

UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO DE TECNOLOGÍAS Y SERVICIOS DE LA
TELECOMUNICACIÓN**

TRABAJO FIN DE GRADO

**DISEÑO E IMPLEMENTACIÓN DE UN
CLASIFICADOR DE TIPOS DE MOVIMIENTOS
HUMANOS MEDIANTE SENSORES INERCIALES**

JAVIER LÓPEZ INIESTA DÍAZ DEL CAMPO

2021

GRADO DE TECNOLOGÍAS Y SERVICIOS DE LA TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

TÍTULO: Diseño e implementación de un clasificador de tipos de movimientos humanos mediante sensores inerciales

AUTOR: D. Javier López Iniesta Díaz del Campo

TUTOR: D. Manuel Gil Martín

DEPARTAMENTO: Departamento de Ingeniería Electrónica

MIEMBROS DEL TRIBUNAL:

Presidente: D.

Vocal: D.

Secretario: D.

Suplente: D.

FECHA DE LECTURA:

CALIFICACIÓN:

UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO DE TECNOLOGÍAS Y SERVICIOS DE LA
TELECOMUNICACIÓN**

TRABAJO FIN DE GRADO

**DISEÑO E IMPLEMENTACIÓN DE UN
CLASIFICADOR DE TIPOS DE MOVIMIENTOS
HUMANOS MEDIANTE SENSORES INERCIALES**

JAVIER LÓPEZ INIESTA DÍAZ DEL CAMPO

2021

Resumen

Hoy en día, existe un gran interés por las redes de sensores para su incorporación en entornos inteligentes. De esta manera, es posible monitorizar y realizar un seguimiento de las personas, permitiendo modelar y predecir su comportamiento mediante sensores incorporados en smartwatches, smartphones o incluso en la ropa (*wearables*).

En este Trabajo de Fin de Grado se aborda el diseño e implementación de un sistema clasificador de tipos de movimientos humanos utilizando sensores inerciales basado en *Deep Learning*. El objetivo es clasificar a alto nivel las actividades en función de si son movimientos repetitivos, gestos o posturas. De esta forma, dada una determinada muestra de movimiento poder saber a qué tipo de movimiento pertenece, y poder aplicarle a posteriori una u otra técnica de procesamiento de señal y modelado a través de redes neuronales.

La arquitectura del sistema completo está formada principalmente por varios módulos: adquisición de los datos, procesamiento de las señales, modelado a través de la extracción de características y la clasificación de tipos de movimientos.

En relación con la adquisición de los datos, se han utilizado las bases de datos de PAMAP2 y OPPORTUNITY que incluyen una amplia variedad de actividades de la vida cotidiana, tales como caminar, subir escaleras, sentarse o abrir y cerrar cajones, grabadas con diferentes tipos de sensores. Asimismo en este proyecto, ha sido necesario preparar y procesar las grabaciones de todos los sujetos para su posterior análisis. En este sentido, se han utilizado tres distintas técnicas: datos sin preprocesamiento (Raw), la transformada de Fourier y la energía por ventana. Después de lo cual, se han analizado los beneficios de cada técnica en función del tipo de movimiento.

Por último, se ha desarrollado una arquitectura de *Deep Learning* utilizando redes neuronales convolucionales para el modelado a través de la extracción de características. Para la clasificación de tipos de movimientos en PAMAP2 se ha diseñado un módulo clasificador de tipos de movimientos con una tasa de acierto del 96,02% y se ha utilizado este módulo como paso previo en un sistema de clasificación de actividades finales, en lugar de clasificar las actividades directamente sin distinción entre tipos de movimientos. De forma análoga, en OPPORTUNITY se ha diseñado un nuevo módulo para clasificar los cuatro tipos de movimientos. Los mejores resultados fueron obtenidos cuando se realiza la clasificación en dos pasos, diferenciando primero los grupos de movimientos repetitivos (con o sin gestos) y posturas (con o sin gestos) y posteriormente distinguiendo los gestos en cada uno de estos grupos.

Palabras clave: Reconocimiento de la actividad humana, tipos de movimientos, redes neuronales convolucionales, transformada de Fourier, Raw, Energía, validación cruzada dejando un sujeto fuera, PAMAP2, OPPORTUNITY.

Abstract

Nowadays, there is an increasing interest in sensor networks and their incorporation into intelligent environments. This way, it is possible monitoring people's live, which enable to model and predict their movement using integrated sensor in smartwatches, smartphones or even in the clothes (wearables).

This work deals with the design and implementation of a Deep Learning system in order to classify types of movements using inertial sensors. The main objective of this project is classify activities at high level according to whether they are repetitive movements, gestures or postures. In a certain way, given a sample of movement we can know which type of movement belong to, and consequently be able to apply an ex-post in one or another signal processing and modelling technique through neural networks.

The whole system architecture is composed mainly of several modules: data acquisition, signal processing, modelling through feature extraction and type of movements classification.

In relation with data acquisition, PAMAP2 and OPPORTUNITY datasets were used, including a wide variety of daily live activities, such as walking, ascending stairs, sitting or open and close few drawers, recorded with different type of sensors. Additionally in this project, it has been necessary to prepare and process the recordings of all the subjects for further analysis. In this sense, three different techniques have been used: data without preprocessing (Raw), module of the Fast Fourier Transform and energy per window. Afterwards, the benefits of each technique have been analysed depending on the type of movement.

Finally, we developed a Deep Learning architecture using convolutional neural networks (CNNs) for modelling through feature extraction. For classifying types of movements in PAMAP2 we have designed a classifying module of type of movements with an accuracy rate of 96,02 % and we have used this module as a preliminary step on a final activity classification system, instead of classifying activities directly without any distinction between type of movements. Similarly, in OPPORTUNITY we designed a new module to classify four types of movements. Best results were obtained when classification is done in two steps, differentiating firstly between repetitive movements (with or without gestures) and postures (with or without gestures) groups, and later distinguishing gestures within these groups.

Keywords: Human Activity Recognition, type of movements, Convolutional Neural Networks, Fast Fourier Transform, Raw, energy, Leave One Subject Out, PAMAP2, OPPORTUNITY.

Agradecimientos

En primer lugar, me gustaría agradecer a mi tutor Manuel por haberme dado la oportunidad de realizar este proyecto con él. Gracias por tu paciencia, conocimientos y dedicación durante todo este proceso.

A la Universidad Politécnica de Madrid, y especialmente a todos los profesores que forman parte de la Escuela Técnica Superior de Ingenieros de Telecomunicación, por la formación recibida durante estos años.

No puedo olvidarme de mi familia, mi apoyo fundamental, no solo durante la realización de este TFG, sino durante toda la carrera. Ellos, que siempre creyeron en mí, que siempre estuvieron al otro lado del teléfono para darme ánimos y que siempre apostaron por mi sueño. Gracias por darme siempre las mejores oportunidades, por vuestro apoyo continuo y por inculcarme la cultura del esfuerzo y el trabajo, que me ha hecho llegar hasta aquí. Lo conseguido es gracias a vosotros.

También, me gustaría dar las gracias a todos mis amigos, que he tenido la oportunidad de conocer durante estos cuatro años en Madrid, tanto en el Colegio Mayor como en la universidad. Con los que he compartido tantos buenos momentos y con los que espero seguir compartiendo otros nuevos. Y como no acordarme de mis amigos de Daimiel, que aunque haya estado en Madrid han estado siempre a mi lado en todo momento.

Gracias a todos aquellos que habéis estado a mi lado durante este tiempo.

Índice general

1. Introducción	1
1.1. Introducción	1
1.2. Objetivos y metodología	2
1.3. Estructura del documento	3
2. Estado del arte	5
2.1. Aplicaciones HAR	5
2.2. Dispositivos utilizados en HAR	6
2.3. Técnicas de Deep Learning en HAR	6
2.4. Clasificación de tipos de movimientos	7
3. Descripción de las bases de datos	8
3.1. Recolección de datos	8
3.2. Formato de los datos	10
4. Procesamiento de la señal	13
4.1. Acondicionamiento de la señal	13
4.2. Segmentación de las muestras en ventanas	13
4.3. Transformación y procesamiento de las señales	14
4.3.1. Datos sin preprocesamiento (Raw)	14
4.3.2. Módulo de la FFT	16
5. Análisis por tipo de movimiento	19
5.1. Módulo de la energía por ventana	19
5.2. Histogramas y umbral de energía	20
6. Deep Learning	26
6.1. Características de la red neuronal	26
6.2. Optimización de la red neuronal	28
7. Experimentos	29
7.1. Métricas de evaluación	29
7.2. Experimentos en PAMAP2	30
7.2.1. Módulo clasificador entre movimientos repetitivos y posturas	31
7.2.2. Sistema de clasificación de actividades finales	34
7.3. Experimentos en OPPORTUNITY	39
7.3.1. Módulo clasificador de cuatro clases de movimientos	40
7.3.2. Módulo clasificador realizando dos pasos	41
8. Conclusiones	44
8.1. Conclusiones	44
8.2. Líneas futuras	45

Referencias	46
Anexos	49
A. Aspectos éticos, económicos, sociales y ambientales	49
A.1. Introducción	49
A.2. Descripción de impactos relevantes relacionados con el proyecto	50
A.3. Análisis detallado de alguno de los principales impactos	51
A.4. Conclusiones	52
B. Presupuesto económico	53
C. Software utilizado	55
D. Código extracción de características	56
E. Código de modelado	58

Índice de figuras

1.1. Diagrama del módulo completo clasificador de movimientos.	3
3.1. Localización de los sensores (a) Sensores IMU (b) Acelerómetros bluetooth.	9
4.1. Ventanas de la señal considerando ventanas de 3 segundos con un paso entre ventanas de 0,25 segundos.	13
4.2. Aceleración del eje X del sensor de la muñeca - Sujeto 101 de PAMAP2.	15
4.3. Aceleración del eje X del sensor de la muñeca en las posturas - Sujeto 101 de PAMAP2.	15
4.4. Acelerómetros del sensor de la mano mientras el Sujeto 1 de OPPOR- TUNITY está caminando.	16
4.5. Módulo de la FFT del eje X del sensor de la muñeca - Sujeto 101 de PAMAP2.	17
4.6. Módulo de la FFT del eje X del sensor de la mano - Sujeto 2 de OPPORTUNITY.	18
5.1. Módulo de la energía por ventana de la muñeca - PAMAP2.	21
5.2. Módulo de la energía por ventana del pecho - PAMAP2.	21
5.3. Módulo de la energía por ventana del tobillo - PAMAP2.	22
5.4. Módulo de la energía por ventana de la mano - OPPORTUNITY. . . .	24
5.5. Módulo de la energía por ventana de la espalda - OPPORTUNITY. . .	24
5.6. Módulo de la energía por ventana del pie - OPPORTUNITY.	24
6.1. Arquitectura CNN inicial.	27
7.1. LOSO “cross-validation” para cuatro sujetos.	29
7.2. Matriz de confusión de test para FFT - PAMAP2.	32
7.3. Tasa de validación para distintas epochs con FFT - PAMAP2.	33
7.4. Arquitectura CNN final para FFT - PAMAP2.	34
7.5. Estrategia de PAMAP2 con un paso.	34
7.6. Estrategia de PAMAP2 con dos pasos.	35
7.7. Matriz de confusión de test del sistema de HAR con todas las actividades - PAMAP2.	36
7.8. Estrategia de OPPORTUNITY clasificando movimientos directamente.	39
7.9. Estrategia de OPPORTUNITY clasificando movimientos en dos pasos.	39
7.10. Matriz de confusión del módulo con 4 clases - OPPORTUNITY. . . .	40

Índice de tablas

3.1. Protocolo de recolección de datos de PAMAP2.	9
3.2. Etiquetas de la base de datos de PAMAP2.	11
3.3. Etiquetas actividades con movimientos no repetitivos de OPPORTUNITY.	12
3.4. Etiquetas actividades con movimientos repetitivos o posturas de OPPORTUNITY.	12
5.1. Estadísticos de la energía de los sensores de PAMAP2 en movimientos repetitivos.	20
5.2. Estadísticos de la energía de los sensores de PAMAP2 en posturas.	20
5.3. Tasas de acierto conseguidas en OPPORTUNITY con un umbral en los histogramas.	22
5.4. Estadísticos de la energía de los sensores de OPPORTUNITY en movimientos repetitivos (con o sin gestos).	23
5.5. Estadísticos de la energía de los sensores de OPPORTUNITY en posturas (con o sin gestos).	23
5.6. Tasas de acierto conseguidas en OPPORTUNITY con un umbral en los histogramas.	25
6.1. Parámetros iniciales de la red neuronal.	27
7.1. Módulo clasificador entre movimientos repetitivos y posturas con ventanas de 3 segundos - PAMAP2.	31
7.2. Parámetros de la red neuronal optimizada.	33
7.3. Sistemas de HAR de movimientos repetitivos y posturas - PAMAP2.	37
7.4. Tasas finales de la estrategia del módulo clasificador de tipos de movimientos combinado con sistemas de HAR con ventanas de 3 segundos - PAMAP2.	38
7.5. Tasas de test finales de las estrategias utilizadas en PAMAP2.	38
7.6. Módulo clasificador de cuatro clases con ventanas de 3 segundos - OPPORTUNITY.	40
7.7. Módulo clasificador mediante dos pasos - OPPORTUNITY.	42
7.8. Tasas finales de la estrategia del módulo clasificador realizando dos pasos con ventanas de 3 segundos - OPPORTUNITY.	42
7.9. Tasas de test finales de las estrategias utilizadas en OPPORTUNITY.	43
A.1. Aspectos relacionados con las aplicaciones en sistemas de HAR.	50
A.2. Descripción de impactos relevantes relacionados con el proyecto.	50
B.1. Costes de personal.	53
B.2. Costes de recursos materiales.	53
B.3. Operaciones contratadas.	53
B.4. Costes totales.	54

Lista de Acrónimos

FFT: Fast Fourier Transform.

IA: Inteligencia Artificial.

CNN: Convolutional Neural Networks.

RNN: Recurrent Neural Networks.

ANN: Artificial Neural Networks.

IMU: Inertial Measurement Unit.

ADL: Activities of Daily Living.

RFID: Radio Frequency Identification.

HAR: Human Activity Recognition.

LOSO: Leave One Subject Out.

MEMS: Micro Electro Mechanical Systems.

PAMAP2: Physical Activity Monitoring Data Set.

ReLU: Rectified Linear Unit.

RMSPProp: Root Mean Square Propagation.

SGD: Stochastic Gradient Descent.

FF-BP: Feed-forward Back Propagation.

Capítulo 1.

Introducción

1.1. Introducción

Desde sus inicios, la inteligencia artificial (IA) ha hecho posible que sistemas de cómputo aprendan a partir de la experiencia y realicen tareas de forma similar a como las hacen los humanos. Una de las claves de la IA avanzada se encuentra en el aprendizaje, puesto que es cada vez más habitual la búsqueda de máquinas que aprendan por sí solas. De esta forma, en los últimos años el uso de herramientas como el *Machine Learning* (Aprendizaje Automático) y el *Deep Learning* (Aprendizaje Profundo) se han extendido bastante y han supuesto una auténtica revolución en el tratamiento de datos.

El *Deep Learning* [1] es una técnica de aprendizaje en la que un algoritmo automático estructurado simula el aprendizaje del ser humano con el fin de obtener ciertos conocimientos. En la actualidad, las aplicaciones del *Deep Learning* [2] son muchas y muy variadas, como por ejemplo en el tratamiento de imágenes para detección de objetos, bordes o reconocimiento facial, en los asistentes virtuales (Siri, Google Assistant o Alexa) o en sistemas de seguridad de fábricas para reconocer defectos de fabricación, entre otros.

Hoy en día, existe un gran interés por las redes de sensores para su incorporación en diversos entornos inteligentes que permitan el modelado y la predicción del movimiento de las personas. Una aplicación de gran importancia dentro del *Deep Learning* es el llamado *Human Activity Recognition* (HAR), conocido en castellano como reconocimiento de la actividad humana [3]. En los últimos años, se ha convertido en uno de los temas más investigados, en parte debido a la presencia de multitud de sensores en el día a día de las personas, así como las mejoras en la transmisión de datos en vivo, *Machine Learning*, inteligencia artificial e *IoT*. Su principal objetivo es controlar, analizar y modelar el comportamiento humano para así interpretar eventos en curso correctamente. Esto es posible gracias a los datos, recogidos a partir de diferentes sensores situados en *wearables*, *smartwatches* o *smartphones*.

El reconocimiento de la actividad humana permite un amplio abanico de posibilidades en la investigación de nuevas interfaces persona-máquina [4], como por ejemplo, aplicaciones de asistencia médica, sistemas de monitorización de la actividad física para mejorar la salud de las personas, sistemas de seguridad o sistemas de vida activa y asistida (*Active and Assisted Living*) en casas inteligentes...

Este Trabajo Final de Grado fue motivado por la línea futura de un artículo de investigación [5] sobre el HAR, adaptado a los diferentes tipos de movimientos humanos. En él, se propone desarrollar un módulo inicial para poder clasificar el tipo de movimiento humano antes de seleccionar la mejor estrategia de *Deep Learning* para diferenciar entre las actividades de cada tipo de movimiento. De este modo, sería posible incluir un módulo inicial destinado a clasificar las actividades en los diferentes tipos de movimiento para que el sistema pudiera a posteriori diferenciar entre las actividades finales.

1.2. Objetivos y metodología

Como se ha comentado previamente, el principal objetivo de este trabajo es el diseño e implementación de un clasificador de tipos de movimientos humanos. En concreto el clasificador deberá diferenciar las actividades entre movimientos repetitivos, movimientos no repetitivos (gestos) y posturas. De manera que dada una muestra de movimiento, poder saber a qué tipo de movimiento pertenece, y poder aplicarle una u otra técnica de procesamiento de señal y modelado a través de redes neuronales. Para cumplir con el objetivo final se contemplarán los siguientes objetivos específicos:

- Analizar las características de los diferentes tipos de movimientos en términos de energía.
- Estudiar y aplicar técnicas de procesamiento de señales.
- Estudiar y optimizar arquitecturas de redes neuronales con el propósito de extraer características de las señales y clasificar el tipo de actividad.
- Diseñar e implementar un módulo clasificador de tipos de movimientos.

Para alcanzar dichos objetivos, se ha seguido la siguiente metodología:

- *Tarea 1*: Instalación de la plataforma.
En esta tarea, se ha instalado el correspondiente software para el correcto diseño e implementación del módulo clasificador de movimientos. Se incluye tanto *Octave* para la parte de procesamiento de la señal, como *Anaconda*, *Python* y *Keras* utilizados en la parte de *Deep Learning*.
- *Tarea 2*: Estudio, implementación y evaluación en la extracción de propiedades a través de técnicas de procesamiento de señales.
En esta segunda tarea, se han estudiado las distintas señales inerciales y sus propiedades. Además, se han analizado varias técnicas de extracción de propiedades. Cada una de estas técnicas ha sido implementada en *Octave*.

- *Tarea 3*: Estudio, implementación, evaluación y optimización de varias técnicas de modelación y reconocimiento de la actividad.

Al igual que en la tarea anterior, esta tarea incluye el estudio, implementación, evaluación y optimización de diferentes módulos para poder clasificar correctamente las actividades dependiendo del tipo de movimiento que esta incluya. Para la implementación de estas técnicas de experimentación se ha utilizado Keras.

- *Tarea 4*: Escritura del documento.

Finalmente, la última tarea ha consistido en la elaboración de la presente memoria. Mientras se realizaba esta tarea, se han llevado a cabo otros experimentos adicionales con el fin de completar y apoyar las conclusiones.

La Figura 1.1 muestra el diagrama completo del módulo clasificador de movimientos que se ha diseñado e implementado en este Trabajo Final de Grado:

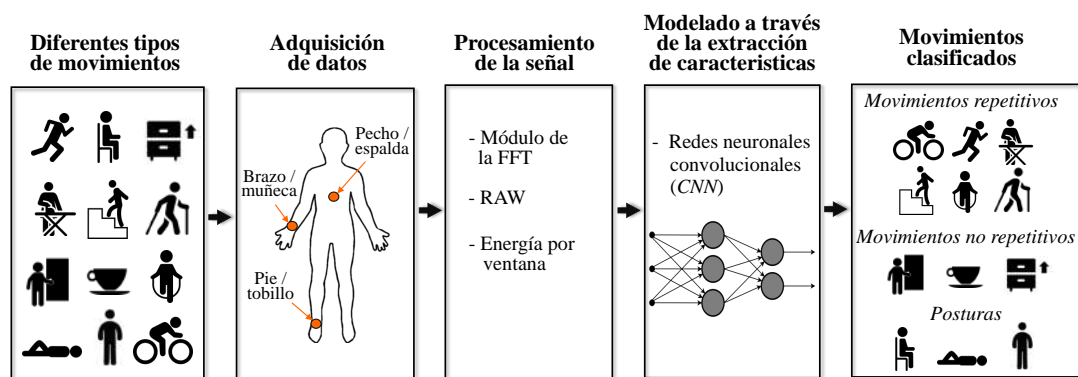


Figura 1.1: Diagrama del módulo completo clasificador de movimientos.

1.3. Estructura del documento

El documento está dividido en varios capítulos:

- *Capítulo 1*: Introducción.

- *Capítulo 2*: Estado del arte.

En este capítulo se hace un resumen de los trabajos recientes publicados que han tratado temas relacionados con el *Human Activity Recognition*, explicando las técnicas y algoritmos usados. Posteriormente, el estudio se ha centrado en los trabajos que clasifican tipos de movimientos.

- *Capítulo 3*: Descripción de las bases de datos.

En este capítulo, se describen las bases de datos utilizadas, incluyendo información sobre los sensores, actividades, sujetos y el protocolo utilizados en la

recolección de datos. Además, se explica con detalle las nuevas etiquetas creadas para el procesado de datos.

– *Capítulo 4: Procesamiento de la señal.*

Este capítulo describe las técnicas de procesamiento de las señales inerciales utilizadas, así como los distintos pasos que seguidos para su transformación.

– *Capítulo 5: Análisis por tipo de movimiento.*

En este capítulo se realiza el análisis por tipo de movimiento. Se calcula el módulo de la energía por ventana y se incluyen una análisis de estadísticos de la energía con el fin de modelar los diferentes tipos de actividades. Además se presentan distintos histogramas de la energía de cada sensor para cada base de datos.

– *Capítulo 6: Deep Learning.*

En este capítulo se describen las características de la red neuronal utilizada durante los diferentes experimentos realizados. Asimismo, se presentan las optimizaciones realizadas de la red para mejorar la tasa de acierto.

– *Capítulo 7: Experimentos.*

En el capítulo de experimentos se describen los distintos módulos implementados para clasificar tipos de movimientos en cada base de datos. Se mostrarán las tasas de acierto de cada experimento, con el objetivo de conseguir el sistema que clasifique mejor cada tipo de movimiento.

– *Capítulo 8: Conclusiones.*

En este capítulo, se incluyen las principales conclusiones obtenidas en este Trabajo de Final de Grado y la propuesta de líneas futuras.

Capítulo 2.

Estado del arte

En este capítulo se exponen los principales trabajos, que otros investigadores han realizado al respecto en el campo de reconocimiento de actividades humanas. En concreto, se pueden encontrar diversos trabajos donde se aplican técnicas de *Deep Learning* y procesamiento de la señal. Estas aplicaciones tan variadas se pueden clasificar atendiendo a diferentes características: por el tipo de sensores utilizados, la ubicación de los sensores, por los de tipos de movimientos, por el tipo de aprendizaje utilizado...

2.1. Aplicaciones HAR

Como se ha comentado previamente, existe multitud de aplicaciones relacionadas con el reconocimiento de la actividad humana, como por ejemplo en el ámbito de la salud con el objetivo concienciar a los usuarios de los recursos disponibles en hospitales o para detectar comportamientos anómalos en la monitorización de actividades físicas ([6], [7]). Conviene destacar trabajos relacionados con la detección precoz del Parkinson, a través de sensores en *smartphones* que permiten recoger continuamente señales [8] o la rehabilitación de pacientes a través de sensores RFID pasivos con redes CNN [9]. Estos estudios se enfocan sobretodo en enfermedades relacionadas con el cambio de ciertos movimientos del cuerpo o funciones, pudiendo ser detectadas a tiempo.

Además, podemos encontrar aplicaciones en casas inteligentes que permiten caracterizar los actividades y comportamientos de personas mayores, siendo posible monitorizar sus hábitos en el entorno doméstico [10]. También lo encontramos en aplicaciones de vida cotidiana asistida en sistemas de reconocimiento de gestos, con el fin de controlar el movimiento de un robot [11] o en la vigilancia y seguridad de hogares permitiendo actuar automáticamente en ellos sin que un centro de control dé la orden [12].

Sin embargo, la mayoría de los trabajos realizados hasta la fecha se centran en el reconocimiento de actividades diarias y en deportes. En gran parte de ellos, se utilizan redes neuronales convolucionales, combinando las características aprendidas de datos de sensores inerciales con información complementaria de características superficiales, para habilitar una clasificación de actividades precisa en tiempo real [13]. En otro estudio previo [14] se utilizan datos recogidos de sensores de *smartphones*, ya que en la actualidad estos incorporan acelerómetros que permiten crear un sistema de HAR basado en la aceleración. El principal objetivo de estos estudios es monitorizar el ejercicio, ayudando a la adaptación de entrenamientos o previniendo la obesidad.

2.2. Dispositivos utilizados en HAR

El reconocimiento de la actividad humana se centra en los datos de movimientos recogidos de distintos sensores tales como acelerómetros, giroscopios, Bluetooth, sensores ambientales y demás sensores. Un ejemplo de aplicación de acelerómetros tri-axiales es [15], donde se proponen usar un esquema jerárquico a través de diferentes niveles para ser capaces clasificar 15 actividades pertenecientes a tres estados distintos (estáticas, con transiciones y dinámicas), consiguiendo tasas de acierto bastante altas.

Actualmente, estos sensores se están incorporando en dispositivos utilizados en el día a día, como pueden ser *smartphones* o *smartwatches*. En algunos trabajos de HAR [13] los datos son recogidos a partir de los sensores que incorporan estos dispositivos, permitiendo infinidad de ventajas en comparación con otros tipos de sensores. Estas ventajas se centran en la portabilidad, su uso extendido y en la alta capacidad para procesar datos en tiempo real.

Adicionalmente los *wearables*, especialmente los *smartwatches*, son más cómodos de llevar continuamente. En cambio, solo permiten almacenar movimientos del brazo, a diferencia de otro tipo de dispositivos que permiten almacenar los datos de movimientos de otros movimientos del cuerpo. Debido a su auge en los últimos años, existen algunos trabajos que incorporan estos relojes para recolectar los datos usados en la red neuronal [16]. Esta publicación centra su investigación en el reconocimiento de la actividad usando *smartwatches* mediante la técnica de *Deep Learning* conocida como “Restricted Boltzmann Machines”.

2.3. Técnicas de Deep Learning en HAR

En la actualidad, existen diversos trabajos que emplean diferentes técnicas de *Deep Learning* dependiendo de su finalidad. En el estudio [17], se analizan los principales métodos de extracción de características usando modelos convolucionales, recurrentes y deep a través de tres bases de datos que contienen datos de movimientos capturados con sensores de *wearables*.

Una de las técnicas más efectivas en el reconocimiento de actividades son las redes neuronales convolucionales (*CNNs*). El objetivo de esta técnica es capturar las dependencias locales y las características que permanecen invariantes de las señales de las diversas actividades. Asimismo, un artículo [18] trata de diferenciar 6 actividades diferentes de varias bases de datos usando *CNNs* con 2 niveles convolucionales. Los autores tratan de extraer primero las características de los datos Raw y después mediante un modelo de predicciones identificar la actividad. Las tasas de acierto conseguidas en este estudio varían dependiendo de la base de datos, variando desde un 96,88 % de ANTITRACKER al 76,83 % conseguido con OPPORTUNITY.

Respecto a los datos utilizados para la extracción de características en *CNN*, encontramos trabajos que usan directamente las señales de acelerómetros (Raw) usando auto-encoders seguido de un clasificador de tipo “K-nearest neighbor” [19]. Otros en cambio, usan señales en el dominio de la frecuencia (*FFT*), puesto que en el estudio [13] se comenta que usando información adicional de la *FFT* en la primera capa convolucional mejora la tasa de acierto hasta en un 1 %. Además, estas capas convolucionales son muy efectivas a la hora de clasificar movimientos dinámicos que son difícil de clasificar. Del mismo modo, también se utilizan diversos estadísticos, como por ejemplo la media y desviación estándar como entrada de las redes neuronales.

Otra técnica utilizada es las redes neuronales artificiales (*ANN*), utilizada en el ámbito de la salud. Reúne los datos de los pacientes y reconoce anomalías en su estado de salud, usando como datos de entrada acelerómetros situados en *smartwatches*, o en el pecho [20]. Este estudio trataba de diferenciar 10 posturas diferentes usando *ANN* con 2 niveles de tipo *FF BP*, con 10 y 6 neuronas, en los niveles ocultos y de salida, respectivamente. En adicción, se usaba “Levenberg-Marquardt” como algoritmo de entrenamiento.

Finalmente, encontramos las redes neuronales recurrentes (*RNN*) [21] donde se reconoce los patrones, separados por algunos intervalos. En esta red neuronal, las entradas son los datos de sensores situados en varias localizaciones del entorno.

2.4. Clasificación de tipos de movimientos

En este sentido, todos los trabajos anteriores no hacen ninguna distinción entre los tipos de movimientos, de forma que las actividades se procesan y modelan utilizando un enfoque único. En cambio, un estudio previo [5] analiza las principales características de los distintos tipos de movimientos usando datos de sensores portátiles. De forma que, se analizan diferentes estrategias basadas en el *Deep Learning* con el fin de conseguir la mejor alternativa para separar movimientos del mismo tipo. En este trabajo, para la base de datos de PAMAP2 se consiguieron tasas del 96,7 % para separar movimientos repetitivos y del 88,7 % para las posturas. En cambio, para la base de datos de OPPORTUNITY se consiguió tasas del 66,9 % y del 97,1 % para movimientos no repetitivos (gestos) y posturas, respectivamente.

Con el fin de poder aplicar a cada tipo de movimiento una u otra técnica de procesado de señal y modelado, existe la necesidad de diseñar un módulo clasificador de diferentes tipos de movimientos. En la actualidad, no existe ningún trabajo al respecto. Por ello resulta necesario la generación de este módulo previo a la elección de la estrategia de *Deep Learning* idónea para diferenciar entre las actividades de cada tipo de movimiento, diseñado e implementado en el presente Trabajo Final de Grado.

Capítulo 3.

Descripción de las bases de datos

En este Trabajo Final de Grado, se han usado dos bases de datos: PAMAP2 (*Physical Activity Monitoring Data Set*) [22] y OPPORTUNITY [23]. Estas contienen una amplia variedad de actividades de la vida cotidiana, tales como caminar, subir escaleras, sentarse o abrir y cerrar cajones, grabadas con diferentes tipos de sensores. PAMAP2 incluye 27 señales grabadas bajo condiciones de laboratorio, mientras que en OPPORTUNITY contiene 113 señales grabadas con plena libertad. Estas bases de datos nos ofrecen una amplia variedad de actividades que permiten realizar una clasificación en tipos de movimientos.

3.1. Recolección de datos

En esta sección se presentan las bases de datos y se describe el hardware utilizado para la recolección de dichos datos, los sujetos involucrados y el protocolo de recolección de datos.

Por un lado, la base de datos de PAMAP2 contiene datos de 18 actividades diferentes, grabados de nueve sujetos. Es preciso señalar que estas actividades únicamente pertenecen a movimientos repetitivos y posturas, es decir, no incluyen actividades con movimientos no repetitivos (gestos). Para la recolección de los datos se utilizaron tres IMUs ("Inertial Measurement Units") y un monitor de frecuencia cardiaca. Cada IMU contiene dos acelerómetros MEMS (Micro Electro Mechanical Systems), un giroscopio MEMS y un magnetómetro (magneto-resistivo), todos ellos de tres ejes y con una frecuencia de muestreo de 100 Hz . Dichos sensores fueron colocados en tres posiciones distintas: sobre la muñeca del brazo dominante (1 IMU), en el pecho (1 IMU + 1 monitor de frecuencia cardiaca) y en el tobillo de la pierna dominante (1 IMU). En la recolección de datos participaron nueve sujetos (ocho hombres y una mujer), todos ellos diestros, excepto uno que era zurdo.

En adición, se definió un protocolo compuesto por 12 actividades, descrito en la Tabla 3.1. Este protocolo estaba compuesto por 12 actividades o posturas (tumbado, sentado, de pie, pasar la aspiradora, subir y bajar escaleras, andar normal, marcha nórdica, ciclismo, correr y saltar a la comba). Los sujetos podían realizar este protocolo estándar, o podían seguir otra secuencia y actividades diferentes. Además, podían realizar otras 6 actividades voluntarias (ver la televisión, trabajo en el ordenador, conducir, doblar ropa, limpiar y jugar al fútbol) con el fin de enriquecer el rango de actividades en la base de datos de PAMAP2.

Actividad	Duración (min)	Actividad	Duración (min)
Tumbado	3	Bajar escaleras	1
Sentado	3	Descanso	2
De pie	3	Caminar	3
Planchar	3	Descanso	1
Descanso	1	Marcha nórdica	3
Pasar la aspiradora	3	Descanso	1
Descanso	1	Ciclismo	3
Subir escaleras	1	Descanso	1
Descanso	2	Correr	3
Bajar escaleras	1	Descanso	2
Descanso	1	Saltar a la comba	2
Subir escaleras	1		

Tabla 3.1: Protocolo de recolección de datos de PAMAP2.

De tal forma que la base de datos de PAMAP2 cuenta con 12 actividades diferentes y aproximadamente 10 horas de datos.

Por otro lado, en la base de datos de OPPORTUNITY los datos fueron recogidos de 12 sujetos distintos usando 15 redes sensoriales, los cuales incluyen 72 sensores de 10 modalidades, integrados en el entorno, en objetos y en el cuerpo. Los sensores incluyen 24 acelerómetros y giroscopios inalámbricos (bluetooth), dos *Sun SPOTs*, dos *InertiaCube3*, el sistema de localización y un sensor de campo magnético a medida. Sin embargo, en este trabajo únicamente se van a utilizar los datos de los sensores *wearable* de los cuatro sujetos. En este caso, los datos fueron muestreados a 32 *Hz*.

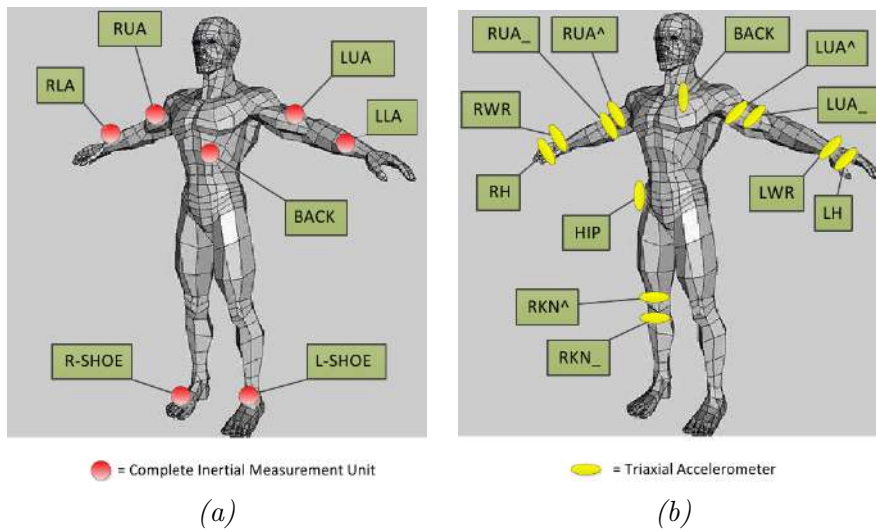


Figura 3.1: Localización de los sensores (a) Sensores IMU (b) Acelerómetros bluetooth.

Esta base de datos incluye actividades humanas naturales grabadas en un rico entorno de reconocimiento, que simulaba un estudio con cocina y salida al exterior donde todos los sujetos grabaron cada día actividades matutinas. Existía dos tipos de sesiones, una sesión de simulacro (*Drill Session*), en la que se realizaban 20 repeticiones de una secuencia de actividades predefinidas y una ADL (*Activities of Daily Living*) con tareas tales como despertarse, arreglarse, relajarse, preparar el desayuno o limpiar la mesa con mayor libertad en la secuencia de actividades. Las actividades existentes en esta base de datos son:

- Abrir y cerrar 2 puertas.
- Abrir y cerrar el frigorífico.
- Abrir y cerrar el lavavajillas.
- Abrir y cerrar tres cajones (a diferentes alturas).
- Limpiar la mesa.
- Beber de pie.
- Beber sentado.
- Pulsar interruptor.

3.2. Formato de los datos

En esta sección se describe el formato de los datos de PAMAP2 y OPPORTUNITY, así como el reetiquetado que se ha realizado.

En PAMAP2, cada fichero de datos contiene los datos de todos los sensores, fusionados en un fichero por sujeto y sesión. Cada fila se ha organizado en 54 columnas, donde cada una contiene los siguientes datos:

- 1: Tiempo (s).
- 2: Identificador de la actividad.
- 3: Ritmo cardiaco (bpm).
- 4-20: IMU de la muñeca.
- 21-37: IMU del pecho.
- 38-54: IMU del tobillo.

Los sensores de tipo IMU contienen a su vez, las siguientes columnas:

- 1: Temperatura ($^{\circ}C$).
- 2-4: Datos de los acelerómetros-3D ($m \cdot s^{-2}$), escala: $\pm 16g$, resolución: 13 *bits*.
- 5-7: Datos de los acelerómetros-3D ($m \cdot s^{-2}$), escala: $\pm 6g$, resolución: 13 *bits*.
- 8-10: Datos de los giróscopos-3D (rad/s).
- 11-13: Datos de los magnetómetros-3D (μT).
- 14-17: Orientación.

Durante el procesado de datos, se han creado dos etiquetas diferentes: una etiqueta para la actividad (1) y otra para el tipo de actividad (2). Esta segunda etiqueta variará en función de si la actividad contiene movimientos repetitivos (1) o posturas (2). Además, se han descartado las actividades que no se encontraban dentro del protocolo de recolección de datos (ver la televisión, trabajo en el ordenador, conducir, doblar ropa, limpiar y jugar al fútbol) y otras actividades temporales (estar esperando al equipamiento o desplazarse a otro sitio para realizar otra actividad). De modo que dependiendo del objetivo del experimento, usaremos una u otra. La Tabla 3.2 muestra el posible valor de las etiquetas en PAMAP2.

Actividad	Etiqueta 1	Etiqueta 2
Tumbado	1	2
Sentado	2	2
De pie	3	2
Caminar	4	1
Correr	5	1
Ciclismo	6	1
Marcha nórdica	7	1
Subir escaleras	8	1
Bajar escaleras	9	1
Pasar la aspiradora	10	1
Planchar	11	1
Saltar a la comba	12	1

Tabla 3.2: Etiquetas de la base de datos de PAMAP2.

En cambio, en OPPORTUNITY los datos se organizan de distinta forma. Cada fichero de datos, esta organizado en 250 columnas, en la que cada una de ellas contiene:

- 1: Tiempo (*ms*).
- 2-134: Sensores corporales.
 - 2-37: Acelerómetros.
 - 38-134: IMUs.
- 135-194: Sensores en objetos.
- 195-207: *Reed Switches*.
- 208-243: Sensores en el entorno.
- 244-250: Etiquetas.

Asimismo, se realizaron las conversiones oportunas para trabajar en ambas bases de datos con las muestras de los acelerómetros en $[m/s^2]$ y las de los giróscopos en $[rad/s]$.

En cambio, en la base de datos de OPPORTUNITY se han utilizado las etiquetas de los distintos modos de locomoción (“Locomotion”) y de los gestos (“ML_Both_Arms”). Del mismo modo que en PAMAP2, se han creado dos etiquetas diferentes. No obstante, durante la creación de estas etiquetas se observó que un sujeto podía estar realizando un movimiento no repetitivo (gesto) mientras realizaba una actividad con movimientos repetitivos (caminar) o se encontraba en una postura (de pie, sentado o tumbado). Por consiguiente, el valor de la etiqueta en función del tipo de actividad (Etiqueta 2) será: actividad con movimiento repetitivo sin gestos (1), gesto durante movimiento repetitivo (2), posturas sin gestos (3) y gestos durante posturas (4). En las Tablas 3.3 y 3.4 se observa las nuevas etiquetas de OPPORTUNITY:

Actividad	“ML_Both_Arms”	Gestos durante Mov. repetitivos (2)	Gestos durante posturas (3)
		Etiqueta 1	Etiqueta 1
Abrir puerta 1	406516	1	2
Abrir puerta 2	406517	3	4
Cerrar puerta 1	404516	5	6
Cerrar puerta 2	404517	7	8
Abrir frigorífico	406520	9	10
Cerrar frigorífico	404520	11	12
Abrir lavavajillas	406505	13	14
Cerrar lavavajillas	404505	15	16
Abrir cajón 1	406519	17	18
Cerrar cajón 1	404519	19	20
Abrir cajón 2	406511	21	22
Cerrar cajón 2	404511	23	24
Abrir cajón 3	406508	25	26
Cerrar cajón 3	404508	27	28
Limpiar mesa	408512	29	30
Beber de la taza	407521	31	32
Pulsar interruptor	405506	33	34

Tabla 3.3: Etiquetas actividades con movimientos no repetitivos de OPPORTUNITY.

Actividad	“Locomotion”	Etiqueta 1	Etiqueta 2
De pie	1	38	4
Caminar	2	37	1
Sentado	4	36	4
Tumbado	5	35	4

Tabla 3.4: Etiquetas actividades con movimientos repetitivos o posturas de OPPORTUNITY.

Capítulo 4.

Procesamiento de la señal

En este capítulo se describe las técnicas utilizadas para procesar las señales inerciales utilizadas en este proyecto. En primer lugar, se han preparado y preprocesado las grabaciones de todos los sujetos, para su posterior análisis. A continuación, estas señales han sido segmentadas en ventanas. Finalmente, se han procesado y transformado al dominio de la frecuencia en algunos experimentos.

4.1. Acondicionamiento de la señal

Las bases de datos utilizadas en este proyecto contienen la información de los distintos de sensores, junto con una etiqueta del tipo de actividad y el instante de muestreo. De esta forma, el primer paso realizado ha sido leer y cargar estos datos en *Octave*. Estos ficheros asociados a los distintos sujetos se han procesado uno por uno realizando el reetiquetado de las muestras cuando era necesario.

4.2. Segmentación de las muestras en ventanas

Con respecto a la división de los datos en distintas ventanas, un artículo de referencia [24] evalúa los datos con ventanas de diferente duración. Este artículo, comenta que para obtener al menos 2 o 3 periodos diferentes de movimientos repetitivos, un tamaño de ventana de 3 – 5 segundos de duración es razonable. No obstante, para conseguir tasas de test mayores a veces es recomendable utilizar ventanas de duración mayor en función del tipo de actividad a reconocer [5]. Para las dos bases de datos usadas en este Trabajo Final de Grado, se usarán ventanas de 3 segundos de duración, con un paso entre ventanas de 0,25 segundos (Figura 4.1).

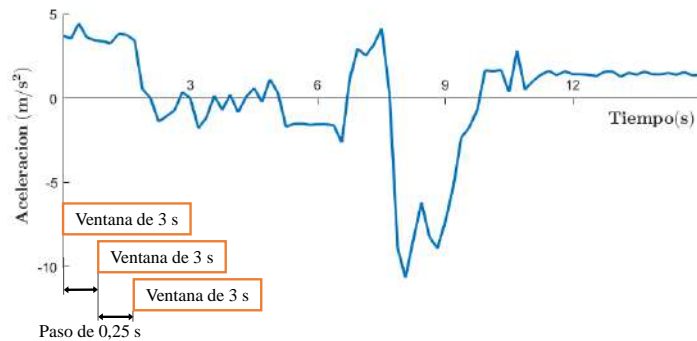


Figura 4.1: Ventanas de la señal considerando ventanas de 3 segundos con un paso entre ventanas de 0,25 segundos.

Dado que si se utilizan ventanas de mayor duración para diferenciar tipos de movimientos, es posible que no se detecten actividades cortas como los gestos.

4.3. Transformación y procesamiento de las señales

A continuación, se seleccionan los datos de los sensores utilizados para cada base de datos. Para el caso de PAMAP2 se han utilizado los sensores de la muñeca, pecho y tobillo. Con el objetivo de poder comparar ambas bases de datos, inicialmente en OPPORTUNITY se eligió el sensor de la mano derecha (*Right Lower Arm*), espalda (*Back*) y pie derecho (*Right Shoe*). Además, como se cuenta en el Capítulo 7 también se han utilizado en algún experimento todos los sensores existentes del brazo derecho.

Como se ha comentado previamente, los datos en PAMAP2 están muestreados cada 0,01 segundos, es decir, la frecuencia de muestreo es de 100 *Hz*. En cambio, en OPPORTUNITY la frecuencia de muestreo es de 32 *Hz*, por lo que se ha realizado una interpolación de los datos para trabajar con la misma frecuencia y poder aplicar la misma transformación.

Posteriormente, los datos almacenados de ambas bases de datos fueron preprocesados y se les realizó un enventanado de la señal del tipo “Hamming”. Como entradas de la red neuronal, se utilizan señales en el dominio del tiempo y de la frecuencia. Usando en cada caso un módulo diferente de preprocesamiento de la señal, dependiendo del dominio de la señal.

4.3.1. Datos sin preprocesamiento (Raw)

En primer lugar, como se ha comentado en el Capítulo 3 las bases de datos contienen datos de tipo Raw. Esta técnica de preprocesamiento de señales, consiste en no aplicar ninguna transformación a las señales originales. De esta forma, las entradas a la redes convolucionales son directamente las ventanas en el dominio del tiempo.

Además, en la Sección 5.1 se detalla cómo se ha calculado el módulo de la energía por ventana, el cual también se ha utilizado como entrada de la red neuronal para comprobar si dicha información era suficiente para distinguir entre tipos de actividad. Para realizar dicho cálculo, se utilizan directamente los datos sin ninguna transformación (Raw) de los acelerómetros.

En la Figura 4.2 se muestra los datos sin procesar del acelerómetro del eje X del sensor de la muñeca. Dicha figura representa todas las actividades realizadas por el sujeto 101 de PAMAP2. En ella, se puede observar que en los movimientos repetitivos existen mayor número de oscilaciones que en las posturas, donde la aceleración en el sensor es constante.

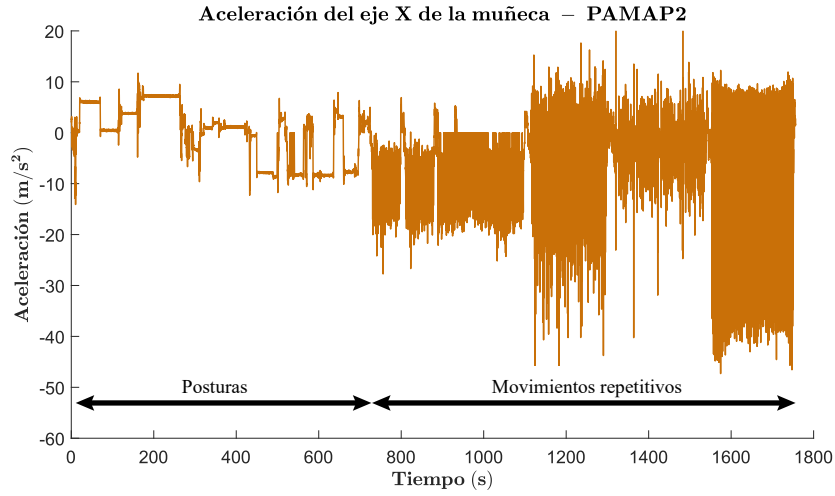


Figura 4.2: Aceleración del eje X del sensor de la muñeca - Sujeto 101 de PAMAP2.

La Figura 4.3 muestra únicamente los datos del acelerómetro del eje X del sensor de la muñeca del sujeto 101 mientras realizaba las tres posturas existentes en PAMAP2 (tumbado, sentado y de pie). Los cambios existentes en la aceleración se deben principalmente a cambios de la orientación del acelerómetro durante las posturas.

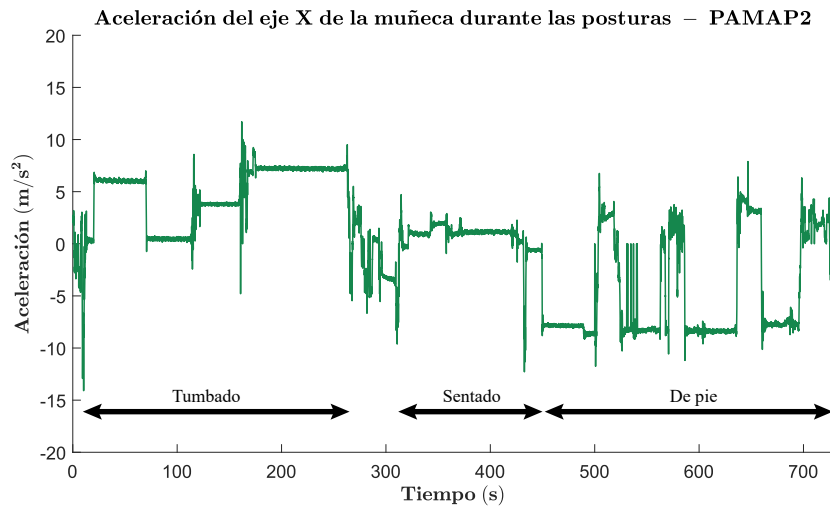


Figura 4.3: Aceleración del eje X del sensor de la muñeca en las posturas - Sujeto 101 de PAMAP2.

En cambio, en OPPORTUNITY existen movimientos repetitivos, y posturas que pueden haberse grabado solos o incluyendo algún tipo de gesto (ver Sección 3.2). De forma que diferenciar directamente estos tipos de movimientos será complejo, debido a la similitud entre algunos movimientos. No obstante, observando los distintos ejes de los sensores utilizados se pueden obtener mejoras con respecto a otras técnicas de procesamiento utilizadas.

La Figura 4.5 muestra un espectrograma del módulo de la FFT del eje X del sensor de la muñeca del sujeto 101 de PAMAP2, representados entre 0 y 25 Hz . Además, teniendo en cuenta las distintas etiquetas de cada actividad, se ha asociado el espectrograma a las diferentes actividades a lo largo del tiempo. Se puede observar la periodicidad existente en algunas actividades, como por ejemplo subir escaleras o correr.

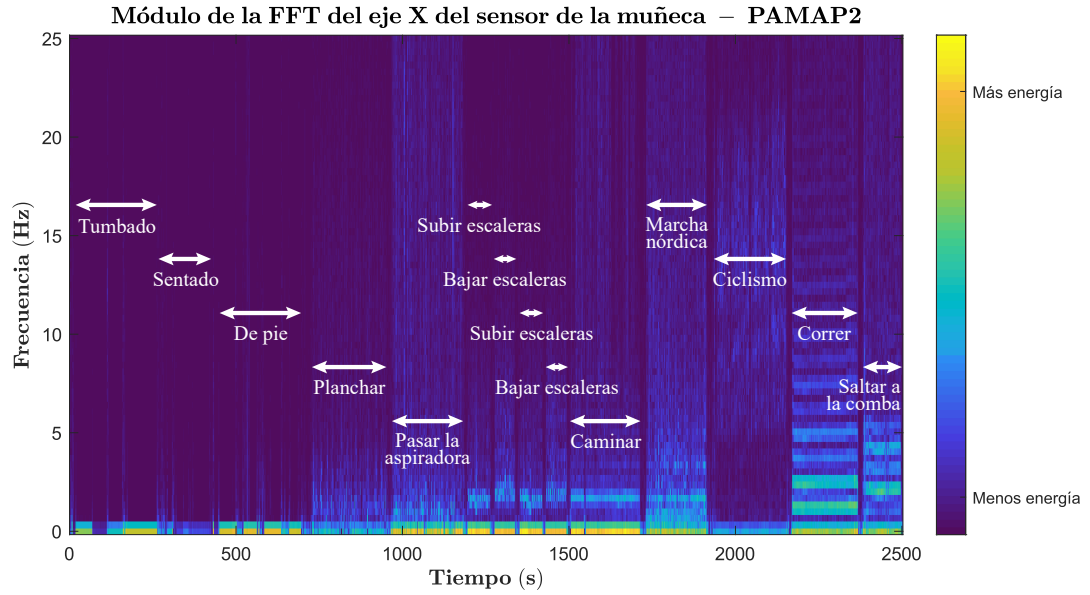


Figura 4.5: Módulo de la FFT del eje X del sensor de la muñeca - Sujeto 101 de PAMAP2.

Esta periodicidad, es mucho más notable en los movimientos repetitivos que en las posturas, debido a la propia naturaleza de dichos movimientos. En consecuencia, diferenciar las posturas a través de la FFT será más difícil que con datos de tipo Raw, ya que como se ha mostrado anteriormente (ver Figura 4.3) la aceleración en el eje X del sensor de la muñeca es constante durante las posturas. Además, destacar que la energía de la señal se concentra en todas las actividades en bajas frecuencias.

Sin embargo, en la base de datos de OPPORTUNITY al calcular el módulo de la FFT, no se aprecia igual de bien que en PAMAP2 esta periodicidad, debido a que únicamente existe un movimiento repetitivo, como es caminar, en la base de datos. No obstante, la energía de la señal se sigue concentrando en bajas frecuencias.

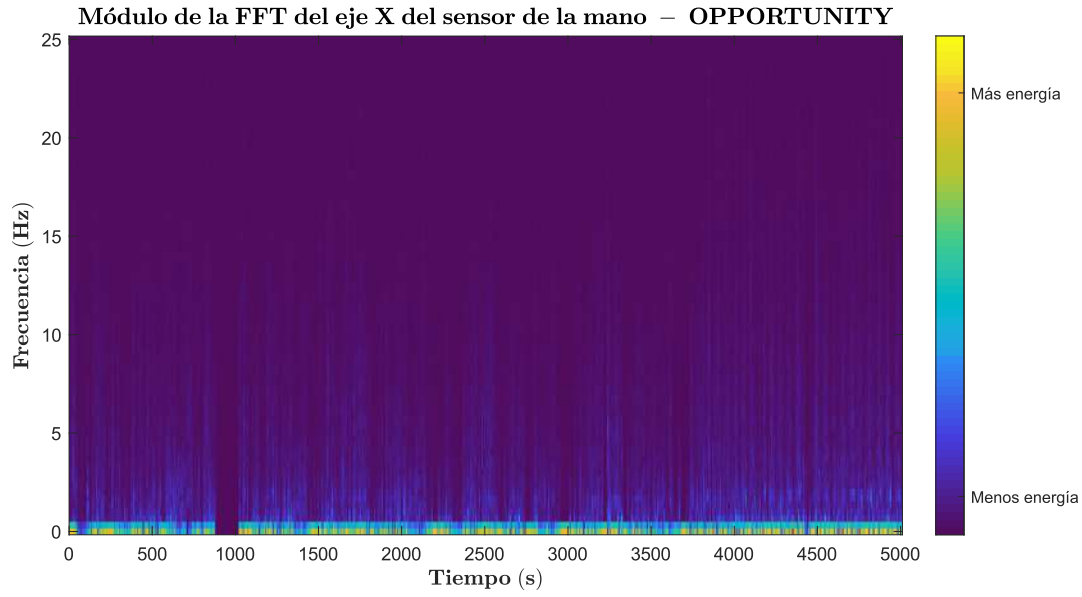


Figura 4.6: Módulo de la FFT del eje X del sensor de la mano - Sujeto 2 de OPPORTUNITY.

La Figura 4.6 muestra un espectrograma del módulo de la FFT del eje X del sensor de la mano (*RLA*) de todas las actividades realizadas por el sujeto 2 de OPPORTUNITY. Se ha representado nuevamente el espectrograma entre 0 y 25 *Hz*. Este espectrograma nos indica que únicamente existe información en la componente continua, por lo que el módulo de la FFT puede que no sea útil para distinguir entre gestos y posturas.

Capítulo 5.

Análisis por tipo de movimiento

En este capítulo, se describe el análisis por tipo de movimiento. En un trabajo previo [5] se propone una tipología de movimientos de actividad física. Por un lado, los movimientos, estos a su vez están divididos en 2 subgrupos, movimientos repetitivos y no repetitivos (gestos). Los movimientos repetitivos son actividades dinámicas repetitivas en el tiempo (andar, correr, subir escaleras...). En cambio, los movimientos no repetitivos son actividades dinámicas que incluyen un movimiento aislado, como por ejemplo, abrir un cajón, cerrar una puerta o beber un café. Normalmente, los movimientos repetitivos tienen mayor duración en el tiempo que los no repetitivos. Por otro lado, las posturas son actividades estáticas mientras se mantiene una posición específica (sentado, tumbado o de pie). Al ser actividades estáticas tendrán menos aceleración del propio movimiento que otros tipos de movimientos, como las actividades dinámicas.

Con el objetivo de poder aplicarle a cada tipo de movimiento una u otra técnica de procesamiento de señal y modelado, existe la necesidad de generar un módulo clasificador de diferentes tipos de movimientos. De forma que, dada una determinada muestra de movimiento, podamos saber con exactitud a qué tipo de movimiento corresponde antes de procesarla y modelarla con técnicas específicas para ese tipo de movimiento.

5.1. Módulo de la energía por ventana

Siguiendo la idea anterior, una primera clasificación podría basarse en la cantidad de energía de cada ventana de una determinada actividad, para tratar de separar los movimientos en función de su energía.

Para ello, se calcula la energía de cada una de las actividades de ambas bases de datos de la red con ventanas de las señales en crudo (Raw) de los sensores de los acelerómetros. Cada señal de los acelerómetros tiene 3 coordenadas distintas (x, y, z). En el cálculo del módulo de la energía de la ventana, $|E|$, se ha usado la siguiente expresión:

$$|E| = \frac{1}{F} \sum_{i=1}^F \left| \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2} \right|^2 \quad (5.1)$$

Siendo, F el número total de puntos de cada ventana. Al realizar el análisis con datos Raw tenemos 300 puntos, ya que, la frecuencia de muestreo es 100 Hz (100 puntos/s) y el tamaño de la ventana es de 3 segundos. Al aplicar la fórmula 5.1 a una señal de un acelerómetro, se obtiene su potencia. Dicha potencia representa la

energía de la señal. Por último, estos datos fueron almacenados en un fichero para el posterior análisis.

5.2. Histogramas y umbral de energía

A continuación, se ha realizado un análisis estadístico de la energía por ventana de cada sensor. En las Tablas 5.1 y 5.2 se muestra la media (η_E) y la desviación estándar (σ_E), que representa el promedio de variación respecto a la media. Estos estadísticos se han calculado para los tres sensores utilizados en PAMAP2 (muñeca, pecho y tobillo) y cada uno de los tipos de movimientos.

Sensor	Media (μ_E)	Desviación estándar (σ_E)
Muñeca	71,65	168,05
Pecho	29,96	64,84
Tobillo	129,71	177,86

Tabla 5.1: Estadísticos de la energía de los sensores de PAMAP2 en movimientos repetitivos.

En la Tabla 5.1 se puede observar que la energía en el sensor del tobillo es superior a la de los sensores de la muñeca o el pecho, debido a la propia naturaleza de movimiento durante las actividades dinámicas como correr o subir escaleras. Asimismo, la desviación estándar de la energía en el tobillo es bastante grande, debido a la diferencia de movimientos existente en algunas actividades en el pie, como por ejemplo caminar o saltar a la comba. Análogamente, la desviación estándar en la muñeca también es grande, puesto que en actividades como planchar existe bastante diferencia de movimientos en el sensor de la muñeca. En último lugar, el sensor del pecho es el que tiene menos desviación estándar, ya que en este sensor la aceleración de movimiento es similar en todas las actividades con movimientos repetitivos de PAMAP2.

Sensor	Media (μ_E)	Desviación estándar (σ_E)
Muñeca	3,40	11,46
Pecho	0,86	3,90
Tobillo	1,09	4,72

Tabla 5.2: Estadísticos de la energía de los sensores de PAMAP2 en posturas.

Por otro lado, mientras se realiza alguna postura la media de la aceleración de movimiento en el sensor de la muñeca es superior al de otros los otros sensores que es muy pequeña. La razón por la que la energía media en el sensor del pecho y el tobillo es tan pequeña es porque cuando el sujeto está tumbado o sentado apenas realiza movimientos con esas partes del cuerpo humano.

Además, el sensor de la muñeca tiene mayor desviación estándar que el resto de los sensores para el caso de las posturas ya que mientras se mantiene una postura estática, los movimientos de las manos son más frecuentes. Comparando ambas tablas, se puede ver como la energía media en todos los sensores de los movimientos repetitivos es superior a la de las posturas, debido a la diferencia de aceleración de movimiento que existe entre las actividades dinámicas y estáticas.

Del mismo modo, se ha calculado mediante *Octave* el umbral donde se consigue una tasa de acierto mayor a la hora de diferenciar entre movimientos repetitivos, movimientos no repetitivos y posturas. Este umbral de energía se calcula simulando el experimento “*oracle*”, donde el sistema se comporta siempre de la misma forma. La tasa de acierto se calcula como la relación entre el número de ventanas clasificadas correctamente y el número de ventanas totales.

En las Figuras 5.1, 5.2 y 5.3 se puede observar para el caso de PAMAP2 diferentes histogramas de la energía por ventana en cada sensor con su correspondiente umbral.

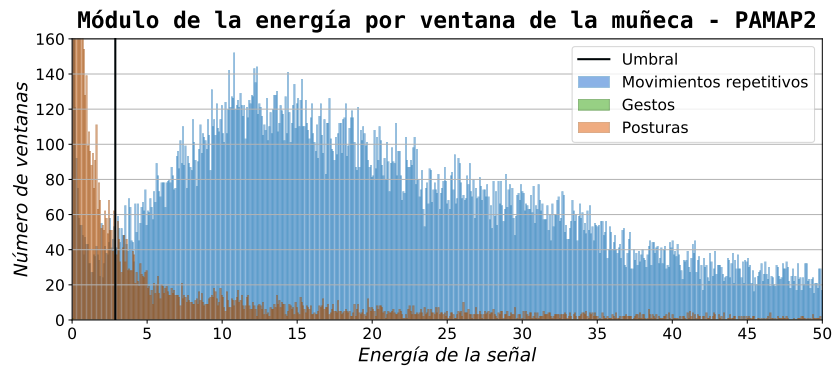


Figura 5.1: Módulo de la energía por ventana de la muñeca - PAMAP2.

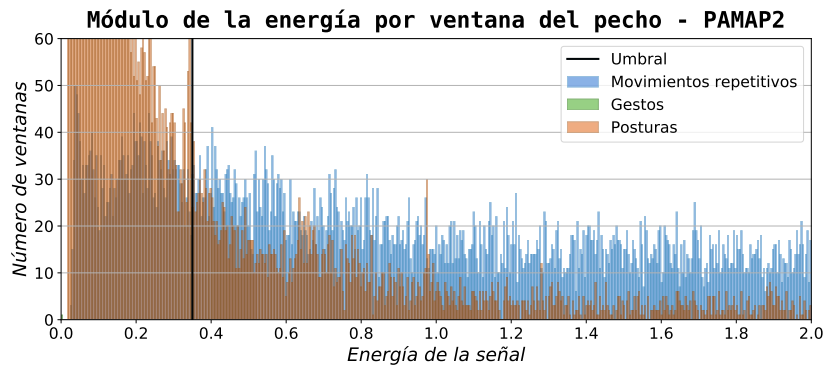


Figura 5.2: Módulo de la energía por ventana del pecho - PAMAP2.

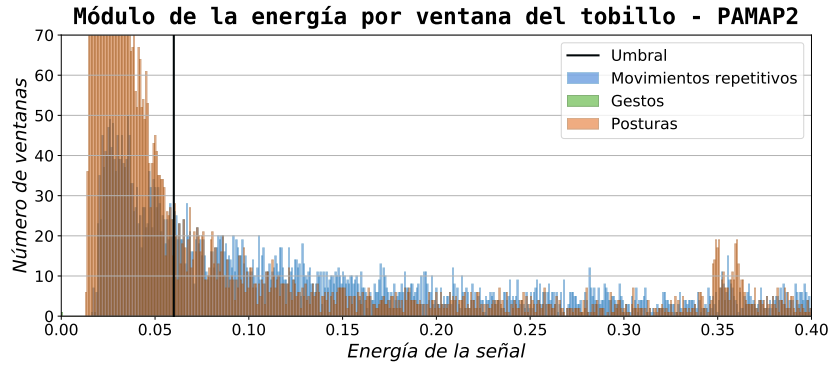


Figura 5.3: Módulo de la energía por ventana del tobillo - PAMAP2.

En la Tabla 5.3 se observan las tasas de acierto máximas conseguidas para PAMAP2 con un umbral en los histogramas:

Sensor	Tasa de acierto (%)	Umbral
Muñeca	92,4	2,88
Pecho	90,8	0,35
Tobillo	89,3	0,06

Tabla 5.3: Tasas de acierto conseguidas en OPPORTUNITY con un umbral en los histogramas.

La energía de las actividades con movimientos repetitivos es mayor que las posturas. Para el caso del sensor de la muñeca esta diferencia es mayor, que la que se obtiene con los sensores del pecho y del tobillo. Por ello, se consigue una tasa de acierto mayor que para los otros sensores.

Al igual que en PAMAP2, inicialmente para OPPORTUNITY se aplica la clasificación de 3 clases (movimientos repetitivos, movimientos no repetitivos y posturas). Sin embargo, a la hora de realizar los histogramas se pudo comprobar que se producía un solapamiento entre los movimientos de dichas actividades, dado que existen actividades con gestos durante movimientos repetitivos y durante posturas. Por este motivo, se propone desde el trabajo diferenciar entre esos dos nuevos tipos de actividad en la base de datos de OPPORTUNITY, tal y como se ha comentado en el Capítulo 3.

De la misma forma, las Tablas 5.4 y 5.5 representan los estadísticos de la energía por ventana, de tres sensores diferentes. En este caso, se han utilizado los sensores de la mano derecha (*Right Lower Arm*), espalda (*Back*) y pie derecho (*Right Shoe*).

Sensor	Media (μ_E)	Desviación estándar (σ_E)
Mano	15,90	14,13
Espalda	6,08	5,20
Pie	72,54	49,75

Tabla 5.4: Estadísticos de la energía de los sensores de OPPORTUNITY en movimientos repetitivos (con o sin gestos).

En la Tabla 5.4 se puede observar como al igual que en el caso anterior, la energía media en el sensor del pie es superior a la del resto de sensores para el caso de movimientos repetitivos realizados solos o mientras se realizaba algún gesto. Además, la energía media en este sensor es inferior al caso de PAMAP2 (Tabla 5.1), ya que en este caso existe una única actividad con movimientos repetitivos, como es caminar. Con esta actividad, existe menos aceleración de movimiento que para otras actividades existentes en PAMAP2 con movimientos repetitivos, como pueden ser correr o ciclismo. Nuevamente, la desviación estándar en los sensores del pie o de la mano es superior a la de la espalda, aunque es menor que con PAMAP2.

Sensor	Media (μ_E)	Desviación estándar (σ_E)
Mano	13,79	17,55
Espalda	4,28	6,90
Pie	8,22	17,60

Tabla 5.5: Estadísticos de la energía de los sensores de OPPORTUNITY en posturas (con o sin gestos).

Con respecto a las posturas realizadas con o sin gestos (Tabla 5.5), se observa como la energía media de todos los sensores es superior a las posturas de la base de datos de PAMAP2. Puesto que gran parte de estas posturas se realizan con algún tipo de gesto (abrir puertas o cajones), donde la aceleración de movimiento es superior al de una postura sin ningún tipo de gesto. De la misma forma, la desviación estándar en la mano y el pie es mayor a la de la espalda, debido a que mientras se mantiene una postura es posible que se esté realizando al mismo tiempo algún tipo de gesto, donde el movimiento de los pies y manos es frecuente.

Cabe notar que la energía media en los sensores de la mano y de la espalda para ambos tipos de movimientos son similares, por lo cual diferenciar entre estos movimientos será complicado. En cambio, se puede observar que para el sensor situado en el pie, la energía de las actividades con movimientos repetitivos, es bastante superior a la de las posturas.

De forma análoga, en las Figuras 5.4, 5.5 y 5.6 se representan los histogramas del módulo de la energía por ventana con una clasificación de 4 clases distintas y el umbral calculado para cada caso.

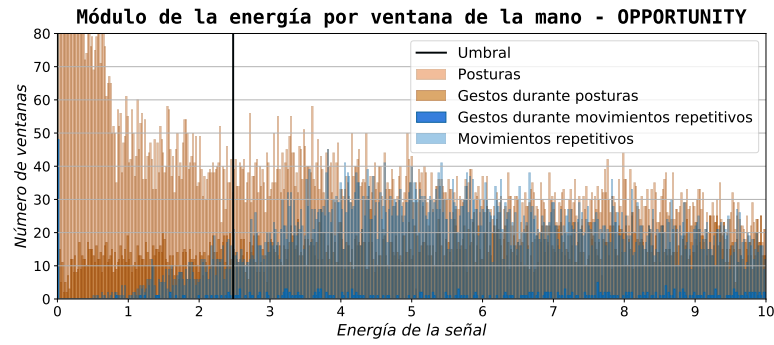


Figura 5.4: Módulo de la energía por ventana de la mano - OPPORTUNITY.

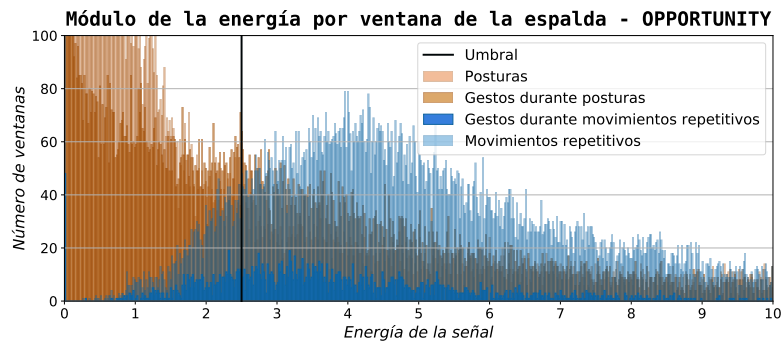


Figura 5.5: Módulo de la energía por ventana de la espalda - OPPORTUNITY.

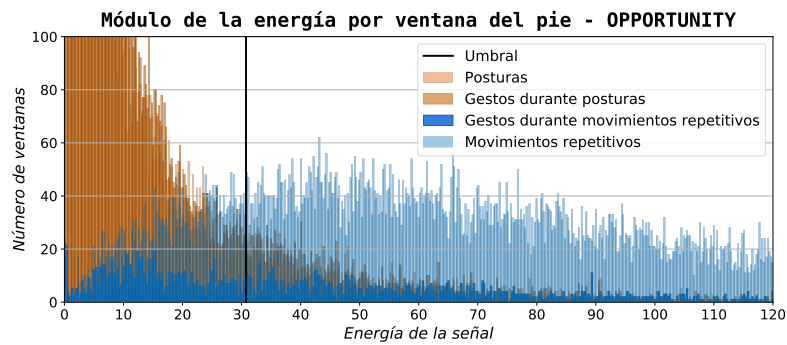


Figura 5.6: Módulo de la energía por ventana del pie - OPPORTUNITY.

En este caso, solamente se han diferenciado entre dos subgrupos: movimientos repetitivos (solos o durante gestos) y posturas (solos o durante gestos). Las tasas de acierto obtenidas en este caso, se observan en la Tabla 5.6:

Sensor	Tasa de acierto (%)	Umbral
Mano	72,9	2,4
Espalda	68,45	2,48
Pie	88,7	30,75

Tabla 5.6: Tasas de acierto conseguidas en OPPORTUNITY con un umbral en los histogramas.

Como se puede apreciar, existe bastante solapamiento en las energías de estos subgrupos (movimientos repetitivos y gestos durante movimientos repetitivos de posturas y gestos durante posturas) en los sensores de la mano y la espalda, por lo que distinguir entre este tipo de actividades es un reto difícil de afrontar en términos de actividad. En cambio, con el sensor del pie se consigue una tasa de acierto superior, del 88,7 %.

Capítulo 6.

Deep Learning

En este capítulo se describen las características de la red neuronal utilizada durante la realización de este Trabajo Final de Grado. De igual forma, se explica las optimizaciones realizadas para conseguir aumentar la tasa de acierto.

En los últimos años, debido a los enormes avances en la capacidad de procesamiento, se están desarrollando una gran cantidad de técnicas de *Deep Learning*, siendo aplicadas en numerosos trabajos relacionados con el reconocimiento. Como se ha visto en el Capítulo 2 mediante el uso de estas técnicas de aprendizaje no supervisado, se pueden conseguir grandes tasas de acierto cuando se usan para el reconocimiento de actividades basadas en señales recogidas de distintos sensores. Siendo las redes neuronales convolucionales (*CNN*), la mejor técnica de *Deep Learning* para el reconocimiento de la actividad humana.

6.1. Características de la red neuronal

Inicialmente, la arquitectura *CNN* usada como referencia en este Trabajo Final de Grado para el reconocimiento de tipos de movimientos está formada por dos subredes. Por una parte, la primera subred se ocupa del aprendizaje de las características de los ejemplos introducidos y a su vez contiene dos niveles convolucionales. Entre estos niveles, se usa un nivel de *MaxPooling* de tamaño (1,3) para poder extraer las diferentes características de la red. Asimismo se ha utilizado 16 kernels convolucionales de tamaño (1,5). Después de introducir los datos en la capa de *MaxPooling*, se introducen en una capa de *dropout* de valor 0,3. Posteriormente, estos datos se aplanan (*Flatten layer*). Por otra parte, la segunda subred se encarga de clasificar dichas características en diferentes clases. Esta subred contiene 3 niveles completamente interconectados, donde cada neurona de la capa se conecta con las neuronas de la siguiente capa. Adicionalmente, en todas las capas excepto en la última, se ha utilizado una función de activación de tipo *ReLU*.

En la capa de salida de la red neuronal se ha usado la función de *softmax* para poder clasificar entre las distintas clases. Finalmente, el optimizador empleado ha sido *RMSProp*, usando los parámetros por defecto. En cuanto a los hiperparámetros, se han utilizado cinco *epochs* y un *batchsize* de 1000 ejemplos.

Los datos de entrada utilizados en la red neuronal convolucional han sido los datos de la transformada de Fourier (FFT), de tipo Raw y de la energía por ventana.

La Figura 6.1 muestra la estructura de la red neuronal convolucional utilizada inicialmente para procesar todos los experimentos realizados. Posteriormente, se han modificado algunos parámetros con el objetivo de mejorar la tasa de acierto (ver Sección 6.2):

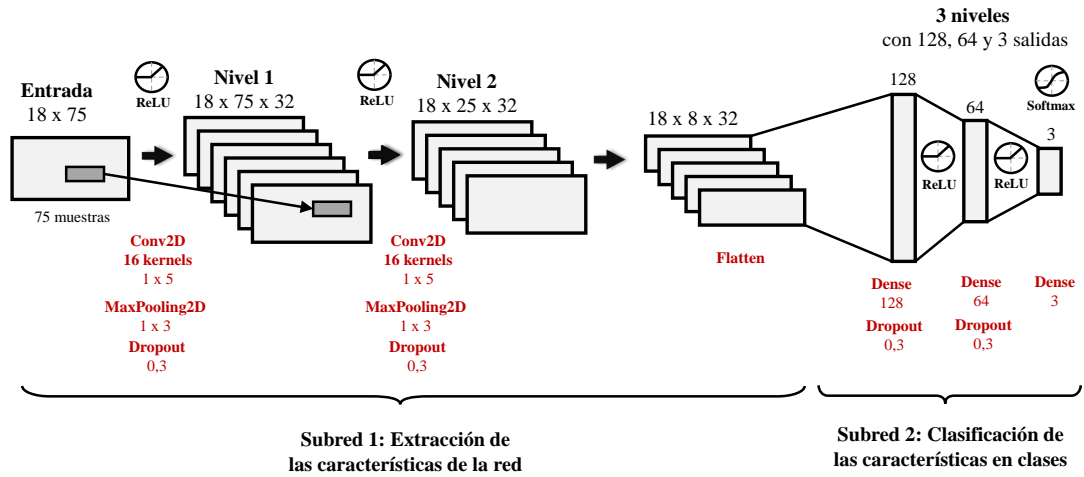


Figura 6.1: Arquitectura CNN inicial.

En la Tabla 6.1 se muestran un resumen de los parámetros utilizados en la red neuronal convolucional usada como base en todos los experimentos realizados.

Parámetro	Valor
<i>Epochs</i>	5
<i>Batchsize</i>	1000
Nº de niveles convolucionales	2
Nº de kernels de los niveles convolucionales	16
Tamaño del kernel convolucional	(1,5)
<i>MaxPooling</i>	(1,3)
<i>Dropout</i>	0,3
Optimizador	<i>RMSPProp</i>
Tasa de aprendizaje	0,001

Tabla 6.1: Parámetros iniciales de la red neuronal.

6.2. Optimización de la red neuronal

A lo largo de los distintos módulos se ha ido optimizando cada red neuronal. Con el fin de obtener mejores resultados, se han modificado los diferentes hiperparámetros y formato de los datos, optimizando así la red sobre la tasa de validación. Los parámetros que se han modificados son:

- *Epochs*: Se han realizado experimentos variando el número de epochs (2, 3, ..., 10).
- *Batchsize*: Este hiperparámetros fue modificado, con 1000 (inicial), 500, 200 y 100 ejemplos.
- Número de niveles convolucionales: Se han realizado experimentos con 1, 2, 3 y 4 niveles convolucionales.
- Número de kernels de los niveles convolucionales: Se realizaron experimentos con 16, 32 y 64 kernels.
- Tamaño del kernel convolucional: El tamaño del kernel se ha optimizado desde (1,5) pasando por (1,10), (1,20) e incluso (1,50) en los experimentos con datos de tipo Raw.
- *MaxPooling*: Se han realizado experimentos variando el *MaxPooling* de (1,2) y (1,3).
- *Dropout*: Este parámetro se ha ajustado con un valor de *dropout* igual a 0,2, 0,3 y 0,4.
- Optimizador: De la misma forma, se modificó el optimizador utilizado (*Adam*, *RMSPProp* y *SGD*).
- Tasa de aprendizaje (*lr*): La mayoría de los experimentos fueron realizados con una tasa de aprendizaje de 0,001 (por defecto), aunque también se utilizó una tasa de aprendizaje de 0,005.

Capítulo 7.

Experimentos

En este capítulo se describen las métricas de evaluación utilizadas en los experimentos de este Trabajo Final de Grado. Asimismo, también se define la forma de calcular la tasa de acierto y el intervalo de confianza. Posteriormente, se analizan los distintos módulos clasificadores implementados en ambas bases de datos, mostrando los resultados conseguidos y el método más eficaz para poder clasificar las actividades en los distintos tipos de movimientos en cada caso.

7.1. Métricas de evaluación

El método de evaluación empleado ha sido LOSO (*Leave One Subject Out*) con “n-fold cross-validation”. Esta técnica consiste en dividir aleatoriamente los distintos sujetos en subconjuntos, donde los datos de un sujeto son usados para test, otro sujeto se usa para validación y el resto son usados para el entrenamiento. De forma que el sujeto que se usa en test, no es utilizado para el entrenamiento. Además, los resultados finales de cada iteración son promediados, con el fin de realizar una estimación siguiendo la estrategia conocida como “round-robin”. Esta estrategia consiste en cambiar en cada iteración los sujetos de test y validación, de forma que un mismo sujeto solo pueda ser validado y usado para test en una única iteración (ver Figura 7.1).

Como se ha descrito en el Capítulo 3 existen nueve sujetos distintos para el caso de PAMAP2, y cuatro para el caso de OPPORTUNITY.

	Sujeto 1	Sujeto 2	Sujeto 3	Sujeto 4
Iteración 1	Test	Validación	Entrenamiento	Entrenamiento
Iteración 2	Entrenamiento	Test	Validación	Entrenamiento
Iteración 3	Entrenamiento	Entrenamiento	Test	Validación
Iteración 4	Validación	Entrenamiento	Entrenamiento	Test

Figura 7.1: LOSO “cross-validation” para cuatro sujetos.

De esta forma, se obtiene una matriz de confusión que describe el comportamiento de un determinado modelo de clasificación para un conjunto de test. Cada elemento de la matriz, P_{ij} es el número de ejemplos de la actividad real (i , *filas*) que ha sido reconocido de la actividad (j , *columnas*).

De manera que la tasa de acierto (acc), se define mediante la suma de todos los ejemplos de la diagonal principal de la matriz dividido por el número total de ejemplos usados en la evaluación (fórmula 7.1):

$$acc = \frac{1}{N} \sum_{i=1}^C P_{ii} \quad (7.1)$$

Donde, N representa el número total de ejemplos de test, C el número diferente de clases y P_{ii} son los elementos de la diagonal. Adicionalmente, para poder comparar correctamente las tasas de acierto de de distintos experimentos, es necesario calcular el intervalo de confianza (IC) al 95 % (fórmula 7.2). De tal forma que se pueda comprobar si dos tasas son significativamente diferentes (si no existe solapamiento entre las tasas incluyendo el intervalo de confianza):

$$IC (95 \%) = \pm 1,96 \cdot \sqrt{\frac{acc \cdot (100 - acc)}{N}} \quad (7.2)$$

Durante el análisis de los distintos experimentos, únicamente se han utilizado los resultados obtenidos de la validación para ajustar los principales parámetros de la red neuronal. Sin embargo, los resultados finales fueron obtenidos de los subconjuntos de test usando la mejor configuración de parámetros sobre los subconjuntos de validación.

7.2. Experimentos en PAMAP2

Por un lado, para la base de datos de PAMAP2 se han desarrollado distintos experimentos con el fin de obtener el módulo clasificador entre movimientos repetitivos y posturas, donde se consigan las mayores tasas de test y validación. Se han utilizado para las redes neuronales los datos de entrada de la FFT, Raw y de la energía por ventana. Conviene destacar que los resultados de los experimentos que se muestran en todas las tablas han sido obtenidos tras la optimización de la red en cuanto a hiperparámetros y formato de los datos como se explica en la Sección 6.2.

De la misma forma, con el objetivo de obtener la mejor estrategia de clasificación de actividades, se han diseñado varios sistemas de HAR. A través de estos sistemas, es posible conocer si es mejor clasificar las actividades directamente sin usar un módulo clasificador de tipos de movimientos. O por el contrario, es mejor clasificar primero en función del tipo de movimiento y posteriormente aplicar la mejor estrategia de *Deep Learning* para diferenciar las actividades de cada tipo de movimiento. Finalmente, se detallará que método es mejor para clasificar entre tipos de movimientos, en términos de tasa de acierto, tiempo de cómputo o tamaño de la red.

7.2.1. Módulo clasificador entre movimientos repetitivos y posturas

En primer lugar, se ha desarrollado un módulo clasificador para intentar diferenciar los movimientos repetitivos de las posturas. La Tabla 7.1 muestra los resultados obtenidos usando ventanas de 3 segundos de duración y un paso entre ventanas de 0,25 segundos. Los sensores utilizados han sido los acelerómetros y giróscopos de la muñeca, pecho y tobillo.

PAMAP - Módulo clasificador entre movimientos repetitivos y posturas			
Número de ejemplos de test: 77707		Número de ejemplos de validación: 86163	
Sensores utilizados: Acelerómetros y giróscopos de la muñeca, pecho y tobillo			
Tamaño de ventana: 3 (s)		Paso de ventana: 0,25 (s)	
Datos de entrada	Tiempo de ejecución	Tasa de test (%)	Tasa de validación (%)
FFT	0:05:57.19	96,02 ± 0,13	96,32 ± 0,12
Raw	0:13:23.00	94,40 ± 0,16	95,14 ± 0,14
Energía	0:00:35.09	92,08 ± 0,19	91,54 ± 0,18

Tabla 7.1: Módulo clasificador entre movimientos repetitivos y posturas con ventanas de 3 segundos - PAMAP2.

Se puede observar, cómo las tasas de validación son similares para los datos de entrada del tipo Raw y FFT. Sin embargo, con el módulo de la FFT se consiguen tasas de test superiores a los conseguidos con los otros datos de entrada. En términos de tiempo de ejecución, el sistema que menos tiempo tarda es para cuando los datos de entrada son la energía por ventana. Para datos de tipo Raw el sistema tarda mucho más que los otros sistemas, debido principalmente a que tenemos más datos por ejemplo. Cabe señalar que, la tasa de acierto conseguida usando los datos de la energía por ventana como entrada de la *CNN* es de 92,08%, similar al resultado obtenido en el experimento “*oracle*” usando el sensor de la muñeca (92,40%). Por consiguiente, la diferencia existente entre ambos resultados no es significativa (ver Sección 5.2).

Además, las redes convolucionales aprenden dependencias. Por lo tanto, como se ha mostrado en el espectrograma de la Figura 4.5 la red neuronal diferencia bastante bien entre estos tipos de movimientos. Usando ventanas de tres segundos de duración tanto los datos tipo Raw como los de la FFT ofrecen resultados competitivos en validación. No obstante, dado que utilizar FFT como datos de entrada a la red neuronal consume menos recursos (tiempo de ejecución) desde mi punto de vista utilizaría la FFT.

La Figura 7.2 muestra la matriz de confusión de test con los datos de entrada de la FFT. En ella se puede observar, el número de ejemplos que se confunden para cada tipo de movimiento, es decir, el número de ejemplos predichos incorrectamente con respecto a la actividad real.

Matriz de confusión de test para FFT - PAMAP2

Mov. repetitivos -	53478	1522
Posturas -	1567	21140
	Mov. repetitivos	Posturas

Figura 7.2: Matriz de confusión de test para FFT - PAMAP2.

De forma similar a este primer experimento, se añadió los datos del sensor del magnetómetro. No obstante, no se consiguió superar las tasas de validación obtenidas mediante únicamente acelerómetros y giróscopos. Debido a que al comparar un magnetómetro con un acelerómetro en un mismo nivel convolucional, el sistema no aprende correctamente.

Como se ha mencionado en la Sección 4.2, a veces es recomendable utilizar ventanas de mayor duración en función del tipo de actividad a reconocer. Por ello, se ha realizado un experimento utilizando ventanas de 15 segundos de duración, junto con los acelerómetros y giróscopos para tratar de conseguir tasas de acierto mayores. Cabe destacar que la tasa de test conseguida, 97,57 % es notablemente superior a la tasa conseguida con ventanas de tres segundos (95,51 %). Sin embargo a pesar de esta mejora, el tiempo de ejecución se multiplica por cinco. Del mismo modo, el módulo final elegido utiliza ventanas de tres segundos porque al utilizar ventanas largas para diferenciar tipos de movimientos, es posible perder algún gesto.

Por último, se ha optimizado la red neuronal de este módulo clasificador entre movimientos repetitivos y posturas para conseguir una tasa de acierto mayor. Primero, se empezó variando el *dropout* y el número de kernels convolucionales. A continuación, se modificó el *MaxPooling*, el tamaño del kernel y el número de niveles convolucionales. A pesar de estas modificaciones en el formato de los datos, no se consiguió ninguna mejora significativa. En consecuencia, se modificó el optimizador usado, así como el “*Learning rate*”. Se optó por utilizar como optimizadores *RMSPProp*, *SGD* y *Adam*, con un “*Learning rate*” de 0,001(por defecto) y 0,005.

Tras estas optimizaciones, el siguiente paso fue modificar el número de *epochs* y el *batchsize*. La Figura 7.3 muestra las tasas de validación conseguidas usando distinto número de *epochs* con los datos de la FFT. Se puede observar como usando 7 *epochs* se consigue una tasa de validación del 95,75 %, superior que con el resto de *epochs*.

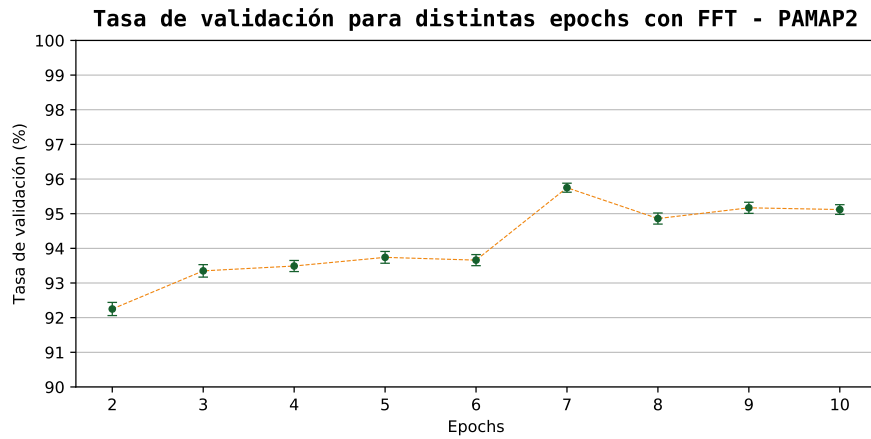


Figura 7.3: Tasa de validación para distintas epochs con FFT - PAMAP2.

Finalmente, se realizaron experimentos con distinto *batchsize* (500, 200 y 100) con 7 *epochs* para tratar de conseguir alguna mejora significativa. Usando *batchsize* igual a 500, se consigue una tasa de validación de 96,02 %. Dicha tasa, es similar a la que se consigue con *batchsize* 200 o 100, pero el tiempo de ejecución es mucho inferior. Con el objetivo de optimizar recursos (tiempo de ejecución) se ha optado por usar un *batchsize* de 500.

Parámetro	Valor
<i>Epochs</i>	7
<i>Batchsize</i>	500
Nº de niveles convolucionales	2
Nº de kernels de los niveles convolucionales	32
Tamaño del kernel convolucional	(1,5)
<i>MaxPooling</i>	(1,3)
<i>Dropout</i>	0,3
Optimizador	<i>RMSPProp</i>
Tasa de aprendizaje	0,001

Tabla 7.2: Parámetros de la red neuronal optimizada.

Mediante esta red optimizada pasamos de una tasa de validación de 94,21 % al 96,32 %. La tasa de test también se ve mejorada del 95,51 % al 96,02 %. En la Figura 7.4 se representa la arquitectura de la red neuronal convolucional final (optimizada) utilizada en el módulo clasificador entre movimientos repetitivos y posturas para procesar los datos del módulo de la FFT de PAMAP2.

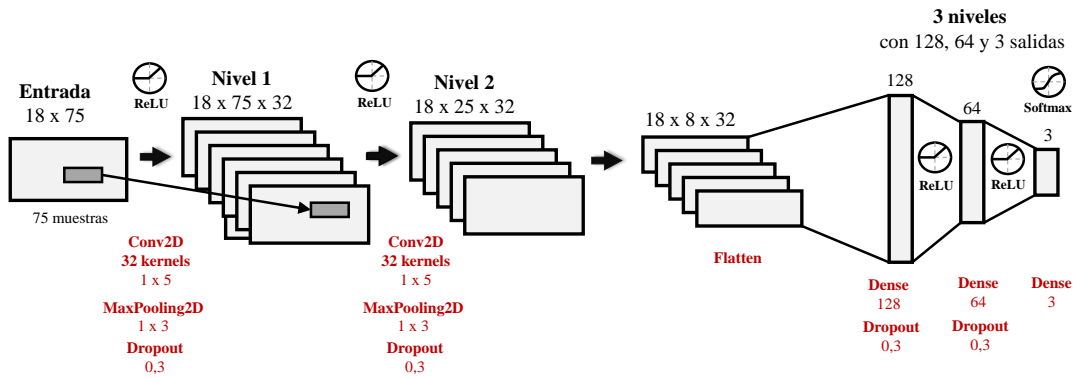


Figura 7.4: Arquitectura CNN final para FFT - PAMAP2.

7.2.2. Sistema de clasificación de actividades finales

El principal objetivo de un sistema de reconocimiento de actividades humanas es la clasificación de actividades. Con el fin de obtener la mejor tasa de acierto a la hora de clasificar actividades, se han analizado dos estrategias diferentes usando datos de tipo FFT y Raw. Estas estrategias han sido:

- Clasificar las 12 actividades directamente usando el sistema de HAR en LOSO con todas las actividades.



Figura 7.5: Estrategia de PAMAP2 con un paso.

- Usar el módulo clasificador de la Sección 7.2.1, y los sistemas de HAR en LOSO de los movimientos repetitivos y posturas.

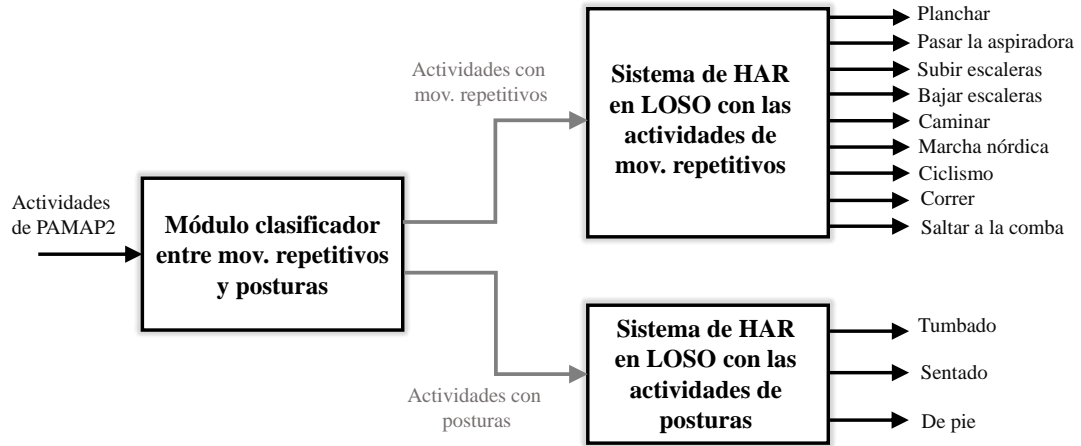


Figura 7.6: Estrategia de PAMAP2 con dos pasos.

En consecuencia, se han diseñado varios sistemas de HAR en LOSO para PAMAP2, en función de las actividades a clasificar. Inicialmente, la arquitectura utilizada en todos ellos ha sido la mostrada en la Figura 6.1, y posteriormente se ha optimizado la red en función de la tasa de validación.

- Sistema de HAR en LOSO con todas las actividades (12)
- Sistema de HAR en LOSO con las actividades de movimientos repetitivos (9)
- Sistema de HAR en LOSO con todas las actividades de posturas (3)

7.2.2.1. Módulo clasificador de las 12 actividades usando directamente un sistema HAR

Para empezar, el primer sistema de HAR planteado ha sido con todas las actividades existentes en la base de datos de PAMAP (12 actividades). Su principal objetivo es clasificar directamente las actividades, sin utilizar ningún tipo de módulo clasificador de tipos de movimientos previo. En este sistema los datos de entrada utilizados han sido de tipo FFT. Dichos datos, usan como etiqueta la del tipo de actividad final, en lugar de la etiqueta del tipo de movimiento.

En este módulo se ha utilizado únicamente los acelerómetros y giróscopos de la muñeca, pecho y tobillo. Para el experimento realizado usando los datos de la FFT se consigue para datos de entrada de la FFT una tasa de test del 87,83 % y una tasa de validación del 87,96 %, mediante ventanas de tres segundos de duración y paso entre ventanas de 0,25 segundos.

La Figura 7.7 muestra la matriz de confusión de test del sistema HAR con todas las actividades existentes en PAMAP2.

Matriz de confusión del sistema de HAR con todas las actividades - PAMAP2

Tumbado	7193	23	271	0	0	1	0	7	0	99	107	0
Sentado	14	5471	1359	4	0	24	25	9	5	98	394	5
De pie	0	563	6422	1	0	13	2	27	8	38	516	8
Caminar	0	0	230	9121	0	0	37	27	39	24	65	8
Correr	12	20	126	26	3608	4	1	14	35	18	41	16
Ciclismo	0	15	70	0	1	6198	2	0	92	95	110	2
Marcha nórdica	0	72	80	446	3	4	6781	4	7	31	91	6
Sub. escaleras	2	26	327	128	12	0	0	3833	188	21	136	16
Baj. escaleras	1	49	330	162	19	44	11	237	2837	83	170	258
Pasar la aspiradora	32	24	163	2	0	9	9	37	42	6301	396	0
Planchar	0	93	357	0	0	4	1	1	0	271	8814	7
Saltar a la comba	0	37	60	72	0	0	5	1	58	3	52	1677
	Tumbado	Sentado	De pie	Caminar	Correr	Ciclismo	Marcha nórdica	Sub. escaleras	Baj. escaleras	Pasar la aspiradora	Planchar	Saltar a la comba

Figura 7.7: Matriz de confusión de test del sistema de HAR con todas las actividades - PAMAP2.

En la matriz de confusión se puede observar como las posturas de estar sentado y de pie, se confunden bastante, ya que el número de predicciones incorrectas en estos casos son bastante altos. Además, algunos movimientos repetitivos como de subir y bajar escaleras o caminar son fácilmente confundibles con la postura de permanecer de pie, ya que el movimiento realizado con manos y pies es similar. Asimismo, la actividad de planchar se confunde en numerosas predicciones con estar de pie. Estas predicciones incorrectas se deben a que mientras un sujeto esta planchando, este permanece de pie realizando la misma postura estática.

De esta forma, utilizando el módulo clasificador entre tipos de movimientos previo se puede evitar esta confusión por tipología, ya que unos son movimientos repetitivos y el otro es una postura.

7.2.2.2. Módulo clasificador de tipos de movimientos combinado con los sistemas de HAR de movimientos repetitivos y posturas

En segundo lugar, se ha analizado el uso del módulo clasificador de movimientos repetitivos y posturas, añadiéndole un sistemas de HAR, dependiendo de que tipo de movimiento se trate. Por consiguiente, se han implementado dos sistemas de HAR, uno con las nueve actividades de movimientos repetitivos y otro con las tres posturas de PAMAP2.

Los resultados de la red neuronal optimizada con ventanas de tres segundos y datos de entrada de la FFT y Raw, se muestran en la Tabla 7.3. En todos ellos, se han utilizado los acelerómetros y giróscopos de la muñeca, pecho y tobillo.

PAMAP2 - Sistemas de HAR de movimientos repetitivos y posturas			
Tamaño de ventana: 3 (s)		Paso de ventana: 0,25 (s)	
Sensores utilizados: Acelerómetros y giróscopos de la muñeca, pecho y tobillo			
Sistema	Datos de entrada	Tasa de test (%)	Tasa de validación (%)
Movimientos repetitivos	FFT	92,41 \pm 0,22	92,45 \pm 0,22
	Raw	87,76 \pm 0,27	92,93 \pm 0,21
Posturas	FFT	90,38 \pm 0,38	91,74 \pm 0,35
	Raw	89,66 \pm 0,39	89,78 \pm 0,39

Tabla 7.3: Sistemas de HAR de movimientos repetitivos y posturas - PAMAP2.

Se puede observar en la Tabla 7.3 que las tasas de test y validación, son notablemente superiores que para el caso del sistema de HAR de las 12 actividades, debido a que estos sistemas de HAR únicamente clasifican actividades con movimientos repetitivos o posturas, no ambos juntos. Para el caso del sistema con únicamente movimientos repetitivos, las tasas de validación son similares en ambos casos. No obstante, para el caso de la FFT se consigue una mayor tasa de test (92,41 %). Por otra parte en las posturas, las tasas de test y validación son inferiores respecto al caso anterior, debido a que como se ha comentado previamente diferenciar entre posturas es una tarea complicada.

Adicionalmente, se han realizado algunos experimentos añadiendo el magnetómetro de la muñeca, pecho y tobillo, a los sensores ya existentes. Sin embargo, con dicho sensor no se consigue una mejora significativa.

La tasa de acierto total de clasificar las actividades directamente usando estos subsistemas será la combinación de ambas tasas de acierto ponderada al número total de ejemplos de test que se utilicen en cada caso. De esta forma, la tasa final tras la optimización de las redes neuronales es de 88,16 \pm 0,38.

En la Tabla 7.4 se muestra la tasa de test para cada subsistema, donde se consigue una tasa de validación mayor.

Subsistema	Tasa de test (%)	Nº de ej. de test
Modulo clasificador de tipos de movimientos	$96,02 \pm 0,13$	77707
Sistema de HAR con las 9 act. de mov. rep	$92,41 \pm 0,22$	55000
Sistema de HAR con las 3 act. de posturas	$90,38 \pm 0,38$	22707
Tasa final tras optimizar:	$88,16 \pm 0,38$	

Tabla 7.4: Tasas finales de la estrategia del módulo clasificador de tipos de movimientos combinado con sistemas de HAR con ventanas de 3 segundos - PAMAP2.

Esta tasa final de test es superior que para el caso de la estrategia con un solo paso (sistema de HAR con todas las actividades) donde se consigue una tasa del 87,83 %. Sin embargo, esta mejora de la tasa de test en los experimentos no es significativa. En la Tabla 7.5 se observa una comparativa entre ambas estrategias, con su respectiva tasa de test. En este sentido, se plantea como línea futura incluir el tamaño de ventana como parámetro de optimización para el clasificador de tipos de movimiento, de tal forma que se pueda obtener mayor tasa de acierto cuando se modelan movimientos repetitivos y/o posturas teniendo en cuenta una duración superior a 3 segundos.

Estrategia	Tasa de test (%)
Clasificando las 12 actividades directamente	$87,83 \pm 0,23$
Utilizar módulo clasificador de tipos de movimientos combinado con sistemas de HAR	$88,16 \pm 0,38$

Tabla 7.5: Tasas de test finales de las estrategias utilizadas en PAMAP2.

7.3. Experimentos en OPPORTUNITY

Por otro lado, para la base de datos de OPPORTUNITY se han implementado distintos módulos para clasificar los cuatro tipos de movimientos existentes en OPPORTUNITY con los datos de entrada de la FFT, Raw y de la energía. La red neuronal utilizada inicialmente ha sido la mostrada en el Capítulo 6, posteriormente la red de cada módulo ha sido optimizada, sobre la tasa de validación. En este caso, únicamente se ha implementado el clasificador para tipos de movimientos, sin actividades finales. Los distintos módulos implementados han sido:

- Módulo clasificador de cuatros clases de movimientos directamente.

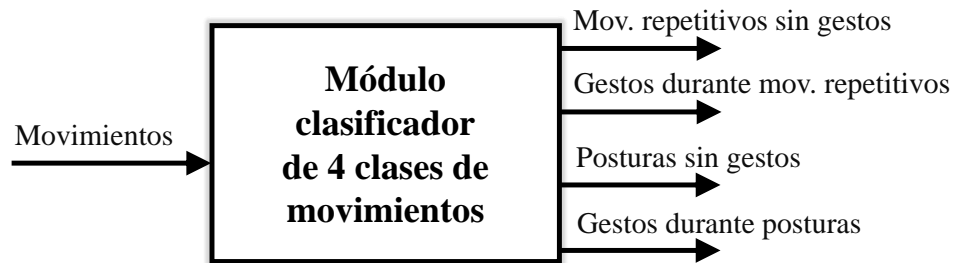


Figura 7.8: Estrategia de OPPORTUNITY clasificando movimientos directamente.

- Módulo clasificador realizando dos pasos.

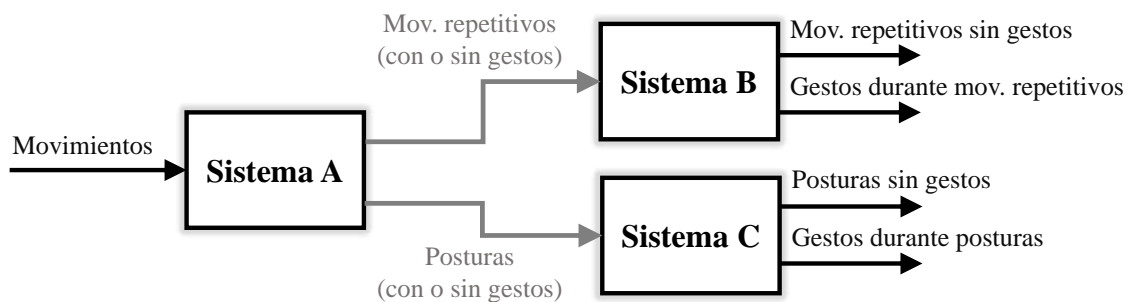


Figura 7.9: Estrategia de OPPORTUNITY clasificando movimientos en dos pasos.

En las Figuras 7.8 y 7.9 se muestra de forma gráfica las dos estrategias utilizadas en OPPORTUNITY para clasificar tipos de movimientos.

7.3.1. Módulo clasificador de cuatro clases de movimientos

En primer lugar, siguiendo la idea de PAMAP2 se ha diseñado un módulo para tratar de clasificar cuatro clases de movimiento. Como se ha explicado en secciones anteriores, se tratará de clasificar entre movimientos repetitivos sin gestos, gestos durante movimientos repetitivos o posturas, y posturas sin gestos.

Se han realizado tres diferentes experimentos, con datos de tipo FFT, Raw y de la energía de las señales. Al igual que en los experimentos anteriores, se han utilizado ventanas de tres segundos de duración con un paso entre ventanas de 0,25 segundos.

OPPORTUNITY - Módulo clasificador de cuatro clases de movimientos			
Número de ejemplos de test: 77814		Número de ejemplos de validación: 76973	
Tamaño de ventana: 3 (s)		Paso de ventana: 0,25 (s)	
Sensores utilizados: Acelerómetros y giróscopos de la mano, espalda y pie			
Datos de entrada	Tiempo de ejecución	Tasa de test (%)	Tasa de validación (%)
FFT	0:48:59.22	73,81 ± 0,30	74,56 ± 0,30
Raw	1:05:17.37	63,67 ± 0,33	64,09 ± 0,33
Energía	0:00:14.27	69,19 ± 0,32	69,87 ± 0,32

Tabla 7.6: Módulo clasificador de cuatro clases con ventanas de 3 segundos - OPPORTUNITY.

En la Tabla 7.6 se observan las tasas de test y validación obtenidas para los experimentos realizados con cuatro clases. En estos experimentos las tasas son bajas, comparadas con experimentos anteriores.

Matriz de confusión con 4 clases - OPPORTUNITY				
Mov. repetitivos sin gestos -	16525	142	1047	1009
Mov. repetitivos con gestos -	932	326	903	139
Posturas sin gestos -	955	146	13304	7356
Posturas con gestos -	1460	30	7354	26186
	Mov. repetitivos sin gestos	Mov. repetitivos con gestos	Posturas sin gestos	Posturas con gestos

Figura 7.10: Matriz de confusión del módulo con 4 clases - OPPORTUNITY.

La matriz de confusión de la Figura 7.10 muestra que en gran parte de los ejemplos las posturas con gestos se confunden con las posturas sin gestos. Conviene destacar la dificultad al clasificar entre estos dos tipos en el dominio de la frecuencia debido a la similitud de los movimientos en esas actividades. Además, los movimientos repetitivos con gestos son incorrectamente predichos en repetidas ocasiones por las posturas con gestos o los movimientos sin gestos, puesto que la actividad de andar puede enmascarar el gesto y se puede confundir con la postura estar de pie.

En este caso, únicamente se han utilizado ventanas de 3 segundos de duración, ya que los gestos son movimientos aislados, cortos en el tiempo. Utilizando ventanas de mayor duración, no se podría diferenciar correctamente los gestos.

7.3.2. Módulo clasificador realizando dos pasos

De forma similar, para tratar de mejorar las tasas conseguidas con el módulo clasificador de 4 clases se propone un nuevo módulo clasificador en dos pasos. Este módulo consiste en clasificar primero en dos clases, para tratar de diferenciar movimientos repetitivos, ya sean durante algún gesto o sin gestos, de las posturas (con o sin gestos). Posteriormente una vez distinguido el tipo de movimiento, se trata de separar las actividades con gestos que se hayan realizado durante dichos movimientos, de las que no.

Por consiguiente se han desarrollado tres sistemas diferentes:

- *Sistema A*: Módulo clasificador entre mov. repetitivos y gestos durante mov. repetitivos de posturas y gestos durante posturas.
- *Sistema B*: Módulo clasificador entre mov. repetitivos sin gestos y gestos realizados durante movimientos repetitivos.
- *Sistema C*: Módulo clasificador entre posturas y gestos durante posturas.

En este sentido, se han realizado experimentos mediante la red neuronal convolucional inicial con los datos de entrada de la FFT, Raw y la energía. Posteriormente, se han optimizado los sistemas donde se conseguían mayores tasas de validación. Por consiguiente, los datos que se observan en la Tabla 7.7 son los obtenidos tras la optimización de la red neuronal. En estos sistemas han sido utilizados los datos de los acelerómetros y giróscopos de los sensores de la mano derecha, espalda y pie derecho. Con ventanas de 3 segundos y paso entre ventanas de 0,25 segundos.

OPPORTUNITY - Módulo clasificador mediante dos pasos			
Tamaño de ventana: 3 (s)		Paso de ventana: 0,25 (s)	
Sensores utilizados: Acelerómetros y giróscopos de la mano, espalda y pie			
Sistema	Datos de entrada	Tasa de test (%)	Tasa de validación (%)
A	FFT	93,35 ± 0,17	92,60 ± 0,18
B	FFT	91,59 ± 0,37	94,87 ± 0,29
C	FFT	74,91 ± 0,35	77,92 ± 0,34

Tabla 7.7: Módulo clasificador mediante dos pasos - OPPORTUNITY.

Se puede observar en la Tabla 7.7 como en todos los sistemas se ha utilizado los datos de la FFT salvo para el sistema B, debido a que como se ha comentado en la Sección 4.3.2 el módulo de la FFT no es útil para tratar de clasificar entre movimientos con o sin gestos. Además, el sistema C es el sistema donde las tasas de acierto son menores. Como se ha comentado en el módulo clasificador de cuatro clases (apartado 7.3.1) clasificar posturas de gestos durante posturas es una tarea complicada a resolver.

Debido a que gran parte de los gestos realizados durante las posturas se desarrollan a través de las manos, se propone utilizar los datos de todos los sensores del brazo, acelerómetros (LUA , LUA^{\wedge} , LUA_{-} , RUA , RUA^{\wedge} , RUA_{-} , LH , RH , LWR , RWR , RLA y RLA) y giróscopos (RLA , RUA , LLA y LUA) existentes en la base de datos de OPPORTUNITY como se ha comentado en el Capítulo 3 para el sistema C. Añadiendo todos estos sensores, se consigue aumentar la tasa de test de este sistema del 74,91 % conseguido con los sensores de la mano, espalda y pie al 79,81 %. Asimismo, la tasa de validación también aumenta de 77,92 % a 80,61 %.

Finalmente, se ha calculado la tasa final de test obtenida con el módulo clasificador realizando dos pasos. Se han utilizado los experimentos donde las tasas de validación eran mayores. Dicha tasa final de test se ha calculado mediante la combinación de las tasas, ponderada al número de ejemplos de test que se utilicen en cada caso. En la Tabla 7.8 se puede observar como la tasa de test final es de 76,51 \pm 0,47.

Sistema	Tasa de test (%)	Nº de ej. de test
A	93,35 \pm 0,17	77814
B	91,59 \pm 0,37	21023
C	79,81 \pm 0,33	56791
Tasa final tras optimizar:	77,47 \pm 0,46	

Tabla 7.8: Tasas finales de la estrategia del módulo clasificador realizando dos pasos con ventanas de 3 segundos - OPPORTUNITY.

Estas tasas de test son superiores que para el caso de la estrategia con un solo paso (módulo clasificador de cuatro clases de movimientos) donde se consigue una tasa del 73,81 %. Por lo tanto, es mejor utilizar el módulo clasificador de movimientos mediante dos pasos. En la Tabla 7.9 se observa una comparativa entre las tasas de test conseguidas con ambas estrategias.

Estrategia	Tasa de test (%)
Módulo clasificador de cuatro tipos de movimientos	73,81 \pm 0,30
Módulo clasificador realizado en dos pasos	77,47 \pm 0,46

Tabla 7.9: Tasas de test finales de las estrategias utilizadas en OPPORTUNITY.

Capítulo 8.

Conclusiones

En este capítulo se valoran los resultados obtenidos tras la realización de este Trabajo Final de Grado en relación con los objetivos marcados al inicio del mismo. Asimismo, se plantean posibles líneas futuras de trabajo que permitan perfeccionar el módulo clasificador de tipos de movimientos.

8.1. Conclusiones

Tras la realización del proyecto, se han evaluado los resultados obtenidos de forma objetiva. En esta sección se pretende verificar uno por uno todos los objetivos marcados al inicio del trabajo, destacando las principales conclusiones obtenidas.

Para empezar, se han analizado las características de los diferentes tipos de movimientos en función de la energía por ventana. Además, se ha observado a partir de diferentes histogramas y estadísticos la energía existente en cada sensor dependiendo del tipo de movimiento realizado. De modo que en el pie existe más energía para los movimientos repetitivos que en el resto de sensores. De forma análoga, en las posturas el sensor de la mano es donde existe mayor energía, tanto en PAMAP2 como en OPPORTUNITY. Asimismo, se ha observado como la energía de las actividades con movimientos repetitivos es mayor que las posturas. Finalmente en este capítulo, se ha simulado mediante los experimentos “oracle” a través del umbral de energía el sistema se comporta siempre de la misma forma. El sensor de la muñeca es el más informativo, ofreciendo una tasa de acierto del 92,40 %, similar a la tasa de test obtenida mediante la red neuronal convolucional usando como datos de entrada la energía en el módulo clasificador de tipos de movimientos en PAMAP2, $92,06 \pm 0,19$ %.

Posteriormente, se han estudiado y analizado varias técnicas de procesamiento de señales. Como resultado, el módulo de la FFT es la técnica donde se obtienen mejores resultados a la hora de clasificar tipos de movimientos. Se ha observado que las posturas ofrecen información relevante en los datos de tipo Raw. Además, los gestos realizados durante movimientos repetitivos en OPPORTUNITY se pueden diferenciar a partir del eje X del sensor de la mano. Por otro lado, a partir del módulo de la FFT se distinguen mejor los movimientos repetitivos debido a los armónicos que presentan los espectrogramas.

Del mismo modo, se han estudiado y optimizado diferentes redes neuronales con el propósito de extraer características de las señales y clasificar los diferentes tipos de movimientos. Se ha comprobado que las redes neuronales convolucionales ofrecen buenos resultados en el reconocimiento de actividades. No obstante, construir una

arquitectura CNN para reconocimiento de actividades no es una tarea fácil, debido a que influyen muchos parámetros.

Finalmente, se ha diseñado un módulo clasificador de tipos de movimientos tanto para la base de datos de PAMAP2, como para OPPORTUNITY. Por un lado, en PAMAP2 la mejor técnica de procesamiento de la señal es el módulo de la FFT, usando una red neuronal con 2 niveles y 32 kernels de tamaño (1,5) se consigue una tasa de test del $96,02 \pm 0,13 \%$. En esta base de datos, utilizando el módulo clasificador de tipos de movimientos ($88,16 \pm 0,38 \%$) no se consigue mejora significativa respecto a clasificar las actividades finales directamente ($87,83 \pm 0,23 \%$). Por otro lado en OPPORTUNITY, se ha visto como utilizando un módulo clasificador en dos pasos con el módulo de la FFT como datos de entrada de la red neuronal, se consigue mejorar la tasa de test del $73,81 \pm 0,30 \%$ al $77,47 \pm 0,46 \%$ conseguido clasificando los cuatro tipos de movimientos existentes directamente. Del mismo modo, se potencia la tasa de acierto entre posturas y gestos durante posturas cuando se usan mayor número de sensores de una misma mano.

8.2. Líneas futuras

En esta sección, se exponen algunas ideas de líneas futuras de trabajo a partir del presente Trabajo Final de Grado. En primer lugar, a partir de los resultados obtenidos en el Capítulo 7, un posible trabajo futuro partiendo de la información de la que ya se dispone podría ser incluir la longitud de la ventana como parámetro de optimización, a la hora de diseño del módulo clasificador de tipos de movimientos. Ya que, como se ha visto a lo largo de este trabajo en ocasiones es recomendable utilizar ventanas de mayor duración, en función de la actividad a reconocer.

En segundo lugar, se puede replantear la tipología de movimiento. Originalmente, existen algunos movimientos repetitivos, como por ejemplo planchar o pasar la aspiradora, que pueden ser considerados actividades dinámicas repetitivas. Sin embargo, algunas de estas actividades incluyen movimientos aislados que podrían ser considerados como gestos.

Finalmente otra posible línea de investigación futura, es utilizar más sensores de diferentes localizaciones, añadiéndolos a los ya utilizados de la mano, pecho/espalda y pie. Así como, introducir otras bases de datos de actividades humanas, de cara a aumentar la información obtenida de los sujetos. También, se podría aumentar el número de sujetos encargados de recolectar los datos, realizando todos ellos las mismas actividades para garantizar el significado de los datos.

Referencias

- [1] SmartPanel, “¿Qué es el Deep Learning?” 2020, [En línea]. Diponible en: <https://www.smartpanel.com/que-es-deep-learning/>. Accedido: 03-mar-2021.
- [2] Atria Innovation, “Deep Learning y sus muchas aplicaciones,” 2019, [En línea]. Diponible en: <https://www.atriainnovation.com/deep-learning-aplicaciones/>. Accedido: 03-mar-2021.
- [3] S. Ranasinghe *et al.*, “A Review on Applications of Activity Recognition Systems with Regard to Performance and Evaluation,” *International Journal of Distributed Sensor Networks*, vol. 12, 07 2016.
- [4] C. Jobanputra *et al.*, “Human Activity Recognition: A Survey,” *Procedia Computer Science*, vol. 155, pp. 698–703, 01 2019.
- [5] M. Gil-Martín *et al.*, “Human activity recognition adapted to the type of movement,” *Computers & Electrical Engineering*, vol. 88, p. 106822, 2020.
- [6] M. Tentori and J. Favela, “Activity-Aware Computing for Healthcare,” *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 51–57, 2008.
- [7] J. L. Carús Candás *et al.*, “An automatic data mining method to detect abnormal human behaviour using physical activity measurements,” *Pervasive and Mobile Computing*, vol. 15, pp. 228–241, 2014, special Issue on Information Management in Mobile Applications Special Issue on Data Mining in Pervasive Environments.
- [8] W. Y. Cheng *et al.*, “Human Activity Recognition from Sensor-Based Large-Scale Continuous Monitoring of Parkinson’s Disease Patients,” in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2017, pp. 249–250.
- [9] X. Li *et al.*, “Deep Neural Network for RFID-Based Activity Recognition,” vol. 2016, 10 2016.
- [10] J. Soulas *et al.*, “Unsupervised discovery of activities of daily living characterized by their periodicity and variability,” *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 90–102, 2015.
- [11] I. Stančić *et al.*, “Gesture recognition system for real-time mobile robot control based on inertial sensors and motion strings,” *Engineering Applications of Artificial Intelligence*, vol. 66, pp. 33–48, 2017.
- [12] J. Qin *et al.*, “Compressive Sequential Learning for Action Similarity Labeling,” *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 756–769, 2016.

- [13] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.
- [14] Y. Chen and Y. Xue, "A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 1488–1492.
- [15] A. Khan *et al.*, "A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, pp. 1166 – 1172, 10 2010.
- [16] S. Bhattacharya and N. D. Lane, "From smart to deep: Robust activity recognition on smartwatches using deep learning," in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, 2016, pp. 1–6.
- [17] N. Y. Hammerla *et al.*, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, ser. IJCAI'16. AAAI Press, 2016, p. 1533–1540.
- [18] M. Zeng *et al.*, "Convolutional Neural Networks for human activity recognition using mobile sensors," in *6th International Conference on Mobile Computing, Applications and Services*, 2014, pp. 197–205.
- [19] L. Zhang *et al.*, "Recognizing Human Activities from Raw Accelerometer Data Using Deep Neural Networks," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, pp. 865–870.
- [20] S. Oniga and J. Sütő, "Human activity recognition using neural networks," in *Proceedings of the 2014 15th International Carpathian Control Conference (ICCC)*, 2014, pp. 403–406.
- [21] D. Singh *et al.*, "Human Activity Recognition Using Recurrent Neural Networks," *Machine Learning and Knowledge Extraction*, p. 267–274, 2017.
- [22] A. Reiss, "PAMAP2 Physical Activity Monitoring Data Set," Department Augmented Vision, DFKI, Germany, 2012, [En línea]. Dponible en: <http://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring>. Accedido: 20-abr-2021.
- [23] D. Roggen *et al.*, "OPPORTUNITY Activity Recognition Dataset," *Wearable Computing Laboratory, ETH Zurich*, 2009, [En línea]. Dponible en: <https://archive.ics.uci.edu/ml/datasets/opportunity+activity+recognition>. Accedido: 11-feb-2021.

-
- [24] A. Reiss and D. Stricker, “Creating and Benchmarking a New Dataset for Physical Activity Monitoring,” *ACM International Conference Proceeding Series*, 2012.
- [25] Organización Mundial de la Salud, “Obesidad y sobrepeso,” 2020, [En línea]. Diponible en: <https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight>. Accedido: 09-may-2021.
- [26] M. Tan *et al.*, “Obesity and covid-19: the role of the food industry,” *BMJ*, vol. 369, 2020, [En línea]. Diponible en: <https://www.bmj.com/content/369/bmj.m2237>. Accedido: 25-may-2021.
- [27] S. Ayuso, “El sobrepeso cuesta a cada español 265 euros adicionales de impuestos al año,” 2019, [En línea]. Diponible en: https://elpais.com/sociedad/2019/10/10/actualidad/1570697278_598482.html?ssm=TW_CC. Accedido: 09-may-2021.
- [28] R. Powers *et al.*, “Smartwatch inertial sensors continuously monitor real-world motor fluctuations in parkinson’s disease,” *Science Translational Medicine*, vol. 13, no. 579, 2021, [En línea]. Diponible en: <https://stm.sciencemag.org/content/13/579/eabd7865>. Accedido: 25-may-2021.
- [29] M. González-Gross *et al.*, “La nutrición en la práctica deportiva: Adaptación de la pirámide nutricional a las características de la dieta del deportista,” *Archivos Latinoamericanos de Nutrición*, vol. 51, pp. 321 – 331, 12 2001.

Anexo A.

Aspectos éticos, económicos, sociales y ambientales

En este anexo se incluye una reflexión sobre el impacto social, económico y ambiental de este trabajo. Así como las responsabilidades profesionales que este implica y el compromiso con la ética profesional.

A.1. Introducción

El reciente aumento de las redes sensoriales y su incorporación a la vida diaria de las personas, está favoreciendo el desarrollo del internet de las cosas (*IoT*). A través de estos sensores inteligentes se está permitiendo la monitorización de personas, siendo posible modelar y predecir su comportamiento.

Hoy en día, gran parte de la población dispone de un smartphone o un smartwatch, los cuales incorporan sensores que permiten realizar multitud de funciones. Además el reconocimiento de la actividad humana está llamando la atención debido al gran número de aplicaciones que se están desarrollando y el interés mostrado por las autoridades administrativas y organizaciones comerciales.

Es importante destacar que este trabajo se llevará a cabo en la fase de diseño, siendo de suma importancia la componente de investigación. En este sentido, el objetivo de este trabajo es crear un módulo que permita clasificar tipos de movimientos, pudiendo elegir la mejor estrategia de *Deep Learning* para poder diferenciar las actividades de cada tipo de movimiento. No obstante, en un futuro tendrá una gran relevancia en la sociedad actual, aunque llevará tiempo alcanzar al determinado público.

De esta forma, este anexo expondrá algunos de los aspectos más importantes relacionados con las futuras aplicaciones comerciales de este Trabajo Final de Grado. Las aplicaciones de este trabajo pueden mejorar la calidad de vida de las personas mientras realizan actividades de la vida diaria (*ADL*), ayudar a prevenir enfermedades o controlar la nutrición dependiendo de la actividad física que se realiza. En la Tabla A.1 se muestran algunos de los aspectos relacionados con las aplicaciones que se pueden realizar mediante este trabajo.

Este trabajo no está destinado a ninguna aplicación directa con el medio ambiente. Sin embargo, hay que destacar los efectos derivados del consumo computacional.

Aspectos	Aplicaciones en sistemas de HAR
Sociales	Reducción de enfermedades relacionadas con la actividad física
	Supervisión de personas en casas inteligentes
	Control de la nutrición dependiendo de la actividad física
Económicos	Reducción de costes en el ámbito de la salud
Ambientales	Efectos derivados del consumo computacional
Profesionales y éticos	Inclusión social
	Incremento de la productividad

Tabla A.1: Aspectos relacionados con las aplicaciones en sistemas de HAR.

A.2. Descripción de impactos relevantes relacionados con el proyecto

La Tabla A.2 muestra una descripción de los impactos más relevantes relacionados con el trabajo.

	Descripción	Grupos objetivos	Leyes regulatorias y compromiso ético	Posibilidades de evaluación	Implicaciones económicas
<i>Impacto 1</i>	Reducción de enfermedades relacionadas con la actividad física	Personas con obesidad y problemas de salud	Privacidad de la gente	Porcentaje de población con obesidad y problemas de salud	Reducción de costes sanitarios en los hospitales
<i>Impacto 2</i>	Supervisión de personas en casas inteligentes	Personas mayores o pacientes con enfermedades crónicas	Privacidad de la gente	Coste de la atención primaria para la supervisión de pacientes	Personas mayores o pacientes con enfermedades crónicas
<i>Impacto 3</i>	Reducción de costes en el ámbito de la salud	Hospitales y centros médicos	Privacidad de la gente	Coste de enfermedades específicas	Reducción de costes en los hospitales
<i>Impacto 4</i>	Inclusión social	Personas con problemas de movilidad	Privacidad de la gente	Satisfacción de los pacientes	-

Tabla A.2: Descripción de impactos relevantes relacionados con el proyecto.

A.3. Análisis detallado de alguno de los principales impactos

En esta sección se incluye un análisis detallado de algunos de los principales impactos comentados en la Tabla A.2.

Impacto 1

Una aplicación muy interesante del reconocimiento de la actividad humana, es la identificación de actividades físicas. La causa fundamental del sobrepeso y la obesidad es el descenso en la actividad física debido principalmente a la naturaleza cada vez más sedentaria de muchas formas de trabajo, nuevos modos de transporte y la creciente urbanización. De ahí, que la obesidad es una de las preocupaciones más importantes de la Organización Mundial de la Salud (*OMS*) [25]. En consecuencia, el Departamento de Salud y Servicios Humanos de Estados Unidos ha encontrado una relación directa entre la actividad física y la disminución del riesgo de sufrir obesidad, problemas en el corazón, Parkinson, diabetes e incluso cáncer. Además, de acuerdo con un artículo [26] publicado por *The British Medical Journal* existe una evidencia científica que indica la relación entre la obesidad y la COVID-19. De esta forma, la posibilidad de morir debido a la COVID-19 aumentan si la persona padece obesidad.

En este sentido, es posible ayudar a prevenir este tipo de enfermedades, incentivando la actividad física a través de la monitorización de las actividades de la vida diaria.

Es preciso resaltar, en relación al código ético, que la monitorización de actividades a través de sensores (acelerómetros, giróscopos...) es menos intrusiva que otros tipos de seguimiento médico. Además, se podrían reducir los costes económicos relacionados que conllevan estas enfermedades. Por poner un ejemplo, la obesidad cuesta a cada español 265 € adicionales de impuestos al año [27], por lo que si se logra reducir el número de personas que sufren esta enfermedad, supondría una reducción de los costes sanitarios en los hospitales.

Impacto 2

La supervisión de personas mayores o pacientes con enfermedades crónicas en casas inteligentes es otra aplicación *HAR* interesante. Esta aplicación ofrece numerosas ventajas a este tipo de personas, pudiendo disfrutar continuamente de un seguimiento médico en su domicilio y en caso de que exista algún problema poder actuar rápidamente. Este tipo de seguimientos es muy útil especialmente en el caso de pacientes con Alzheimer, Parkinson o con problemas de visión. Un artículo [28] trata de detectar problemas relacionados con el movimiento y los temblores en pacientes con Parkinson, mediante un sistema de monitorización que usa los sensores inerciales

de *smartwatches*. El estudio muestra que las evaluaciones del sistema coincidían con la estimaciones de un médico en el 94 % de los sujetos evaluados. Con respecto al código ético, el continuo seguimiento en los hogares puede afectar a la privacidad de la gente.

En relación con los aspectos económicos, el uso de esta aplicación podría disminuir el coste de la atención primaria relacionado con la supervisión de estos pacientes quienes tienen que ir continuamente al hospital para realizar seguimientos médicos.

Impacto 3

Otro impacto importante que puede tener este trabajo es el relacionado con el control de la nutrición dependiendo de la actividad física. Una dieta sana depende de muchos factores, como pueden ser la edad, sexo, metabolismo, actividad física... Por ello no se debe generalizar, sino que hay que individualizar dicha dieta.

Por consiguiente, dependiendo de algunos factores de la actividad física como pueden ser su intensidad, duración o musculatura implicada puede variar el gasto energético. De esta forma, un artículo [29] muestra que se pueden clasificar las actividades físicas en función del gasto energético, aeróbicas, mixtas o anaeróbicas, en vez de por tipo de movimiento.

Del mismo modo, mediante el uso de estas técnicas se puede ver afectado la privacidad de la gente en lo referido a los aspectos éticos. Por último, como se ha comentado en el Impacto 1 se podría lograr reducir los costes sanitarios.

A.4. Conclusiones

Para finalizar, es importante resaltar que este trabajo se ha centrado en el diseño de un módulo clasificador de movimientos. En este sentido, tiene un gran impacto. Sin embargo, conlleva bastante tiempo hasta que alcance al público general, por lo que los aspectos considerados en este anexo están enfocados con aplicaciones futuras.

En este sentido, la mayoría de los aspectos donde este trabajo podría tener impacto están relacionados con la actividad física, la reducción de enfermedades y de los costes sanitarios.

Anexo B.

Presupuesto económico

En este anexo, se detalla el presupuesto económico de este trabajo. Se ha dividido el presupuesto en diferentes costes: de personal, de recursos materiales, operaciones contratadas y material fungible.

- **Costes de personal:** La realización de este trabajo ha sido llevada a cabo por un estudiante del Grado de Tecnologías y Servicios de Telecomunicación. El ingeniero ha trabajado una cantidad de 330 horas durante 6 meses con un salario de 15 €/h (ver Tabla B.1).

	Coste horario (€)	Horas	Total (€)
Estudiante de ingeniería	15	330	4 950
TOTAL			4 950

Tabla B.1: Costes de personal.

- **Costes de recursos materiales:** En cuanto a los costes de recursos materiales de este proyecto, únicamente se ha usado un PC. Se ha considerado una amortización de 5 años, con un tiempo de uso de 6 meses (ver Tabla B.2).

	Coste (€)	Amortización (años)	Uso (meses)	Total (€)
PC	1 200	5	6	120
TOTAL				120

Tabla B.2: Costes de recursos materiales.

- **Operaciones contratadas:** Los costes generales (costes indirectos) ascienden al 15 % sobre los costes directos (costes de personal y recursos materiales) que son de 5070 €. Asimismo, el beneficio industrial es del 6 % sobre los costes directos e indirectos:

		Total (€)
Gastos generales	15 %	760,5
Beneficio industrial	6 %	349,83
TOTAL		1 110,33

Tabla B.3: Operaciones contratadas.

- **Material fungible:** En relación a los gastos del material fungible, se estima un coste de 50 €, relacionado con los gastos de impresión y encuadernación del presente documento.

Finalmente en la Tabla B.4 se muestra el coste total del proyecto. Se ha calculado el subtotal del presupuesto, y a continuación se le ha aplicado el 21 % de IVA. El coste final de este Trabajo Final de Grado ha sido de **7478,20 €**.

	Total
Costes de personal	4 950 €
Costes de recursos materiales	120 €
Operaciones contratadas	1 110,33€
Material fungible	50 €
Subtotal	6 230,33 €
IVA (21 %)	1 308,37 €
TOTAL PRESUPUESTO	7 538,70 €

Tabla B.4: Costes totales.

Anexo C.

Software utilizado

En este anexo, se muestra una breve descripción del software utilizado para la realización de este trabajo.

- **Octave**

GNU Octave es un programa y lenguaje de programación orientado para la realización de cálculos matemáticos. Tiene un interprete propio, el cual permite la ejecución de órdenes en modo interactivo. Además, es compatible con Matlab. En este proyecto, se ha usado para la extracción de características en la parte de procesamiento de la señal.

- **Anaconda**

Anaconda es una distribución de código abierto de Python, utilizado en *Machine Learning* permitiendo el procesamiento de grandes cantidades de datos. Se ha usado *Anaconda* con el fin de utilizar Python como lenguaje de programación en la parte de *Deep Learning*.

- **Keras**

Keras es una API escrita en Python usada para reducir la carga computacional. Este software es muy útil cuando se usa en redes neuronales.

- **Matlab**

Matlab es un sistema de cómputo numérico que ofrece un IDE junto con un lenguaje de programación propio. Entre sus principales características se encuentran la manipulación de matrices y la representación de datos. En este trabajo se ha utilizado para representar algunas figuras, como por ejemplo los espectrogramas.

- **Spyder**

Spyder es un entorno de desarrollo interactivo de código abierto para el lenguaje Python, diseñado para el ámbito científico, ingeniería y análisis de datos. Permite funciones avanzadas de edición, pruebas y depuración. Además, a través de distintas librerías (*pandas*, *matplotlib*, *numpy*...) permite la representación de datos 2D y 3D. Se ha utilizado, para la realización de las figuras de los distintos histogramas y matrices de confusión, entre otras.

- **Visual Studio Code**

El software de *Visual Studio Code* es un editor de código fuente. Se ha utilizado para la realización de los distintos ficheros Python (.py) y los ficheros .bat.

Anexo D.

Código extracción de características

En este anexo, se muestra varios fragmentos del código para la extracción de características de *Octave*. En primer lugar, se muestra el código relativo al enventanado aplicado a las señales, como se muestra en el Capítulo 4.

```
1  % Extracion de caracteristicas
2
3  fs = 100; % Frecuencia de muestreo
4  step_t = steptime; % Segundos
5  window_t = wintime; % Segundos
6
7  win_size = round(fs*window_t); % Ventana
8  step = round(fs*step_t); % Paso entre ventanas
9  final_frequency = 25; % Hz
10
11 label = Label(1:win_size);
12 FrameLabel(1,1) = 0;
13
14 maximum_label = 24; % Numero maximo de actividades
15
16 for j = 1:maximum_label % Enventanado
17     if (sum(label == j) >= win_size/2)
18         FrameLabel(1,1) = j;
19     endif
20 endfor
21
22 for i = 1:floor((columns(Label')-win_size-1)/step)
23     label = Label((i*step)+1:(i*step)+win_size);
24     FrameLabel(i+1,1) = 0;
25
26     for j = 1:maximum_label
27         if (sum(label == j) >= win_size/2)
28             FrameLabel(i+1,1) = j;
29         endif
30     endfor
31
32 endfor
```

Código D.1: Enventanado.

Posteriormente, se muestra la función *spectra*, donde se calcula el módulo de la transformada de Fourier (FFT).

```

1 function spectra = spectra(z,sr,wintime,steptime,flimit)
2
3     winpts = round(wintime*sr); % 3 s * 100 Hz = 300 puntos
4     steppts = round(steptime*sr); % 0.25 s * 100 Hz = 25 puntos
5     NFFT = winpts; % 300 puntos
6
7     % Ventana de Hamming
8     % hanning(m) -> Return the filter coefficients of a hanning window of
9     % length m
10    WINDOW = [hanning(winpts)'];
11    NOVERLAP = winpts - steppts; % 300 puntos - 300 puntos = 0
12    SAMPRATE = sr;
13
14    pspectrum_orig = abs(specgram(z,NFFT,SAMPRATE,WINDOW',NOVERLAP)).^2;
15
16    X=pspectrum_orig';
17    X=X.^0.33;
18    final = round(wintime*flimit); % 3 s * 25 Hz = 75 puntos
19    spectra = [X(:,1:final)];
20 endfunction

```

Código D.2: Módulo de la transformada de Fourier.

Finalmente, se muestra la función *interpolation* implementada para la interpolación de los datos de OPPORTUNITY para poder trabajar con la misma frecuencia (100 Hz).

```

1 function interpolation = interpolation(data)
2
3     fs_Opportunity=32; % 32 Hz
4     fs_PAMAP2=100; % 100 Hz
5     x_adapted = [0:1/fs_Opportunity:(rows(data)-1)*(1/fs_Opportunity)];
6     x_new = [0:1/fs_PAMAP2:(rows(data)-1)*(1/fs_Opportunity)];
7
8     interpolation = interp1(x_adapted, data', x_new)';
9
10 endfunction

```

Código D.3: Interpolación OPPORTUNITY.

Anexo E.

Código de modelado

En este anexo, se muestra parte del código de modelado a través de la extracción de características. En este fragmento, se muestra la definición de la red neuronal.

```
1
2 np.random.seed(2020)
3
4 output_path = sys.argv[1]
5 output_file = sys.argv[1] + "output.out" # Salida
6 sys.stdout=open(output_file,"w") # Se guardan los datos en el fichero
   especificado
7
8 num_channels = 18 # Número de señales totales
9 batch_size = int(sys.argv[6]) # Número de ejemplos que se utiliza antes de
   actualizar los pesos
10 epochs = int(sys.argv[5]) # Número de iteraciones que hace para entrenar
11 dropout = 0.3 # Porcentaje de neurona que se desconectan en las capas de
   dropout
12 cnn_num=16 # Número de filtros
13 kernel_num=3 # Dimensión del kernel)
14 num_classes=3 # Mov. repetitivas, gestos y posturas
15 ws = sys.argv[3] # Tamaño de ventana (segundos)
16 ss = sys.argv[4] # Paso de ventana (segundos)
17 dataset = sys.argv[7] # PAMAP2 o OPPORTUNITY
18 strategy = sys.argv[8] # Estrategia utilizada
19
20 ...
21
22 input_arch = Input(shape=(num_channels,num_points,1))
23
24 # Subred 1: Extracción de características de la red
25
26 # Nivel convolucional 1
27 x = Conv2D(cnn_num, (1, 5), padding='same',input_shape=(num_channels,
   num_points,1),activation='relu')(input_arch)
28 x = MaxPooling2D(pool_size=(1, 3))(x)
29 x = Dropout(dropout)(x)
30
31 # Nivel convolucional 2
32 x = Conv2D(cnn_num, (1, 5), padding='same', activation='relu')(x)
33 x = MaxPooling2D(pool_size=(1, 3))(x)
34 x = Dropout(dropout)(x)
35 x = Flatten()(x) # Aplanar (coge los datos y los pasa a una columna entera)
36
37 # Subred 2: Clasificación de las características en clases
38
39 x = Dense(128, activation='relu')(x) # Capa densa
40 x = Dropout(dropout)(x)
41 x = Dense(64, activation='relu')(x) # Capa densa
42 x = Dropout(dropout)(x)
43
44 # Dense de 3 neuronas (num_classes) y la función de activación es una softmax
45 end_arch = Dense(num_classes, activation='softmax')(x)
46
47 model = Model(inputs=[input_arch], outputs=end_arch) # Se guarda el modelo
```

```
48
49 model.summary() # Muestra la arquitectura completa de la red
50
51 # Se compila mediante el optimizador RMSProp con una tasa de aprendizaje de
    0,001 (default)
52 model.compile(loss='categorical_crossentropy',optimizer='rmsprop', metrics=['
    accuracy'])
53
54 ...
```

Código E.1: Red neuronal.