



Universidad de Valladolid



Escuela Técnica Superior de Ingenieros de Telecomunicación

Universidad de Valladolid

GRADO EN INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN

Un entorno para el procesado de señales
actigráficas y la construcción de un sistema
experto para la determinación del TDAH en la
infancia

Alumno/a: Jorge García Hospital

Tutor/es/as: Carlos Alberola López



A mis padres

AGRADECIMIENTOS

Desde estas líneas me gustaría expresar mi más sincero agradecimiento:

A mis padres, por haberme proporcionado la mejor educación y lecciones de vida, sin olvidar el apoyo de todo tipo que he recibido día a día a lo largo de esta etapa.

En especial a mi padre, por haberme enseñado que con esfuerzo, trabajo y constancia todo se consigue, y que en esta vida nadie regala nada.

En especial a mi madre, por hacerme ver cada día la vida de una forma diferente y confiar en mis decisiones.

A mi hermano, ese granuja que, quizás sin saberlo, tanto me ha ayudado a seguir adelante y nunca rendirme.

A mi tutor Carlos Alberola, por ayudarme y guiarme durante todo el proyecto.

Al resto de profesores y personal de la escuela, quienes dentro y fuera de clase han aportado tanto a mi evolución desde el primer día.

A todos mis familiares por haberme apoyado y animado.

A mis compañeros de clase, con los que he compartido momentos tanto buenos como malos, pero sin duda inolvidables.

A mis amigos, por estar siempre a mi lado dispuestos a soportar los cabreos y monólogos relacionados con la carrera, una y otra vez, además de darme ánimo y fuerza para que el trabajo llegara a su fin.

A todos aquellos que, de una manera u otra han aportado lo máximo posible para que hoy sea quien soy.

A todos aquellos que siguen estando cerca de mí y me regalan parte de su vida.

A todos vosotros, **GRACIAS**.

RESUMEN

Con este proyecto se pretende simplificar las tareas de estudio y desarrollo de 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. Para ello, se ha desarrollado una interfaz gráfica fácilmente manejable y adaptable que incluye funciones de extracción y selección de características o patrones a partir de registros de actividad y el diseño de un sistema de clasificación que permita discernir entre pacientes sanos y pacientes con sospechas de TDAH.

ABSTRACT

This project is meant to simplify the study and development of an objective, non-intrusive and low cost system for the detection of Attention-Deficit Hyperactivity Disorder (ADHD) based on actimetry. That purpose is achieved by developing an user-friendly graphical interface that includes the extraction and selection of characteristics or patterns from activity registers and the design of a classification system that allows discerning between healthy patients and patients with suspected ADHD.

ÍNDICE GENERAL

Agradecimientos	II
Resumen	III
Abstract	IV
Lista de figuras	X
1. Introducción	1
1.1. Contexto	1
1.1.1. El Trastorno de Déficit de Atención e Hiperactividad	2
1.1.2. El estudio de la actimetría	3
1.2. Objetivos	4
1.3. Fases	5
1.4. Medios empleados	6
2. Diseño de interfaces de usuario	7
2.1. Características GUI/UX	7
2.2. Diseño de interfaces en MATLAB: Herramienta GUIDE	10
2.2.1. Diseño gráfico de la interfaz	10
2.2.2. Programación de la interfaz	12

3. Bases del estudio	14
3.1. Registros actigráficos. Validación y preprocesado.	14
3.1.1. Detección automática de los intervalos de sueño	15
3.1.2. Fragmentado de los registros	15
3.1.3. Diezmado	16
3.2. Registros actigráficos. Extracción de características.	17
3.2.1. Características estadísticas de la señal	17
3.2.2. Características no lineales de la señal	17
3.2.3. Características del ritmo circadiano	18
3.2.4. Parámetros del sueño	19
3.3. Desestacionalización	20
3.4. Selección de características	21
3.5. Análisis de prestaciones de los clasificadores. Validación cruzada	22
3.5.1. Introducción a los sistemas clasificadores	22
3.5.2. Tipos de clasificadores utilizados	22
3.5.3. Dimensión de los clasificadores	23
3.5.4. Validación Cruzada	24
4. Adición de funcionalidades y flexibilidad del sistema	26
4.1. Apertura de registros actigráficos	26
4.2. Extracción de características	27
4.2.1. Identificación de características	28
4.3. Parámetros de procesamiento	29
4.4. Guardado de datos en etapas intermedias	31
4.5. Seguimiento de los procesos	32
4.6. Clasificación de pacientes	33

5. Resultados: Ejecución completa del sistema	34
5.1. Procesado de registros	35
5.1.1. Apertura de ficheros	35
5.1.2. Elección de generadores de características	37
5.1.3. Generación y guardado	37
5.2. Selección de características discriminantes y análisis de las prestaciones de los clasificadores	39
5.2.1. Apertura de registros	39
5.2.2. Grupos de población	40
5.2.3. Elección de Características	41
5.2.4. Selección de características discriminantes	42
5.2.5. Manejo de la lista de características	44
5.2.6. Análisis de las prestaciones de los clasificadores	46
5.3. Lectura de resultados y selección de los mejores conjuntos clasificadores	47
5.4. Detección de TDAH sobre registros sin clasificar	48
6. Conclusiones y líneas futuras	50
6.1. Conclusiones	50
6.2. Líneas futuras	51
6.2.1. Lenguaje de programación empleado	51
6.2.2. Ampliación de características	51
6.2.3. Adaptación de nuevos clasificadores	52
A. Manual de referencia	53
A.1. Estructura de directorios	53
A.2. Creación de scripts para la extracción de nuevas características	55
A.2.1. Variables de entrada y salida	55
A.2.2. Lectura del fichero de datos correspondiente	55

A.2.3. Creación de la estructura de datos	56
A.2.4. Cálculo iterativo de los valores y concatenación	58
A.2.5. Guardado en fichero temporal	58
A.3. Identificación de características según su estructura	59
A.4. Estructura de los ficheros de resultados de las pruebas de clasificadores	61
Bibliografía	64

ÍNDICE DE FIGURAS

2.1. Ejemplo de entorno de apertura de ficheros	8
2.2. Herramienta GUIDE para la creación de interfaces	10
2.3. Parámetros disponibles para los elementos <i>button</i>	11
2.4. Múltiples líneas dentro del mismo elemento <i>tabla</i>	13
3.1. Resultados de prueba de clasificación	25
4.1. Filtrado de individuos y selección de registros	27
4.2. Ejemplo de estructura de características	28
4.3. Elección de características	30
4.4. Barras de selección de individuos para validación cruzada	31
4.5. Seguimiento del progreso y estimación de tiempo restante	32
4.6. Resultados de una prueba de diagnóstico	33
5.1. Menú para la apertura de las diferentes partes del programa completo	34
5.2. Apertura de registros del tipo casos.	35
5.3. Apertura de registros del tipo controles.	36
5.4. Elección de scripts generadores de características	37
5.5. Generación de características y ventana de progreso	38
5.6. Resultados de el procesado de registros.	39

5.7. Apertura de registros procesados	40
5.8. Elección de registros en función de edad y sexo	41
5.9. Elección de características a procesar	42
5.10. Selección de características por Test P-Valor	43
5.11. Resultados de prueba de selección	44
5.12. Reducción del número de características	45
5.13. Prueba de clasificadores	46
5.14. Resultados validación cruzada	48
5.15. Detección de TDAH en pacientes	49
A.1. Estructura de directorios del sistema	53
A.2. Estructura de la variable ‘senales’	56
A.3. Ejemplo de estructura de características	57
A.4. Estructura de la característica estudiada	60
A.5. Resultados para ciertos grupos clasificadores	61

CAPÍTULO 1

INTRODUCCIÓN

En el presente capítulo se contextualiza el problema planteado. Se detallan los antecedentes del estudio y se justifica la necesidad de disponer de una herramienta adaptable y configurable que facilite el trabajo de extracción y análisis de resultados, así como que permita el seguimiento continuo del procesado y resulte en todo momento amigable de cara al usuario.

1.1. CONTEXTO

El presente proyecto se enmarca en el estudio de datos actigráficos para su utilización en el diagnóstico de enfermedades, en concreto, la detección del Trastorno de Déficit de Atención e Hiperactividad a edad temprana.

El Trastorno por Déficit de Atención-Hiperactividad (TDAH) es el problema de salud mental más prevalente en la población infanto-juvenil en los países desarrollados. Se caracteriza por un patrón mantenido de inatención y/o hiperactividad-impulsividad, que es más frecuente y grave que el observado en sujetos de un nivel de desarrollo similar. Los síntomas deben presentarse en dos o más ambientes, comenzar antes de los siete años de edad y causar un deterioro clínicamente significativo de la actividad social, académica o laboral del individuo [1].

Actualmente, la detección de este tipo de patologías implica una serie de entrevistas periódicas durante meses con diversos profesionales de la salud y 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 [2]. Este tipo de métodos resultan intrusivos y poco adecuados para niños en edad temprana, por lo que se hace conveniente el desarrollo de métodos alternativos que permitan realizar un diagnóstico de manera fiable y sin interferir en el desarrollo diario de los pacientes.

Ante este problema se encuentra la actigrafía, un método incruento y no invasivo que se utiliza para registrar y medir la actividad corporal de forma objetiva. La actigrafía es reconocida como un método válido para el estudio del insomnio, desórdenes en el ritmo circadiano y la somnolencia [3], además del uso de sus datos como marcadores precoces del desarrollo neurológico [4]. Por ello, se considera el uso de la actimetría como alternativa no intrusiva y de bajo coste en la detección precoz del TDAH [5].

1.1.1. EL TRASTORNO DE DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD

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 [6].

Sus características se dividen en las dos dimensiones que son afectados por el trastorno [7]:

- **Déficit de atención:** Dificultad para mantener la concentración y cuidar los detalles en entornos escolares, trabajo y otras actividades diarias, así como el descuido y falta de atención para con sus interlocutores y una marcada desobediencia de las reglas en juegos y actividades grupales. Suelen aparecer cambios de actividad frecuentes, saltando entre tareas incompletas o bien abandonando éstas debido a un estímulo irrelevante al que una persona sana no respondería. Normalmente viene acompañado de falta de constancia y numerosos olvidos en las tareas cotidianas.
- **Hiperactividad/Impulsividad:** Agitación (*fidgiting*) en manos y pies e intranquilidad, movimiento continuo en situación sentada, además de incapacidad para permanecer sentado por largos periodos de tiempo. Se presentan acciones inadecuadas para ciertas situaciones, como salir corriendo o escalar objetos así como hablar en exceso o antes de ser preguntado. En ocasiones se refiere a esta patología como *estar en marcha* o *estar empujado por un motor*, solo limitado por el sentimiento de fatiga.

En la actualidad es uno de los problemas de salud más importantes en términos de morbilidad y disfuncionalidad, abarcando desde la infancia a la adolescencia y hasta la vida adulta; por ello este trastorno despierta un gran interés tanto en el ámbito clínico como en el de la investigación en neurociencias y es un motivo frecuente por el que la población infantil es remitida a pediatras, psiquiatras infantiles y psicólogos clínicos. Esto supone un enorme consumo de recursos para el Sistema Nacional de Salud, ocasionando un elevado número de consultas tanto en Atención Primaria como en Atención Especializada, además de un elevado gasto farmacéutico al requerir tratamiento con fármacos caros y durante un periodo de tiempo prolongado, ocupando el primer lugar en porcentaje de gasto en medicamentos en algunos cupos de pediatría de Atención Primaria.

La parte de la población afectada se establece normalmente entre 1,2 - 8 % en la población española en edad de escolarización, aunque varía en función de los criterios de análisis, el origen de las muestras y el desarrollo metodológico empleado. En un estudio reciente realizado en Castilla y León[8], la tasa global de prevalencia fue del 6,7 % (IC 95 %: 5,1-8,1 %), con el

1 % del subtipo hiperactivo, el 1,3 % del subtipo inatento y el 4,4 % del subtipo combinado. La tasa de prevalencia en el sexo masculino (9 %) es significativamente superior a la del sexo femenino (4 %).

Entre los síntomas característicos del TDAH más relevantes se encuentra desinhibición y la pobre regulación de la actividad e impulsividad, siendo los que más discriminan al niño con TDAH de otras psicopatologías y de los niños sanos. Estos factores son relevantes en el análisis y detección del trastorno y serán los que determinarán las características a estudiar para su diagnóstico.

1.1.2. EL ESTUDIO DE LA ACTIMETRÍA

La actimetría es una técnica no invasiva utilizada para evaluar y definir los ciclos de actividad y descanso a lo largo del día a través de la cuantificación continua del movimiento en los tres ejes cartesianos. Su importancia en el campo de la medicina ha experimentado un gran crecimiento en las dos últimas décadas gracias al gran número de estudios en los que se aplica esta técnica, la mayor parte de ellos relacionados con el seguimiento de patrones de sueño y ritmo circadiano [9, 10, 11, 12].

La versatilidad de este método permite su uso clínico en diversos campos además de los anteriormente citados. Como ejemplo cabe destacar los diferentes estudios realizados por el Laboratorio de Procesado de Imagen de la Universidad de Valladolid, en el que también se desarrolla el presente estudio. Entre sus contribuciones destacan [13, 14, 15].

El uso de la actigrafía permite disponer de un método económico, poco intrusivo y objetivo para el diagnóstico de diferentes afecciones. El método más común de recogida de datos consta de un acelerómetro que cuantifica el movimiento en los tres ejes del espacio y un sistema de registro de memoria, donde se graban las medidas de actividad. Se coloca normalmente en la muñeca de la mano dominante, pero también puede hacerse en la cintura, el brazo o el tobillo, dependiendo del tipo de estudio que se pretenda realizar y de la población con la que se trabaje.

En relación con el diagnóstico del TDAH, la ventaja fundamental del uso de la actimetría para la obtención de datos radica en su facilidad de uso, ya que hace posible que el niño siga con su vida normal durante el registro y permite una libertad que no se puede mantener en el laboratorio de sueño, por la rigidez de las condiciones que se imponen en ellos. Además, la utilización del actímetro no distorsiona la dinámica familiar y el entorno social y escolar del niño. El aparato es ligero y fácil de llevar, así como resistente a los golpes que acompañan a la vida normal de los niños a estas edades.

1.2. OBJETIVOS

El objetivo de este trabajo es construir un sistema integrado que unifique el procesado y estudio de señales actigráficas para su utilización como sistema experto de detección del TDAH, así como completar sus funcionalidades.

La construcción de sistemas clasificadores implica un análisis previo de los datos a aportar para realizar la clasificación de los pacientes. Tomando como referencia los estudios presentados en [16] y [17], es posible observar que el número de posibilidades en cuanto a características de señal que se pueden utilizar para detectar el trastorno es de más de cincuenta mil, con previsiones de aumentar al añadir nuevos tipos de parámetros. El manejo de tal cantidad de datos para cada uno de los individuos analizados supone un gran reto de cara a la programación del sistema, ya que es necesario utilizar técnicas matriciales e identificar en todo momento los parámetros con los que se trabaja, lo que complica las tareas de modificación de parámetro o expansión de funcionalidades. Como consecuencia, se presenta la necesidad de disponer de un sistema con parámetros de ejecución configurables sin necesidad de modificar el código interno y que permita su manejo por parte de cualquier tipo de usuario, incluyendo personal clínico.

Para ello se plantea el desarrollo de una interfaz gráfica que permita el manejo de los registros actigráficos existentes, su selección individualizada o dentro de grupos de población tales como edad y sexo y su posterior procesado para obtener resultados estadísticos acerca de la utilidad práctica de las diferentes herramientas que pueden ser utilizadas para el diagnóstico, tanto en lo relativo al tratamiento de los datos actigráficos como en lo relativo a la construcción de elementos clasificadores, su entrenamiento y prueba. El trabajo se completará con el estudio de nuevas características discriminantes y el comportamiento de clasificadores basados en el aprendizaje neuronal y bosques aleatorios.

1.3. FASES

Para la consecución del objetivo marcado en la sección anterior se determinan diferentes fases que permitan desarrollar cada subobjetivo planteado:

1. Revisión de trabajos anteriores, búsqueda de bibliografía complementaria y adquisición de conocimientos acerca del estudio tratado.
2. Validación de los recursos disponibles y análisis crítico del código existente.
3. Adquisición de conocimientos en lo relativo a la creación de interfaces gráficas (GUI's) en entorno Matlab.
4. Análisis de requisitos funcionales y no funcionales de interfaces de usuario y obtención de la lista de funcionalidades concretas a desarrollar.
5. Diseño, implementación, optimización y validación de la interfaz interactiva, así como inclusión de nuevas funcionalidades.
6. Estudio de resultados procedentes de la ejecución completa del sistema desarrollado.

1.4. MEDIOS EMPLEADOS

La realización del presente proyecto implica el uso de lenguajes de programación específicos, así como ciertas herramientas y equipos. También cabe destacar dentro de este apartado los registros médicos disponibles, sin los cuales sería imposible diseñar el sistema y comprobar su eficiencia:

- Registros actigráficos pertenecientes a 74 pacientes que padecen TDAH y 75 pacientes sanos utilizados como referencia. Los registros han sido identificados como casos (padecen TDAH) y controles (no padecen TDAH) gracias al test DSM-IV [6].
- 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).
- Servidor en red para acceso remoto y cálculo.
- Software de cálculo numérico: MATLAB R2016b, con toolbox de cálculo estadístico y Machine Learning, así como el toolbox de Redes Neuronales y Deep Learning

CAPÍTULO 2

DISEÑO DE INTERFACES DE USUARIO

En este capítulo se presentarán las principales características necesarias en cualquier interfaz de usuario para permitir su correcta utilización, proporcionando una experiencia de uso amigable. Además se señalarán las particularidades del diseño GUI sobre el lenguaje matemático MATLAB.

2.1. CARACTERÍSTICAS GUI/UX

El diseño GUI o diseño de Interfaces Gráficas de Usuario se refiere al proceso de creación de herramientas destinadas a comunicar al usuario con una máquina, equipo, computadora o dispositivo, incluyendo todas las relaciones posibles dentro de la comunicación usuario-máquina.

En los últimos tiempos esa definición ha evolucionado para centrarse en lo que se conoce como UX o experiencia de usuario. El diseño centrado en la experiencia implica ciertos requerimientos y buenas prácticas destinadas a conseguir que el sistema o servicio pueda ser utilizado por cualquier usuario de forma efectiva, eficiente y que resulte satisfactoria dentro del contexto específico al que está destinado el sistema. La experiencia de usuario incluye las percepciones y respuestas que se producen como resultado del uso del sistema, ya sea durante la ejecución del mismo o durante el proceso de aprendizaje[18].

El diseño de cualquier sistema destinado al usuario hace necesario tener en cuenta tanto las necesidades presentes como futuras a las que se verá expuesto el programa. Esto implica una primera etapa de captación de requerimientos, sintetizado de los mismos y propuesta de posibles actuaciones futuras. Las principales variables que definen a un sistema como válido competen tanto la lista de requerimientos como las características que aportarán una experiencia de uso adecuada para el usuario, incluyendo flexibilidad, facilidad de uso y aprendizaje, compatibilidad y capacidad global[19].

Habitualmente se toman como referencia una lista de ocho características críticas que sirven como base para el diseño correcto de interfaces centradas en la experiencia de usuario, aunque la evolución de los entornos gráficos y las necesidades asociadas hacen necesario mantener los sistemas continuamente actualizados para asegurar la mejor experiencia.

1. **Claridad:** La claridad es una de las características más importantes dentro del diseño de sistemas destinados al usuario. Los destinatarios de la interfaz deberán ser capaces de entender en todo momento qué está ocurriendo en la ejecución, además de conseguir esa información de manera fácil sin llegar a resultar confusa ni frustrante.
2. **Concisión:** Aunque la claridad es esencial para que el usuario comprenda el funcionamiento del sistema, se debe mantener siempre presente que la saturación de información no es una solución válida. Contrariamente a lo esperado, añadir demasiada información puede llegar a resultar confuso e incluso provocar que el usuario abandone la ejecución del programa debido al tiempo necesario para adquirir toda la información. Un contenido conciso, con la información estrictamente necesaria resultará más agradable que un programa con demasiadas etiquetas, mensajes y avisos.
3. **Familiaridad:** Uno de los principales objetivos del diseño es resultar intuitivo, es decir, que la ejecución del sistema resulte natural para los usuarios. La mejor manera de conseguir este propósito es hacerlo familiar, hasta tal punto que resulte reconocible para el usuario incluso en su primera ejecución, dando la sensación de ya haberse utilizado antes. Para ello es necesario que cada acción sea seguida por una reacción esperable, como por ejemplo la conocida interfaz de apertura de ficheros (figura 2.1) o los mensajes de alerta, presentes en cualquier sistema operativo.

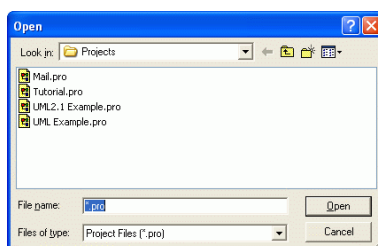


Figura 2.1: Ejemplo de entorno de apertura de ficheros

4. **Receptividad:** El diseño receptivo o *responsive*, junto con la familiaridad, tiene gran importancia en cuanto a las sensaciones experimentadas por el usuario. Un programa *responsive* implica que la experiencia sea lo suficientemente fluida y rápida, referido a las transiciones entre pantallas, la aparición de información o la respuesta a acciones del usuario. Independientemente de los tiempos de ejecución propios del programa en los que el diseñador no tiene responsabilidad, es importante que la interfaz responda a cada acción del usuario de manera rápida, evitando que el usuario desconozca si ha completado o no la acción requerida. De esta manera es imprescindible que cada acción por parte del usuario desemboque en una reacción visible del sistema.
5. **Consistencia:** La cualidad de consistencia se refiere a la importancia de mantener un contexto común a lo largo de la ejecución del programa. En todo momento se

deben reconocer los elementos como propios, sin incluir ningún elemento que se pudiese identificar como externo al programa. Esto permite al usuario aprender conforme avanza en la ejecución e incluso llega a prever acciones en base a situaciones pasadas.

6. **Atractivo:** Dentro de las posibilidades del programa, la interfaz debe atraer a los usuarios. Esto no implica la inclusión de figuras, colores, etc. por el simple hecho de mejorar estéticamente el programa, sino hacer una estructuración lógica y equilibrada de los objetos dentro del espacio disponible, mostrar los mensajes de manera consistente con tamaños de letra adecuados y mantener la sobriedad para adecuar el programa a toda clase de contextos y ámbitos de ejecución.
7. **Eficiencia:** A diferencia de la velocidad y fluidez de la que se habló en el apartado de la receptividad, la eficiencia se centra en la facilidad que tienen los usuarios para efectuar los procedimientos para los que está creada la interfaz. Para ello es necesario reducir el número de acciones requeridas para cada procedimiento y añadir funcionalidades y atajos que faciliten la rápida ejecución de las tareas. No se debe olvidar nunca el criterio de concisión, de manera que saturar el programa de funcionalidades hará que sea menos eficiente y más confuso para el usuario.
8. **Indulgencia:** La cualidad de la indulgencia se refiere a la capacidad del programa para perdonar errores al usuario. Cualquier acción cometida durante la ejecución debe poder ser rectificadas, sin penalizar al usuario por las acciones que tome. Durante el diseño se deben tener en cuenta las posibles acciones que el usuario puede realizar en cualquier momento, permitiendo prever los posibles errores y corrigiéndolos sin penalizar al usuario.

Tomando las anteriores características como una referencia, la mejor práctica posible debe contener un compromiso o balance entre todas ellas y su consiguiente tiempo de desarrollo, factor muy relevante en la mayor parte de los proyectos. Durante el desarrollo se deben tener en cuenta las necesidades funcionales tomándolas como referencia principal, sin descuidar en ningún momento las buenas prácticas mencionadas anteriormente que permiten constituir una interfaz amigable de cara al usuario, consistente en toda su extensión y adaptable y eficiente a las necesidades de cada usuario.

2.2. DISEÑO DE INTERFACES EN MATLAB: HERRAMIENTA GUIDE

El entorno de programación MATLAB (*MATrix LABoratory*) es bien conocido en entornos educativos y profesionales por sus capacidades de computación matemática, su eficiencia en cálculos matriciales y la cantidad de herramientas de las que dispone para su utilización en diversos campos como el procesado de imágenes y audio, control de sistemas, gestión de bases de datos y modelado. Una de esas herramientas es la conocida como GUIDE, una funcionalidad básica que permite la creación de interfaces para el control interactivo de los scripts de MATLAB.

El uso de esta herramienta resulta de gran utilidad para adaptar la enorme capacidad computacional del lenguaje a un entorno más amigable, de cara a permitir que cualquier usuario pueda trabajar con MATLAB sin tener conocimientos de programación. Para la creación de nuevas interfaces con Guide es necesario crear la propia interfaz con la herramienta visual y programar las funcionalidades en el script asociado a dicha interfaz.

2.2.1. DISEÑO GRÁFICO DE LA INTERFAZ

Para la creación del apartado gráfico del sistema, MATLAB dispone de una utilidad simple e intuitiva en la que se pueden añadir y modificar los diferentes elementos que conforman la interfaz gráfica del programa a desarrollar, tales como botones, textos, listas, gráficas, etc. La figura 2.2 muestra la herramienta en su estado inicial con una plantilla vacía.

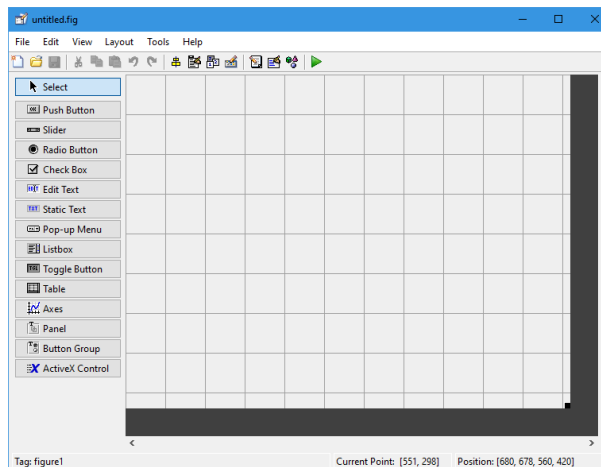


Figura 2.2: Herramienta GUIDE para la creación de interfaces

Es posible crear diferentes elementos dentro de la pantalla, adaptarlos al espacio disponible y alinearlos de la forma requerida. Gracias a la utilización de pestañas y su superposición es posible crear múltiples pantallas dentro del mismo programa, aumentando el espacio disponible de forma ilimitada.

Los elementos añadidos disponen de diversos parámetros de configuración, tales como los mostrados en la figura 2.3 correspondientes a la modificación de un botón. Entre estos parámetros cabe destacar la posición, el valor, la habilitación y la referencia, esta última indispensable en el apartado de la programación.

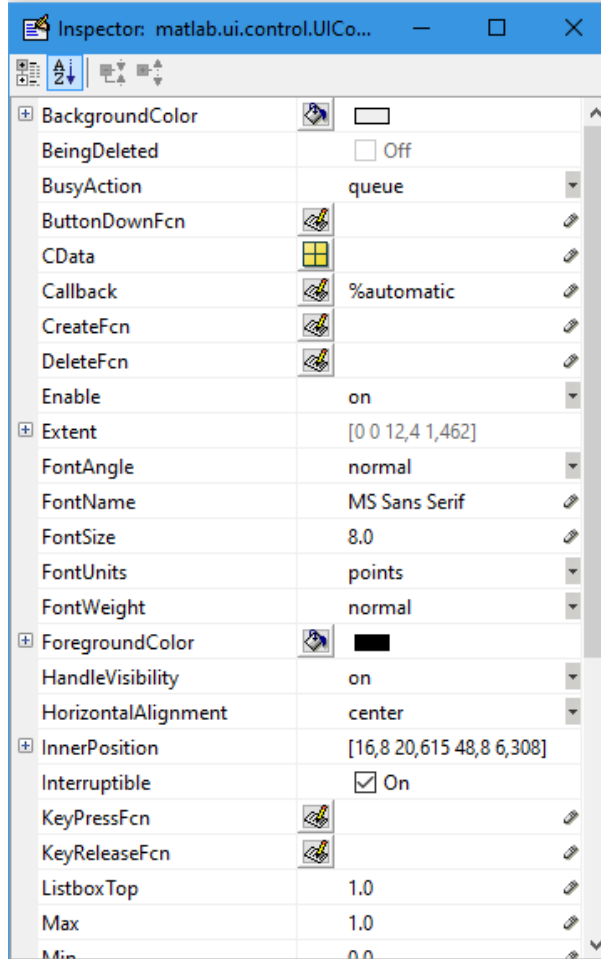


Figura 2.3: Parámetros disponibles para los elementos *button*

Los elementos disponen de parámetros comunes entre sí, pero en cada uno de ellos aparecen diferentes particularidades y configuraciones especializadas, como el campo que permite a las listas de objetos disponer de varios elementos seleccionados al mismo tiempo o la disposición de selectores o botones agrupados para el funcionamiento como radio botones, de forma que solo se permita activar una de las posibilidades del grupo cada vez.

Como parte de la creación de la interfaz gráfica se incluye la configuración del comportamiento de la ventana del programa en cuanto a su redimensionado o el tamaño inicial que tendrá al abrirse, detalles que se deberán tener en cuenta para poder obtener una experiencia agradable por parte de usuarios de diferentes plataformas.

2.2.2. PROGRAMACIÓN DE LA INTERFAZ

Como ya se dijo anteriormente, la creación de una interfaz gráfica de usuario en MATLAB implica tanto su diseño gráfico como la programación de sus funcionalidades.

Una vez creado el apartado gráfico, se encuentra que cada elemento añadido está emparejado a uno o varios métodos conocidos como *callbacks*, los cuales se corresponden con subrutinas que se ejecutan cuando se realiza una determinada acción sobre el elemento al que se refieren. Para ello la propia herramienta de creación proporciona una plantilla con las funciones esenciales de cada elemento creado en la parte gráfica, como por ejemplo las rutinas que se ejecutan al pulsar botones o seleccionar elementos de una lista. Cada objeto dispone de varios *callbacks* propios referidos a su comportamiento particular, y es posible modificarlos permitiendo añadir funcionalidades según las necesidades.

También se incluye en el anterior grupo a aquellos métodos que se ejecutan en el arranque y cierre del programa. Estas subrutinas, ya sean referidas al programa completo o a cada elemento, permiten controlar la apariencia y el valor que presentan los objetos de la interfaz por defecto, esto es, justo después de abrir el programa así como la lista de acciones a realizar al efectuarse un cierre del mismo. Estos métodos permiten deshabilitar ciertas partes de la interfaz para impedir su uso hasta que, por ejemplo, se cargue un fichero y los datos a procesar estén disponibles, designar las ventanas de la interfaz que aparecen en primer lugar, dejando las demás ocultas o ejecutar una interfaz externa al acabar la ejecución de la que se encuentra activa, pudiendo pasar de un programa a otro de manera transparente para el usuario.

Sin entrar demasiado en la parte interna de la programación GUI de MATLAB, cabe destacar que los métodos de cada elemento gráfico son independientes entre sí, lo que implica el paso continuo de parámetros y datos entre cada uno de ellos para poder trabajar con un entorno completo. Para ello existe una estructura de datos conocida como *handles*, una variable tipo *struct* que contiene tanto los valores y parámetros de cada elemento gráfico como los parámetros del sistema completo, y que permite además almacenar en ella los datos con los que se trabaja durante la ejecución del programa. Accediendo a cada elemento es posible consultar su estado, incluyendo su valor, su posición, sus elementos actuales, etc.

El paso de la estructura de valores y elementos se realiza en la llamada de cada función, actuando como entrada. Cada método de la interfaz dispone de tres elementos básicos de entrada: la estructura *handles*, la variable *eventdata* que contiene la información de las acciones que realiza el usuario con el ratón o el teclado y la componente *hObject*, la cual es encargada de realizar internamente el disparo de cada *callback*. En aquellos casos en los que se modifique cualquier valor de la estructura se deberá incluir la sentencia *guidata(hObject,handles)* que permite actualizar los valores modificados en la interfaz. Al crear nuevos métodos además de los propios de cada elemento, será necesario pasar como argumento la estructura *handles* para poder acceder a los valores de sus elementos.

La asignación u obtención de valores de los elementos de la interfaz se debe realizar con sentencias **get** y **set**, dependiendo de si se quiere obtener el valor o si se quiere modificar. Estos métodos necesitan como entrada la referencia del elemento al que se refieren, el nombre del parámetro con el que se quiere trabajar y, en caso de modificación, el nuevo valor a asignar.

Por ejemplo, la sentencia `set(handles.botonGenerar,'enable','off')` deshabilita el botón con referencia 'botonGenerar' al introducir el valor 'off' en su parámetro 'enable'. Como ejemplo de consulta de valores, la sentencia `get(handles.textSaveCasos,'Position')` obtiene el valor de posición del elemento con referencia 'textSaveCasos'.

Gracias a la posibilidad de cambiar los parámetros de los elementos dinámicamente a lo largo de la ejecución, es posible mostrar u ocultar partes de la interfaz en función del estado de ejecución, así como disponer de diferentes pestañas haciendo visibles las partes requeridas y ocultando las no necesarias. De igual manera se permite bloquear los elementos que no deben ser manipulados por el usuario para evitar fallos de ejecución, pudiéndose habilitar en el momento en que sean necesarios.

32158->desestacionalizado CTM rho = 17 diezmados = 60s intervalos = dia canales = z	SVM_Q	55	PCA	0.72803
45790->desestacionalizado CTM rho = 78 diezmados = 30s intervalos = dia canales = z				
53580->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = y				
32382->desestacionalizado CTM rho = 18 diezmados = 60s intervalos = dia canales = z	SVM	76	PCA	0.72652
45790->desestacionalizado CTM rho = 78 diezmados = 30s intervalos = dia canales = z				
53580->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = y				
32158->desestacionalizado CTM rho = 17 diezmados = 60s intervalos = dia canales = z				
45118->desestacionalizado CTM rho = 75 diezmados = 30s intervalos = dia canales = z	SVM	52	PCA	0.72348
53582->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = comp				
32382->desestacionalizado CTM rho = 18 diezmados = 60s intervalos = dia canales = z				
45118->desestacionalizado CTM rho = 75 diezmados = 30s intervalos = dia canales = z	SVM	73	PCA	0.72197
53582->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = comp				

Figura 2.4: Múltiples líneas dentro del mismo elemento *tabla*

Como último detalle cabe mencionar que las posibilidades de configuración de parámetros son muy limitadas en cuanto a la personalización de los elementos, como por ejemplo en el caso de las tablas gráficas o los elementos emergentes. Ante la necesidad de disponer de opciones avanzadas en varios elementos, se descubrió la posibilidad de extraer dichos objetos directamente del código interno para obtener la referencia al objeto JAVA asociado con el que trabaja internamente el entorno MATLAB. De esta manera se pueden obtener innumerables posibilidades de personalización de cada elemento, tantas como las que disponga su homólogo en JAVA. Por ejemplo, se utilizó este método ante la necesidad de disponer de campos multilínea en los objetos *tabla* (figura 2.4), cuya resolución solo es posible con la modificación del objeto JAVA asociado al homólogo *uitable* de MATLAB.

CAPÍTULO 3

BASES DEL ESTUDIO

En este capítulo se explicarán las bases de trabajo, las cuales forman parte del entorno desarrollado. Se detallarán las diferentes secciones del entorno de procesado, lo que permitirá obtener una idea general de la composición del programa final. Partiendo del procesado de los ficheros de datos, se explicarán los diferentes parámetros obtenibles a partir de ellos y cómo se consigue determinar su efectividad en el diagnóstico del trastorno. Por último se presentará el método utilizado para obtener una cuantificación de dicha efectividad a partir de los registros disponibles.

3.1. REGISTROS ACTIGRÁFICOS. VALIDACIÓN Y PREPROCESADO.

Como ya se mencionó en el capítulo 1, el estudio se realiza en base a una serie de registros actigráficos convenientemente clasificados como pacientes sanos y pacientes con TDAH (en adelante, «controles» y «casos»).

La información actigráfica pertenece a 74 pacientes que padecen TDAH y 75 pacientes de los que no se tiene constancia de que padezcan hiperactividad ni ningún trastorno mental, neurológico o de sueño, los cuales actuarán como individuos de control. Dichos registros se obtuvieron gracias a los pediatras y psicólogos del Complejo Asistencial Universitario y del Centro de Salud Jardinillos, ambos ubicados en la ciudad de Palencia, España. Los pacientes analizados comprenden edades entre 6 y 15 años e incluyen tanto niñas como niños.

Los registros (uno por paciente) consisten en ficheros de datos obtenidos directamente del aparato actigráfico, en este caso, pulseras cuantificadoras de la marca Actigraph [20] colocadas en la mano dominante del sujeto y con las que se toman datos durante un tiempo de 24 horas. Cada registro incluye datos de aceleración de cada uno de los ejes cartesianos (x y z) obtenidos a razón de una muestra por segundo, además de podómetro e inclinómetro.

Se utilizan los datos de aceleración de los 3 canales y se añade un canal complementario calculado como la actividad total ($comp = \sqrt{x^2 + y^2 + z^2}$).

Dichos registros, en estado *crudo*, necesitan ser sometidos a un procesamiento preliminar que descarte aquellos no válidos y los adecúe a las necesidades de los procesos posteriores. La primera de estas tareas es descartar los registros no válidos o erróneos. Los registros inválidos son aquellos que cumplen alguna de estas condiciones:

- Los ficheros cuyos registros contienen en su mayoría valores de saturación (valores por encima del límite máximo de lectura del dispositivo captador).
- Registros con valores no válidos por ser no numéricos (*NaN*).
- Registros con valores extremos fuera de los valores habituales (*outliers*).
- Registros con longitud insuficiente, esto es, con registros de menos de 22 horas.
- Registros con formato erróneo o no soportado.

Los registros válidos se someten entonces al procesamiento inicial, que incluye la detección de los intervalos de sueño, la fragmentación de los registros en los diferentes intervalos del día y el diezmado de las muestras.

3.1.1. DETECCIÓN AUTOMÁTICA DE LOS INTERVALOS DE SUEÑO

La separación de los registros en sueño-vigilia permite trabajar con determinados parámetros relacionados con cada una de las etapas del día y extraer características aisladas de cada uno de esos intervalos. El método utilizado, desarrollado en [13], emplea el canal de datos complementario asignado como actividad total para determinar el intervalo correspondiente al sueño, pudiendo de esta manera dividir las muestras de cada registro en cada uno de los dos intervalos de sueño o de vigilia. Esta detección automática se basa en un suavizado de la señal seguido de la comparación con un umbral, llevándose a cabo la selección de dicho umbral de forma iterativa.

Esta técnica permite prescindir de la metodología clásica que requería información complementaria por parte de los tutores de los pacientes, además de conseguir una división completamente objetiva de los intervalos.

3.1.2. FRAGMENTADO DE LOS REGISTROS

Una vez obtenida la máscara discriminadora de los intervalos sueño-vigilia es posible subdividir las muestras en diferentes segmentos a lo largo de la longitud total del día. En concreto se obtienen fragmentos correspondientes al día completo, vigila, vigilia matutina, vigilia vespertina, intervalo de sueño completo, tramo inicial de sueño, tramo de sueño medio y tramo de sueño final.

Esta división en diferentes tramos permite obtener características discriminantes relativas a cada etapa del día, diferentes de las que se conseguirían con los registros completos de 24 horas. Se añade además la posibilidad de obtener las características particulares de cada etapa de sueño.

3.1.3. DIEZMADO

Los registros incluyen lecturas de datos a razón de una muestra por segundo. El proceso de diezmado reduce el número de valores leídos por unidad de tiempo de manera que es posible capturar actividades e intervalos de diferente longitud, como por ejemplo las diferencias de actividad en periodos cortos de tiempo de 1 segundo ó 1 minuto, tareas más largas de entre 5 y 30 minutos o espacios más largos de hasta varias horas. El diezmado también permite aplicar un compromiso entre la gran cantidad de información originalmente registrada y los datos que se quieren extraer, adecuando la señal para su extracción posterior.

3.2. REGISTROS ACTIGRÁFICOS. EXTRACCIÓN DE CARACTERÍSTICAS.

Los registros actigráficos se componen de una sucesión de valores correspondientes a las aceleraciones instantáneas registradas por el actígrafo. A partir de dichos valores es necesario obtener diferentes características que actuarán como identificadores de la actividad registrada.

Existe una variedad infinita de posibilidades en cuanto a las características extraíbles y es indispensable disponer de un gran número de ellas para poder encontrar aquellas que más se adecúen a las necesidades de detección del sistema.

3.2.1. CARACTERÍSTICAS ESTADÍSTICAS DE LA SEÑAL

Los parámetros estadísticos de la señal utilizados son los encontrables en cualquier referencia de estadística [21, p. 219], tales como la media, mediana, rango intercuartil, asimetría estadística, curtosis y coeficiente de variación, este último obtenido como el cociente entre la desviación típica y la media.

3.2.2. CARACTERÍSTICAS NO LINEALES DE LA SEÑAL

Dentro de los parámetros no lineales se incluyen la medida de tendencia central, la dinámica simbólica, la complejidad binaria de Lempel Ziv, y la entropía aproximada, cuyos cálculos [17] se detallan brevemente a continuación.

- El parámetro CTM, o medida de tendencia central, se calcula a partir de las diferencias de primer orden de tres muestras consecutivas de señal. Para ello se determina un parámetro ρ que actúa como radio de circunferencia y se calcula el valor CTM como la proporción de muestras que caen en su interior, entendiendo como muestras a la representación del diagrama de dispersión que representa la serie de diferencias de primer orden $x[n+2] - x[n+1]$ frente a $x[n+1] - x[n]$. De esta manera se obtiene que las señales regulares tendrán un valor alto de CTM, mientras que las señales que presenten una alta variabilidad tendrán menores valores de CTM.
- El parámetro DS, o dinámica simbólica, se calcula en base a la frecuencia de cambio de la señal entre periodos de diferente actividad. Para ello se definen 2 o más niveles de actividad (3 en este caso) conocidos como símbolos, con los que construir palabras mediante la unión de dos o más de estos símbolos (se utilizan palabras de 2 símbolos de longitud) de forma que cada una de esas palabras se identifique con el cambio entre dos o más niveles de actividad, dependiendo de la longitud elegida. De esta manera se pueden obtener las probabilidades de aparición de dichas palabras (se utilizan $3^2 = 9$ palabras diferentes correspondientes a las formas de ordenar los 3 símbolos de 2 en 2)

y calcular la entropía de Shannon [22] de dicho alfabeto. El parámetro calculado será mayor cuanto mayor sea la variabilidad en la señal.

- La complejidad binaria de Lempel-Ziv se basa en la conversión de la señal analizada en una secuencia binaria a partir de un umbral preestablecido. Esta secuencia se examina de izquierda a derecha y se incrementa el contador de complejidad en una unidad cada vez que se encuentra una nueva secuencia de caracteres consecutivos. Este contador refleja la aparición de nuevos patrones dentro de la señal, y será mayor cuanto mayor variabilidad presente la señal.
- La entropía aproximada (ApEn) se basa en la aparición de patrones, al igual que el parámetro anterior. En ese caso se trata de una medida logarítmica que refleja la regularidad de señales discretas sin tener en cuenta el modelo de sistema que las generó. Para su obtención [23] se mide la frecuencia de aparición de patrones similares pertenecientes a las señal original y se calcula su algoritmo promediado. Esta operación se repite para diferentes valores de longitud de los patrones buscados y se calcula la diferencia entre los resultados obtenidos. De igual forma que en otros parámetros, la entropía aproximada obtiene valores mayores cuanto mayor variabilidad de señal exista, mientras que valores cercanos a cero se atribuyen a señales regulares con muchos patrones repetidos.

3.2.3. CARACTERÍSTICAS DEL RITMO CIRCADIANO

Para el cálculo de los parámetros del ritmo circadiano se han utilizado diferentes métodos, tanto los recogidos en [24] como en [4]:

- **Parámetros clásicos:** Se realiza un ajuste de mínimos cuadrados sobre la secuencia original en el dominio del tiempo de una senoide, obteniéndose los tres parámetros correspondientes a este tipo de señal: amplitud, fase y frecuencia.
- **Variabilidad intradiaria (IV):** Este parámetro estudia la frecuencia y extensión de las transiciones entre periodos de valores altos y bajos de actividad dentro del día, cuantificando la fragmentación del ritmo. Para su calculo se promedian los cuadrados de la diferencia de actividad entre muestras distanciadas 1 hora y se divide entre la varianza global de la señal.
- Valor **M5/M10:** Los parámetros M5 y M10 se refieren al máximo valor medio de actividad durante un intervalo de 5 ó 10 horas a lo largo del día. Para su cálculo se procede a obtener los diferentes intervalos de 5 ó 10 horas que existen en el periodo de 24 horas y se determina cuál de ellos posee una media mayor.
- Valor **L5/L10:** Al contrario que los parámetro anteriores, L5 y L10 se refieren a los valores mínimos de media en los intervalos de 5 ó 10 horas. Su cálculo es equivalente al de los parámetros antes comentados.
- Amplitud relativa **RA:** Este parámetro mide, por un lado, la diferencia entre M5 y L10 y, por otro, la diferencia entre M10 y L5.

- Índice dicotómico **IO** (In bed - Out of bed): El índice dicotómico mide la concordancia entre el sueño nocturno de los niños y el periodo de sueño estándar, establecido entre las 23:00 y las 7:00. De esta manera se obtiene un valor comprendido entre 0 y 100 % en función de la similitud entre el sueño medido y el sueño estándar de los progenitores establecido según mediciones empíricas entre los valores antes mencionados.

3.2.4. PARÁMETROS DEL SUEÑO

Los parámetros de puntuación del sueño o SSA permiten estudiar las diferentes etapas de sueño en cuanto a duración, nivel de actividad y diferencia con los intervalos anterior y posteriores a dichas etapas. Diferenciando la probabilidad de encontrarse en la etapa de sueño silencioso o PQS (Probability of Quiet Sleep), en la etapa de sueño activo o PAS (Probability of Active Sleep) y en la etapa de vigila o PAW (Probability of Active Waking), el método Sadeh-Acebo estima los diferentes parámetros en base a la resolución de un sistema lineal construido a partir de observaciones empíricas. En [25] se puede observar con más detenimiento el desarrollo empleado.

PARÁMETROS ESPECTRALES

Dentro de los posibles parámetros espectrales se ha trabajado con la dimensión fractal de Maragos-Sun [26], una técnica que consiste en la medición de la presencia de irregularidades en un entorno de la frontera de un fractal. Entendiendo como fractal aquellos elementos o estructuras aparentemente irregulares que se repiten a diferentes escalas, este método analiza el área que cubren a lo largo de la señal y determina un ajuste de mínimos cuadrados cuya pendiente se tomará como resultado.

3.3. DESESTACIONALIZACIÓN

Una vez obtenidas las diferentes características de la señal, se realiza un procesado adicional sobre todos ellos conocido como desestacionalización.

Este proceso permite eliminar el efecto de las variaciones cíclicas estacionales que los parámetros experimentan a lo largo del año, mediante la aplicación de cierto procedimiento estadístico. Esto permite eliminar la dependencia existente entre los parámetros de actividad de los pacientes y la fecha del año en que han sido tomados.

El proceso se basa en la premisa de que tanto la media como la desviación típica de las señales estudiadas siguen un ritmo circanual, por tanto es posible realizar un ajuste lineal obteniendo la media y desviación de cada mes del año y aplicándola sobre los valores relativos a dicho periodo. Para ello se resta la media a los valores correspondientes y se divide entre la desviación típica, obteniéndose una distribución de parámetros equivalente a la original pero sin la influencia estacional.

3.4. SELECCIÓN DE CARACTERÍSTICAS

Ante el gran número de características obtenidas en los procesos anteriores, aparece la necesidad de determinar cuales de ellas tienen posibilidades de ser efectivas en la detección del trastorno. Con ello se consigue reducir considerablemente el número de características posibles y se facilita el trabajo sucesivo, trabajando solamente con aquellos parámetros con poder discriminante.

El método empleado es la prueba de p-valor de Wilcoxon, tal y como se expone en [21]. En ella se obtienen los p-valores de cada parámetro en relación a los dos tipos de individuos (sanos y con TDAH), permitiendo cuantificar la verosimilitud entre el parámetro de un tipo de individuo y el otro. Se establece que la característica es discriminante si su p-valor es inferior a 0.05, ya que este valor será menor cuanto mayor diferencia exista entre las muestras comparadas.

Debido al pequeño tamaño muestral disponible resulta necesario realizar un proceso de aleatorización durante la realización de esta prueba. Para ello se realizan múltiples iteraciones del método de Wilcoxon utilizando en cada una de ellas los parámetros obtenidos de un subconjunto aleatorio de casos y controles, por lo que para cada iteración de cada característica se utilizan orígenes de datos diferentes. Por tanto las características tendrán que superar el test en pruebas sucesivas, haciendo que sea más probable encontrar aquellas con alto poder discriminante después de la corrección con comparaciones múltiples.

Como consecuencia de la gran restrictividad de esta prueba se plantea la posibilidad de reducir el número total de iteraciones a superar por cada característica, imponiendo un porcentaje de corrección del total. Así se evita la eliminación de una característica por el resultado de una sola iteración, obteniéndose un resultado más abierto.

3.5. ANÁLISIS DE PRESTACIONES DE LOS CLASIFICADORES. VALIDACIÓN CRUZADA

3.5.1. INTRODUCCIÓN A LOS SISTEMAS CLASIFICADORES

La utilidad final de las características es la de entrenar, con datos de ambos tipos de individuos, un determinado sistema que después sea capaz de diferenciar las muestras de cada origen. Así se consigue que una vez configurado el sistema (conocido como clasificador) éste sea capaz de determinar la naturaleza de paciente sano o paciente con TDAH a partir de los datos actigráficos procesados provenientes de un paciente sin clasificar.

De manera esquemática, los clasificadores actuarán como *cajas negras*, las cuales requerirán de dos acciones: entrenamiento y prueba.

- El entrenamiento consiste en la introducción de datos para la configuración del sistema clasificador. Estos datos deberán provenir de los dos orígenes a discriminar y deberán estar convenientemente etiquetados para que el sistema sea capaz de diferenciarlos. En el contexto tratado, se aportarán al clasificador características de pacientes sanos y pacientes con TDAH. Después de este proceso se obtendrá un clasificador entrenado.
- La prueba o test se refiere a la acción final a desempeñar por el sistema. A partir de un clasificador adecuadamente entrenado, la introducción de datos determinantes provenientes de un origen sin clasificar provocará una salida binaria cuyo valor determina el tipo de dato analizado. En este caso, ante la introducción de características de un paciente, el sistema detectará si tiene o no el trastorno.

3.5.2. TIPOS DE CLASIFICADORES UTILIZADOS

Existen numerosas posibilidades en cuanto a sistemas clasificadores y el uso de uno u otro depende fundamentalmente de la naturaleza de los datos a analizar y de las necesidades de selección requeridas. En el presente estudio se ha trabajado con los sistemas de predicción K Nearest Neighbors, Support Vector Machine, Decision Tree y Random Forest, además de una aproximación a las Redes Neuronales [27, 28].

- K Nearest Neighbors (KNN): El método de los K vecinos más cercanos o KNN se basa en encontrar la distribución más cercana (parecida) a la de referencia a partir de los datos de entrada. Ante una entrada de datos, el sistema examina los K vecinos más cercanos y decide la salida como la etiqueta mayoritaria en esos elementos.
- Support Vector Machine (SVM): Las máquinas de soporte de vectores o SVM se basan en la representación de las características de entrada en un plano cuya dimensión es igual a la dimensión de los datos aportados (número de características), de forma que se pueda hallar un hiperplano (un subespacio con una dimensión menos que el espacio

continente) que sea capaz de separar los puntos de las dos regiones a diferenciar. Para ello hacen uso de funciones conocidas como *kernels*, los cuales se pueden configurar en función de las necesidades requeridas. Se han utilizado *kernels* cuadráticos o **quad**, *radial basis function* o **rbf** y funciones lineales o **linear**.

- Árboles de decisión: Los árboles de decisión son sistemas multietapa organizados en una estructura con forma de árbol, de ahí su nombre. El sistema se divide en nodos de los que surgen ramas, cada una de las cuales asociada a una posible decisión de salida desde el nodo anterior. A partir de una entrada unificada se realiza el paso de nodo a nodo mediante la toma de decisiones hasta llegar a las hojas, las cuales actúan como salidas. En cada nodo se realiza una decisión binaria basada en diferentes parámetros, los cuales se eligen en función de los datos introducidos. En este estudio se han realizado pruebas tanto con árboles de decisión clásicos como con Random Forest (**RF**), una visión más completa de este tipo de sistemas. Los bosques aleatorios utilizan árboles de decisión individuales, a los que se les aporta una porción aleatoria de la información total que se dispone como entrada. El resultado general se determina en función de las decisiones de cada uno de los árboles empleados, eligiéndose la opción más 'votada'.
- Redes Neuronales: Las redes neuronales intentan imitar el comportamiento de las estructuras biológicas de las neuronas. Para ello, el sistema se divide en múltiples partes o neuronas, cada una de las cuales se encarga de procesar parte de la información. Este procesado intenta simular la respuesta natural de las neuronas ante diferentes estímulos de su entorno, y se adapta dinámicamente al medio circundante cambiando los pesos que asigna a cada estímulo o entrada.

3.5.3. DIMENSIÓN DE LOS CLASIFICADORES

En lo referente a los datos a aportar al sistema clasificador, existe la posibilidad de utilizar una o más características diferentes como entrada. La elección de más de una característica supondrá utilizar varios parámetros de cada individuo tanto en la fase de entrenamiento como en la fase de prueba, y a priori cabe esperar que los resultados obtenidos sean mejores que con el uso de una sola característica diferenciadora. De acuerdo con los resultados obtenidos en [17] el uso de entradas de tamaño mayor que dos no implica una mejora sustancial de los resultados a pesar del gran aumento de carga computacional que implica su procesamiento, por lo que se decide utilizar clasificadores bidimensionales en las pruebas realizadas en el presente estudio. Sin embargo, el tamaño de los clasificadores es uno de tantos parámetros configurables dentro del sistema durante su ejecución.

3.5.4. VALIDACIÓN CRUZADA

Una vez introducidos los métodos de clasificación, resulta necesario explicar el proceso seguido para evaluar el desempeño de los diferentes conjuntos clasificadores.

Tomando como sistema clasificador al conjunto formado por las características diferenciadoras introducidas (una o varias) así como el tipo de clasificador y sus parámetros, es fácil apreciar la cantidad de diferentes combinaciones posibles a tener en cuenta. La primera reducción de dicha cantidad se realizó en el proceso anterior, en el cual se seleccionaron aquellas características con mayor poder diferenciador de pacientes, aunque el número total sigue siendo muy elevado. Para poder cuantificar las prestaciones de los mismos será necesario probarlos, dicho de otra manera, entrenarlos con ciertos datos de entrada y comprobar el número de aciertos ante una prueba con individuos previamente clasificados. Este proceso es el conocido como validación cruzada.

Para cada una de las agrupaciones o sistemas clasificadores se realizará la prueba de validación cruzada y se cuantificará su validez. La prueba consiste en realizar diversas iteraciones de test-train (entrenamiento-prueba) sobre los clasificadores utilizando en cada repetición diferentes conjuntos de individuos del total de población disponible.

Dado un sistema clasificador, es decir, un tipo de clasificador con sus correspondientes parámetros y una o varias características que se tomarán como entradas, se elige un subconjunto de casos y controles para el entrenamiento del clasificador y se comprueba su funcionamiento con el resto de individuos adecuadamente etiquetados, lo que permite calcular el número de aciertos y errores cometidos. En cada iteración se elegirán aleatoriamente los subconjuntos a utilizar haciendo que el tamaño muestral efectivo aumente y consiguiendo resultados similares a los que se encontrarían en entornos reales.

El desempeño del clasificador en cada iteración se calcula en términos de precisión, sensibilidad y especificidad:

- **Precisión:** Se calcula como la proporción de aciertos entre el total de individuos clasificados.

$$Precision = \frac{Aciertos}{TotalIndividuos}$$

- **Sensibilidad:** Se calcula como la proporción de pacientes con TDAH correctamente identificados entre el número total de enfermos.

$$Sensibilidad = \frac{VerdaderosPositivos}{TotalCasos}$$

- **Especificidad:** Se calcula como la proporción de pacientes sanos correctamente identificados entre el número total de pacientes sin el trastorno.

$$Especificidad = \frac{VerdaderosNegativos}{TotalControles}$$

3.5. ANÁLISIS DE PRESTACIONES DE LOS CLASIFICADORES. VALIDACIÓN CRUZADA

Con los valores anteriores se calcula la media y mediana del conjunto de iteraciones para obtener valores manejables que distingan los resultados de cada sistema. La figura 3.1 muestra un ejemplo de resultados en donde se aprecian los valores mencionados anteriormente, así como las particularidades de cada sistema clasificador asociado a ellos.

Grupo de Características		Guardar clasificadores							
		Precisión		Sensibilidad		Especificidad			
		MEAN	MEDIAN	MEAN	MEDIAN	MEAN	MEDIAN		
10779->estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM	9819	PCA	0.75076	0.75	0.77353	0.79412	0.72656	0.71875
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	8805	PCA	0.74773	0.74242	0.79706	0.79412	0.69531	0.71875
10331->estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM	8802	PCA	0.74621	0.75758	0.78088	0.76471	0.70937	0.71875
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM_Q	8802	PCA	0.74545	0.74242	0.75441	0.76471	0.73584	0.73438
10331->estacionalizado CTM rho = 36 diezmados = 30s intervalos = dia canales = z	SVM_Q	9819	PCA	0.74545	0.74242	0.73676	0.75	0.75469	0.71875
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	Arbol	19665	PCA	0.7447	0.75	0.72059	0.69118	0.77031	0.75
33470->desestacionalizado CTM rho = 23 diezmados = 30s intervalos = dia canales = z	SVM_RBF	8802	PCA	0.74242	0.75	0.76876	0.79412	0.69531	0.70313
53581->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = z	SVM_RBF	9819	PCA	0.74242	0.75	0.72059	0.73529	0.76583	0.75
10331->estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM_RBF	10325	PCA	0.74015	0.74242	0.75882	0.76471	0.72031	0.71875
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	10822	PCA	0.73939	0.74242	0.76176	0.73529	0.71582	0.70313
11003->estacionalizado CTM rho = 43 diezmados = 30s intervalos = dia canales = z	SVM	10325	PCA	0.73788	0.74242	0.75294	0.73529	0.72188	0.70313
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	5681	PCA	0.73712	0.73485	0.78382	0.79412	0.6875	0.70313
11227->estacionalizado CTM rho = 44 diezmados = 30s intervalos = dia canales = z	SVM	10325	PCA	0.73788	0.74242	0.75294	0.73529	0.72188	0.70313
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	10325	PCA	0.73788	0.74242	0.75294	0.73529	0.72188	0.70313
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	5681	PCA	0.73712	0.73485	0.78382	0.79412	0.6875	0.70313
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	10325	PCA	0.73788	0.74242	0.75294	0.73529	0.72188	0.70313
11003->estacionalizado CTM rho = 43 diezmados = 30s intervalos = dia canales = z	SVM	10325	PCA	0.73788	0.74242	0.75294	0.73529	0.72188	0.70313
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	5681	PCA	0.73712	0.73485	0.78382	0.79412	0.6875	0.70313
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	5681	PCA	0.73712	0.73485	0.78382	0.79412	0.6875	0.70313
98150->desestacionalizado CTM rho = 35 diezmados = 30s intervalos = dia canales = z	SVM	1707	PCA	0.73712	0.74242	0.70441	0.72059	0.77187	0.76563
6747->estacionalizado CTM rho = 24 diezmados = 30s intervalos = dia canales = z	Arbol	1707	PCA	0.73712	0.74242	0.70441	0.72059	0.77187	0.76563
53581->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = z	SVM	3473	PCA	0.73636	0.73485	0.73971	0.73529	0.73281	0.75
7898->estacionalizado CTM rho = 29 diezmados = 60s intervalos = dia canales = z	SVM	10331	PCA	0.73409	0.72727	0.77647	0.77941	0.68906	0.6875
54946->desestacionalizado CTM rho = 39 diezmados = 60s intervalos = dia canales = z	SVM	12454	PCA	0.73333	0.72727	0.84118	0.85294	0.61875	0.625
11003->estacionalizado CTM rho = 43 diezmados = 30s intervalos = dia canales = z	SVM	10331	PCA	0.73409	0.72727	0.77647	0.77941	0.68906	0.6875
37054->desestacionalizado CTM rho = 41 diezmados = 30s intervalos = dia canales = z	SVM	10331	PCA	0.73409	0.72727	0.77647	0.77941	0.68906	0.6875
11099->estacionalizado CTM rho = 47 diezmados = 30s intervalos = dia canales = z	SVM	12454	PCA	0.73333	0.72727	0.84118	0.85294	0.61875	0.625
53905->desestacionalizado Percentiles intervalos nulos percentiles = 75% diezmados = 300s intervalos = vigilia canales = z	SVM	12454	PCA	0.73333	0.72727	0.84118	0.85294	0.61875	0.625

Figura 3.1: Resultados de prueba de clasificación

Por último cabe mencionar que las características empleadas en cada iteración se pueden aplicar directamente o previo procesado con el método PCA. El análisis de componentes principales o PCA se refiere al proceso por el cual se obtienen muestras incorreladas a partir de componentes dependientes entre sí [29]. Este proceso se podrá emplear para adecuar los datos disponibles a la prueba de validación.

CAPÍTULO 4

ADICIÓN DE FUNCIONALIDADES Y FLEXIBILIDAD DEL SISTEMA

En esta sección se detallarán las diferentes características que permiten al sistema adaptarse a las necesidades del usuario, tanto en lo referente a la entrada de registros como a su procesado. Se mencionarán además ciertos ejemplos de funcionalidades añadidas que se utilizarán como referencia en ampliaciones futuras. En cuanto al procesado, la elección manual de parámetros permite modificar su funcionamiento durante la ejecución, sin tener que acceder al código interno del programa.

4.1. APERTURA DE REGISTROS ACTIGRÁFICOS

La primera etapa del sistema, y una de las que más complejidad soporta, es la apertura de los registros actigráficos sin procesar procedentes de los aparatos de medida. Resulta de suma importancia poder elegir cualquier fichero ante la posibilidad de aumentar la base de datos con nuevos pacientes.

Los registros provenientes de los actígrafos presentan una estructura predefinida y estable, con un formato de fichero *.dat* fácilmente procesable por MATLAB. El proceso de apertura actual está simplificado de manera que sólo se necesite disponer de los registros de ambos tipos (casos y controles) en sendos directorios diferenciados, así como proporcionar un fichero de información adicional para cada tipo. Durante la ejecución se solicitará al usuario que seleccione los directorios y los correspondientes ficheros de información. Estos ficheros, en formato *.csv* estándar, contendrán la información de edad y sexo de cada individuo en formato tabla, permitiendo su modificación en función de los registros a procesar. Cabe destacar que el sistema detectará cualquier deficiencia en los ficheros aportados y cualquier problema en relación a su procesado, ignorando aquellos registros erróneos o que no cumplan los requerimientos oportunos.

4.2. EXTRACCIÓN DE CARACTERÍSTICAS

Una vez procesados los registros, existe la posibilidad de seleccionar aquellos que resulten de interés en función del grupo de población al que pertenecen. En la figura 4.1 se observan las diferentes posibilidades de elección de individuos, tanto en función de sus rasgos como de forma manual.

The screenshot shows a software interface for filtering individuals. It is titled "Individuos" and is divided into two main sections: "Edad" (Age) and "Sexo" (Sex).
The "Edad" section features a dropdown menu with values from 6 to 15. Below the dropdown is a list of records with IDs: c04.dat, c05.dat, c06.dat, c13.dat, c18.dat, c21.dat, c22.dat, c26.dat, c28.dat, c35.dat, c36.dat, c38.dat, c40.dat, c42.dat, c43.dat, and c44.dat. Below the list, it shows "Total = 5" and a checkbox labeled "Todos".
The "Sexo" section has two checkboxes: "Niños" (checked) and "Niñas" (unchecked). Below the checkboxes is a list of records with IDs: x01.dat, x03.dat, x04.dat, x05.dat, x12.dat, x14.dat, x15.dat, x20.dat, x22.dat, x23.dat, x26.dat, x35.dat, x36.dat, x41.dat, x43.dat, and x44.dat. Below the list, it shows "Total = 4" and a checkbox labeled "Todos".
At the bottom of the interface, there are three buttons: "Atras", "Añadir Todos", and "Siguiete".

Figura 4.1: Filtrado de individuos y selección de registros

4.2. EXTRACCIÓN DE CARACTERÍSTICAS

Continuando con el procesado de los registros, la extracción de características presenta también múltiples opciones de ampliación. Existe numerosa bibliografía sobre las características de actividad relevantes para la detección de diferentes patologías, y resulta indispensable poder ampliar las funciones de extracción existentes con nuevos parámetros, lo que permitirá comprobar su posible validez para el estudio del TDAH.

La solución propuesta es aportar el código de procesado de características a través de scripts independientes fácilmente programables por cualquier usuario. Estos ficheros se solicitarán durante la ejecución y será el propio sistema el encargado de clasificar las diferentes características extraíbles por ellos, así como comprobar el correcto funcionamiento de los mismos. Dada la enorme cantidad de diferentes posibilidades en este campo y sus consecuentes errores de cálculo, el sistema escogerá al azar un registro actigráfico disponible y procederá a extraer todas las características que permitan los scripts seleccionados, comprobando cuales de ellos son capaces de procesar correctamente la información necesaria o informando sobre los posibles defectos de programación encontrados en cada fichero probado. En el anexo A.2 se detallarán con más detalle los pasos a seguir para construir los scripts de generación.

Como ejemplo práctico se creó una nueva característica denominada "Percentil de intervalos nulos", en la cual se mide la longitud de los intervalos en los que el valor de actividad

es nulo. A partir de estas longitudes se hallan los percentiles 25 y 75 correspondientes a los valores de longitud por debajo de los cuales se encuentra el 25 o 75 % del resto de valores. El código necesario es de apenas unas líneas y permite obtener valores para los diferentes diezmos, intervalos y canales disponibles, con lo que se demuestra la facilidad de ampliación del sistema.

Además de esta característica también se añadieron una serie de nuevos métodos correspondientes al estudio [4] tales como la Variabilidad Intradía o la Amplitud Relativa, parámetros relacionados con el ritmo circadiano de los que ya se habló en el capítulo anterior.

4.2.1. IDENTIFICACIÓN DE CARACTERÍSTICAS

Ante la complejidad aportada a la generación de características surge la necesidad de identificar cada una de ellas dentro del conjunto total, de manera que sea posible conocer a qué características se refieren los resultados obtenidos o, por ejemplo, a la hora de elegir con cuáles se quiere trabajar en las sucesivas pruebas.

La cantidad de características disponibles para cada individuo procesado se situó en torno a las cincuenta mil, aunque varía en función de las decisiones tomadas por el usuario en la etapa de generación al elegir uno o más ficheros de extracción. Dicha cantidad total de datos resulta excesiva para un procesamiento iterativo realizado elemento por elemento, lo que hace necesario utilizar técnicas de cálculo matricial. MATLAB es capaz de manipular datos a gran velocidad siempre y cuando se presenten en forma matricial, no individualmente, y las diferencias de rendimiento entre los dos métodos son más que importantes, resultando críticas cuando se trabaja con un número de datos tan grande como el del presente estudio. De esta manera, al manipular el total de las características de forma conjunta resulta indispensable disponer de un método adecuado de identificación de las mismas independientemente de cuántas o cuáles hayan sido extraídas de los registros.

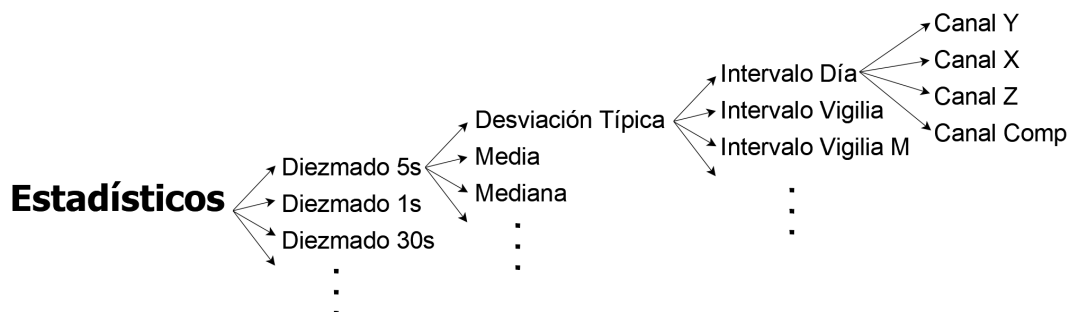


Figura 4.2: Ejemplo de estructura de características

La figura 4.2 muestra la estructuración de una de las características disponibles. Se observa que el parámetro **Estadísticos** dispone de diferentes diezmos de señal, de los que se calculan los diversos tipos de estadísticos. Estos se calculan a su vez en diferentes intervalos de tiempo y para los 4 canales de actividad disponibles. En total, la característica **Estadísticos** contiene 1528 valores diferentes. A este número se le dobla el valor después de considerar

4.3. PARÁMETROS DE PROCESAMIENTO

el proceso de desestacionalización, el cual genera una estructura paralela a la original, por lo que aparecen 3056 valores para un solo tipo de característica. En el caso del parámetro CTM esta cantidad es aún mayor, disponiendo de $22624 * 2 = 45248$ valores diferentes. El total actual de características disponibles es de cerca de 54300, lo que hace imprescindible disponer de un control sobre ellas.

Para conseguir mantener las características clasificadas, el sistema maneja información sobre la estructura actual de parámetros, la cual extrae de los ficheros generadores de características. Esta estructura se propaga de sección en sección y varía en función de los ficheros de generación elegidos al procesar los registros de individuos. La existencia de esta estructura permite disponer de ‘traductores’, unos pequeños scripts que obtienen la posición de la característica dentro de la matriz a partir de las opciones a las que pertenece o, al contrario, especifican qué tipo de característica es de acuerdo a la posición de la matriz completa. La identificación se realiza a través de un código numérico de seis dígitos, los cuales se corresponden con las diferentes particularidades de cada característica. En el anexo A.3 se realiza una aproximación al método empleado para la localización de características.

Campo		Valor	Significado
Estacionalizada/Desestacionalizada		1	Desestacionalizada
Tipo de característica		2	Grupo CTM
Variable en función del tipo de característica	Parámetro rho	57	rho = 57
	Diezmado utilizado	4	60 segundos
	Intervalo del día estudiado	5	Descanso Completo
	Canal de datos	2	Canal Y

Tabla 4.1: Identificación de características por código

El significado de cada dígito depende del tipo de característica a la que se refiera, pudiéndose utilizar los seis dígitos del código o algún elemento menos en función de la estructura del parámetro. Por ejemplo, una determinada característica tendría la codificación [1, 2, 57, 4, 5, 2] que identifica cada una de sus particularidades, tal y como se detalla en la tabla 4.1.

Los traductores, así como diferentes partes gráficas del sistema, son variables y se adaptan en función de la estructura de características con las que se trabaje. Una de esas secciones es la que permite elegir las características con las que trabajar en el análisis de las prestaciones de los clasificadores. Mediante la interfaz mostrada en la figura 4.3 es posible elegir una a una, o por grupos funcionales, aquellas características que se quieren tener en cuenta en los procesamientos siguientes, de manera que se pueda determinar la efectividad de cada grupo disponible según convenga.

4.3. PARÁMETROS DE PROCESAMIENTO

De la misma forma que es posible elegir los individuos y las características con las que trabajar, resulta también necesario poder modificar los diferentes parámetros de cada etapa

Estacionalizados Todos Estacionalizados

Tipo

Dinamica Simbolica Todos Dinamica Simbolica

probPalabra00
probPalabra01
probPalabra02
probPalabra10
probPalabra11
probPalabra12
probPalabra20

Todos

1s
5s
30s
60s
300s
600s
900s

Todos

dia
vigilia
vigiliaM
vigiliaT
descanso
descansoL
descansoM
descansoF

Todos

x
y
z
comp

Todos

Añadir

Figura 4.3: Elección de características

de procesado. La interfaz permite elegir el valor de cada parámetro disponible en las diferentes pruebas o procesos del sistema como, por ejemplo, la selección de características con el test de P-Valor o la validación cruzada.

En el caso de la selección de características diferenciadoras con el método de Wilcoxon, los parámetros modificables son el número de iteraciones a realizar, el coeficiente o p-valor necesario para superar la prueba y la cantidad de individuos a utilizar como referencia en cada iteración. También se permite elegir si es necesario que las características superen todas las iteraciones del test o si solo se necesita superar un porcentaje de las mismas, valor también modificable.

En cuanto a la validación cruzada, los valores alterables se refieren tanto a las propiedades de la prueba como a los parámetros internos de los clasificadores a utilizar. Es posible modificar la dimensión de los clasificadores y el número de repeticiones a realizar, así como la cantidad de casos y controles a emplear en el entrenamiento. Estos últimos parámetros se relacionan directamente con el número de individuos activos, es decir, con aquellos que han sido elegidos por el usuario y que constituirán el total disponible. A partir de ellos es posible seleccionar el número de individuos de cada tipo que se usarán tanto para entrenar

los clasificadores como para probarlos, teniendo en cuenta que estos dos parámetros están relacionados por ser su suma igual al total disponible. En la figura 4.4 se observan los selectores encargados de modificar estos valores, con los que se puede elegir fácilmente cualquier combinación. Entre ellas destaca la conocida como *leave-one-out* en la cual se utilizan todos los individuos disponibles menos uno para el entrenamiento del clasificador y el restante es el que se emplea en la prueba.

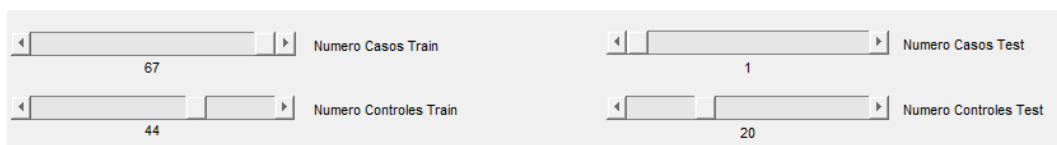


Figura 4.4: Barras de selección de individuos para validación cruzada

Además de los valores anteriores también es posible elegir la realización o no del procesado PCA antes de la validación cruzada. Al habilitar esta opción se consigue que el sistema realice la simulación con PCA y sin él, lo que permite comprobar la efectividad de este proceso sobre las muestras estudiadas.

Los parámetros a modificar dentro de los clasificadores dependen del tipo al que pertenezcan. El sistema permite elegir los tipos de clasificadores a utilizar, además de ciertos valores internos, como por ejemplo el número de vecinos del clasificador KNN o el tipo de kernel empleado en el clasificador SVM. Los valores que implican entrada numérica se han programado de manera que se puedan introducir múltiples valores simultáneos, lo que permite estudiar las diferencias entre ellos.

4.4. GUARDADO DE DATOS EN ETAPAS INTERMEDIAS

Cada una de las etapas de ejecución permite realizar un guardado de los resultados obtenidos hasta ese punto para continuar con la ejecución en el futuro.

El primer punto de guardado se realiza en la extracción de características a partir de los registros actigráficos. La pantalla de extracción permite elegir un directorio en el que guardar las características obtenidas, y en dicho lugar será donde se almacenarán los diferentes ficheros de guardado de las demás secciones. De esta manera se consigue tener reunidos en un mismo directorio todos los datos referidos al mismo fichero de características, ya que este actuará como base para el resto del procesamiento.

La segunda posibilidad de guardado se refiere a la lista de características elegidas para la ejecución de la prueba de validación cruzada. Esta lista se puede crear añadiendo características de forma manual o realizando una prueba de validez de p-valor. En ambos casos se generará un fichero con la lista de características disponibles, lo que facilita la repetición de pruebas sin necesidad de elegir de nuevo las características.

Después de la ejecución de la prueba de validación cruzada, los datos quedan guardados en un fichero de resultados, el cual puede ser visualizado desde la función correspondiente

del programa (sección Análisis de resultados) o de manera manual cargando las variables al entorno Matlab. En el anexo A.4 se detalla la estructura de los datos y cómo interpretarlos.

Por último, la sección de Análisis de Resultados permite guardar la configuración y parámetros de los mejores clasificadores en función de los resultados obtenidos en la validación cruzada, lo que permite utilizar dichos clasificadores para la detección del trastorno en registros sin clasificar.

4.5. SEGUIMIENTO DE LOS PROCESOS

La realización de determinados procesos del sistema implica un gran número de operaciones, como por ejemplo la extracción de las características, el test de validez del p-valor o el cálculo de las prestaciones de los clasificadores. Su ejecución conlleva largos tiempos de espera en los que el usuario debe esperar a su finalización, y en ningún caso precisan de acciones o interacción externa.

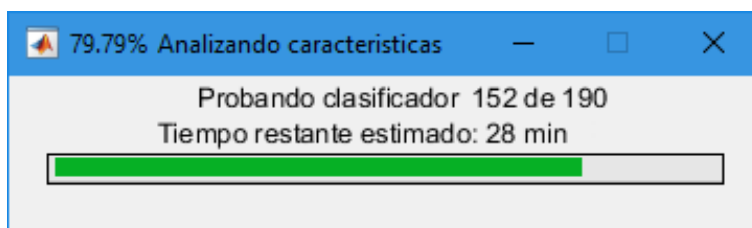


Figura 4.5: Seguimiento del progreso y estimación de tiempo restante

Para poder realizar el seguimiento se han utilizado barras de progreso, como la que se muestra en la figura 4.5. En estas ventanas emergentes se especifica la tarea que se está realizando, el progreso en forma de porcentaje completado y una estimación del tiempo restante, calculada en función de las iteraciones ya realizadas. Este tiempo varía en función del estado instantáneo del equipo en el que se está ejecutando el programa, ya que la realización de otras tareas en paralelo puede ocasionar un descenso del rendimiento destinado a cada tarea.

Para el cálculo de la estimación del tiempo se emplea un compromiso entre la última iteración realizada y las estimaciones anteriores, aplicando una ponderación diferente según el procesado que se esté monitorizando. Por ejemplo, en el caso de la extracción de características a partir de los registros actigráficos se da prioridad al valor global de la estimación en detrimento del último cálculo realizado. Se utiliza este método ya que en caso contrario la estimación tendría demasiadas variaciones en los momentos que el sistema detecta un registro erróneo y lo descarta, acción que requiere mucho menos tiempo que una extracción completa. En los procesos cuya realización compete iteraciones de la misma naturaleza (misma cantidad de cálculos), el sistema estima el tiempo dando prioridad al último elemento analizado, ya que es el más parecido a las próximas iteraciones en lo relativo al estado momentáneo del rendimiento disponible en el equipo.

4.6. CLASIFICACIÓN DE PACIENTES

Además de la monitorización del progreso las ventanas emergentes permiten cancelar la ejecución en cualquier momento, ya sea para poder rectificar los valores de entrada o por falta de tiempo en comparación con lo estimado en el proceso.

4.6. CLASIFICACIÓN DE PACIENTES

La última sección del programa permite clasificar un paciente como sano o como poseedor de TDAH en función de sus datos actigráficos. Para ello se hace uso de la base de datos de individuos procesados y de la configuración de los mejores clasificadores, aquellos que mejores resultados han mostrado después de la prueba de validación cruzada.

Esta etapa es completamente flexible en tanto que es posible utilizar cualquier base de datos de registros preprocesados como referencia, así como realizar la clasificación con diferentes sistemas. La elección de los conjuntos clasificadores se hace acorde a los resultados obtenidos en pruebas de validación cruzada, eligiendo aquellos que obtengan mejores valores de precisión, sensibilidad o especificidad en las pruebas realizadas, medida como la media o mediana de los resultados.

La utilización de varios clasificadores diferentes permite obtener múltiples predicciones, de manera que el usuario pueda decidir el diagnóstico final basándose en la información aportada por más de una predicción. La figura 4.6 muestra un ejemplo de diagnóstico en el que se aprecian diferentes clasificadores con su correspondiente predicción.

Modelo	Clasificación	Control
SVM_Q	NoPCA	CONTROL
SVM	NoPCA	CONTROL
SVM_RBF	NoPCA	CONTROL
SVM_Q	PCA	CASO
SVM_RBF	PCA	CONTROL
Arbol	NoPCA	CONTROL
SVM	NoPCA	CONTROL
SVM	NoPCA	CONTROL
SVM	NoPCA	CONTROL
SVM	NoPCA	CONTROL
SVM_Q	NoPCA	CONTROL
SVM_Q	NoPCA	CONTROL

Figura 4.6: Resultados de una prueba de diagnóstico

CAPÍTULO 5

RESULTADOS: EJECUCIÓN COMPLETA DEL SISTEMA

Este capítulo pretende servir como referencia durante el manejo del entorno de procesado. Utilizando una ejecución típica como hilo conductor, se explicarán cada una de las partes que componen el entorno completo, identificando tanto sus características y particularidades como las posibilidades que presenta cada componente del programa.



Figura 5.1: Menú para la apertura de las diferentes partes del programa completo

La subdivisión en secciones diferenciadas permite realizar tareas con fuerte carga computacional de manera separada en diferentes etapas y diferentes momentos, de manera que no es necesario procesar toda la información de una sola vez sino que las tareas se adaptan a la exigencias del usuario. La ventana inicial mostrada en la figura 5.1 permite acceder a las diferentes partes del sistema. Esta ventana aparece al iniciar el programa y se oculta al iniciar una de las secciones disponible, volviendo a aparecer en caso de cerrar dicha sección.

5.1. PROCESADO DE REGISTROS

En esta primera etapa del procesado se pretende obtener una base de datos con registros actigráficos adecuadamente procesados. Para ello se leerán ficheros actigráficos correspondientes tanto a controles (individuos calificados como sanos) como a casos (individuos en los que se ha detectado TDAH), se procesarán y se extraerán las características necesarias. Este conjunto de resultados serán la base del procesado futuro y serán los datos tomados como referencia a la hora de construir los sistemas clasificadores.

5.1.1. APERTURA DE FICHEROS

El primer paso para el procesado de registros es la apertura de los ficheros actigráficos obtenidos desde los dispositivos de captura. Estos ficheros deben estar previamente calificados como casos/controles y divididos en sendos directorios independientes. En la figura 5.2 se observa la pantalla de carga de los ficheros de casos.

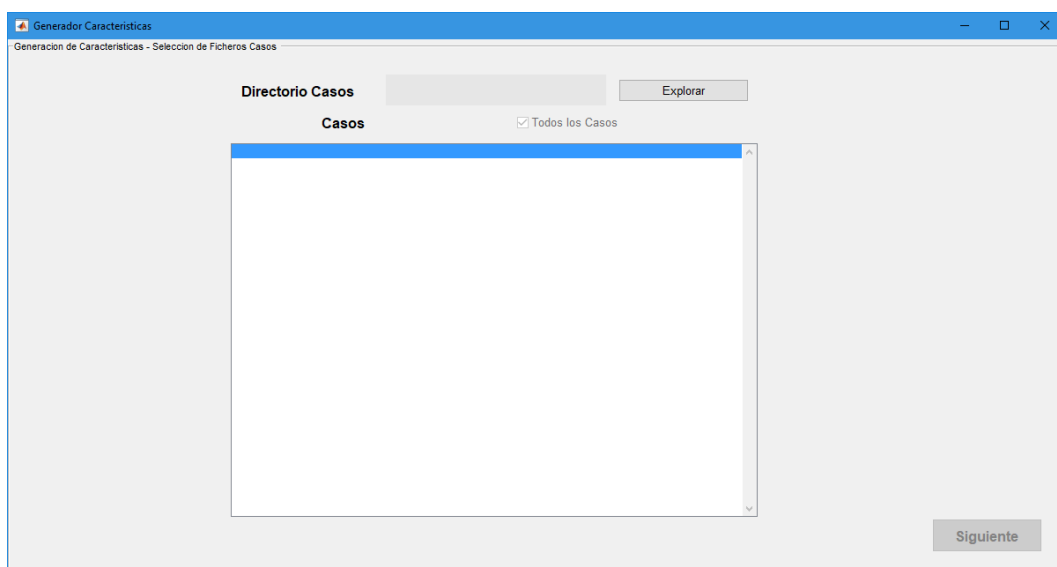


Figura 5.2: Apertura de registros del tipo casos.

Al pulsar en el botón **Explorar** se solicitará la apertura del directorio en donde se encuentran los ficheros actigráficos de casos. Seguidamente se solicita el fichero .csv que contiene la información de edad y sexo correspondiente a los registros leídos en la pantalla anterior.

El fichero de información debe contener las columnas ‘Nombre’, ‘Sexo’ y ‘Edad’ y debe estar relleno con los datos de cada registro a procesar. Sin esta información el programa no puede procesar los registros actigráficos.

Una vez seleccionados tanto el directorio como el fichero de información, el programa comprobará la concordancia entre los registros existentes y la información leída, mostrando en pantalla la lista de ficheros de registro con información válida. Por defecto el sistema procesará todos los registros existentes, pero es posible seleccionar uno a uno los que se deseen procesar. Bastará con deshabilitar la casilla **Todos** y pulsar sobre los nombres en la lista disponible, pudiéndose realizar selecciones múltiples con la actuación conjunta de las teclas **Ctrl** y **Alt**.

El mismo procedimiento se deberá repetir en el apartado de controles, el cual contiene la misma estructura de procedimientos que con el tipo anterior, tal y como se muestra en la figura 5.3.

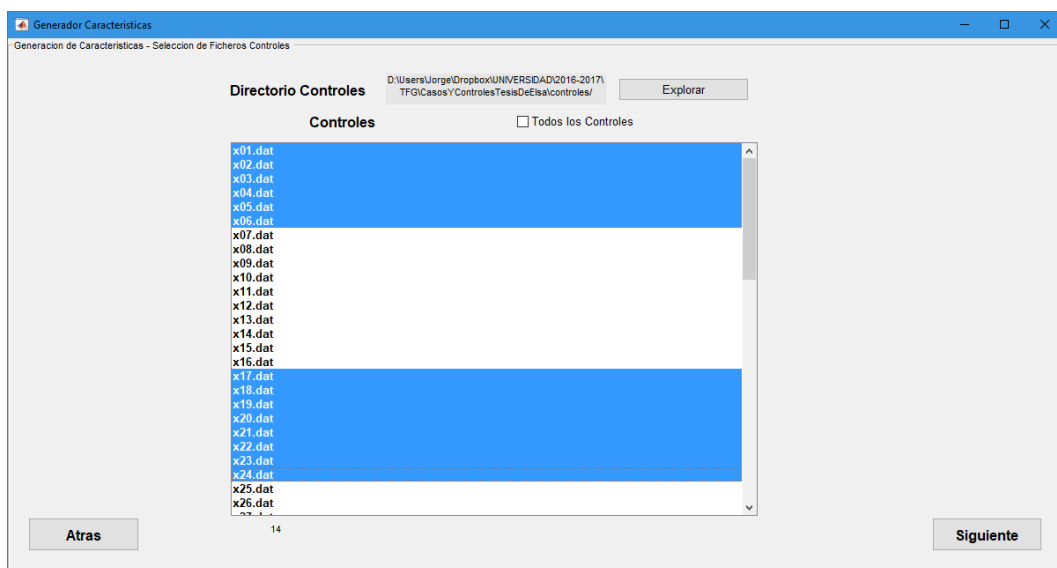


Figura 5.3: Apertura de registros del tipo controles.

Se establece un número mínimo de 2 individuos para cada conjunto, necesarios para el proceso de extracción de características. Como se verá más adelante, la selección de los ficheros no implica que resulten válidos y se pueda realizar su procesamiento, por lo que se recomienda procesar 10 o más individuos de cada tipo para obtener una muestra consistente.

5.1.2. ELECCIÓN DE GENERADORES DE CARACTERÍSTICAS

Una vez cargados los datos actigráficos, el siguiente paso será seleccionar los ficheros conocidos como «generadores de características». Estos ficheros contienen un conjunto de métodos que recogen los valores actigráficos y extraen diferentes características. Es posible añadir nuevos generadores mediante la inclusión de un fichero de métodos en el directorio correspondiente, procedimiento detallado en las secciones del anexo A.1 y A.2, y será en este paso del programa donde se comprobará el correcto funcionamiento de dichos ficheros y se habilitarán para ser utilizados en el procesamiento de los registros.

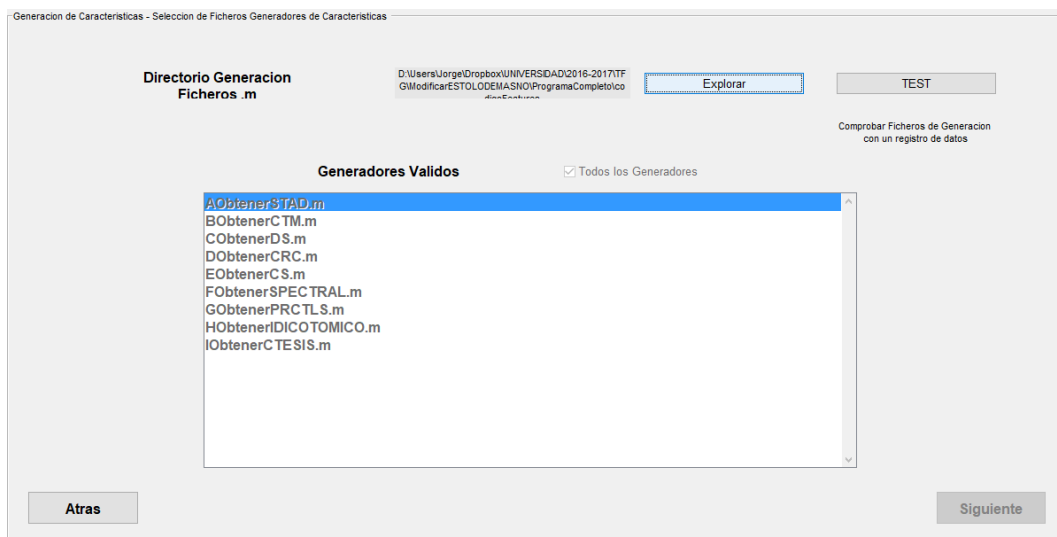


Figura 5.4: Elección de scripts generadores de características

La pantalla mostrada en la figura 5.4 permite utilizar los ficheros generadores por defecto o seleccionar unos nuevos. Para ello bastará con elegir el directorio en el que se encuentran pulsando el botón **Explorar** y realizar la prueba de comportamiento con el botón **Test**. La lista disponible mostrará los ficheros procesadores válidos, pudiéndose elegir cuáles se utilizarán de manera individual.

5.1.3. GENERACIÓN Y GUARDADO

Por último, la pantalla de generación solicitará el directorio de guardado donde se almacenará la base de datos obtenida. Una vez seleccionada bastará con pulsar en **Generar** para que el procesamiento comience. La duración total dependerá de la cantidad de registros a procesar y del número de características a extraer de cada uno de ellos, así como la potencia computacional del equipo utilizado. La ventana de seguimiento permite conocer en todo momento el estado de la ejecución así como una estimación del tiempo restante. También es posible cancelar el proceso cerrando la ventana de progreso visible en la figura 5.5, generando el correspondiente mensaje que indica la cancelación de la ejecución por parte del usuario.

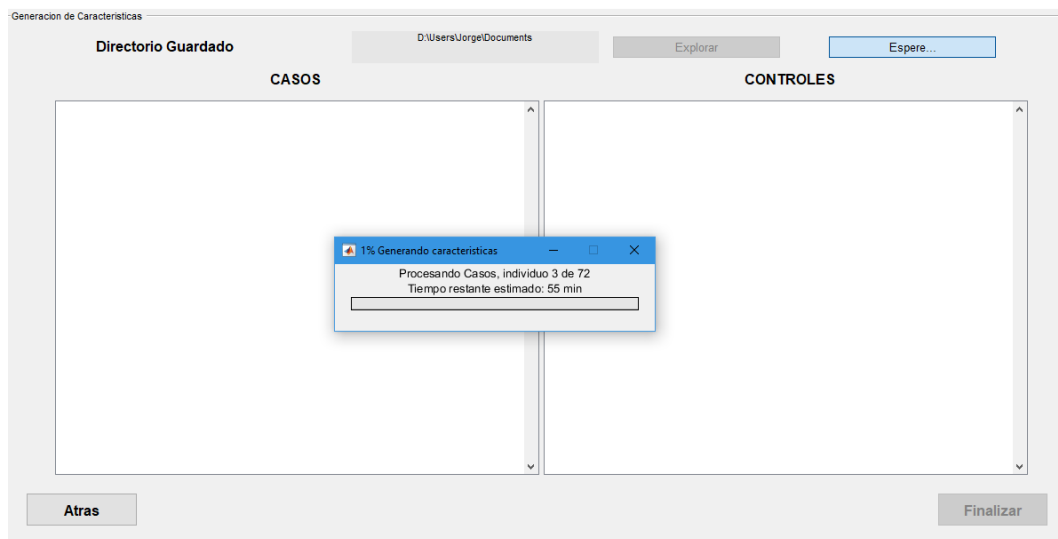


Figura 5.5: Generación de características y ventana de progreso

Una vez comenzado el proceso, el sistema analizará cada uno de los ficheros de datos actigráficos y extraerá sus características siguiendo el procedimiento marcado por los ficheros de generación. Este proceso puede no completarse correctamente en determinados ficheros de registro debido a varios factores, como por ejemplo la longitud insuficiente de los datos, un formato de fichero desconocido o no válido u otros errores de procesamiento relacionados con los ficheros de extracción de características. En caso de no poderse procesar, el fichero de registro quedará marcado como erróneo y no se almacenarán sus características.

La pantalla final visible en la figura 5.6 muestra la información obtenida al finalizar el procesado de ciertos registros, tanto para casos como para controles. Además de los campos de información acerca del registro procesado, tales como la fecha, el nombre del fichero o la edad y sexo del individuo, se muestran otros campos relativos al resultado del procesamiento. El valor del campo «Procesado» indica si se ha conseguido procesar o no, indicando con 1 el caso satisfactorio, -1 si ha habido errores de ejecución o el fichero es erróneo y 0 si el registro era demasiado corto. Además se indica el lugar que ocupa el fichero procesado en el conjunto de datos con el campo «Posición» y se adjuntan una serie de comentarios relativos al procesamiento, de manera que resulta sencillo localizar cualquier posible error.

Una vez visualizada la información es posible pasar a la siguiente sección del programa mediante el botón **Finalizar** o volver al menú inicial cerrando la ventana actual de trabajo. En cualquier caso el trabajo ha quedado guardado y los registros procesados se encuentran almacenados para futuros usos.

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

Generación de Características

Directorio Guardado: D:\Users\Jorge\Pictures\Greenshots

Explorar Espere...

CASOS							CONTROLES						
Fichero	Edad	Sexo	Fecha	Horalnicio	Procesado	Posicion	Fichero	Edad	Sexo	Fecha	Horalnicio	Procesado	Posicion
c01.dat	15	1	30/06/2010	11:00:00	1	1	x01.dat	6	1	21/06/2010	10:00:00	1	1
c02.dat	14	1	04/03/2010	14:30:00	1	2	x02.dat	7	2	04/08/2010	11:00:00	1	2
c03.dat	9	2	29/06/2010	09:00:00	1	3	x03.dat	6	1	13/09/2010	10:30:00	1	3
c04.dat	13	1	19/05/2010	15:00:00	1	4	x04.dat	6	1	21/02/2011	11:00:00	1	4
c05.dat	13	1	25/02/2010	14:00:00	1	5	x05.dat	6	1	11/04/2011	11:00:00	1	5
c06.dat	13	1	13/12/2010	11:00:00	1	6	x06.dat	11	1	08/02/2012	14:00:00	1	6
c07.dat	14	1	28/09/2010	11:00:00	1	7	x07.dat	7	2	08/02/2010	21:30:00	0	7
c08.dat	12	2	11/06/2010	12:00:00	1	8	x08.dat	7	1	10/02/2010	09:30:00	1	7
c09.dat	11	1	22/02/2010	13:00:00	1	9	x09.dat	7	2	17/02/2010	10:00:00	1	8
c10.csv	11	1	-	-	-1	0	x10.dat	7	1	23/02/2010	10:00:00	1	9
c11.dat	11	1	22/03/2010	14:30:00	1	10	x11.dat	7	1	24/02/2010	10:00:00	1	10
c12.dat	8	1	06/07/2010	09:00:00	0	0	x12.dat	6	1	02/03/2010	10:30:00	1	11
c13.dat	9	1	10/09/2010	09:00:00	1	11	x13.dat	7	1	15/03/2010	09:00:00	1	12
c14.dat	12	2	08/11/2010	11:00:00	1	12	x14.dat	6	1	16/03/2010	09:30:00	1	13
c15.dat	10	1	28/06/2010	12:00:00	1	13	x15.dat	6	1	05/04/2010	10:30:00	1	14
c16.dat	12	1	14/09/2010	22:00:00	1	14	x16.dat	12	2	09/04/2010	14:00:00	1	15
c17.dat	10	1	06/04/2010	14:00:00	1	15	x17.dat	12	2	09/04/2010	09:30:00	1	16
c18.dat	9	1	09/03/2010	13:00:00	1	16	x18.dat	13	1	09/04/2010	13:00:00	-1	0
c19.dat	9	2	23/03/2010	12:00:00	1	17	x19.dat	14	1	09/04/2010	14:00:00	1	17
c20.dat	8	1	09/08/2010	14:00:00	1	18	x20.dat	6	1	19/04/2010	10:00:00	1	18
c21.dat	9	1	08/03/2010	10:00:00	1	19	x21.dat	8	2	20/04/2010	10:00:00	1	19
c22.dat	9	1	25/05/2010	14:00:00	1	20	x22.dat	6	1	27/04/2010	10:30:00	1	20
c23.dat	8	1	18/06/2010	12:00:00	1	21	x23.dat	6	1	05/05/2010	09:30:00	1	21
c24.dat	8	1	13/09/2010	10:30:00	1	22	x24.dat	7	1	04/05/2010	10:30:00	1	22
c25.dat	11	1	19/05/2010	08:30:00	1	23	x25.dat	7	1	11/05/2010	10:00:00	1	23
c26.dat	6	1	19/04/2010	14:00:00	1	24	x26.dat	13	1	28/05/2010	07:30:00	1	24

Atras Finalizar

Figura 5.6: Resultados de el procesado de registros.

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

La existencia de un número tan grande de diferentes características hace necesaria una evaluación de las más validas para la detección del trastorno, así como la evaluación del de las prestaciones de los clasificadores creados a partir de ellas. En esta sección del programa se cargará la base de registros actigráficos previamente procesado y con las características extraídas, pudiéndose seleccionar con qué individuos y con cuáles de sus características se quiere trabajar.

La selección de características se realizará con el test p-valor, el cual extraerá las características más válidas de cara a la detección del trastorno. El siguiente paso será comprobar la validez de dichas características sobre la base de registros existente realizando pruebas de entrenamiento y test de clasificadores a partir de los resultados anteriores, obteniéndose una aproximación de la efectividad clasificadora del sistema en situaciones reales.

5.2.1. APERTURA DE REGISTROS

El proceso de generación de características crea un fichero que contiene los registros analizados, así como diferente información adicional necesaria para el procesado posterior. Para continuar con el trabajo es posible acceder al siguiente apartado pulsando en el botón **Finalizar** de la pantalla anterior, o cargando los datos previamente guardados a través de la pestaña de carga de la misma sección. La figura 5.7 muestra la pantalla a la que se accederá en caso de abrir directamente la sección de selección de características y prueba

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

de clasificadores, desde la que es posible cargar cualquier fichero de registros procesados mediante el botón **Explorar**. Una vez leídos los registros, su información aparecerá en las listas correspondientes a modo de referencia, pudiéndose volver a ella en cualquier momento en caso de necesitar consultar su información.

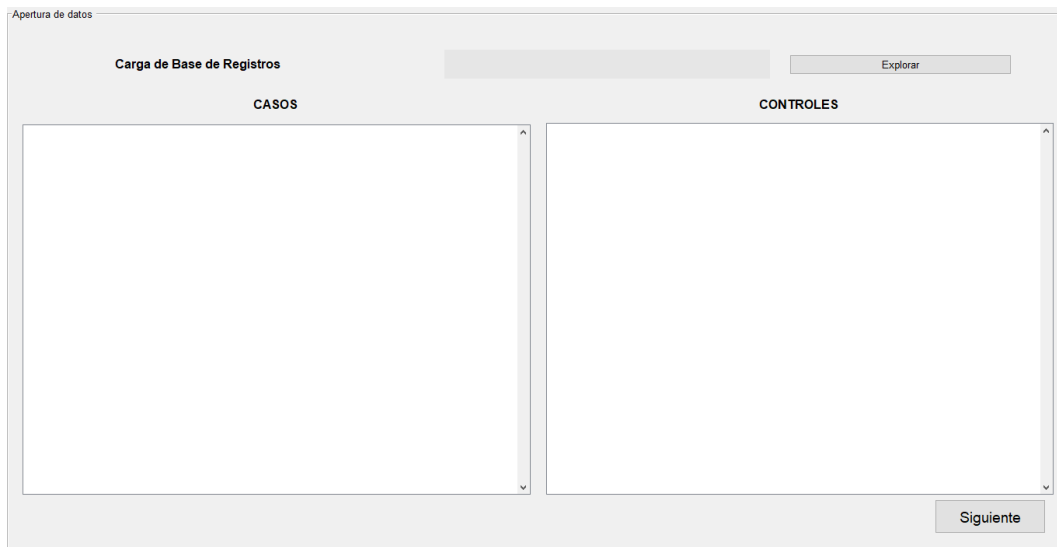


Figura 5.7: Apertura de registros procesados

5.2.2. GRUPOS DE POBLACIÓN

Una vez cargados los datos, ya sea de manera directa al venir desde la sección anterior o de manera manual, la siguiente pantalla permitirá elegir los diferentes casos y controles con los que se quiere trabajar en el resto de la ejecución. La selección se hace por una parte en base a datos demográficos de edad y sexo, pudiéndose elegir diferentes grupos de edad lo que permite analizar los datos de un conjunto de la población total. Por otra parte, es posible elegir manualmente cada fichero de registro, pensado para descartar ciertos individuos de un determinado conjunto.

La figura 5.8 muestra un ejemplo de selección de individuos de sexo femenino con edades entre 6 y 9 años, observándose la existencia de 12 registros de casos y 9 de controles para este grupo de población.

La selección manual de ficheros se realizaría deshabilitando la casilla **Todos** correspondiente y eligiendo uno o varios ficheros de la lista de disponibles, la cual se actualiza con cada modificación del grupo de población mediante las opciones de edad y/o sexo.

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

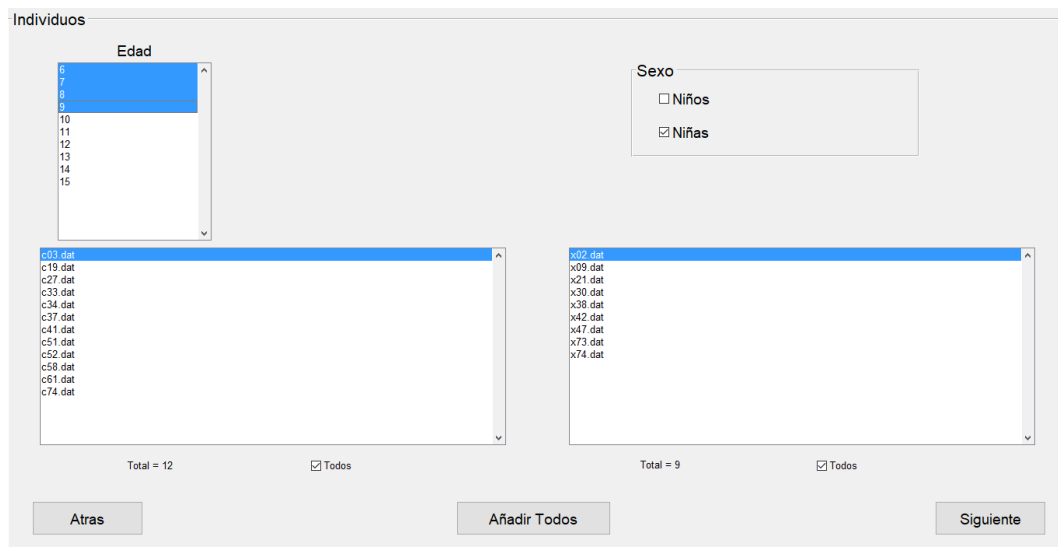


Figura 5.8: Elección de registros en función de edad y sexo

5.2.3. ELECCIÓN DE CARACTERÍSTICAS

Después de elegir los individuos cuyos registros queremos tener disponibles para trabajar, el siguiente paso será elegir las diferentes características que se quieren procesar de entre las disponibles en los registros cargados. La cantidad de características disponibles variará en función de los ficheros de generación empleados en la etapa de procesado de registros actigráficos, un detalle importante a tener en cuenta ya que todo el procesado posterior se realizará en función de esa estructura de características, incluyendo los consiguientes ficheros de guardado de progreso, los cuales se relacionarán directamente con el fichero de registros a partir del cual se generaron.

La opciones disponibles se muestran en la pantalla (figura 5.9) a modo de diferentes menús de selección. Por un lado, el menú 1 permite elegir entre las características estacionalizadas o las que han sufrido el proceso de desestacionalización. El menú 2 selecciona el tipo de características a mostrar, entendiendo como tipo a cada conjunto de características extraídas por un fichero de generación diferente, por tanto existirán tantos tipos como fichero se emplearon. El resto de menús son variables y dependen del tipo elegido en el menú 2, llegando a existir menús de selección sin uso en el caso de que no lo requiera el tipo elegido.

Existen múltiples opciones para añadir características, tanto de una en una seleccionando una sola opción con menú como selecciones múltiples eligiendo varias opciones de cada menú o habilitando las opciones '**Todos**' disponibles junto a cada menú. Además existe el botón **Añadir Todas** que, como su nombre indica, selecciona todas las características disponibles en el programa.

Una vez elegidas las opciones necesarias, pulsando el botón **Añadir** se seleccionará la característica o características correspondientes y aparecerá en la lista situada a la derecha de

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

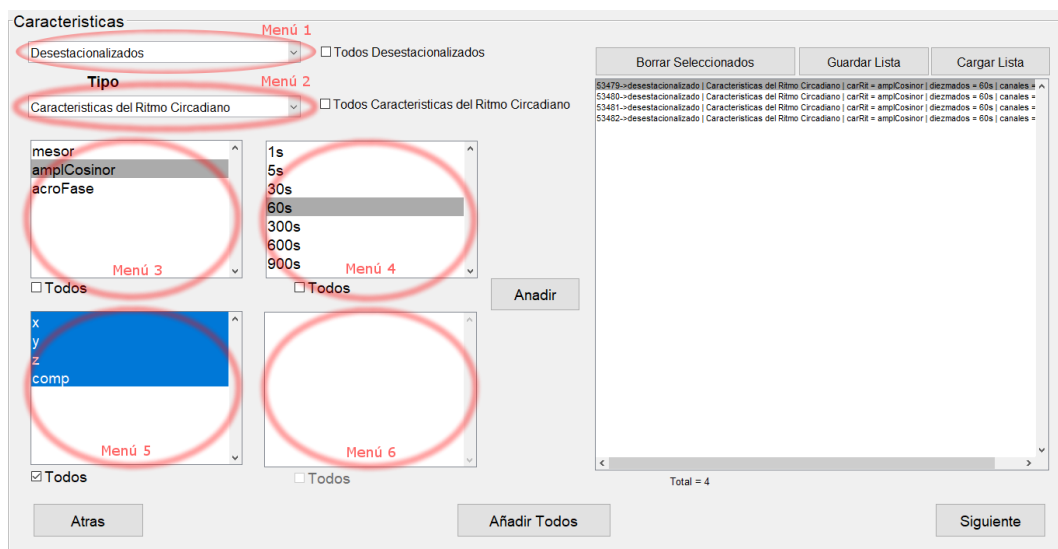


Figura 5.9: Elección de características a procesar

la pantalla, en la que se muestran todos los parámetros seleccionados con los que se trabajará. Esta lista permite además borrar manualmente ciertas características mediante su selección y el botón **Borrar**, guardar la lista actual de características para futuras ejecuciones o cargar una lista previamente guardada, pudiéndose sustituir la lista actual o añadir las nuevas características al conjunto actual. En la correspondiente sección se verán más detalladamente estas opciones de carga y guardado.

En la figura 5.9 se puede ver un ejemplo en el que se han elegido las características desestacionalizadas pertenecientes al grupo de «Características del Ritmo Circadiano», en concreto la «Amplitud del Cosinor» para el diezmado de 60s y todos los canales disponibles. Como se puede observar, dichas características han sido añadidas a la lista de la parte derecha, previa pulsación del botón **Añadir**.

Cabe destacar en este punto que el proceso de extracción de características elimina aquellas que han resultado nulas en todos los ficheros procesados, de manera que puede darse el caso de que se intente añadir una característica y ésta no aparezca en la lista. De este modo se evitan errores de selección y se mantiene la exclusión de aquellas características irrelevantes para el sistema, incluyéndose un mensaje de aviso en el caso de que ninguna característica elegida esté disponible.

5.2.4. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES

Como se ha mencionado anteriormente, el número total de características disponibles es muy grande, cerca de 54300 con los ficheros de extracción disponibles. Este número es demasiado grande para las pruebas de clasificadores siguientes por lo que resulta indispensable reducirlo hasta una cantidad manejable, en torno a las 1000 características. Como se verá

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

en la sección correspondiente, la inclusión de demasiadas características puede suponer un procesamiento de varios días o semanas, o incluso ser imposible de realizar por problemas de memoria en el equipo.

La reducción del número de características se realiza a través de la prueba sucesiva de los diferentes parámetros a estudiar. En la figura 5.10 se pueden ver las diferentes opciones disponibles para la realización de dicha prueba. Será necesario elegir el número de repeticiones aleatorizadas que se llevarán a cabo, el umbral requerido para considerar una característica como válida y el porcentaje de registros o individuos que se utilizarán como referencia. Por último existe la posibilidad de requerir que las características superen todas y cada una de las repeticiones o que sea solamente un porcentaje de ellas, modificable en el campo correspondiente.

Selección de características significativas

Numero de permutaciones: 20

Umbral P-Valor: 0.05

% de Individuos para Test: 50 %

Pasarse Todos Pasarse %

95 %

Realizar Selección

Borrar Seleccionados | Guardar Lista | Cargar Lista

1-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = x

2-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = y

3-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = z

4-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = comp

5-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = x

6-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = y

7-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = z

8-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = comp

9-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = x

10-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = y

11-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = z

12-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = comp

13-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = x

14-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = y

15-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = z

16-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = vigilia | canales = comp

17-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = x

18-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = y

19-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = z

20-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = comp

21-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = x

22-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = y

23-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = z

24-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = comp

25-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = x

26-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = y

27-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = z

28-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = comp

29-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = x

30-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = y

31-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = z

32-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = descanso | canales = comp

33-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = x

34-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = y

35-estacionalizado | Estadísticos | diezmados = 1s | estadísticos = media | intervalos = día | canales = z

Total = 54288

Atras | Reducir Caracteristicas | Siguiete

Figura 5.10: Selección de características por Test P-Valor

Con los parámetros correctamente establecidos, el botón **Realizar Selección** comenzará la prueba. Es posible seguir el proceso desde la ventana emergente, así como detener la ejecución en cualquier momento cerrando dicha ventana. Una vez concluido el test los resultados se mostrarán en pantalla, como queda reflejado en la figura 5.11. En esta ventana se muestran los parámetros que han superado el test ordenados de mejor a peor según su validez para el sistema y acompaña a cada uno de ellos su resultado en la prueba. Cerrando esta ventana aparecerá un mensaje emergente que permitirá decidir entre aplicar la selección de características realizada por la prueba o ignorar los resultados, manteniendo la lista anterior. En cualquiera de las dos circunstancias la lista quedará guardada para poder ser cargada en cualquier momento.

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

Índice	ValoresPTEST
36836--deestacionalizado CTM rho = 37 diezmadros = 60s intervalos = dia canales = comp	7.0745e-06
36612--deestacionalizado CTM rho = 36 diezmadros = 60s intervalos = dia canales = comp	7.5361e-06
38404--deestacionalizado CTM rho = 44 diezmadros = 60s intervalos = dia canales = comp	8.8122e-06
37508--deestacionalizado CTM rho = 40 diezmadros = 60s intervalos = dia canales = comp	9.0936e-06
36308--deestacionalizado CTM rho = 35 diezmadros = 60s intervalos = dia canales = comp	9.6756e-06
37956--deestacionalizado CTM rho = 42 diezmadros = 60s intervalos = dia canales = comp	9.6779e-06
33160--deestacionalizado CTM rho = 43 diezmadros = 60s intervalos = dia canales = comp	9.6779e-06
32770--deestacionalizado CTM rho = 19 diezmadros = 30s intervalos = dia canales = y	1.0298e-05
36164--deestacionalizado CTM rho = 34 diezmadros = 60s intervalos = dia canales = comp	1.0298e-05
32322--deestacionalizado CTM rho = 17 diezmadros = 30s intervalos = dia canales = y	1.0956e-05
37732--deestacionalizado CTM rho = 41 diezmadros = 60s intervalos = dia canales = comp	1.0956e-05
39940--deestacionalizado CTM rho = 51 diezmadros = 30s intervalos = dia canales = comp	1.1297e-05
39716--deestacionalizado CTM rho = 50 diezmadros = 30s intervalos = dia canales = comp	1.1654e-05
40164--deestacionalizado CTM rho = 52 diezmadros = 30s intervalos = dia canales = comp	1.1654e-05
46660--deestacionalizado CTM rho = 81 diezmadros = 30s intervalos = dia canales = comp	1.2016e-05
32573--deestacionalizado CTM rho = 18 diezmadros = 60s intervalos = dia canales = y	1.2391e-05
39452--deestacionalizado CTM rho = 49 diezmadros = 30s intervalos = dia canales = comp	1.2394e-05
53765--deestacionalizado Percentiles intervalos nulos percentiles = 25% diezmadros = 300s intervalos = vigiliaT canales = y	1.2514e-05
32994--deestacionalizado CTM rho = 20 diezmadros = 30s intervalos = dia canales = y	1.3176e-05
41316--deestacionalizado CTM rho = 57 diezmadros = 60s intervalos = dia canales = comp	1.3176e-05
32546--deestacionalizado CTM rho = 18 diezmadros = 30s intervalos = dia canales = y	1.3179e-05
40308--deestacionalizado CTM rho = 53 diezmadros = 30s intervalos = dia canales = comp	1.3179e-05
40836--deestacionalizado CTM rho = 55 diezmadros = 30s intervalos = dia canales = comp	1.3179e-05
39268--deestacionalizado CTM rho = 48 diezmadros = 30s intervalos = dia canales = comp	1.3506e-05
23775--deestacionalizado CTM rho = 11 diezmadros = 60s intervalos = vigilia canales = z	1.3661e-05
33218--deestacionalizado CTM rho = 21 diezmadros = 30s intervalos = dia canales = y	1.4011e-05
45316--deestacionalizado CTM rho = 75 diezmadros = 30s intervalos = dia canales = comp	1.4011e-05
12796--deestacionalizado CTM rho = 51 diezmadros = 30s intervalos = dia canales = comp	1.4877e-05
41092--deestacionalizado CTM rho = 56 diezmadros = 60s intervalos = dia canales = comp	1.4896e-05
38596--deestacionalizado CTM rho = 45 diezmadros = 30s intervalos = dia canales = comp	1.4894e-05
41540--deestacionalizado CTM rho = 58 diezmadros = 60s intervalos = dia canales = comp	1.4894e-05
28776--deestacionalizado CTM rho = 1 diezmadros = 60s intervalos = vigilia canales = comp	1.5296e-05
11260--deestacionalizado CTM rho = 44 diezmadros = 60s intervalos = dia canales = comp	1.5797e-05
13020--deestacionalizado CTM rho = 52 diezmadros = 30s intervalos = dia canales = comp	1.5816e-05
38820--deestacionalizado CTM rho = 46 diezmadros = 30s intervalos = dia canales = comp	1.5822e-05
5402--deestacionalizado CTM rho = 18 diezmadros = 30s intervalos = dia canales = y	1.5825e-05
5626--deestacionalizado CTM rho = 19 diezmadros = 30s intervalos = dia canales = y	1.5825e-05
38628--deestacionalizado CTM rho = 45 diezmadros = 60s intervalos = dia canales = comp	1.5829e-05

Figura 5.11: Resultados de prueba de selección

Además de esta primera selección es posible reducir aún más el número de características de la lista actual. Para ello, pulsando el botón **Reducir Características** se accede a una nueva pantalla observable en la figura 5.12 que permite elegir el número final de características con las que trabajar. Introduciendo un número final de características inferior al total disponible en la lista es posible elegir aleatoriamente las características con las que trabajar (botón **Seleccionar Aleatoriamente**) o elegir aquellas con un mejor resultado en la prueba anterior (botón **Seleccionar por Ranking**). Esta última opción solo estará disponible después de realizar un test de validez o cuando se trabaje con una lista cargada de manera manual, siempre y cuando la lista haya sido guardada con información de una prueba de validez anterior. En la siguiente sección se verán más en profundidad estos detalles.

Al finalizar esta sección pulsando en **Siguiente** se mostrarán diferentes mensajes de aviso en caso de encontrar alguna sugerencia. Se avisará si no se ha realizado una prueba de selección, o si se ha realizado pero se han incluido nuevas características a la lista resultante. También aparecerá un aviso en caso de haber elegido un número demasiado elevado de características ya que esto puede ocasionar que el siguiente proceso se demore demasiado o que no se pueda realizar debido a problemas de memoria insuficiente.

5.2.5. MANEJO DE LA LISTA DE CARACTERÍSTICAS

Como ya se ha visto, las características elegidas se muestran en formato lista tanto en la pantalla de Elección de características como en la pantalla de Selección de características discriminantes. Esta lista se mantiene en el paso entre pestañas, y permite tener siempre la referencia de características elegidas así como modificarla.

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

Reducir numero de características

Numero de características a elegir entre las disponibles en la lista.

1000

Seleccionar Aleatoriamente

Seleccionar por Ranking

Cancelar

Figura 5.12: Reducción del número de características

Las opciones que permite la lista son las siguientes:

- **Borrar:** Esta opción elimina de la lista las características elegidas de la misma. Es posible eliminar una o más características, eligiendo en la lista diferentes elementos con los botones **Ctrl** y **Alt**.
- **Guardar:** La opción guardar almacena la lista actual en un fichero externo para poder ser utilizado en futuras ejecuciones del sistema. Si la lista contiene información de una prueba de validez, ésta será también guardada.
- **Guardar:** Mediante la opción cargar se puede elegir una lista guardada anteriormente, ya sea fruto de una prueba de validez o por un guardado manual. Si la lista cargada incluye información de una prueba de validez, ésta será cargada con la lista. Es posible elegir entre sustituir la lista actual con la cargada del fichero o añadir las características cargadas a las ya presentes en la lista.

En lo referente a la información de una prueba de validez, ésta se mantendrá paralelamente a la lista a pesar de realizar una reducción de su número, ya sea de manera aleatoria, de manera manual con el botón **Borrar** o con la reducción por ranking. La información se perderá solamente si se añade una o varias características nuevas de forma manual o si se carga y añade una lista externa y ésta no contiene la información necesaria.

5.2. SELECCIÓN DE CARACTERÍSTICAS DISCRIMINANTES Y ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

5.2.6. ANÁLISIS DE LAS PRESTACIONES DE LOS CLASIFICADORES

La última pantalla de esta sección corresponde con la prueba de clasificadores para hallar sus prestaciones. En ella se realizan sucesivas repeticiones de entrenamiento y prueba con los datos seleccionados en las pantallas anteriores, tanto de registros de individuos como de sus características. Se muestra la pantalla de prueba en la figura 5.13.

The screenshot shows a web-based configuration interface for testing classifiers. At the top, there are several input fields: 'Dimension Clasificadores' with the value 2, 'Repeticiones' with 20, 'Numero Casos Train' with 58, 'Numero Casos Test' with 10, 'Numero Controles Train' with 16, and 'Numero Controles Test' with 48. Below these is a checkbox for 'Hacer PCA' which is checked. A central panel contains a list of classifier options: 'Clasificador KNN' (checked), 'Clasificador Arbol' (unchecked), 'Clasificador SVM' (checked), 'Clasificador Random Forest' (checked), and 'Clasificador Red Neuronal' (unchecked). To the right of this list, there are settings for the KNN classifier: 'Vecinos' is set to 23, 'Metodo Lineal' is checked, 'Metodo RBF' is unchecked, and 'Metodo Quadratic' is checked. A 'Realizar Prueba' button is located on the right side of the interface, and an 'Atras' button is at the bottom left.

Figura 5.13: Prueba de clasificadores

La prueba de desempeño consiste en entrenar diferentes clasificadores con las múltiples características elegidas, ya sea individualmente o en dimensiones mayores según lo introducido en el campo **Dimensión**. Los datos de entrenamiento se escogerán aleatoriamente de entre los individuos disponibles, tanto casos como controles y se probará con los restantes. Es posible seleccionar cuántos casos y controles se elegirán en cada repetición para entrenar y probar los clasificadores mediante los controles deslizantes, existiendo un límite mínimo de un individuo para cada opción.

Como ya se ha comentado, el campo **Dimensión** permite elegir el tamaño de los clasificadores en cuanto a características diferentes a utilizar en cada ejecución. Eligiendo un número mayor que 1 habilita la formación de grupos, ya sean parejas, tríos, etc. los cuales se formarán a partir de las características disponibles. Como cabe suponer el número de agrupaciones diferentes crece rápidamente con el aumento de la dimensión, pasando por ejemplo de 1000 agrupaciones individuales a $\binom{1000}{2} = 499500$ parejas, $\binom{1000}{3} = 166167000$ tríos o $\binom{1000}{4} = 1,417e + 10$ cuartetos. Por supuesto, un mayor número de opciones implica un mayor tiempo de procesado.

La opción **PCA** permite elegir si se realizará o no este procesado de los datos antes de realizarse la prueba. En caso de habilitarse la función, el procesado se hará tanto con la opción como sin ella, permitiendo la observación de diferencias entre los dos métodos.

5.3. LECTURA DE RESULTADOS Y SELECCIÓN DE LOS MEJORES CONJUNTOS CLASIFICADORES

En cuanto a los clasificadores disponibles, es posible habilitar aquellos que se quieran utilizar así como introducir los parámetros requeridos en el caso de que existan. En los clasificadores que dispongan de entrada manual de parámetros es posible introducir valores unitarios o un conjunto de valores, separando estos con comas o eligiendo un intervalo completo. Por ejemplo, para introducir los valores desde 20 hasta 50 con saltos de 5 se escribiría la secuencia «20:5:50». Al introducir más de un valor en un mismo campo el sistema los interpretará como clasificadores diferentes y realizará la prueba para tantos clasificadores como valores se introduzcan.

El botón **Realizar Prueba** comenzará la ejecución del procesado, apareciendo una ventana emergente en la que observar el progreso y el tiempo restante estimado, así como cancelar la ejecución cerrando esta ventana. La carga computacional de este procesado es muy grande y puede implicar grandes cantidades de tiempo, del orden de días, por lo que se recomienda realizar la prueba con un número de características no demasiado grande, de manera que el conjunto de agrupaciones de características no sea mayor de quinientos mil.

Al finalizar la prueba se permite visualizar los resultados o volver a la ventana de introducción de parámetros de la prueba para poder realizar un nuevo procesado. En cualquier caso, los resultados de la prueba quedarán guardados para cualquier futura revisión, la cual se realiza a través de la siguiente sección del programa.

5.3. LECTURA DE RESULTADOS Y SELECCIÓN DE LOS MEJORES CONJUNTOS CLASIFICADORES

Una vez realizadas las pruebas de validez de los clasificadores es posible consultar los resultados obtenidos. Para ello se cargará una serie de estadísticos a través de los cuales es posible determinar los mejores sistemas clasificadores dentro de la gran variedad existente. Para ello se muestra tanto la composición del clasificador, referido a las características que lo componen, el tipo de clasificador empleado y la realización o no de PCA, como los estadísticos calculados en base a la serie de pruebas realizadas sobre cada uno. Estos estadísticos se refieren a la media y mediana de la precisión, sensibilidad y especificidad, pudiéndose ordenar en función de cada una de ellas. Se podrán además guardar los primeros de la lista para poder utilizarlos en la detección del trastorno sobre pacientes sin clasificar.

En la figura A.5 se observa la interfaz de lectura de resultados, la cual permite cargar un fichero de resultados y calcular los distintos datos de referencia que evalúan las prestaciones de los conjuntos clasificadores.

El botón **Explorar** permite elegir el fichero de resultados a visualizar y, una vez que su información ha sido cargada y procesada, se mostrará en la tabla disponible. Los campos representados se refieren a la información del clasificador, en este caso el tipo de sistema clasificador, parámetros internos del mismo, características utilizadas en su entrenamiento y prueba, uso o no de PCA en la prueba y una información sobre la posición que ocupa en el registro. Además se muestran los estadísticos calculados a partir de los resultados obtenidos, es decir, la media y mediana de la precisión, especificidad y sensibilidad, expresados en tanto

5.4. DETECCIÓN DE TDAH SOBRE REGISTROS SIN CLASIFICAR

Grupo de Características		Guardar clasificadores							
		Precisión		Sensibilidad		Especificidad			
		MEAN	MEDIAN	MEAN	MEDIAN	MEAN	MEDIAN		
10779--estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM	9819	PCA	0.75076	0.75	0.77353	0.79412	0.72656	0.71875
36606--desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	8805	PCA	0.74773	0.74242	0.79706	0.79412	0.69531	0.71875
10331--estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM	8802	PCA	0.74621	0.75758	0.78088	0.76471	0.70937	0.71875
36382--desestacionalizado CTM rho = 36 diezmados = 30s intervalos = dia canales = z	SVM_Q	8802	PCA	0.74545	0.74242	0.75441	0.76471	0.73584	0.73438
10331--estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM_Q	9819	PCA	0.74545	0.74242	0.73676	0.75	0.75469	0.71875
33470--desestacionalizado CTM rho = 23 diezmados = 30s intervalos = dia canales = z	Arbol	19665	PCA	0.7447	0.75	0.72059	0.69118	0.77031	0.75
53531--desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = z	SVM_RBF	8802	PCA	0.74242	0.75	0.78676	0.79412	0.69531	0.70313
36382--desestacionalizado CTM rho = 36 diezmados = 30s intervalos = dia canales = z	SVM_RBF	9819	PCA	0.74242	0.75	0.72059	0.73529	0.76563	0.75
10779--estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM_RBF	10325	PCA	0.74015	0.74242	0.75882	0.76471	0.72031	0.71875
11227--estacionalizado CTM rho = 44 diezmados = 30s intervalos = dia canales = z	SVM	10822	PCA	0.73939	0.74242	0.76176	0.73529	0.71562	0.70313
37054--desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM_RBF	8805	PCA	0.73864	0.75758	0.77353	0.79412	0.70156	0.71875
10331--estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM	10325	PCA	0.73788	0.74242	0.75294	0.73529	0.72188	0.70313
36606--desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM_RBF	9824	PCA	0.73788	0.72727	0.74265	0.76471	0.73281	0.75
11003--estacionalizado CTM rho = 43 diezmados = 30s intervalos = dia canales = z	SVM	5661	PCA	0.73712	0.73485	0.78382	0.79412	0.6875	0.70313
37054--desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	Arbol	1707	PCA	0.73712	0.74242	0.70441	0.72059	0.77187	0.76563
10779--estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM	3473	PCA	0.73636	0.73485	0.73971	0.73529	0.73281	0.75
9887--estacionalizado CTM rho = 34 diezmados = 30s intervalos = dia canales = z	SVM	10331	PCA	0.73409	0.72727	0.77647	0.77941	0.68906	0.6875
36158--desestacionalizado CTM rho = 35 diezmados = 30s intervalos = dia canales = z	SVM	12454	PCA	0.73333	0.72727	0.84118	0.85294	0.61875	0.625

Figura 5.14: Resultados validación cruzada

por uno.

Los botones que aparecen encima de los estadísticos tienen la función de ordenar los resultados en función del parámetro seleccionado que corresponda, de forma descendente. Con esto se consigue distinguir los mejores clasificadores considerando los criterios de precisión, sensibilidad o especificidad.

Una vez consultados los resultados y ordenados según el criterio elegido, los mejores conjuntos se pueden almacenar para ser utilizados en la última sección del programa, es decir, para hacerlos funcionar en situaciones reales de detección de TDAH con registros de individuos sin clasificar. Para guardar estos registros es necesario ordenarlos por el parámetro que se requiera, introducir en el campo superior el número de registros diferentes a almacenar y pulsar el botón **Guardar**. Con esto quedará extraído un registro de las particularidades de cada conjunto elegido y su funcionamiento podrá ser replicado en el futuro.

5.4. DETECCIÓN DE TDAH SOBRE REGISTROS SIN CLASIFICAR

El último apartado del programa se ha destinado al empleo de la herramienta como medio de diagnóstico, es decir, a la clasificación de registros de datos actigráficos en pacientes sanos o patológicos.

En la figura 5.15 se observan las diferentes partes de esta sección. En primer lugar se encuentra la carga de la base de datos de registros actigráficos, los cuales se utilizarán como

5.4. DETECCIÓN DE TDAH SOBRE REGISTROS SIN CLASIFICAR

Clasificación Individuos - Selección Features y Clasificación

Fichero Datos Clasificación Features-05-31-2017_11-40.mat

Clasificadores MejoresClasificadores-06-18-2017_17-40.mat

Mejores 20 clasificadores ordenado por la mediana de la precision

Individuo (Datos Actigraficos) x05.dat

10331->estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
36382->desestacionalizado CTM rho = 36 diezmados = 30s intervalos = dia canales = z	SVM_RBF	PCA	CONTROL
10331->estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
33470->desestacionalizado CTM rho = 23 diezmados = 30s intervalos = dia canales = z	Arbol	PCA	CONTROL
53581->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = z	SVM	PCA	CONTROL
10779->estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
10331->estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM_RBF	PCA	CASO
36382->desestacionalizado CTM rho = 36 diezmados = 30s intervalos = dia canales = z	SVM_RBF	PCA	CASO
10779->estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
6747->estacionalizado CTM rho = 24 diezmados = 30s intervalos = dia canales = z	Arbol	PCA	CONTROL
53581->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = z	Arbol	PCA	CONTROL
33023->desestacionalizado CTM rho = 21 diezmados = 30s intervalos = dia canales = z	Arbol	PCA	CONTROL
53506->desestacionalizado Percentiles intervalos nulos percentiles = 75% diezmados = 300s intervalos = vigilia canales = comp	SVM	PCA	CONTROL
39294->desestacionalizado CTM rho = 49 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
53582->desestacionalizado Percentiles intervalos nulos percentiles = 25% diezmados = 300s intervalos = vigilia canales = comp	SVM	PCA	CONTROL
10331->estacionalizado CTM rho = 40 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
36506->desestacionalizado CTM rho = 37 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
10779->estacionalizado CTM rho = 42 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
11903->estacionalizado CTM rho = 43 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
37054->desestacionalizado CTM rho = 39 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL
11227->estacionalizado CTM rho = 44 diezmados = 30s intervalos = dia canales = z	SVM	PCA	CONTROL

Figura 5.15: Detección de TDAH en pacientes

referencia. Se trata del fichero generado en la primera parte del sistema, tratada en el capítulo 5.1.

En segundo lugar se encuentra la carga de información acerca de los clasificadores a utilizar, la cual ha sido obtenida en la sección del programa del capítulo 5.3. Una vez cargados, los clasificadores se entrenarán o configurarán con los registros de referencia de la base de datos.

Por último queda la carga de datos del paciente. A través de esta función se selecciona un fichero actigráfico sin procesar, es decir, el registro que se obtiene directamente del medidor de actividad. El sistema procesará dicho registro para obtener las características necesarias para su clasificación.

Una vez cargados todos los datos, pulsando el botón **Realizar Prueba** se iniciará el proceso de clasificación, mostrándose el resultado en la ventana de información. En ella aparecerán tanto los datos de los diferentes clasificadores utilizados como la predicción que realiza cada uno. La existencia de varios conjuntos clasificadores (la cantidad elegida en el momento de su guardado) hace posible tener una visión más abierta en cuanto a la predicción obtenida, pudiéndose observar diferentes predicciones para una misma prueba y pudiendo decidir el resultado en función de diferentes puntos de vista.

En este punto de ejecución es posible repetir la prueba tanto con otros datos actigráficos, es decir, con otro individuo o modificando los clasificadores con los que trabajar, cargando un nuevo registro de resultados.

CAPÍTULO 6

CONCLUSIONES Y LÍNEAS FUTURAS

Este capítulo recoge las conclusiones extraídas durante el desarrollo del sistema. Como punto final se mencionan diferentes modificaciones o ampliaciones que tienen cabida en el sistema desarrollado, en lo referente al tipo de programación, adición de funcionalidades o adecuación de la interfaz.

6.1. CONCLUSIONES

El propósito del presente trabajo es la implementación de un entorno gráfico para el estudio y construcción de sistemas detectores de la patología TDAH en la infancia a partir de datos actigráficos. Se ha desarrollado un sistema integrado que reúne las funciones de procesado, calificación y prueba y que es completamente configurable e interactivo, adecuándose a las necesidades del usuario.

Tomando como referencia las funcionalidades requeridas y los criterios de diseño detallados en el capítulo 2.1, la herramienta final resulta intuitiva y fácilmente manejable con una curva de aprendizaje adecuada que permite la realización de las tareas básicas dentro del contexto en el que se encuentra.

En cuanto a las funciones añadidas, por un lado se han estudiado diferentes características o patrones extraíbles de las señales de actividad, obteniéndose resultados diversos. Las características referidas en [4] tales como la Variabilidad Intradiaria o el Índice Dicotómico presentan una implementación sencilla con el sistema desarrollado, tomando como base para su cálculo los tiempos de inicio y fin del sueño o los registros de actividad de diferentes intervalos. De igual manera se han creado las características conocidas como Percentiles de Intervalos Nulos, los cuales presentan una gran sensibilidad en las pruebas realizadas con ellos.

Por otro lado, los clasificadores Random Forest y Redes Neuronales han presentado resul-

tados inferiores a los esperados debido a un factor determinante: el tiempo de procesado. La carga computacional de estos sistemas de clasificación hace muy difícil su implementación en pruebas como la validación cruzada, en la cual se debe realizar un número elevado de repeticiones. Debido a esto ha sido necesario disminuir el número de árboles en el caso del Random Forest, y de capas en las Redes Neuronales, haciendo que las pruebas no tuviesen gran valor por no trabajar con las mejores configuraciones posibles de cada sistema. En cuanto a los resultados, su eficiencia se encuentra por debajo de los sistemas SVM, los cuales presentan precisiones de hasta 85 %.

6.2. LINEAS FUTURAS

6.2.1. LENGUAJE DE PROGRAMACIÓN EMPLEADO

El sistema se ha desarrollado en el lenguaje matemático MATLAB, lo que permite el uso de infinidad de métodos especializados en el procesamiento matricial con un rendimiento muy alto. Por contra, se trata de un lenguaje propietario que requiere disponer de la correspondiente licencia de uso e impide la utilización del sistema fuera de su propio marco de trabajo.

Actualmente existen múltiples lenguajes de uso general con los que resulta sencillo diseñar programas destinados al usuario final y que son accesibles desde cualquier plataforma, incluso con versiones de acceso vía web. Esto permite tener una mayor flexibilidad en la ejecución a costa de perder rendimiento en el procesado de datos.

El lenguaje Python es uno de los mejores candidatos al respecto. Se trata de un entorno altamente utilizado en la actualidad con una gran cantidad de soporte y librerías complementarias. Su flexibilidad unida a la nueva tendencia del estudio de Big Data hacen que existan múltiples métodos matemáticos en este lenguaje que tratan de emular el funcionamiento matemático de MATLAB y que podrían permitir la implementación de un sistema como el del presente estudio en este tipo de lenguajes.

6.2.2. AMPLIACIÓN DE CARACTERÍSTICAS

Como se mencionó en el capítulo 4, el sistema permite extraer nuevos parámetros o características de las señales actigráficas mediante la inclusión de sus métodos en un script. En este sentido resulta claramente visible la necesidad de añadir nuevas características con posibilidades de resultar útiles para la detección de la patología, las cuales se procesarán por el sistema y se obtendrán sus capacidades de diagnóstico. En el presente trabajo se han añadido una serie de nuevos parámetros con diferentes resultados, aunque el número de posibilidades de ampliación al respecto es prácticamente infinito.

6.2.3. ADAPTACIÓN DE NUEVOS CLASIFICADORES

La existencia de múltiples modelos de clasificación especializados en el procesado de diferentes estructuras de datos hace necesario disponer de un método sencillo para ampliar el sistema en este sentido. Su implementación permitiría probar nuevos sistemas de clasificación sin la necesidad de cambiar el código interno del programa, sencillamente aportando los métodos relacionados con el entrenamiento y prueba de cada clasificador.

Su implementación sería similar a la empleada en la adición de nuevas características, es decir, habría que adaptar el sistema para que leyese diferentes scripts con el código adecuado para cada clasificador. En este caso sería necesario codificar dos métodos diferentes, uno para el entrenamiento de los clasificadores y otro para la prueba, identificando en cada caso los parámetros de configuración necesarios. El sistema se deberá modificar para que sea capaz de adaptar la propia interfaz gráfica adecuándola a los requerimientos de cada clasificador, permitiendo la introducción de los parámetros configurables en cada uno de ellos, así como posibilitando la lectura de esas opciones para poder replicarlas a partir de una lista de resultados.

APÉNDICE A

MANUAL DE REFERENCIA

En este anexo se presenta la estructura interna del sistema así como los métodos a emplear para añadir nuevas características. Se realiza una aproximación a la técnica utilizada para la identificación de características dentro de la matriz total y se especifica el formato de ciertos ficheros de guardado para poder efectuar su estudio de manera externa al programa.

A.1. ESTRUCTURA DE DIRECTORIOS

El sistema completo consta de cinco directorios básicos en los que se encuentra el código necesario para su funcionamiento, a los que hay que sumar el directorio donde se decidan almacenar los diferentes ficheros de guardado, el cual se seleccionará dentro de la propia interfaz en tiempo de ejecución. La figura A.1 muestra el contenido del directorio raíz del sistema.

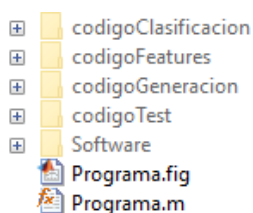


Figura A.1: Estructura de directorios del sistema

El menú de inicio desde el que se accede a las demás funciones es el conocido como **Programa.m** y se encuentra en la raíz, acompañado de su correspondiente fichero de interfaz.

El resto del sistema está estructurado en las diferentes funcionalidades que lo componen. Cada una de las secciones está compuesta por dos ficheros principales (los relativos a la interfaz) así como por un conjunto de ficheros auxiliares que se utilizan en ese apartado.

Por una parte se encuentra el directorio **codigoGeneración**, en el que se encuentran las funciones de análisis de registros actigráficos y extracción de características. En este directorio se encuentra el fichero **GeneracionFeats.m** acompañado de su fichero de interfaz, así como otros ficheros auxiliares. El directorio **codigoFeatures** es dónde se encuentran los scripts de generación de características propiamente dichos, los cuales se podrán modificar y añadir en función de las necesidades del usuario, según se detallará en la sección A.2. En este directorio aparecerán también los ficheros de *log* consecuencia de cualquier fallo de programación en los scripts de extracción, en los que se podrá ver el error encontrado.

El directorio **codigoClasificacion** reúne las funcionalidades de prueba de clasificadores y visualización de los resultados. Por una parte el fichero **SistemaClasificacion.m** es el encargado de realizar la selección de características y la prueba de eficiencia de los clasificadores, y por otra parte el fichero **AnalisisResultados.m** es el encargado de mostrar los resultados obtenidos.

La última parte del programa se encuentra en el directorio **codigoTest** y se trata de la herramienta de uso diagnóstico que permite clasificar pacientes a partir de su registro actigráfico.

Por último, el directorio **Software** contiene diferentes ficheros de código utilizados por el programa y en su mayoría corresponden a ficheros de cálculo de variables o procesamiento de registros actigráficos que se utilizan de forma auxiliar durante el procesamiento y generación de características.

A.2. CREACIÓN DE SCRIPTS PARA LA EXTRACCIÓN DE NUEVAS CARACTERÍSTICAS

La inclusión de nuevas características se ha simplificado para poderse realizar sin necesidad de entrar en el código interno del programa, sencillamente añadiendo un nuevo fichero de extracción a los ya existentes. Este fichero deberá aportar como salida el nombre del fichero temporal donde se ha guardado el vector de característica extraídas así como la estructuración de las mismas.

Se va a detallar cada elemento que debe contener el fichero de generación utilizando como referencia el script encargado de obtener las características *percentiles de intervalos nulos*.

A.2.1. VARIABLES DE ENTRADA Y SALIDA

La sentencia inicial incluye las variables que se tomarán como salidas y las variables de entrada que el sistema aporta al script.

$$\text{function } [outputFeat,estructura] = \text{ObtenerPRCTLs}(ficheroFragmentos, tipo, origin, saveDirFeat)$$

Todos los scripts siguen el mismo esquema de variables. Por un lado, como entrada se reciben los nombres de los ficheros de datos actigráficos preprocesados ('ficheroFragmentos'), el tipo de paciente al que pertenecen los datos, ya sea 'Caso' o 'Control', el nombre del fichero de registro del que se obtienen los datos ('origin') y el directorio temporal donde se deben almacenar las características extraídas ('saveDirFeat'). Por otro lado, como salidas se tiene el nombre del fichero temporal el que se han guardado las características calculadas ('outputFeat') y la estructura de las mismas ('estructura'). En los sucesivos apartados se detallará la composición de las variables mencionadas.

Cabe destacar que el nombre del fichero de generación deberá seguir el formato (*Obtener*.m) ya que será la secuencia que buscará el sistema a la hora de localizar los ficheros disponibles en el directorio. Esto se utiliza como medida de seguridad para evitar que el sistema ejecute por error cualquier fichero del directorio elegido, fruto de una confusión del usuario. El sistema prueba uno a uno los ficheros encontrados ejecutándolos, lo que puede ocasionar acciones no esperadas al utilizar scripts con código que, por ejemplo, elimine ficheros. La ejecución de código no correspondiente haría que el sistema lo detectase como erróneo y no lo tendría en cuenta para obtener características, pero podrían producirse acciones desconocidas que el usuario no quiera realizar.

A.2.2. LECTURA DEL FICHERO DE DATOS CORRESPONDIENTE

Existen tres ficheros de los que se pueden obtener diferentes datos. La variable que los contiene es un vector de tipo *cell* con tres elementos correspondientes a los tres nombres

A.2. CREACIÓN DE SCRIPTS PARA LA EXTRACCIÓN DE NUEVAS CARACTERÍSTICAS

de ficheros. El primero de ellos contiene todas las muestras obtenidas del registro con sus cuatro canales diferentes (tamaño $n \times 4$) en la variable llamada ‘Casos’ y la hora de inicio de la lectura de datos en formato texto en la variable ‘HoraIniCasos’. El segundo fichero contiene la variable ‘FechaCasos’ en la que se encuentra la fecha en la que se capturaron los datos, en formato texto. Cabe destacar que aunque el nombre de las variables anteriores se refiere a ‘Casos’ su contenido corresponde a un fichero de caso o controles en función del individuo al que pertenece, usándose el mismo nombre en ambos casos.

Por último, el tercer fichero contiene la variable ‘senales’ de tipo *struct* en la que se incluyen múltiple datos, algunos de los cuales ya contenidos en los anteriores ficheros, cuya estructura se muestra en la figura A.2.

Dia	86400x4 double
Vigilia	47165x4 double
VigiliaManana	23583x4 double
VigiliaTarde	23582x4 double
Descanso	39235x4 double
DescansoIni	13078x4 double
DescansoMed	13077x4 double
DescansoFin	13079x4 double
Inicio	'09:30:00'
Fecha	'10/02/2010'
inicioSueno	1x1 datetime
finSueno	1x1 datetime

Figura A.2: Estructura de la variable ‘senales’

Por una parte, la variable ‘senales’ contiene los datos actigráficos correspondientes a los diferentes intervalos de tiempo en los que se divide la señal original (‘Dia’, ‘Vigilia’, ‘VigiliaManana’, ‘VigiliaTarde’, ‘Descanso’, ‘DescansoIni’, ‘DescansoMed’ y ‘DescansoFin’) con los correspondientes canales. También incluye la hora de inicio de captura y la fecha en formato texto (‘Inicio’ y ‘Fecha’) y los instantes de inicio y fin del sueño en formato *datetime*, los cuales permiten obtener características relativas a la etapa de sueño.

Sabiendo qué datos se van a necesitar, el primer paso será cargar el fichero correspondiente, por ejemplo con la sentencia `load(ficheroFragmentos{3})` se cargaría en memoria la variable ‘senales’.

A.2.3. CREACIÓN DE LA ESTRUCTURA DE DATOS

La variable de salida ‘estructura’ de tipo *cell* debe contener los diferentes parámetros de las características extraídas, ordenados en función de cómo se dispongan estas características en el vector final. El ejemplo seguido recorre primero los diferentes canales, después los intervalos, luego los diezmados y por último los parámetros específicos ‘25 %’ y ‘75 %’. Para entender mejor cómo se colocan las características se presenta a continuación la posición que ocupan en el vector:

A.2. CREACIÓN DE SCRIPTS PARA LA EXTRACCIÓN DE NUEVAS CARACTERÍSTICAS

```

1→tipo: 25 % | diezmado: 1s | intervalo: día | canal: X
2→tipo: 25 % | diezmado: 1s | intervalo: día | canal: Y
3→tipo: 25 % | diezmado: 1s | intervalo: día | canal: Z
4→tipo: 25 % | diezmado: 1s | intervalo: día | canal: comp
5→tipo: 25 % | diezmado: 1s | intervalo: vigilia | canal: X
...
9→tipo: 25 % | diezmado: 1s | intervalo: vigiliaM | canal: X
...
13→tipo: 25 % | diezmado: 1s | intervalo: vigiliaT | canal: X
...
33→tipo: 25 % | diezmado: 5s | intervalo: día | canal: X
...
65→tipo: 25 % | diezmado: 30s | intervalo: día | canal: X
...
225→tipo: 75 % | diezmado: 1s | intervalo: día | canal: X
...
448→tipo: 75 % | diezmado: 900s | intervalo: descansoF | canal: comp

```

Se puede ver cómo se realiza el recorrido de los parámetros y éste se deberá tener en cuenta a la hora de crear la estructura y al organizar los valores en forma vectorial.

La estructura identificadora sigue un esquema predefinido, tal y como se muestra en la figura A.3. El primer valor corresponde con el nombre del grupo de características, en este caso 'Percentiles intervalos nulos'. El resto de elementos se distribuyen en orden como el nombre del parámetro, por ejemplo 'diezmados', y un objeto *cell* que contiene las diferentes opciones de dicho parámetro. El orden seguido para la inclusión de los tipos de parámetros es el inverso al que se recorren en el vector, en este caso primero los tipo de percentiles, después los diezmados, los intervalos y por últimos los canales.

'Percentiles intervalos nulos'	'percentiles'	1x2 cell	'diezmados'	1x7 cell	'intervalos'	1x8 cell	'canales'	1x4 cell
--------------------------------	---------------	----------	-------------	----------	--------------	----------	-----------	----------

Figura A.3: Ejemplo de estructura de características

Siguiendo el ejemplo, los valores de los diferentes parámetros y la estructura final se formarán con las sentencias siguientes:

```

canales={ 'x', 'y', 'z', 'comp' };
intervalos={ 'dia', 'vigilia', 'vigiliaM', 'vigiliaT', ...
'descanso', 'descansoI', 'descansoM', 'descansoF' };
diezmados={ '1s', '5s', '30s', '60s', '300s', '600s', '900s' };
percentiles={ '25 %', '75 %' };
estructura={ 'Percentiles intervalos nulos', 'percentiles', percentiles, ...
'diezmados', diezmados, 'intervalos', intervalos, 'canales', canales };

```

A.2.4. CÁLCULO ITERATIVO DE LOS VALORES Y CONCATENACIÓN

Una vez definida la estructura el siguiente paso es calcular iterativamente cada uno de los valores posibles y asignarlos a la variable correspondiente. En algunos casos los parámetros se podrán calcular vectorialmente, por ejemplo, sobre los cuatro canales a la vez o, como en este caso, se obtienen a la vez los dos tipos de percentil ('25%' y '75%'), y esto se deberá tener en cuenta a la hora de componer el vector de salida.

En el ejemplo que se sigue, los valores se guardan en una variable estructura llamada 'Percentil'. Ésta se compone de cada intervalo de señal de la forma 'Percentil.Descanso' o 'Percentil.Vigilia', y dentro de cada uno de ellos se encontrarán todos los posibles valores restantes en cada una de sus dimensiones. Los diferentes intervalos se podrían haber reunido en una dimensión como ocurre con el resto de parámetros, pero se realiza por separado para entender mejor la forma de concatenación. En este caso la dimensión cuarta corresponde a los tipos de percentil, la tercera a los diezmados y la segunda a los canales. Así, para acceder al canal **c** del diezmado **d** del tipo 1 ('25%' en este caso) del intervalo de **DescansoIni** se utilizará la sentencia *Percentil.DescansoIni(1,c,d,1)*. La primera dimensión se mantiene por retrocompatibilidad pero no tiene función en este método.

La concatenación de los intervalos se realiza de forma ordenada según la estructura de datos, en este caso con la sentencia:

```
PRCTLS = [Percentil.Dia, Percentil.Vigilia,...  
Percentil.VigiliaManana, Percentil.VigiliaTarde,...  
Percentil.Descanso, Percentil.DescansoIni,...  
Percentil.DescansoMed, Percentil.DescansoFin];
```

De esta manera se concatenan los canales de cada intervalo con los del siguiente manteniéndose el resto de dimensiones. Para completar la formación del vector se utiliza el método *reshape* el cual pasa la matriz multidimensional a un vector concatenando sus dimensiones, de la forma que se necesita.

```
PRCTLS = reshape(PRCTLS, 1, numel(PRCTLS));
```

A.2.5. GUARDADO EN FICHERO TEMPORAL

Una vez se han obtenido las características y se les ha aplicado el formato vectorial oportuno, el siguiente y último paso es guardarlas en un fichero temporal, el cual será leído por el sistema para poder conformar la agrupación total de características.

El nombre elegido en el proceso de guardado se almacenará en la variable de salida 'outputFeat'. La variable deberá ser una cadena de caracteres que contiene en primer lugar el directorio de guardado temporal, disponible en la variable de entrada 'saveDirFeat', así como las particularidades que se quieran añadir. Por ejemplo, en el script analizado como referencia se utiliza el nombre del tipo de característica 'PRCTLS', el tipo de individuo que se procesa y el fichero de datos del que proviene:

```
outputFeat=strcat(saveDirFeat, 'PRCTLS', tipo, origin);  
save(outputFeat, 'PRCTLS');
```

A.3. IDENTIFICACIÓN DE CARACTERÍSTICAS SEGÚN SU ESTRUCTURA

Las características, una vez obtenidas, se agrupan de forma vectorial en una misma matriz, haciendo imposible su identificación sin ayuda externa. Esta identificación se realiza en los ficheros traductores, ya sea para obtener las propiedades de la característica a partir de su posición en la matriz o determinando dicha posición en función de los parámetros de la característica a la que se refiere.

La estructura de características se repite para el grupo de estacionalizadas y para el de desestacionalizadas, de manera que la pertenencia a uno u otro grupo se determina directamente por situarse en la primera o segunda mitad de conjunto de características, ya que el conjunto final se forma uniendo ambos tipos. Para determinar la pertenencia a un grupo concreto, como CTM o Dinámica Simbólica, se realiza en función de la longitud de cada uno de esos grupos, calculada a partir del número de elementos de la estructura de identificación. Por ejemplo, la estructura del grupo PRCTLS consta de 448 elementos, fruto de la multiplicación de la longitud de sus componentes: $4(\text{canales}) * 8(\text{intervalos}) * 7(\text{diezmados}) * 2(\text{tiposdepercentil}) = 448$. Una vez identificado el tipo de característica, el resto de parámetros se obtienen de forma similar teniendo en cuenta el número de características que contiene cada uno de los grupos de rango inferior.

Para entender el proceso se toma como ejemplo una posible estructura de datos con $4658 * 2$ características (estacionalizadas y desestacionalizadas) entre las que se quiere hallar la correspondiente a la posición 8515.

1. El primer paso es determinar si es una característica estacionalizada o desestacionalizada. Para ello, tomando como referencia el tamaño total de la estructura se ve que se trata del segundo tipo, ya que se encuentra en la segunda mitad del vector total. Una vez sabido esto, el valor de posición de la característica se toma con referencia a la estructura, es decir, entre 1 y 4658, para lo cual es necesario restar de su valor la longitud total de características: $posicion = 8515 - 4658 = \mathbf{3857}$
2. El siguiente paso es determinar el tipo de característica al que pertenece. La estructura del ejemplo contiene 8 tipos de características cuyas longitudes son [1568, 2240, 84, 9, 112, 448, 1, 196], las cuales se obtienen al multiplicar los elementos de los parámetros que lo componen. Se puede apreciar entonces que la característica pertenece al tercer grupo, ya que los dos anteriores suman $1568 + 2240 = 3808$ dejando un valor de $3857 - 3808 = \mathbf{49}$. Se obtiene entonces que la característica pertenece al tercer grupo y ocupa la posición 49 de las 84 que contiene dicho tipo de característica.

A.3. IDENTIFICACIÓN DE CARACTERÍSTICAS SEGÚN SU ESTRUCTURA

3. La obtención del resto de la información depende del tipo de característica en la que se encuentre. En el caso del ejemplo seguido, ese tipo de característica sigue la siguiente estructura:

'Características del Ritmo Circadiano'	'carRit'	1x3 cell	'diezmados'	1x7 cell	'canales'	1x4 cell
--	----------	----------	-------------	----------	-----------	----------

Figura A.4: Estructura de la característica estudiada

Tomando la estructura como referencia y recordando cómo se ordenan las características de los parámetros, es posible obtener las posiciones en cada subgrupo hasta completar los parámetros posibles. Los cuatro primeros datos del vector total corresponden a los 4 canales del primer diezmado y, de igual manera, las $4 * 7 = 28$ primeras posiciones se refieren a los posibles valores de canales y diezmados del primer tipo de parámetro llamado 'carRit'. De esta forma se obtiene que la característica pertenece al segundo valor de 'carRit', ya que deja atrás los 28 valores del primer tipo y se sitúa en la posición $49 - 28 = 21$ de los 28 posibles valores del segundo tipo. Repitiendo el mismo razonamiento con el resto de parámetros se tiene que la característica pertenece al grupo 6 de diezmados y al tipo 2 de canal.

Los índices de grupo obtenidos corresponden con un valor de cada parámetro, como por ejemplo 'diezmado 1s' o 'canal Z' y ese significado es fácilmente obtenible a partir de la estructura de referencia, ya que contiene todos los valores con su significado correspondiente.

En caso de realizarse el proceso inverso, se parte de los diferentes índices y se calcula el total correspondiente a la posición. Para ello se van sumando los posibles valores que deja atrás de la forma siguiente:

$$1(\text{canales que deja atrás}) + 5(\text{canales}) * 4(\text{grupos de diezmados que deja atrás}) + 28(\text{diezmados y canales}) * 1(\text{grupos del parámetro 'carRit' que deja atrás}) + 3808(\text{características previas al tipo 3}) + 4658(\text{características estacionalizadas}) = 8515$$

Para finalizar cabe destacar que el proceso anterior se refiere a una matriz de datos que contenga todos los elementos especificados por la estructura. Este caso no se suele presentar, ya que durante el procesado de los registros se realiza una eliminación de aquellas características nulas en todos los individuos analizados. Como consecuencia, la matriz final no contiene todos los valores posibles ya que algunos de los intermedios han sido eliminados. La información de las características eliminadas se mantiene de forma paralela a la estructura de características ya que resulta necesaria para la identificación. Esta información es un vector, de longitud el número total de características de la estructura, que contiene en sus primeros valores las posiciones de las características que no han sido eliminadas, completando con valores nulos hasta la longitud total. Así, si se han eliminado las características número 3 y 7 el vector contendrá $[1, 2, 4, 5, 6, 8, \dots]$. Esto permite obtener fácilmente la posición actual de cualquier característica consultando la posición que ocupan en el vector, o determinar qué característica original se encuentra en una determinada posición de la matriz consultando el valor del vector en dicha posición.

A.4. ESTRUCTURA DE LOS FICHEROS DE RESULTADOS DE LAS PRUEBAS DE CLASIFICADORES

Es posible que resulte necesario comprobar de forma manual los resultados de la prueba de validación cruzada para ver con precisión los datos de cada una de las iteraciones que se realizan. Para realizar esta revisión será necesario cargar el fichero de resultados con formato ‘*.dat’ en el entorno MATLAB con el método *load*.

El fichero contiene diferentes variables con información necesaria para el cálculo de los resultados y para la obtención de los parámetros utilizados por los clasificadores, además de las variables de resultados. Los resultados se almacenan en matrices separadas para cada clasificador y para cada estadístico obtenido. Por ejemplo, la variable ‘SensibilidadkNN_3’ contiene los estadísticos conocidos como sensibilidad correspondientes a las ejecuciones sobre el clasificador kNN cuya variable (número de vecinos en este caso) toma el valor 3. En cada una de estas variables se almacenan las diferentes ejecuciones de cada grupo de características disponible, así como las dos ejecuciones sin PCA y con PCA en caso de que se habilite dicha opción al realizar la prueba. Una de estas variables tendría el tamaño **2000x20x2**, correspondiente a 2000 grupos de características, 20 ejecuciones de cada una de ellas y las ejecuciones sin PCA y con PCA.

10444->estacionalizado CTM rho = 40 diezmos = 600s intervalos = descanso canales = comp	SVM_RBF 9128	NoPCA	0.74924	0.75	0.82941	0.85294	0.66406	0.65625
43107->desestacionalizado CTM rho = 65 diezmos = 60s intervalos = dia canales = z	SVM_RBF 9128	PCA	0.74621	0.75	0.82353	0.82353	0.66406	0.65625
10444->desestacionalizado CTM rho = 40 diezmos = 600s intervalos = descanso canales = comp	SVM_RBF 9128	PCA	0.74621	0.75	0.82353	0.82353	0.66406	0.65625
43107->desestacionalizado CTM rho = 65 diezmos = 60s intervalos = dia canales = z	SVM_RBF 9128	PCA	0.74621	0.75	0.82353	0.82353	0.66406	0.65625
9548->estacionalizado CTM rho = 36 diezmos = 600s intervalos = descanso canales = comp	SVM_Q	8045	NoPCA	0.74621	0.75758	0.74412	0.75	0.74844
35779->desestacionalizado CTM rho = 32 diezmos = 600s intervalos = dia canales = z	SVM_Q	8045	NoPCA	0.74621	0.75758	0.74412	0.75	0.74844

Figura A.5: Resultados para ciertos grupos clasificadores

Viendo los resultados que muestra la figura A.5 correspondientes a la ejecución del apartado Análisis de Resultados del programa, es posible acceder a cada ejecución de los grupos clasificadores que corresponden. Por ejemplo, para revisar las especificidades del primer grupo, correspondientes al clasificador tipo ‘SVM_RBF’, se deberá acceder a la variable de resultados cargada con el nombre *EspecificidadSVM_RBF*. Esta variable contiene todos los grupos clasificadores utilizados, pero el que se refiere al resultado mostrado en la interfaz es el que se encuentra en la posición 9128, tal y como se muestra en la tabla de resultados. De esta manera, para obtener las diferentes ejecuciones se accedería al valor *EspecificidadSVM_RBF(9128, :, 1)*, ya que se refiere al grupo de características 9128, a todas sus ejecuciones y a la realización sin PCA (número 1 de la tercera dimensión). De esta forma se obtienen las especificidades de todas las ejecuciones correspondientes a los parámetros ya mencionados.

BIBLIOGRAFÍA

- [1] S. Alberola López, D. Martín Martínez, P. Casaseca de la Higuera, I. Casares Alonso, I. Pérez García, A. Cano Garcinuño, JA López Villalobos, FC Ruíz Sanz, JM Andrés de Llano, C. Alberola López, and J. Ardura Fernández. Tdah en atención primaria: registro de la actividad mediante actimetría ambulatoria. In *61 Congreso de la Asociación Española de Pediatría Granada*, Junio 2012.
- [2] Subcommittee on Attention-Deficit. Adhd: Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics*, 2011.
- [3] Littner M, Kushida CA, Anderson WM, Bailey D, Berry RB, Davila DG, Hirshkowitz M, Kapen S, Kramer M, Loubé D, Wise M, and Johnson SF. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. May 2003.
- [4] Silvia Fuentes Hernández. *Estudio de la maduración de los ritmos circadianos de temperatura y movimiento en prematuros como marcadores del desarrollo neurológico*. PhD thesis, Universidad de Murcia, Facultad de Medicina, 2017.
- [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). 2012.
- [6] Robert L Spitzer and Janet BW Williams. Diagnostic and statistical manual of mental disorders. In *American Psychiatric Association*. Citeseer, 1980.
- [7] Robert L Spitzer and Janet BW Williams. Diagnostic and statistical manual of mental disorders. *American Psychiatric Association, 4th ed, text rev, Washington, DC: Author; Burket, RC, Schramm, LL, Therapists' attitudes about treating patients with eating disorders (1995) Southern Medical Journal*, 88:813–818, 2000.
- [8] L. Rodríguez Molinero, J.A. Lopez Villalobos, M. Garrido Redondo, A.M. Sacristán Martín, M.T. Martínez Rivera, and F. Ruiz Sanz. Estudio psicométrico-clínico de prevalencia y comorbilidad del trastorno por déficit de atención con hiperactividad en castilla y león (españa). *Pediatría Atención Primaria*, 11:251 – 270, 06 2009.

- [9] Sonia Ancoli-Israel, Roger Cole, Cathy Alessi, Mark Chambers, William Moorcroft, and Charles P Pollak. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26(3):342–392, 2003.
- [10] Clete A Kushida, Arthur Chang, Chirag Gadkary, Christian Guilleminault, Oscar Carrillo, and William C Dement. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Medicine*, 2(5):389 – 396, 2001.
- [11] Avi Sadeh, Peter J. Hauri, Daniel F. Kripke, and Peretz Lavie. The role of actigraphy in the evaluation of sleep disorders. *Sleep*, 18(4):288, 1995.
- [12] Roger J. Cole, Daniel F. Kripke, William Gruen, Daniel J. Mullaney, and J. Christian Gillin. Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461, 1992.
- [13] 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.
- [14] D. Martín-Martínez, P. Casaseca de-la Higuera, G. Vegas-Sánchez-Ferrero, L. Cordero-Grande, J. M. Andrés de Llano, J. R. Garmendia-Leiza, and J. Ardura-Fernández. Characterization of activity epochs in actimetric registries for infantile colic diagnosis: Identification and feature extraction based on wavelets and symbolic dynamics. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2383–2386, Aug 2010.
- [15] 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. 5th World Congress on Sleep Medicine, 28 September to 2 October 2013, Valencia, Spain.
- [16] Diego Martín Martínez. *Contribuciones en Machine Learning y Modelado Estocástico Orientadas al Análisis de Señales Biomédicas*. PhD thesis, Universidad de Valladolid, E.T.S. de Ingenieros de Telecomunicación, 2015.
- [17] Christian Fortich Jassan. Procesado de señales de actigrafía: Extracción, selección de características y diseño de un sistema experto para el diagnóstico del trastorno por el deficit de atención e hiperactividad. 2016.
- [18] Nigel Bevan. What is the difference between the purpose of usability and user experience evaluation methods. In *Proceedings of the Workshop UXEM*, volume 9, pages 1–4, 2009.
- [19] Fethi Calisir and Ferah Calisir. The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (erp) systems. *Computers in Human Behavior*, 20(4):505 – 515, 2004.
- [20] Actigraph: Activity monitors: Gt3x. <http://actigraphcorp.com/support/activity-monitors/gt3x/>. Accessed: 2017-06-27.
- [21] B. Rosner. *Fundamentals of Biostatistics*. Cengage Learning, 2010.

- [22] Abraham Lempel and Jacob Ziv. On the complexity of finite sequences. *IEEE Transactions on information theory*, 22(1):75–81, 1976.
- [23] María G Rodríguez, Miguel Altuve, Bhakti Guerra, Adriano Adrianza, and Sara Wong. Análisis de la entropía aproximada de la variabilidad de la frecuencia cardiaca y la variabilidad de la despolarización y repolarización ventricular en sujetos diabéticos.
- [24] W Nelson. Methods for cosinor-rhythmometry. *Chronobiologia*, 6:305–323, 1979.
- [25] Avi Sadeh, Christine Acebo, R Seifer, Semra Aytur, and Mary A Carskadon. Activity-based assessment of sleep-wake patterns during the 1st year of life. *Infant Behavior and Development*, 18(3):329–337, 1995.
- [26] Petros Maragos and Fang-Kuo Sun. Measuring the fractal dimension of signals: morphological covers and iterative optimization. *IEEE Trans. Signal Processing*, 41(1):108–121, 1993.
- [27] Georgios Paliouras, Vangelis Karkaletsis, and Constantine D Spyropoulos. *Machine learning and its applications: Advanced Lectures*, volume 2049. Springer, 2003.
- [28] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Elsevier Science, 2003.
- [29] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

