análoga al epígrafe anterior, aleatorizar y haber usado varios *fold*s provoca un efecto de aumento de tamaño muestral efectivo, pues se hace más improbable que un mismo subconjunto sea escogido para entrenar y testear los clasificadores lo cual sobreestimaría las características de los mismos, o que se emplee un subconjunto poco representativo que los subestime.

EVALUACIÓN DEL DESEMPEÑO DE LOS CLASIFICADORES

Tras generar las predicciones en el apartado anterior, se evalúa de manera cuantitativa el desempeño de los clasificadores, lo cual es posible dado que los datos están etiquetados, como se mencionó en la sección 1.4. Esta evaluación se hace mediante una comparación entre las decisiones tomadas por los clasificadores y la realidad. El desempeño de cada clasificador se evaluará en términos de la precisión (fracción de casos y controles correctamente clasificados), sensibilidad (fracción de casos correctamente clasificados como tal) y especificidad (cantidad de controles debidamente clasificados como tal), y si además se introduce la consideración de evaluar los parámetros de desempeño de los clasificadores en términos de su valor mediano, con lo cual se favorece que los resultados se vean menos afectados por los valores extremos.

RESULTADOS Y DISCUSIÓN

En esta sección se discutirán los resultados de las pruebas realizadas, descritas en el capítulo 2. Las discusiones se centrarán en los siguientes aspectos:

- dimensión del clasificador.
- el desempeño de los clasificadores, tanto en términos de precisión, sensibilidad y especificidad como en términos de su distancia al clasificador ideal.
- la utilidad del proceso de desestacionalización.
- los mejores clasificadores bidimensionales.
- otros *features* reseñables.
- el uso de los *features* espectrales.

3.1 CLASIFICADORES DE MÁS DE DOS DIMENSIONES

El clasificador es una función del vector de datos \bar{x} , que mapea los mismos a un espacio discreto de dos valores, 0 y 1, asociados en el contexto de este trabajo a la ausencia o presencia de la patología, respectivamente. Dicha función se puede expresar como $f(\bar{x}) : \mathbb{R}^n \rightarrow \{0, 1\}$ y como se mencionó en la sección 2.2.5, uno de los objetivos del presente trabajo es determinar la dimensión óptima del clasificador, n . En la sección 3.2 se indica que la dimensión del clasificador debe estar a lo sumo un orden de magnitud por debajo del tamaño muestral, de modo que, restringiendo al espacio muestral del que se dispone (39 casos y 42 controles), teóricamente se puede considerar la implementación de clasificadores de hasta 4 dimensiones. Sin embargo, un análisis como el que se muestra en la tabla 3.1 revela que el desempeño de los clasificadores de más de dos dimensiones no mejora sustancialmente con respecto a los clasificadores bidimensionales; ésta es una conclusión a la que hemos llegado mediante un experimento en el que mantenemos aproximadamente constante el tiempo de análisis del clasificador, es decir, probamos un número n -tuplas de entrada aproximadamente constante. Para esta prueba se emplea el clasificador *SVM-RBF*, que es el que ofrece mejores resultados.

N_{ind}	Dimensión	n -tuplas	Mejor valor de precisión mediana
750	2	$\binom{750}{2} = 280875$	0.8333
120	3	$\binom{120}{3} = 280840$	0.8095
51	4	$\binom{51}{4} = 249900$	0.8095

TABLA 3.1: Breve comparativa de las prestaciones de los clasificadores 3D y 4D vs clasificador 2D

Es importante tener en cuenta que, aunque se están comparando los clasificadores en condiciones de similar carga computacional, los clasificadores n -dimensionales tienen muchísimas más posibilidades de combinaciones de *features* (con 1000 índices se pueden conseguir 499500 combinaciones de 2 *features*; $1,66 \times 10^8$ combinaciones de 3 *features* y $4,14 \times 10^{10}$ combinaciones de 4 *features*) que no se están explorando en su totalidad. Pero, como un estudio preliminar a igualdad de carga computacional no ofrece mejores resultados, entonces se decide abordar únicamente el problema de la clasificación desde una perspectiva bidimensional, de manera que todos los clasificadores a los que se hace referencia en el presente trabajo son bidimensionales. Aún así, se remarca la parcialidad de la exploración en más de dos dimensiones, lo cual requeriría de un análisis más exhaustivo para lograr resultados plenamente concluyentes.

3.2 DESEMPEÑO DEL CLASIFICADOR

3.2.1 DESEMPEÑO EN TÉRMINOS DE PRECISIÓN, SENSIBILIDAD Y ESPECIFICIDAD

En el capítulo anterior se expuso que el número de índices de *features* supervivientes a las pruebas de significatividad estaba acotado superiormente por un valor que estaba en torno a mil índices, pudiendo alcanzar hasta mil doscientos. Usar todas las posibles parejas de todos los posibles índices asegura cierto nivel de exhaustividad en cuanto al uso de los *features* disponibles, si bien es cierto que el paso de mil a mil doscientos aumenta sensiblemente el tiempo de cómputo; por este motivo, todos los resultados que se muestren estarán basados en un análisis de mil índices a menos que se especifique lo contrario. Como se mencionó en la sección 3.1, no es necesario comparar n -tuplas de $n > 2$, porque el desempeño de los clasificadores n -dimensionales no parece claramente mejor para $n > 2$ en comparación con el caso $n = 2$. Para estos índices, los mejores valores medianos (en términos de precisión) que ofrece el clasificador SVM RBF (que es el que alcanza los mejores resultados) se exponen en la tabla 3.2. El desempeño del resto de sistemas de clasificación, mencionados en el capítulo anterior se muestra en la sección A.1 del apéndice.

Precisión	Sensibilidad	Especificidad
0.8571	0.9500	0.7727
0.8571	0.9000	0.8181
0.8571	0.9500	0.7727
0.8333	0.7000	0.9545
0.8095	0.8000	0.8181

TABLA 3.2: Desempeño del clasificador SVM con *kernel RBF*

A la vista de estos resultados, (y los que se recogen en el apéndice) se pueden hacer dos consideraciones:

- Con respecto a los valores obtenidos en el sistema, se puede afirmar que es bastante sensible. En la mayoría de los casos (el clasificador kMN es una excepción, como se verá en el apéndice) disminuye la probabilidad de diagnosticar como sanos a pacientes enfermos, con el consecuente coste de aumentar la probabilidad de falsas alarmas. En el ámbito de detección de un trastorno, como el que se aborda en este trabajo, se deben favorecer las detecciones puesto que lo más importante es que los pacientes enfermos sean diagnosticados como tal lo antes posible, mientras que puede ser admisible emitir una falsa alarma, lo que comportará que únicamente tengan que ser llevadas a cabo más pruebas para determinar con certeza el estado clínico del paciente.
- En relación al tiempo de ejecución requerido para obtener dichos resultados, se debe indicar que las pruebas se han hecho en la misma máquina, (cuyas características hardware se mencionaron en la sección 1.5) en situaciones aproximadamente iguales de carga de trabajo, y que la ejecución para 250 índices tarda aproximadamente 5 horas, la ejecución para 500 índices tarda aproximadamente 18 horas y la ejecución para 1000 índices tarda unos 6 días. En este sentido, se podría establecer un compromiso entre la cantidad de índices que se procesan y las prestaciones del clasificador resultante. En todo caso, claro está, una vez construido el clasificador, el tiempo empleado en la decisión es despreciable.

3.2.2 DESEMPEÑO DEL CLASIFICADOR EN TÉRMINOS DE LA FUNCIÓN DE DISTRIBUCIÓN DE LA PRECISIÓN

Habiendo discutido el desempeño del clasificador en términos medianos, resulta interesante analizar qué tan parecidos son los clasificadores construidos al clasificador ideal. Para hacer este análisis, es necesario definir de alguna manera un criterio de *distancia* entre los mismos. Una posibilidad es aproximar la función de distribución de la variable aleatoria “precisión del clasificador” con los datos de que se dispone. Para ello se hace un cálculo de varios percentiles de la precisión de los clasificadores. Sabiendo que la función de distribución del clasificador ideal es un escalón que conmuta en 1 (puesto que el clasificador ideal tiene el 100 % de precisión, no

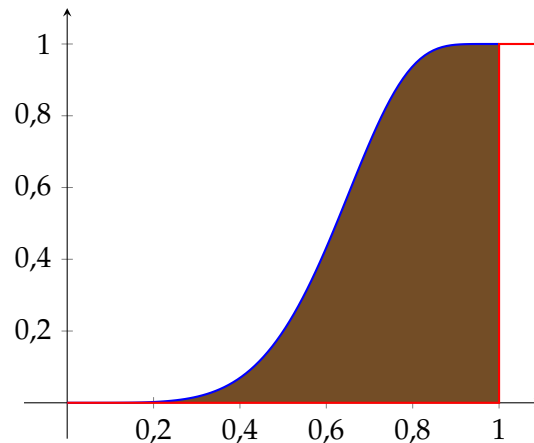


FIGURA 3.1: Función de distribución aproximada de la precisión (azul) vs Función de distribución ideal de la precisión (roja)

falla y todas las observaciones caen exclusivamente en el percentil 100) y cualquier clasificador tiene errores que ubican al menos una observación bajo los demás percentiles, una medida de la distancia del clasificador considerado al ideal es calcular dicha área, la cual se muestra en la figura 3.1.

Intuitivamente, se puede ver que el área marrón resaltada en la figura 3.1 es una medida de qué tan parecido es el clasificador real construido al clasificador ideal: cuanto más pequeña sea dicha área, mejor será el desempeño del clasificador. Para calcular dicha área con el conjunto de datos de los que se dispone, se puede usar un método de integración numérica, como la regla del trapecio. En las tablas de desempeño de los clasificadores recogidas en el apéndice, se encuentra el parámetro AuC, que se puede usar como otra medida de desempeño.

3.3 MEJORES CLASIFICADORES BIDIMENSIONALES

En la tabla 3.3 se recogen las 5 parejas de *features* que producen el mejor desempeño en términos de mejor precisión mediana y de distancia al clasificador ideal (ambos criterios han resultado ser equivalentes). La tabla 3.3 recoge los resultados del clasificador SVM con el *kernel RBF*. En la sección A.2 del apéndice se muestran los resultados para los demás clasificadores.

	Características				Prec	Sens	Espec	AuC
	Tipo	Parám	Fragm	Diezm				
f_1	CTM [†]	$\rho = 18$	día	60s	0.8571	0.9500	0.7727	0.1643
f_2	CTM	$\rho = 69$	descansoF	600s				
f_3	CTM [†]	$\rho = 11$	día	60s	0.8571	0.9500	0.7727	0.1643
f_4	CTM	$\rho = 70$	descansoF	900s				
f_5	CTM	$\rho = 69$	descansoF	600s	0.8571	0.9000	0.8181	0.1643
f_6	CTM	$\rho = 16$	día	5s				
f_7	CTM	$\rho = 22$	día	1s	0.8333	0.7000	0.9545	0.1875
f_8	CTM	$\rho = 73$	descansoF	30s				
f_9	CTM	$\rho = 21$	día	900s	0.8095	0.8000	0.8181	0.2107
f_{10}	CTM	$\rho = 23$	día	5s				

TABLA 3.3: Mejores 5 parejas de características para el clasificador SVM RBF

Las características señaladas con † son desestacionalizadas.

Se puede apreciar que la principal característica discriminante es la CTM; de ello es en buena medida responsable que de todo el espacio de 26.386 *features* empleados, el 85,74 % de ellos se corresponden con la CTM. Por otra parte, si bien hay otras características reseñables (véase sección 3.5 y sección A.3 del apéndice), nuestro algoritmo ha tenido preferencia por la CTM frente a otras por un mero motivo de implementación.

Por otra parte, en todas las parejas mostradas en la tabla 3.3, el fragmento en el cual se encuentran los *features* más discriminantes es bien el día completo, bien la parte final del descanso. Esto permite concluir, en una primera aproximación, que dichos fragmentos son más discriminantes con lo cual a nivel clínico, se puede hacer un análisis más detallado de lo que ocurre con los pacientes cuando están a punto de despertar. Además, cuando el clasificador está construido por dos *features* que son CTM de radios muy distintos (uno muy grande y uno muy pequeño) el radio grande está asociado al fragmento de parte final del descanso, y el radio pequeño está asociado a todo el fragmento en el que hay actividad.

3.4 UTILIDAD DEL PROCESO DE DESESTACIONALIZACIÓN

El estudio hecho en [24] sobre la variabilidad de los niveles de actividad en los pacientes, atribuible a la fecha del año en que se midan los mismos, abre un debate importante sobre si es preciso hacer una comparación acerca de la versión estacionalizada y desestacionalizada de los registros de actividad de los pacientes.

En nuestro caso, los resultados mostrados en la tabla 3.3 indican que las dos mejores parejas de *features* tienen un componente desestacionalizado; bien es cierto que el tercer *feature*, con el que el clasificador tiene un desempeño similar —aunque con menor sensibilidad— no hace uso de *features* desestacionalizados. Con ello podemos concluir que a partir de evidencia de que disponemos, no puede descartarse como opción el empleo de *features* desestacionalizados. Similares conclusiones pueden extraerse a partir de los resultados de otros clasificadores (véase sección A.2 del apéndice).

3.5 OTRAS CARACTERÍSTICAS RESEÑABLES

En los resultados anteriores se puede ver que la característica que destaca es la CTM, lo que indica que es una característica con gran poder discriminante. Se lleva a cabo un análisis de las mejores características, extendiéndolo más allá de lo que ocurre en las primeras 10 posiciones, y se pueden encontrar otras características como DS, estadísticos y parámetros del ritmo circadiano. Lo que es más, los niveles de precisión mediana de estos resultados están entre el 79% y el 81%, lo cual indica que, si bien no son de los mejores, tienen cierto poder discriminante. En la sección A.3 del apéndice se encuentran los 5 mejores resultados de características estadísticas, no lineales y de ritmo circadiano que se encontraron en el análisis llevado a cabo.

Al respecto de la SD, en la que se detecta que el único parámetro de SD que alcanzan estos niveles de significatividad es la probabilidad de la palabra 21. Como se vio en 2.2.3, esto significa el tránsito entre un nivel de actividad alto (2) a medio (1). Esta distinción entre el comportamiento de los pacientes cuando aumentan o disminuyen sus niveles de actividad es también una consideración a tener en cuenta.

Con respecto al ritmo circadiano, se encuentra que las características de este tipo se tienen cuando se analiza el canal z del mesor (el valor medio del ajuste de mínimos cuadrados sinusoidal al nivel de actividad, como se vio en 2.2.3). Esto puede verse como una distinción entre el movimiento en el eje z de casos y controles a lo largo del día. Además, la predilección de este parámetro con respecto al diezmado está en los valores altos de diezmado tanto a nivel de actividades (en torno a 60s) como de tarea (en torno a 900s).

Los parámetros estadísticos que mejor puntúan son la media y la mediana. Es natural pensar que estos parámetros tengan poder discriminante, sobre todo después de las consideraciones hechas con respecto a la desestacionalización. Los parámetros estadísticos no tienen predilección por un tipo específico de diezmado, pues barren aproximadamente todas las posibles tasas.

3.6 USO DE CARACTERÍSTICAS ESPECTRALES

En la sección 2.2.3 se hizo mención particular a *features* que trabajan en el dominio de la frecuencia. Cabe la pena discutir (tal como se hizo con la desestacionalización) si es provechoso hacer un análisis espectral de los datos. Con el *feature* que se empleó (la dimensión fractal de Maragos-Sun), los resultados obtenidos en cuanto a desempeño del clasificador fueron peores en comparación con el caso en el que únicamente se usan los *features* temporales. Además, un

análisis de los índices supervivientes en varias pasadas a lo largo de varias iteraciones revela que dicho *feature* no es significativo. Con estos dos hechos en conjunción se concluye que, aparte de no ser significativo, este *feature* ‘confunde’ al clasificador, haciendo que su desempeño global empeore, razón por la cual se desaconseja su uso.

CONCLUSIONES Y LÍNEAS FUTURAS

En este trabajo se propuso el desarrollo de un sistema de detección del TDAH mediante el uso de la actimetría. Destaca por su bajo costo, poca intrusividad a la vida del paciente, rapidez en obtención del diagnóstico y la precisión del mismo. A lo largo del planteamiento de las características del sistema, y el estudio de su desempeño, hemos podido extraer algunas conclusiones así como identificar algunas líneas de investigación que se dejan abiertas, las cuales pueden ser abordadas convenientemente en el futuro.

4.1 DESESTACIONALIZACIÓN

En [24] se hace un estudio detallado de la influencia de la estación del año. Aquí se hizo un estudio preliminar que, basándose en que la estación del año hace que los niveles de actividad varíen (siendo posible, por ejemplo, que un caso tenga menos actividad en invierno que un control en verano), pretendía independizar los niveles de actividad, a fin de remover esta dependencia estacional y tratar de manera homogénea todos los registros de actimetría procedentes de casos y controles de cualquier fecha del año. Para ello, se planteó un modelo circanual de variación de la actividad y se normalizó la actividad (como se especifica en la sección 2.2.4). Los resultados mostrados en el apéndice revelan que se pueden encontrar características desestacionalizadas entre los clasificadores con mejores resultados. En consecuencia, la desestacionalización provee un número no despreciable de características significativas en términos tanto de precisión como de sensibilidad, que impulsa a continuar la ejecución de pruebas con un tamaño muestral mayor, a fin de poder analizar con mayor nivel de detalle las posibles ventajas que ofrece ésta técnica.

4.2 CARACTERÍSTICAS DISCRIMINANTES

En el capítulo 3 se estudiaron los *features* más discriminantes. Ya se ha hecho mención de que la CTM constituye un muy elevado porcentaje de los *features* considerados. Sin embargo, como se puede apreciar en la sección A.3 del apéndice, hay unos *features* que alcanzan niveles de precisión de hasta 81 %, pero que no se reflejan en las tablas por un motivo de implementación,

que tiene que ver con que se ha calculado primero la CTM, y a igualdad de valores de precisión mediana, se ordenan según fueron calculados.

En todo caso CTM y SD parecen que juegan los papeles más relevantes en materia de discriminación de casos y controles. En un peldaño más abajo se encuentran las características procedentes del ritmo circadiano y los estadísticos básicos. En todos ellos, hemos encontrado patrones repetitivos y/o clínicamente interpretables, lo cual proporciona a esta herramienta un valor añadido adicional.

En todo caso, y como línea futura, merece la pena tratar de aumentar el tamaño de la cohorte para comprobar cuáles de los parámetros analizados en esta muestra relativamente reducida muestra mejores resultados de una forma más concluyente.

4.3 CLASIFICADORES

En la sección 2.2.5 se mencionaron los clasificadores empleados, que son un subconjunto muy reducido del conjunto de clasificadores que se pueden encontrar. Aunque se han obtenido resultados tan buenos como precisión del 85,71 %, se propone que se empleen otros tipos de clasificadores, que pueden ser capaces de elevar las prestaciones de este sistema de detección. En concreto, los autores plantean emplear redes neuronales (*neural networks*) debido a que, por construcción, son capaces de aprender y adaptarse a señales cuyas propiedades estadísticas cambian con el tiempo (no estacionarias) [26] como las que se emplean en este trabajo. Así mismo, se plantea el uso de la regresión logística como apoyo al sistema de clasificación, para que el sistema de decisión que se ofrezca no sea estrictamente binario sino que permita establecer una probabilidad de ser caso o control, dejando el diagnóstico final en manos del profesional de la salud encargado, que podrá someter al paciente a nuevas pruebas a fin de determinar con un menor nivel de incertidumbre su estado clínico.

4.4 CARACTERÍSTICAS ESPECTRALES

El empleo de *features* espectrales en este trabajo fue anecdótico. El sistema de clasificación está basado en su inmensa mayoría en *features* calculados en el dominio del tiempo: de 16 tipos distintos de *features* que se usan, solo la dimensión fractal de Maragos-Sun hace un análisis espectral y los resultados obtenidos a partir de este *feature* no parece avalar su uso. Sin embargo, en la literatura se pueden encontrar muchos *features* que se calculan en el dominio de la frecuencia y que pueden ser significativos, razón por la cual se incita a que en un futuro, sean empleados otros tipos de *features* espectrales, en aras de encontrar *features* que permitan esbozar mejor la distribuciones teóricas que subyacen a los niveles de actividad de casos y controles.

4.5 SEGUIMIENTO

Por último, se propone que el sistema de detección presentado en este trabajo no solo sea empleado para diagnosticar a un paciente, sino que, como mencionan Martínez *et al* [5], se pueda usar el sistema para el seguimiento de los pacientes, con el fin de proporcionar información relevante en cuanto a la evolución del trastorno, que permita al profesional de la salud personalizar la medicación que reciben los pacientes.

Apéndice A

TABLAS DE RESULTADOS

En el presente apéndice se mostrarán pruebas de funcionamiento más detalladas del sistema experto de detección del TDAH que se ha presentado en los capítulos anteriores. En concreto, se comentará acerca de:

- el desempeño de todos los sistemas de clasificación empleados en términos medianos y en términos de su función de distribución.
- parejas de *features* con mejor desempeño.
- otras características reseñables.

A.1 DESEMPEÑO DE LOS CLASIFICADORES

En las tablas que se muestran a continuación se evalúa cuantitativamente el desempeño de los clasificadores mencionados en el capítulo 2 en términos de la mediana y en términos de la distancia entre las funciones de distribución ideal y real, siguiendo el método especificado en 3.2.2. Este anexo extiende los resultados descritos en la sección 3.2.

A.1.1 DESEMPEÑO DEL CLASIFICADOR kNN

Precision	Sensibilidad	Especificidad
0.7857	0.7000	0.8636
0.7857	0.6500	0.9090
0.7619	0.7000	0.8181
0.7619	0.7000	0.8181
0.7619	0.7000	0.8181

TABLA A.1: Desempeño del clasificador kNN

A.1.2 DESEMPEÑO DEL CLASIFICADOR SVM CON *kernel* LINEAL

Precision	Sensibilidad	Especificidad
0.8571	0.9000	0.8181
0.8333	0.8000	0.8636
0.8333	0.8500	0.8181
0.8333	0.9500	0.7272
0.8333	1.0000	0.6818

TABLA A.2: Desempeño del clasificador SVM con *kernel* lineal

A.1.3 DESEMPEÑO DEL CLASIFICADOR SVM CON *kernel* CUADRÁTICO

Precision	Sensibilidad	Especificidad
0.8571	0.8000	0.9090
0.8333	0.9500	0.7272
0.8095	0.7500	0.8636
0.8095	0.8000	0.8181
0.8095	0.8500	0.7727

TABLA A.3: Desempeño del clasificador SVM con *kernel* cuadrático

A.2 PAREJAS DE CARACTERÍSTICAS QUE OFRECEN MEJOR DESEMPEÑO AL CLASIFICADOR

Las tablas que se muestran a continuación contienen las 10 parejas de *features* que ofrecieron mejor desempeño en los clasificadores, para cuando se consideraron $N = 1000$ índices significativos. Se muestran los *features* que alcanzaron tales valores, así como los valores de sensibilidad y especificidad asociados y se distinguen los *features* estacionalizados de los desestacionalizados. Esta distinción se hace para evaluar la utilidad del proceso de desestacionalización, que se comenta en la sección 3.4. Se excluye de este apéndice la tabla 3.3 correspondiente al SVM-RBF, mostrada en la sección 3.3.

A.2.1 CLASIFICADOR KNN

	Características				Prec	Sens	Espec	AuC
	Tipo	Parám	Fragm	Diezm				
f_1	CTM	$\rho = 33$	día	300s	0.7857	0.7000	0.8636	0.2339
f_2	CTM	$\rho = 73$	descansoF	30s				
f_3	CTM	$\rho = 23$	día	5s	0.7857	0.6500	0.9090	0.2339
f_4	CTM [†]	$\rho = 24$	día	30s				
f_5	CTM [†]	$\rho = 11$	día	60s	0.7619	0.7000	0.8181	0.2571
f_6	CTM	$\rho = 21$	día	900s				
f_7	CTM	$\rho = 22$	día	1s	0.7619	0.7000	0.8181	0.2571
f_8	CTM	$\rho = 15$	día	1s				
f_9	CTM	$\rho = 22$	día	1s	0.7619	0.7000	0.8181	0.2571
f_{10}	CTM	$\rho = 73$	descansoF	30s				

TABLA A.4: Mejores 5 parejas de características para el clasificador kNN

Las características señaladas con † son desestacionalizadas.

A.2.2 CLASIFICADOR SVM LINEAL

	Características				Prec	Sens	Espec	AuC
	Tipo	Parám	Fragm	Diezm				
f_1	CTM [†]	$\rho = 18$	día	60s	0.8571	0.9000	0.8181	0.1643
f_2	CTM	$\rho = 69$	descansoF	600s				
f_3	CTM	$\rho = 69$	descansoF	600s	0.8333	1.0000	0.6818	0.1875
f_4	CTM	$\rho = 16$	día	5s				
f_5	CTM [†]	$\rho = 17$	día	30s	0.8333	0.9500	0.7272	0.1875
f_6	CTM	$\rho = 70$	descansoF	900s				
f_7	CTM [†]	$\rho = 44$	día	5s	0.8333	0.8500	0.8181	0.1875
f_8	CTM	$\rho = 78$	descansoF	1s				
f_9	CTM	$\rho = 17$	día	30s	0.8333	0.8000	0.8636	0.1875
f_{10}	CTM	$\rho = 73$	descansoF	30s				

TABLA A.5: Mejores 5 parejas de características para el clasificador SVM lineal

Las características señaladas con † son desestacionalizadas.

A.2.3 CLASIFICADOR SVM CUADRÁTICO

	Características				Prec	Sens	Espec	AuC
	Tipo	Parám	Fragm	Diezm				
f_1	CTM [†]	$\rho = 18$	día	60s	0.8571	0.8000	0.9090	0.1643
f_2	CTM	$\rho = 69$	descansoF	600s				
f_3	CTM [†]	$\rho = 44$	día	5s	0.8333	0.9500	0.7272	0.1875
f_4	CTM	$\rho = 78$	descansoF	1s				
f_5	CTM	$\rho = 69$	descansoF	600s	0.8095	0.8500	0.7727	0.2107
f_6	CTM	$\rho = 16$	día	5s				
f_7	CTM [†]	$\rho = 17$	día	30s	0.8095	0.8000	0.8181	0.2107
f_8	CTM	$\rho = 70$	descansoF	900s				
f_9	CTM	$\rho = 17$	día	30s	0.8095	0.7500	0.8636	0.2107
f_{10}	CTM	$\rho = 73$	descansoF	30s				

TABLA A.6: Mejores 5 parejas de características para el clasificador SVM cuadrático

Las características señaladas con † son desestacionalizadas.

A.3 OTRAS CARACTERÍSTICAS RESEÑABLES

A continuación se muestran las mejores parejas de *features* que exhiben el mejor comportamiento en términos de mediana de precisión (o equivalentemente en términos de la distancia del clasificador ideal al real) y que incluyen al menos un *feature* que no sea CTM. Debido a que el clasificador SVM RBF es el que alcanza los mejores valores de precisión, será el único del que se muestren resultados, a efectos ilustrativos.

A.3.1 DINÁMICA SIMBÓLICA

	Feat	Parám	Fragm	Diezm	Prec
f_1	CTM	$\rho = 78$	descansoF	1s	0.8095
f_2	DS	probPalabra21	día	1	
f_3	CTM	$\rho = 73$	descansoF	30s	0.8095
f_4	DS	probPalabra21	día	1s	
f_5	DS	probPalabra00	descansoI	600s	0.7857
f_6	CTM	$\rho = 78$	descansoF	1s	
f_7	CTM	$\rho = 78$	descansoF	1s	0.7619
f_8	DS	probPalabra21	día	1s	
f_9	CTM	$\rho = 78$	descansoF	1s	0.7619
f_{10}	DS [†]	probPalabra21	día	1s	

TABLA A.7: Mejores 5 parejas de características de las que al menos una es dinámica simbólica

Las características señaladas con † son desestacionalizadas.

A.3.2 CARACTERÍSTICAS DEL RITMO CIRCADIANO

	Feat	Parám	Fragm Canal	Diezm	Prec
f_1	CarRitmoCircadiano	mesor	canalZ	900s	0.7857
f_2	CTM	$\rho = 78$	descansoF	1s	
f_3	CarRitmoCircadiano	mesor	canalZ	900s	0.7857
f_4	CTM	$\rho = 69$	descansoF	600s	
f_5	CarRitmoCircadiano	mesor	canalZ	900s	0.7857
f_6	CTM	$\rho = 76$	descansoF	600s	
f_7	CarRitmoCircadiano	mesor	canalZ	900s	0.7619
f_8	CTM	$\rho = 76$	descansoF	600s	
f_9	CarRitmoCircadiano	mesor	canalZ	900s	0.7619
f_{10}	CTM	$\rho = 73$	descansoF	30s	

TABLA A.8: Mejores 5 parejas de características de las que al menos una es de ritmo circadiano

Ninguna de las características mostradas arriba son desestacionalizadas

A.3.3 ESTADÍSTICOS

	Feat	Parám	Fragm	Diezm	Precisión
f_1	Stad	media	día	900s	0.7857
f_2	CTM	$\rho = 76$	descansoF	600s	
f_3	Stad [†]	media	día	30s	0.7857
f_4	CTM	$\rho = 76$	descansoF	600s	
f_5	Stad	mediana	día	30s	0.7857
f_6	CTM	$\rho = 73$	descansoF	30s	
f_7	Stad [†]	media	día	30s	0.7857
f_8	CTM	$\rho = 76$	descansoF	600s	
f_9	Stad	media	día	300s	0.7619
f_{10}	CTM	$\rho = 76$	descansoF	600s	

TABLA A.9: Mejores 5 parejas de características de las que al menos una es un estadístico

Las características señaladas con † son desestacionalizadas

BIBLIOGRAFÍA

- [1] Samuel B Guze. Diagnostic And Statistical Manual Of Mental Disorders, (DSM-IV). *American Journal of Psychiatry*, 152(8):1228–1228, 1995.
- [2] Hye Jin Kam, Kiyoung Lee, Sun-Mi Cho, Yun-Mi Shin, and Rae Woong Park. High-Resolution Actigraphic Analysis Of ADHD: A Wide Range Of Movement Variability Observation In Three School Courses-A Pilot Study. *Healthcare informatics research*, 17(1): 29–37, 2011.
- [3] Avi Sadeh. The Role And Validity Of Actigraphy In Sleep Medicine: An Update. *Sleep medicine reviews*, 15(4):259–267, 2011.
- [4] D Martín Martínez, P Casaseca de la Higuera, JM Andrés de Llano, JR Garmendia Leiza, S Alberola-López, and C Alberola López. Automatic Detection Of Wakefulness And Rest Intervals In Actigraphic Signals: A Data-Driven Approach. *Medical Engineering & Physics*, 36(12):1585–1592, 2014.
- [5] D Martín Martínez, P Casaseca de la Higuera, S Alberola López, J Andrés de Llano, JA López Villalobos, J Ardura Fernández, and C Alberola López. Nonlinear Analysis Of Actigraphic Signals For The Assessment Of The Attention-Deficit/Hyperactivity Disorder (ADHD). *Medical Engineering & Physics*, 34(9):1317–1329, 2012.
- [6] J Garmendia Leiza, M Aguilar Garcia, J Andres De Llano, D Martin Martinez, P Casaseca De La Higuera, and C Alberola Lopez. Influence of institutionalization on the sleep pattern in elderly population. *Sleep Medicine*, 14:e181–e182, 2013.
- [7] D Martin Martinez, P Casaseca De La Higuera, C Alberola Lopez, J Garmendia Leiza, J Andres De Llano, and S Alberola Lopez. Utility of the statistical and nonlinear analysis for the actigraphic sleep pattern characterization. *Sleep Medicine*, 14:e181, 2013.
- [8] Subcommittee on Attention-Deficit et al. ADHD: Clinical Practice Guideline For The Diagnosis, Evaluation, And Treatment Of Attention-Deficit/Hyperactivity Disorder In Children And Adolescents. *Pediatrics*, 2011.
- [9] Guilherme Polanczyk, Maurício Silva de Lima, Bernardo Lessa Horta, Joseph Biederman, and Luis Augusto Rohde. The Worldwide Prevalence Of ADHD: A Systematic Review And Metaregression Analysis. *American Journal of Psychiatry*, 164(6):942–948, 2007. doi: 10.1176/ajp.2007.164.6.942. URL <http://ajp.psychiatryonline.org/doi/abs/10.1176/ajp.2007.164.6.942>. PMID: 17541055.

- [10] Maaïke M Van Veen, JJ Sandra Kooij, A Marije Boonstra, Marijke CM Gordijn, and Eus JW Van Someren. Delayed Circadian Rhythm In Adults With Attention-Deficit/Hyperactivity Disorder And Chronic Sleep-Onset Insomnia. *Biological Psychiatry*, 67(11):1091–1096, 2010.
- [11] Moshe Zeidner. *Test Anxiety: The State of Art*. Kluwer Academic Publishers, 1998.
- [12] Bernard Rosner. *Fundamentals Of Biostatistics*. Duxbury, quinta edición, 2000. pág-219.
- [13] Maurice E Cohen, Donna L Hudson, and Prakash C Deedwania. Applying continuous chaotic modeling to cardiac signal analysis. *IEEE Engineering in Medicine and Biology Magazine*, 15(5):97–102, 1996.
- [14] A Voss, J Kurths, HJ Kleiner, A Witt, N Wessel, P Saperin, KJ Osterziel, R Schurath, and R Dietz. The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death. *Cardiovascular research*, 31(3): 419–433, 1996.
- [15] M Baumert, T Walther, J Hopfe, H Stepan, R Faber, and A Voss. Joint symbolic dynamic analysis of beat-to-beat interactions of heart rate and systolic blood pressure in normal pregnancy. *Medical and Biological Engineering and Computing*, 40(2):241–245, 2002.
- [16] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.
- [17] Abraham Lempel and Jacob Ziv. On the complexity of finite sequences. *IEEE Transactions on information theory*, 22(1):75–81, 1976.
- [18] Steve Pincus. Approximate entropy (apen) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):110–117, 1995.
- [19] Gregory L Baker and Jerry P Gollub. *Chaotic Dynamics: An Introduction*. Cambridge University Press, 1996.
- [20] W Nelson, Y Liang Tong, Jung-Kuen Lee, and F Halberg. Methods for cosinorhythmometry. *Chronobiologia*, 6(4):305, 1979.
- [21] Avi Sadeh, Katherine M Sharkey, and Mary A Carskadon. Activity-based sleep—wake identification: An empirical test of methodological issues. *Sleep*, 17(3):201–207, 1994.
- [22] Avi Sadeh, Christine Acebo, R Seifer, Semra Aytur, and Mary A Carskadon. Activity-based assessment of sleep-wake patterns during the first year of life. *Infant Behavior and Development*, 18(3):329–337, 1995.
- [23] Petros Maragos and Fang-Kuo Sun. Measuring The Fractal Dimension Of Signals: Morphological Covers And Iterative Optimization. *IEEE Transactions on Signal Processing*, 41(1): 108–121, 1993.
- [24] Patricia Tucker and Jason Gilliland. The Effect Of Season And Weather On Physical Activity: A Systematic Review. *Public Health*, 121(12):909–922, 2007.

- [25] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Academic Press, quinta edition, 2003. ISBN 0126858756.
- [26] Simon Haykin and Neural Network. A comprehensive foundation. *Neural Networks, 2* (2004), 2004.