

**UNIVERSIDAD POLITÉCNICA DE MADRID**

**ESCUELA TÉCNICA SUPERIOR  
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO EN INGENIERÍA Y SISTEMAS DE  
DATOS**

**Diseño e implementación de un sistema de  
reconocimiento de actividades humanas  
con señales inerciales**

**Javier Cruz Fonseca**

**2026**

## GRADO EN INGENIERÍA Y SISTEMAS DE DATOS

### TRABAJO FIN DE GRADO

**Título:** Diseño e implementación de un sistema de reconocimiento de actividades humanas con señales inerciales

**Autor:** Javier Cruz Fonseca

**Supervisión técnica:** Rubén San Segundo Hernández

**Supervisión académica:** Rubén San Segundo Hernández

**Departamento:** Departamento de Ingeniería Electrónica

### MIEMBROS DEL TRIBUNAL

**Presidente:** D. .....

**Vocal:** D. .....

**Secretario:** D. .....

**Suplente:** D. .....

Los miembros del tribunal arriba nombrados acuerdan otorgar la calificación de:

.....

Madrid, a ..... de ..... de 20...

**UNIVERSIDAD POLITÉCNICA DE MADRID**

**ESCUELA TÉCNICA SUPERIOR  
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO EN INGENIERÍA Y SISTEMAS DE  
DATOS**

**Diseño e implementación de un sistema de  
reconocimiento de actividades humanas con  
señales iniciales**

**Javier Cruz Fonseca**

**2026**

## RESUMEN

La monitorización automática de la actividad física se ha convertido en un área de creciente interés dentro del ámbito de la salud digital, especialmente cuando se orienta hacia poblaciones vulnerables como las personas mayores. Poder reconocer qué está haciendo una persona permite analizar situaciones como la detección temprana de caídas, el seguimiento de su movilidad o comprobar si un programa de rehabilitación está funcionando bien.

En este Trabajo Fin de Grado el objetivo es implementar un sistema de reconocimiento de actividades humanas (HAR, por sus siglas en inglés *Human Activity Recognition*) capaz de reconocer actividades cotidianas mediante señales iniciales procedentes de acelerómetros. A diferencia de enfoques tradicionales que requieren el diseño manual de características a extraer de las señales iniciales, las redes neuronales pueden aprender automáticamente patrones relevantes directamente de los datos en bruto, adaptándose mejor a la variabilidad del movimiento humano.

El conjunto de datos utilizado, denominado HAR70+, contiene registros de aceleración triaxial (eje  $x$ ,  $y$ ,  $z$ ) de 18 personas con edades comprendidas entre los 70 y 95 años. Estos datos fueron capturados mediante dos sensores colocados en el muslo derecho y la zona lumbar mientras los participantes realizaban 7 actividades diferentes: caminar, arrastrar los pies, subir y bajar escaleras, estar de pie, estar sentado y estar acostado. En este trabajo se ha desarrollado y evaluado el comportamiento de un sistema de reconocimiento con diferentes números de ventanas temporales. La mejor solución ha consistido en una red neuronal con una capa de atención capaz de combinar información de varias ventanas consecutivas consiguiendo un 0,9268 de acierto sobre el conjunto de entrenamiento.

Este documento comienza abordando la evolución histórica del reconocimiento de actividades a través de la inteligencia artificial, poniendo especial énfasis en el aprendizaje profundo. Despues se describe la implementación del sistema tomando como punto de partida un código existente que se ha adaptado y modificado para ajustarse a las características de la base de datos HAR70+. Posteriormente, se han efectuado múltiples pruebas experimentales para determinar la configuración del sistema que ofrece mejor rendimiento. Además se explican los fundamentos matemáticos del modelo clasificador utilizado, junto con los aspectos teóricos de las técnicas implementadas, con el objetivo de poder justificar los resultados obtenidos. Para concluir, se ofrece una perspectiva general del trabajo desarrollado, incluyendo una reflexión sobre los resultados conseguidos y las posibles direcciones que podría tomar esta línea de investigación en un futuro.

## PALABRAS CLAVE

Reconocimiento de actividades humanas, aprendizaje profundo, inteligencia artificial, sensores iniciales, mecanismos de atención.

## SUMMARY

Automatic monitoring of physical activity has become an area of growing interest within the field of digital health, especially when geared towards vulnerable populations such as the elderly. Being able to recognize what a person is doing allows for the analysis of situations such as early fall detection, mobility monitoring, or verifying the effectiveness of a rehabilitation program.

In this Bachelor's Thesis, the objective is to implement a Human Activity Recognition (HAR) system capable of recognizing everyday activities using inertial signals from accelerometers. Unlike traditional approaches that require the manual design of features to be extracted from inertial signals, neural networks can automatically learn relevant patterns directly from the raw data, adapting better to the variability of human movement.

The dataset used, called HAR70+, contains triaxial acceleration records (x, y, z axes) from 18 people aged between 70 and 95 years. These data were captured using two sensors placed on the right thigh and lumbar region while participants performed seven different activities: walking, shuffling, climbing stairs, standing, sitting, and lying down. This work developed and evaluated the performance of a recognition system with varying numbers of time windows. The best solution consisted of a neural network with an attention layer capable of combining information from several consecutive windows, achieving a success rate of 0,9268 on the test set.

This document begins by addressing the historical evolution of activity recognition through artificial intelligence, with particular emphasis on deep learning. It then describes the system's implementation, starting with existing code that was adapted and modified to fit the characteristics of the HAR70+ database. Subsequently, multiple experimental tests were conducted to determine the system configuration that offers the best performance. Furthermore, the mathematical foundations of the classifier model used are explained, along with the theoretical aspects of the implemented techniques, in order to justify the results obtained. In conclusion, an overview of the work carried out is offered, including a reflection on the results achieved and the possible directions that this line of research could take in the future.

## KEYWORDS

Human activity recognition, deep learning, artificial intelligence, inertial sensors, attention mechanisms.

## ÍNDICE DEL CONTENIDO

<b>1. Introducción y objetivos-----</b>	<b>1</b>
1.1 Introducción-----	1
1.2 Objetivo y subobjetivos-----	2
1.3 Descripción de los capítulos-----	3
<b>2. Estado del arte-----</b>	<b>4</b>
2.1 Inteligencia artificial y aprendizaje automático en reconocimiento de actividades-----	4
2.1.1 Limitaciones de los enfoques tradicionales de aprendizaje automático-----	5
2.1.2 El cambio de paradigma: Aprendizaje de Representaciones-----	5
2.1.3 Arquitecturas neuronales para datos temporales-----	5
2.1.4 Mecanismos de atención-----	8
2.2 Reconocimiento de actividades humanas: evolución y aplicaciones-----	8
2.3 Artículos y proyectos de reconocimiento de actividades-----	9
<b>3. Material y Métodos-----</b>	<b>12</b>
3.1 Bibliotecas y dependencias necesarias-----	12
3.2 Base de datos Har70+-----	13
3.3 Diagrama general del sistema-----	14
3.4 Extracción de características (octave)-----	15
3.5 Entrenamiento de la red neuronal (python)-----	18
3.5.1 Carga de datos-----	19
3.5.2 Arquitectura de la Red-----	19
3.5.3 Metodología de evaluación-----	22
3.5.4 Intervalos de confianza-----	25
3.5.5 Agrupación de ventanas-----	25
<b>4. Experimentos-----</b>	<b>29</b>
4.1 Adaptación a la base de datos Har70+ y optimización inicial del sistema-----	29
4.1.1 Optimización del número de épocas-----	29
4.1.2 Optimización del dropout-----	31
4.1.3 Optimización del Learning rate-----	32
4.1.4 Optimización del batch size-----	33
4.1.5 Exploración de arquitecturas de red neuronal-----	34
4.2 Evaluación de las diferentes estrategias de agrupación de ventanas-----	35
4.2.1 Experimento I: optimización de la arquitectura con mecanismo de atención-----	35
4.2.1.1 Fundamentación teórica de la experimentación (I)-----	35
4.2.1.2 Resultados experimentales (I)-----	35
4.2.2 Experimento II: estrategia de votación multinivel-----	37
4.2.2.1 Fundamentación teórica de la experimentación (II)-----	37
4.2.2.2 Resultados experimentales (II)-----	38
4.2.3 Experimento III: exploración de parámetros de segmentación temporal-----	39
4.2.3.1 Fundamentación teórica de la experimentación (III)-----	39
4.2.3.2 Resultados experimentales (III)-----	40
4.2.3.3 Análisis comparativo de las 3 estrategias de agrupación de ventanas-----	41
4.2.4 Experimento IV: arquitecturas profundas con ventana deslizante-----	42

4.2.4.1 Fundamentación teórica de la experimentación (IV)-----	42
4.2.4.2 Optimización del dropout para la red ampliada 1-----	43
4.2.4.3 Resultados experimentales (IV)-----	44
4.2.4.4 Análisis comparativo de las arquitecturas profundas-----	46
4.3 Análisis comparativo de los experimentos-----	47
4.4 Análisis comparativo por sensores (IV - red ampliada 2)-----	48
4.5 Comparación con el modelo de referencia Har70+-----	50
4.5.1 Resultados comparativos-----	50
<b>5. Conclusiones y líneas futuras-----</b>	<b>52</b>
5.1 Conclusiones -----	52
5.2 Limitaciones y líneas futuras -----	53
<b>6. Aspectos, éticos, económicos, sociales y ambientales-----</b>	<b>54</b>
6.1 Descripción de impactos relevantes relacionados con el proyecto -----	54
6.2 Descripción detallada de un impacto-----	55
6.3 Ejemplos de aplicaciones en la vida real -----	56
6.4 Conclusiones-----	58
<b>7. Presupuesto económico-----</b>	<b>59</b>
<b>8. Bibliografía-----</b>	<b>60</b>

## 1. INTRODUCCIÓN Y OBJETIVOS

El presente Trabajo de Fin de Grado (TFG) ha sido desarrollado dentro del Grupo de Tecnología del Habla y Aprendizaje Automático del Departamento de Ingeniería Electrónica (DIE), perteneciente a la Escuela Técnica Superior de Ingenieros de Telecomunicación (ETSIT) de la Universidad Politécnica de Madrid (UPM). Esta práctica se sitúa en el área de la Inteligencia Artificial, más concretamente del uso de Aprendizaje Profundo (*Deep Learning* en inglés) para llevar a cabo reconocimiento de actividades.

### 1.1 INTRODUCCIÓN

La irrupción de la tecnología en el ámbito de la salud ha marcado una transformación paradigmática en la forma en que se aborda el bienestar físico y mental de las personas. Un ejemplo muy claro lo vemos en las aplicaciones y sistemas que nos ayudan a monitorizar nuestra actividad física. La clave está en la combinación de la Inteligencia artificial con los sensores, ya que gracias a esto podemos supervisar nuestra actividad de manera mucho más precisa y objetiva, además de obtener conclusiones personalizadas [1].

Dentro de la IA, el Reconocimiento de la Actividad Humana (HAR) se ha convertido en un área esencial. El HAR se encarga de la clasificación de las actividades que hacemos (caminar, correr, estar sentado, etc.) a partir de datos que son recogidos por los sensores. Aunque al principio se usaban mucho las cámaras y en análisis de vídeo, la tecnología ha avanzado a pasos agigantados, haciendo que los sensores sean más pequeños y que estén integrados en dispositivos que llevamos puestos como smartwatches o teléfonos, lo que ha abierto un mundo de posibilidades [2].

El HAR tiene un rango de aplicaciones muy amplio, pasando por ejemplo por la fisioterapia donde ayuda a corregir posturas y ejercicios para asegurar la efectividad del tratamiento y reducir el riesgo de lesiones. También es fundamental en la seguridad y vigilancia, al identificar comportamientos inusuales o sospechosos en tiempo real. Además, juega un papel importante en el desarrollo de hogares inteligentes, donde el reconocimiento de la actividad permite automatizar tareas como la iluminación o la climatización. No obstante, su aplicación más crítica se centra en el bienestar y la salud, permitiendo el seguimiento de persona con condiciones sensibles y la detección temprana de accidentes domésticos, como caídas, lo que lo convierte en un apoyo social y sanitario de gran valor [3].

Para lograr esta monitorización, sensores como los acelerómetros son fundamentales en este campo [4]. Estos dispositivos iniciales, que miden la aceleración en tres ejes (x,y,z), se han convertido en el estándar fundamental para la evaluación de la actividad física. Debido a su pequeño tamaño y bajo consumo, se pueden integrar fácilmente en cualquier dispositivo portátil para el registro del movimiento de forma continua. Esto nos permite analizar patrones de manera objetiva, desde el simple caminar hasta actividades más complejas. Los sesgos que aparecen cuando la supervisión es manual desaparecerían ya que ahora tenemos datos fiables que nos ayudan para la toma de decisiones clínicas o de entrenamiento [5].

Manejar la enorme cantidad de datos que generan estos sensores no es tarea fácil, ya que estamos hablando de flujos de información que deben ser procesados, limpiados e interpretados. Aquí es donde entra en juego el fenómeno del Aprendizaje Automático (*Machine Learning (ML)*). Las técnicas de ML, especialmente el aprendizaje profundo (*Deep Learning*), han revolucionado el HAR, permitiendo la creación de modelos capaces de clasificar actividades con una alta precisión, incluso en entornos ruidosos y variables [6].

Para ello utilizamos técnicas como el Aprendizaje por Refuerzo (*Reinforcement Learning, RL*), para procesar e interpretar esos datos de movimiento y clasificar las actividades.

El Aprendizaje por Refuerzo (RL) es un paradigma de ML particularmente prometedor en el contexto de la salud personalizada. A diferencia del aprendizaje supervisado tradicional, el RL se centra en cómo un agente de software debe tomar decisiones secuenciales en un entorno para maximizar una recompensa. En nuestro contexto, esto se traduce en crear un sistema que aprenda a tomar decisiones basadas en los datos para dar recomendaciones de forma dinámica y actuar como si fuese una especie de entrenador en tiempo real. Las aplicaciones de estos sistemas son, por lo tanto, muy amplias y tienen un impacto directo en nuestra vida. No solo sirven para que la gente se mantenga activa, sino que son herramientas muy valiosas para monitorear a personas de avanzada edad, personas con enfermedades crónicas y para optimizar tratamientos de rehabilitación, así como rutinas de entrenamiento deportivo. La capacidad de estos sistemas para proporcionar retroalimentación en tiempo real y adaptar las intervenciones es lo que marca la diferencia, permitiendo enfoques proactivos y personalizados en salud [7].

## 1.2 OBJETIVO Y SUBOBJETIVOS

El objetivo principal de este Trabajo Fin de Grado es diseñar, implementar y evaluar un sistema de Reconocimiento de Actividades Humanas (HAR) de alto rendimiento, basado en una arquitectura de Aprendizaje Profundo que incorpore una capa de atención [8] que permita integrar el conocimiento procedente de varias ventanas temporales consecutivas, con el fin de lograr una clasificación precisa y contextual de las actividades físicas diarias en el conjunto de datos HAR70+ [9].

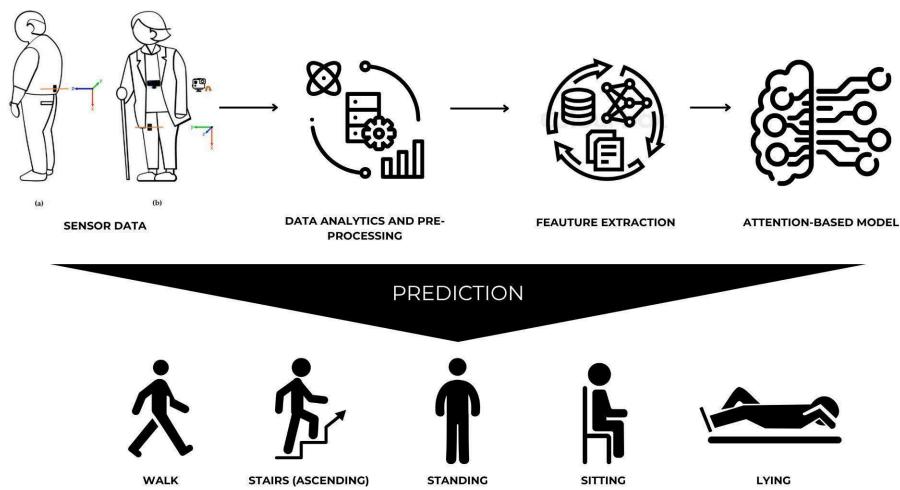


Figura 1. Idea del sistema propuesto

Para alcanzar este objetivo principal, se han definido los siguientes subobjetivos que han guiado todo el proceso de desarrollo:

- Estudio del estado del arte: Investigar y comprender los fundamentos del Reconocimiento de Actividades Humanas (HAR), las arquitecturas de Aprendizaje Profundo, y el funcionamiento de los mecanismos de atención.
- Adaptación e implementación del modelo Base: Partiendo del proyecto de referencia [10], adaptar la arquitectura de la red neuronal para que sea compatible con la

estructura y las dimensiones del conjunto de datos HAR70+. Esto incluye ajustar la red al completo, pasando por las capas de entrada y acabando por las capas de clasificación.

- Diseño e integración de la capa de atención: Implementar y añadir al modelo una capa de atención capaz de combinar múltiples ventanas temporales consecutivas, con el fin de mejorar la comprensión del contexto y la precisión a la hora de clasificar las actividades.
- Experimentación del modelo: Realizar una serie de experimentos para evaluar el rendimiento del sistema. Se analizará cómo varía la precisión del modelo al modificar tanto los hiperparámetros, como el número de ventanas.
- Evaluación y análisis de resultados: Evaluar el rendimiento del modelo final analizando la matriz de confusión para identificar las clases que se clasifican con mayor o menor acierto y así interpretar las conclusiones en el contexto del problema

### 1.3 DESCRIPCIÓN DE LOS CAPÍTULOS

Este Trabajo de Fin de Grado se ha organizado en 6 capítulos que serán detallados seguidamente:

1. Introducción y objetivos: Se introduce el campo del Reconocimiento de Actividades Humanas (HAR) y se definen tanto el objetivo principal del trabajo como los subobjetivos que han guiado su desarrollo.
2. Estado del arte: Se realiza un análisis de los estudios y publicaciones más importantes sobre HAR y Aprendizaje Profundo, situando el contexto tecnológico y científico de este proyecto.
3. Material y métodos: Se describe el conjunto de datos HAR70+ utilizado, las herramientas de software y hardware empleadas, y la metodología para implementar el modelo.
4. Experimentos: Se detallan cronológicamente todas las pruebas realizadas explicando las decisiones tomadas y los resultados intermedios obtenidos.
5. Conclusiones y líneas futuras: Se presentan las conclusiones del trabajo en base a los resultados y se proponen posibles vías de investigación para dar continuidad al proyecto.
6. Aspectos de impacto: Se analiza el impacto del proyecto desde una perspectiva ética, económica, social y ambiental, reflexionando sobre sus implicaciones en el mundo real.

La estructura de este documento ha sido diseñada para proporcionar una visión completa y progresiva del trabajo realizado. De esta forma, se busca que el lector pueda comprender tanto los logros técnicos alcanzados como el proceso de razonamiento que ha dado forma al sistema desarrollado.

## 2. ESTADO DEL ARTE

El reconocimiento automático de actividades humanas mediante sensores inerciales representa un punto de encuentro entre el desarrollo tecnológico en dispositivos portátiles y los avances en técnicas de aprendizaje profundo. En este capítulo se analiza cómo han evolucionado ambos campos y explora los trabajos de investigación que han servido de base para este Trabajo de Fin de Grado.

### 2.1 INTELIGENCIA ARTIFICIAL Y APRENDIZAJE AUTOMÁTICO EN RECONOCIMIENTO DE ACTIVIDADES

El reconocimiento automático de actividades humanas ha experimentado una transformación radical con el desarrollo de la Inteligencia Artificial (IA) y, en particular, del aprendizaje automático. La IA ha permitido que los sistemas pasen de requerir programación explícita de reglas para cada escenario a ser capaces de aprender patrones complejos directamente de los datos.

Los primeros sistemas inteligentes para HAR surgieron en la década de 1980, cuando investigadores comenzaron a aplicar modelos ocultos de Markov (*Hidden Markov Models*, HMM) para reconocer secuencias de actividades [11]. Estos modelos probabilísticos podrían capturar transiciones entre estados, lo que los hace adecuados para modelar la naturaleza secuencial de las actividades humanas. Sin embargo, requerían definir manualmente tanto la estructura del modelo como las características de las señales que se iban a procesar.

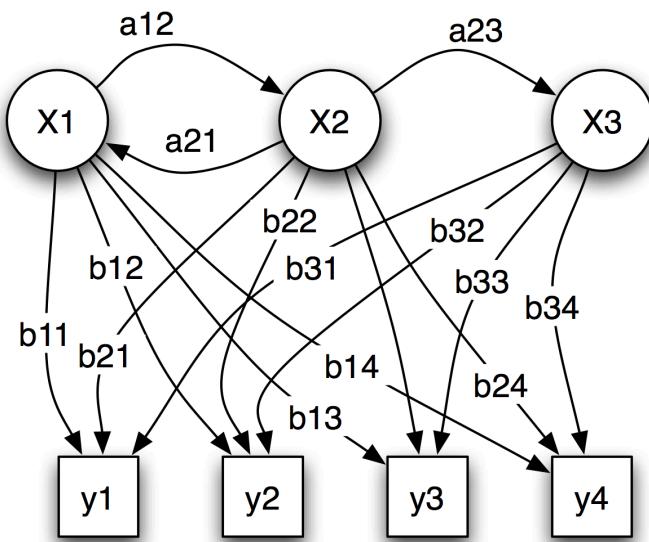


Figura 2. Modelos ocultos de Markov (*HMM: Hidden Markov Models*) [12]

El verdadero cambio llegó con el auge del aprendizaje automático a principios de los años 2000. Trabajos pioneros demostraron que algoritmos como *Support Vector Machines (SVM)* y *Random Forest* podían clasificar actividades con tasas de acierto superiores a 0,90 en entornos controlados. Estos métodos constituyen la primera generación de sistemas

inteligentes para reconocimiento de actividades, aunque todavía dependían de la intervención humana en el diseño de características.

### 2.1.1 LIMITACIONES DE LOS ENFOQUES TRADICIONALES DE APRENDIZAJE AUTOMÁTICO

Durante más de dos décadas, el reconocimiento de actividades se ha abordado principalmente con métodos de aprendizaje automático que requerían procesar las señales en múltiples etapas diferenciadas. El enfoque tradicional necesitaba que expertos en procesamiento de señales diseñaran manualmente las características que se iban a extraer de los datos de los acelerómetros. En este proceso, conocido como ingeniería de características, consistía en calcular diferentes estadísticos sobre ventanas temporales de las señales: la media, desviación estándar, energía, coeficientes de autocorrelación, y transformadas de Fourier para capturar información en el dominio de la frecuencia.

Aunque estos métodos funcionaban razonablemente bien en entornos controlados, tenían una serie de limitaciones importantes. El movimiento humano es muy variable, pues depende de factores como la edad, la condición física, posibles patologías o el estilo personal de cada uno. Cada vez que se trabajaba con una nueva aplicación o grupo diferente de personas, había que repensar qué características usar, lo que hacía el proceso poco escalable.

Los algoritmos de clasificación que se usaban entonces (*Support Vector Machine (SVM)*, *Random Forest* o *K-Nearest Neighbors*) [13] dependían completamente de lo buenas que fueran esas características. Si el conjunto de características estaba mal diseñado, el clasificador no podía distinguir bien entre actividades diferentes. Y si se incluían características irrelevantes, se añadía ruido que empeoraba los resultados.

### 2.1.2 EL CAMBIO DE PARADIGMA: APRENDIZAJE DE REPRESENTACIONES

El aprendizaje profundo cambió radicalmente las cosas al eliminar la necesidad de diseñar características manualmente. Las redes neuronales profundas aprenden automáticamente representaciones de los datos organizadas de forma jerárquica, es decir, extrayendo patrones básicos en las primeras capas y los van combinando en capas más profundas para formar conceptos abstractos.

Este cambio fue posible gracias a varios factores que coincidieron en el tiempo. Por un lado, empezaron a estar disponibles conjuntos de datos grandes con muchas muestras etiquetadas, lo que permitía entrenar modelos con millones de parámetros sin que se produjeran problemas graves de sobreajuste. Adicionalmente, se desarrollaron algoritmos capaces de entrenar un número ilimitado de capas aumentando considerablemente la capacidad de aprendizaje. Finalmente, el desarrollo de hardware especializado hizo viable el entrenamiento de redes con muchas capas, por lo que finalmente hubo innovaciones en los algoritmos como las funciones de activación ReLU, técnicas de normalización por lotes y métodos de regularización como *dropout*.

### 2.1.3 ARQUITECTURAS NEURONALES PARA DATOS TEMPORALES

Las señales de los acelerómetros tienen características que las diferencian de otros tipos de datos. A diferencia de las imágenes, las series temporales de sensores inerciales capturan cómo evoluciona el movimiento continuamente a lo largo del tiempo. Esta naturaleza

secuencial necesita arquitecturas especializadas que puedan modelar dependencias temporales.

### **Redes Convolucionales para Series Temporales**

Aunque las Redes Neuronales Convolucionales (CNN) se desarrollaron inicialmente para visión por computador, han resultado muy efectivas para trabajar con datos inerciales. La clave está en adaptar las operaciones: en lugar de aplicar filtros sobre el espacio 2D de una imagen, se aplican sobre la dimensión temporal de una señal.

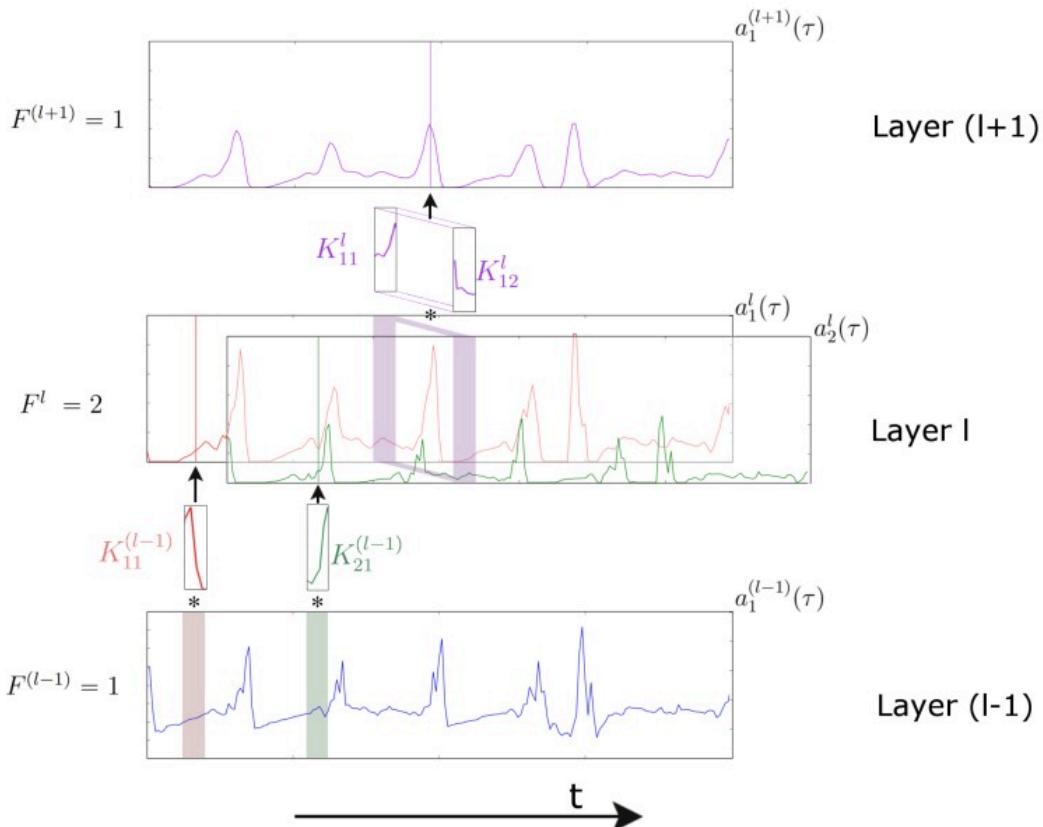


Figura 3. Representación de una convolución temporal de señales [15]

Las CNN ofrecen varias ventajas para HAR. Al compartir los mismos pesos en las convoluciones consiguen ser invariantes a desplazamientos temporales: un patrón característico de dar un paso se reconoce independientemente de cuando ocurra dentro de la ventana analizada.

Trabajos recientes han demostrado que usar arquitecturas con múltiples ramas de convoluciones con diferentes tamaños de *kernel*, puede capturar simultáneamente patrones que ocurren a velocidades distintas: desde impactos del pie al caminar hasta la frecuencia general de los pasos.

### **Redes recurrentes y memoria a largo plazo**

Una limitación de las CNN es que procesan cada ventana de forma independiente, sin guardar información sobre lo que pasó en ventanas anteriores. Esto se soluciona gracias a las Redes

Neuronales Recurrentes (*Recurrent Neural Networks*, RNN) ya que se va actualizando en cada paso temporal, funcionando como una especie de memoria del sistema.

El problema es que las RNN tradicionales tienen dificultades numéricas cuando se entrenan con secuencias largas. Las redes *Long Short-Term Memory* (LSTM) resuelven estos problemas con una arquitectura de celda que tiene puertas que regulan el flujo de información [16]. Hay una puerta de olvido que decide qué información antigua descartar, una puerta de entrada que determina qué información nueva almacenar, y una puerta de salida que controla que se transmite al siguiente paso.

Para HAR, las LSTM son especialmente útiles cuando las actividades duran más que una sola ventana o cuando el contexto previo es importante para clasificar correctamente. Por ejemplo, para distinguir entre “subir escaleras” y “bajar escaleras” puede ser necesario observar la secuencia de varios ciclos de movimiento. Las *Bidirectional LSTM* (BiLSTM) procesan la secuencia en ambas direcciones temporales, lo que permite que cada predicción tenga en cuenta lo que pasó antes como lo que viene después.

Una arquitectura que ha funcionado especialmente bien en HAR combina CNN y LSTM de forma secuencial [17]. Las capas convolucionales del principio procesan cada ventana temporal de forma independiente, aprendiendo a extraer características robustas de los patrones inerciales. Estas representaciones luego se pasan a capas LSTM que modelan las dependencias temporales entre ventanas consecutivas. Esta arquitectura híbrida consigue capturar tanto los patrones instantáneos como su evolución en el tiempo.

Este diseño también es computacionalmente eficiente porque las CNN reducen la dimensionalidad antes de llegar a las operaciones recurrentes, que son más costosas. Además, permite entrenar ambas componentes de forma coordinada: las CNN aprehenden qué características temporales son más útiles para clasificar secuencias, mientras que las LSTM aprenden cómo combinar información de múltiples ventanas.

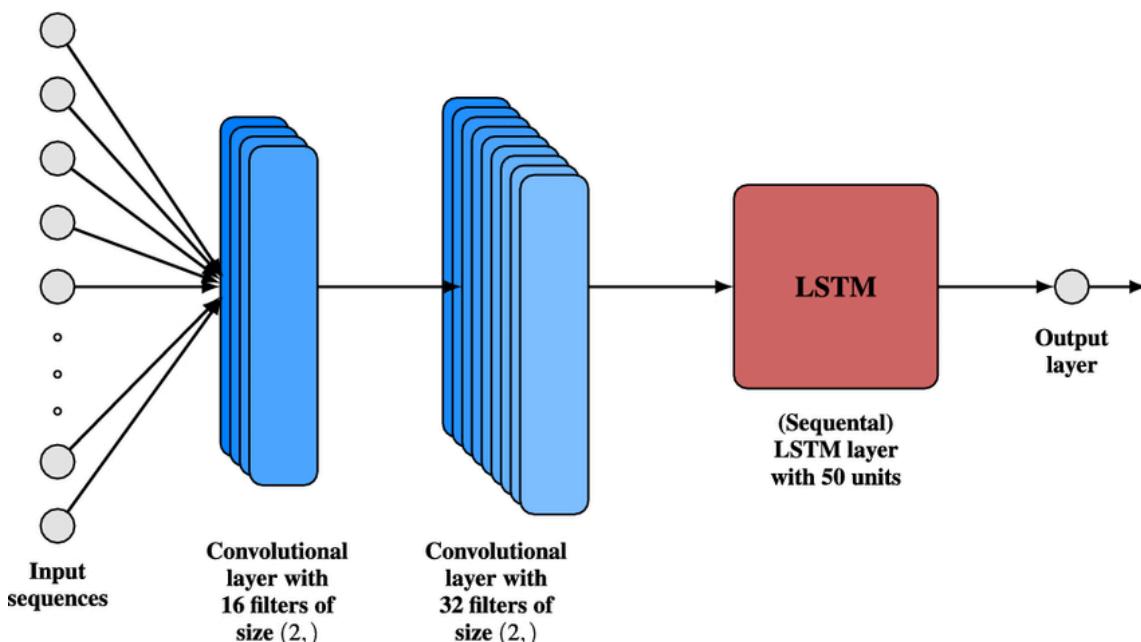


Figura 4. Arquitectura CNN-LSTM [18]

#### 2.1.4 MECANISMOS DE ATENCIÓN

Los mecanismos de atención han transformado el aprendizaje profundo al permitir que los modelos se enfoquen en las partes más relevantes de una secuencia. Aunque surgieron en traducción automática, hoy se aplican ampliamente en tareas con datos secuenciales.

##### ***Fundamentos conceptuales***

La atención asigna pesos a cada elemento de la secuencia según su relevancia para la predicción. Matemáticamente, funciona mediante consultas, claves y valores, calculando la importancia relativa con un producto escalar normalizado

Porcentaje de la ventana compartida.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Figura 5. Ecuación de las capas de atención

Esta normalización estabiliza los gradientes y permite al modelo combinar la información de forma ponderada.

##### ***Aplicación al reconocimiento de actividades con múltiples ventanas***

En el reconocimiento de actividades humanas (*Human Activity Recognition*, HAR), la atención permite priorizar las ventanas más informativas. Por ejemplo, una ventana que contiene el inicio de una actividad puede ser más útil que una situada en una fase estable.

## 2.2 RECONOCIMIENTO DE ACTIVIDADES HUMANAS: EVOLUCIÓN Y APLICACIONES

El reconocimiento de actividades humanas (HAR) es una disciplina en continua evolución que combina técnicas de inteligencia artificial, sensorial y procesamiento de señales para identificar patrones de comportamiento humano a partir de datos capturados por sensores. Su objetivo principal es permitir que los sistemas informáticos comprendan e interpreten las acciones humanas en distintos contextos.

##### ***Evolución general del campo***

En los últimos años, la investigación en HAR ha experimentado una transición significativa, de los sistemas basados en reglas y características manuales hacia enfoques automáticos y adaptativos capaces de aprender directamente de los datos. Esta evolución ha sido impulsada tanto por la disponibilidad de sensores portátiles como acelerómetros o giroscopios.

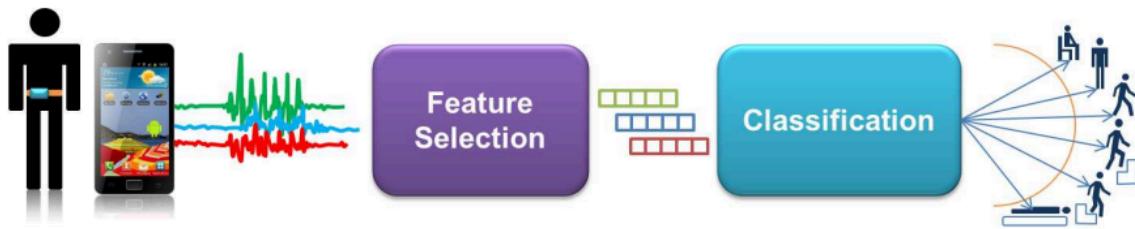


Figura 6. Estructura del sistema HAR

El desarrollo de bases de datos públicas, como UCI HAR [19], WISDM [20] o PAMAP2 [21] también ha sido clave para estandarizar las comparaciones y facilitar la evaluación de nuevas propuestas. Gracias a ellas, ha sido posible medir de manera objetiva la precisión de los algoritmos, acelerar la experimentación y fomentar la reproducibilidad de los resultados.

### ***Tipos de sistemas HAR***

Actualmente, los sistemas HAR pueden clasificarse en tres grupos:

1. Basados en sensores portátiles: Utilizan acelerómetros y giroscopios integrados en dispositivos como teléfonos, relojes inteligentes o pulseras de actividad
2. Basados en visión por computadora: Emplean cámaras RGB o de profundidad para analizar el movimiento corporal mediante detección de poses o análisis de secuencias de vídeo.
3. Multimodales: Combinan varias fuentes de datos para obtener una representación más completa del entorno.

### ***Aplicaciones actuales y líneas de investigación***

Las aplicaciones del reconocimiento de actividades humanas se han expandido significativamente en diversos sectores. En salud digital, los sistemas HAR permiten la monitorización continua de pacientes con enfermedades crónicas y personas mayores, detectando situaciones de riesgo como caídas o patrones anómalos de movilidad. En rehabilitación física, estos sistemas evalúan el progreso de los pacientes y verifican la correcta ejecución de ejercicios terapéuticos. El ámbito deportivo también se beneficia mediante el análisis de lesiones a través de la detección temprana de patrones de movimiento inadecuados.

El futuro del HAR se centra en una mayor integración de técnicas de inteligencia artificial, especialmente en el desarrollo de modelos de aprendizaje profundo más sofisticados. Además, se espera un mayor uso de técnicas de explicabilidad (XAI) que permitan comprender qué características de las señales iniciales son más relevantes para cada predicción.

## **2.3 ARTÍCULOS Y PROYECTOS DE RECONOCIMIENTO DE ACTIVIDADES**

El desarrollo de este Trabajo Fin de Grado se ha apoyado en una investigación previa que abarca desde trabajos fundacionales hasta aplicaciones recientes del aprendizaje profundo en

el reconocimiento de actividades humanas. Esta sección revisa los artículos, proyectos y conjuntos de datos más relevantes que han influido en el diseño e implementación del sistema propuesto.

### **Artículos de investigación**

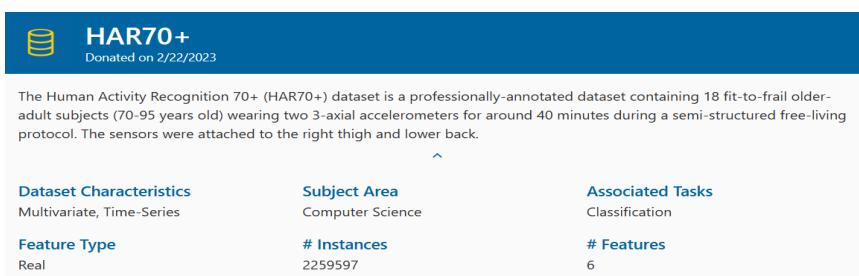
Uno de los artículos más destacados es “*An Improved Human Activity Recognition Technique Based on Convolutional Neural Network (Raj & Kos, 2023)*” [22], cuyo enfoque se basa en la mejora del rendimiento de las redes neuronales en tareas de HAR. La principal contribución del trabajo radica en el diseño de una arquitectura CNN que incorpora capas de *dropout* adaptativo y técnicas de normalización por lotes mejoradas. Los experimentos realizados con el dataset UCI HAR demostraron una precisión de 0,9780, superando en un 2.3% a las arquitecturas CNN estándar de la época.

Otro artículo reciente es “*Human Activity Recognition Using Attention-Mechanism-Based Deep Learning Feature Combination (Akter et al., 2023)*” [23] elaborado por Akter, Ansary, Khan y Kim cuyo objetivo era el de combinar múltiples características extraídas de diferentes etapas convolucionales mediante un mecanismo de atención. La arquitectura propuesta conecta las salidas de múltiples capas convolucionales a un módulo de atención que aprende automáticamente qué características son más relevantes para cada tipo de actividad. Los resultados en los datasets UCI HAR y WISDM mostraron precisiones de 0,9820 y 0,9650 respectivamente, demostrando la efectividad del mecanismo de atención para combinar información multi-escala.

Como último artículo está “*A New Deep-Learning Method for Human Activity Recognition (Vrskova et al., 2023)*” [24], que presenta una arquitectura híbrida que combina redes convolucionales 3D con capas convolucionales LSTM. Aunque muchos trabajos previos habían explorado la combinación de CNN y LSTM, Vrkova y sus colaboradores proponen utilizar convoluciones tridimensionales que procesan simultáneamente las dimensiones espaciales (ejes x,y,z del acelerómetro) y la dimensión temporal. Los resultados demostraron que esta combinación supera consistentemente a las arquitecturas 2D tradicionales, especialmente en la detección de actividades anómalas.

### **Artículo HART70+: Validation of an Activity Type Recognition Model Classifying Daily Physical Behavior in Older Adults**

Para la realización de este TFG, se ha utilizado la base de datos pública “*HAR70+: Validation of an Activity Type Recognition Model Classifying Daily Physical Behavior in Older Adults*”. En cuanto a su información más relevante tenemos la siguiente:



The screenshot shows the HAR70+ dataset landing page. At the top, there is a logo and the text "HAR70+" followed by "Donated on 2/22/2023". Below this, a brief description states: "The Human Activity Recognition 70+ (HAR70+) dataset is a professionally-annotated dataset containing 18 fit-to-frail older-adult subjects (70-95 years old) wearing two 3-axial accelerometers for around 40 minutes during a semi-structured free-living protocol. The sensors were attached to the right thigh and lower back." A small upward arrow icon is positioned below the text. The page then lists several statistics in a grid format:

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate, Time-Series	Computer Science	Classification
Feature Type	# Instances	# Features
Real	2259597	6

Figura 7. Información general y estadísticas de la base de datos

Los datos se recogieron utilizando dos acelerómetros triaxiales que actúan en los ejes X, Y, Z. Estos sensores se situaron en el muslo derecho y la parte baja de la espalda de 18 participantes con edades entre 70 y 95 años, mientras realizaban un protocolo de actividades cotidianas en un entorno controlado que simula condiciones de la vida diaria. Los datos se presentan en un formato CSV, incluyendo la posición en el espacio de ambos sensores y la hora de registro del dato, de la siguiente manera:

	timestamp,back_x,back_y,back_z,thigh_x,thigh_y,thigh_z,label
1	2021-03-29 14:42:07.460,-0.97168,-0.072266,-0.175781,-1.712158,-0.120117,1.502686,6
2	2021-03-29 14:42:07.480,-1.364746,0.182861,-0.377197,-1.580322,-0.170166,-0.145508,6
3	2021-03-29 14:42:07.500,-1.2495120000000002,0.1821289999999999,-0.466553,-1.052734,-0.261719,-0.783691,6
4	2021-03-29 14:42:07.520,-0.8415530000000001,-0.026855,-0.445557,-0.863281,-0.132568,-0.416992,6
5	2021-03-29 14:42:07.539,-0.669189,-0.068115,-0.380371,-0.7202149999999999,-0.083496,-0.291504,6
6	2021-03-29 14:42:07.559,-0.811523,0.0283199999999999,-0.358398,-0.912598,-0.149902,-0.391846,6
7	2021-03-29 14:42:07.559,-0.811523,0.0283199999999999,-0.358398,-0.912598,-0.149902,-0.391846,6

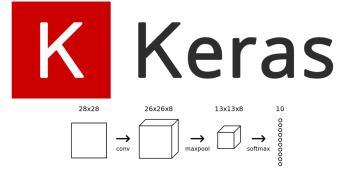
Tabla 1. Formato de las muestras en la base de datos

### 3. MATERIAL Y MÉTODOS

Este capítulo aborda el material y métodos desarrollados para la implementación del sistema de reconocimiento de actividades humanas desarrollado en este Trabajo de Fin de Grado. Se describe la base de datos HAR70+ empleada en el estudio, el código en Octave para la exploración y análisis de los datos, y finalmente el código en Python que detalla en el que se detalla el modelo de aprendizaje profundo. Este último incluye la arquitectura de red neuronal desarrollada, las bibliotecas utilizadas, la metodología de entrenamiento y la evaluación del sistema.

#### 3.1 BIBLIOTECAS Y DEPENDENCIAS NECESARIAS

El desarrollo del sistema requiere la importación de múltiples bibliotecas especializadas que proporcionan las funcionalidades necesarias para cada etapa del proceso:

Librería	Descripción	Logo
<b>Numpy</b>	Es una biblioteca que proporciona estructuras de datos eficientes para el manejo de matrices multidimensionales y operaciones matemáticas vectorizadas. Es esencial para el procesamiento de las señales iniciales que se representan como tensores de múltiples dimensiones.	
<b>Tensorflow</b>	Tensorflow [30] es el framework de aprendizaje profundo desarrollado por Google que proporciona las operaciones de bajo nivel para la construcción y entrenamiento de redes neuronales.	
<b>Keras</b>	Keras [31] es una API de alto nivel que funciona sobre TensorFlow, facilitando enormemente la implementación de arquitecturas neuronales complejas mediante una interfaz intuitiva y modular.	
<b>Scikit-learn</b>	Es una biblioteca estándar para aprendizaje automático que aporta herramientas robustas para la evaluación del rendimiento. Incluye funciones para calcular métricas como la matriz de confusión, exactitud, F1-score y curvas ROC.	

### 3.2 BASE DE DATOS HAR70+

El conjunto de datos HAR70+ [9] fue desarrollado para abordar una limitación crítica en el campo del reconocimiento de actividades humanas: la escasez de datos representativos de personas de avanzada edad. El objetivo es entrenar modelos de aprendizaje automático para el reconocimiento de la actividad humana a partir de datos de acelerómetros anotados profesionalmente de adultos con distintos grados de salud. Esta base de datos se centra exclusivamente en personas mayores cuyas características de movimiento difieren de las personas jóvenes. HAR70+ incluye registros de 18 participantes (9 hombres y 9 mujeres) con edades comprendidas entre los 70 y 95 años.

Los participantes fueron seleccionados siguiendo criterios de inclusión que garantizaban capacidad cognitiva para comprender y seguir instrucciones, capacidad de caminar sin ayuda durante al menos 10 metros, y ausencia de condiciones médicas agudas. Los datos fueron capturados utilizando dos acelerómetros triaxiales validados para la investigación que registran aceleración en tres ejes perpendiculares (x,y,z) con una frecuencia de muestreo de 100 Hz. La colocación de estos sensores fue en la zona lumbar y en la cara anterior del muslo derecho.

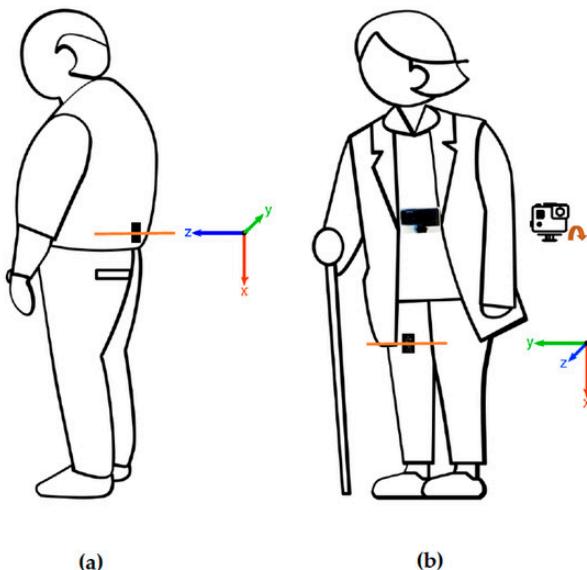


Figura 8. Posición de los sensores en el cuerpo [9]

El protocolo consistió en la realización de las siguientes 7 actividades durante aproximadamente 15 segundos: caminar, arrastrar los pies, subir escaleras, bajar escaleras, estar de pie, estar sentado y estar acostado. Cada secuencia se repitió 3 veces por participante generando múltiples instancias de clase que permiten evaluar la consistencia intra-sujeto. Incluyendo los descansos, la duración de la grabación fue de aproximadamente 40 minutos por sujeto, en la que el acelerómetro captura información con una frecuencia de 50Hz.

La orientación de los sensores fue meticulosamente definida: el sensor lumbar se colocó con el eje Z apuntando hacia adelante, mientras que el sensor del muslo se orientó con el eje Z apuntando hacia atrás. Esta estandarización en la orientación es vital para la reproducibilidad y el análisis de los datos, ya que permite a los investigadores interpretar las componentes de aceleración (ejes x,y,z) de manera consistente. Finalmente, la validación de la verdad fundamental mediante grabaciones de video sincronizadas y la anotación de los dos expertos

asegura que las etiquetas de actividad sean lo más precisas posibles, sentando una base sólida para el entrenamiento y la evaluación del modelo.

Dos expertos anotaron manualmente las actividades usando grabaciones de video sincronizadas, logrando entre ellos una alta concordancia. El preprocesamiento siguió los pasos descritos en el trabajo original de Bach [25], es decir, primero se descargaron los datos en bruto de la memoria del acelerómetro y luego se aplicó un filtro paso-banda Butterworth de cuarto orden a las seis señales, y por último tanto los datos de la espalda como los del muslo se sincronizan y almacenan en formato CSV. Posteriormente, las señales se segmentaron en ventanas temporales no solapadas de cinco segundos, lo que permitió la extracción de características (*features*) para cada segmento, que pueden ser utilizadas para entrenar clasificadores de aprendizaje automático. Se computaron las mismas 161 características derivadas del conjunto de datos HARTH [26]. Las 10 características que proporcionan más información para la predicción del tipo de actividad se muestran en la figura correspondiente:

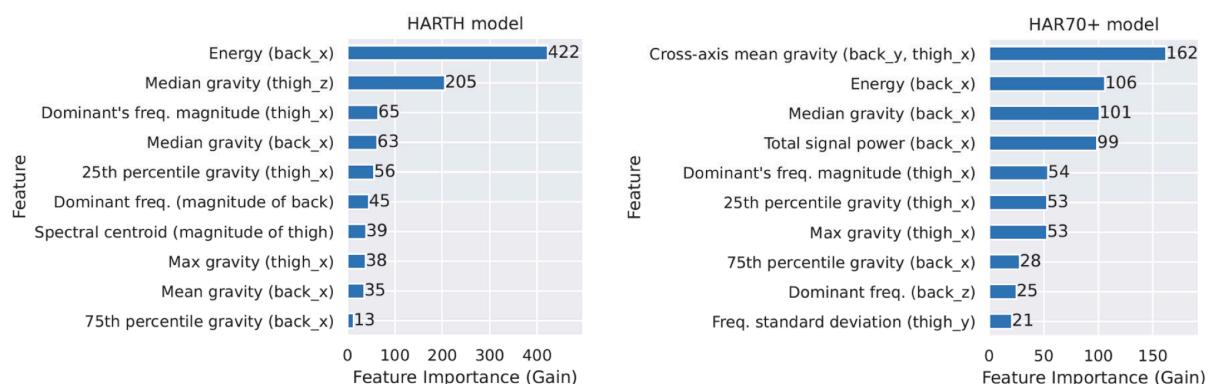


Figura 9. Las 10 características que proporcionan la mayor cantidad de información para la predicción del tipo de actividad en el modelo HARTH (izquierda) y el modelo HAR70+ (derecha). [9]

Para este Trabajo de Fin de Grado, se ha seguido una estrategia de validación *Leave-One-Subject-Out (LOSO)* en la que se entrena y valida el modelo con 16 participantes (repitiendo el proceso para todos los individuos y promediando los resultados) y se evalúa finalmente con los 2 restantes

Este TFG contribuye al estado del arte aplicando arquitecturas de Aprendizaje Profundo al HAR en personas mayores para abordar la escasez de datos en esta población. La contribución clave es la validación de la robustez del modelo mediante la estrategia LOSO, asegurando su capacidad de generalización para aplicaciones en salud digital.

### 3.3 DIAGRAMA GENERAL DEL SISTEMA

A continuación, se presenta un diagrama que resume el flujo de trabajo completo implementado en el script de Python:

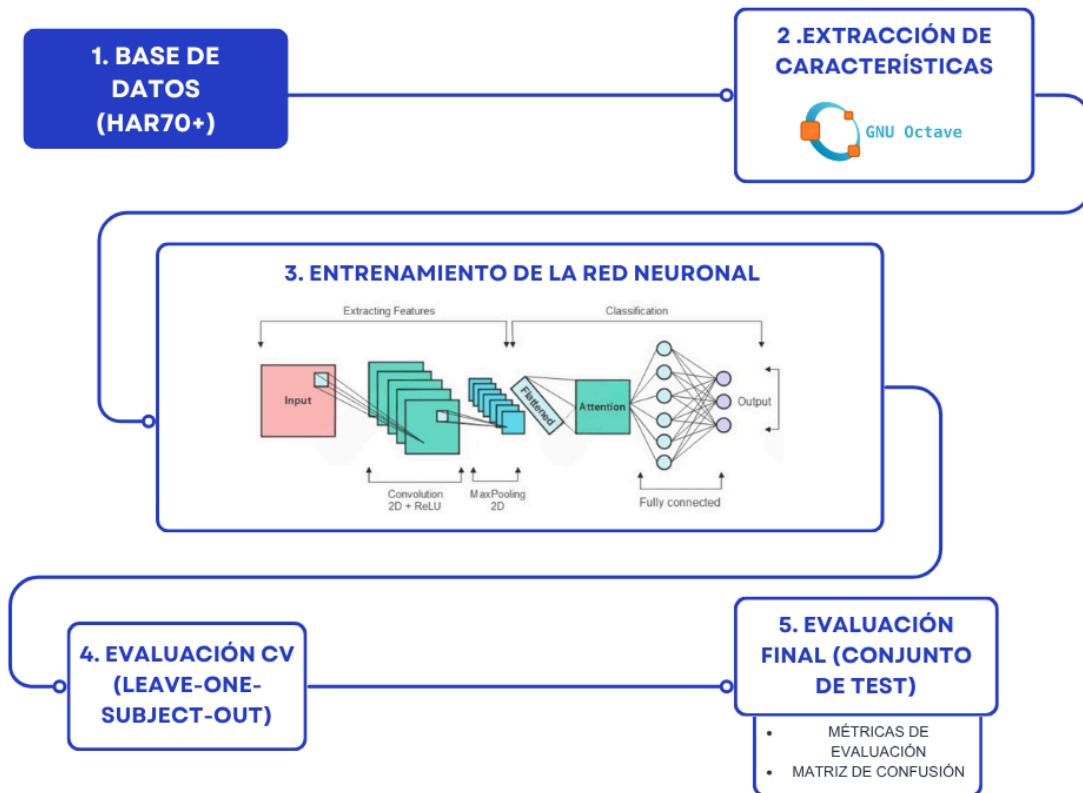


Figura 10. Diagrama general del sistema

### 3.4 EXTRACCIÓN DE CARACTERÍSTICAS (OCTAVE)

El código desarrollado en Octave [27] constituye la primera fase fundamental en el sistema de Reconocimiento de Actividades Humanas (HAR), encargándose del preprocesamiento de las señales iniciales y la extracción de características espectrales. El objetivo de este script es transformar los datos crudos de los acelerómetros, registrados en series temporales, en un formato estructurado y rico en información que pueda ser utilizado por el modelo de aprendizaje profundo para la clasificación de actividades.

Este proceso de preprocesamiento se articula en torno a la Transformada Q Constante [28], una técnica de análisis de frecuencia especialmente adecuada para el estudio de movimientos humanos al ofrecer una resolución logarítmica similar a la percepción auditiva. La CQT transforma las series temporales de aceleración en un vector de 144 características por cada ventana de 2 segundos, combinando 25 características espectrales por cada uno de los seis ejes de los dos sensores.

A continuación, se presenta un diagrama que resume el flujo de trabajo completo implementado en el script de octave:

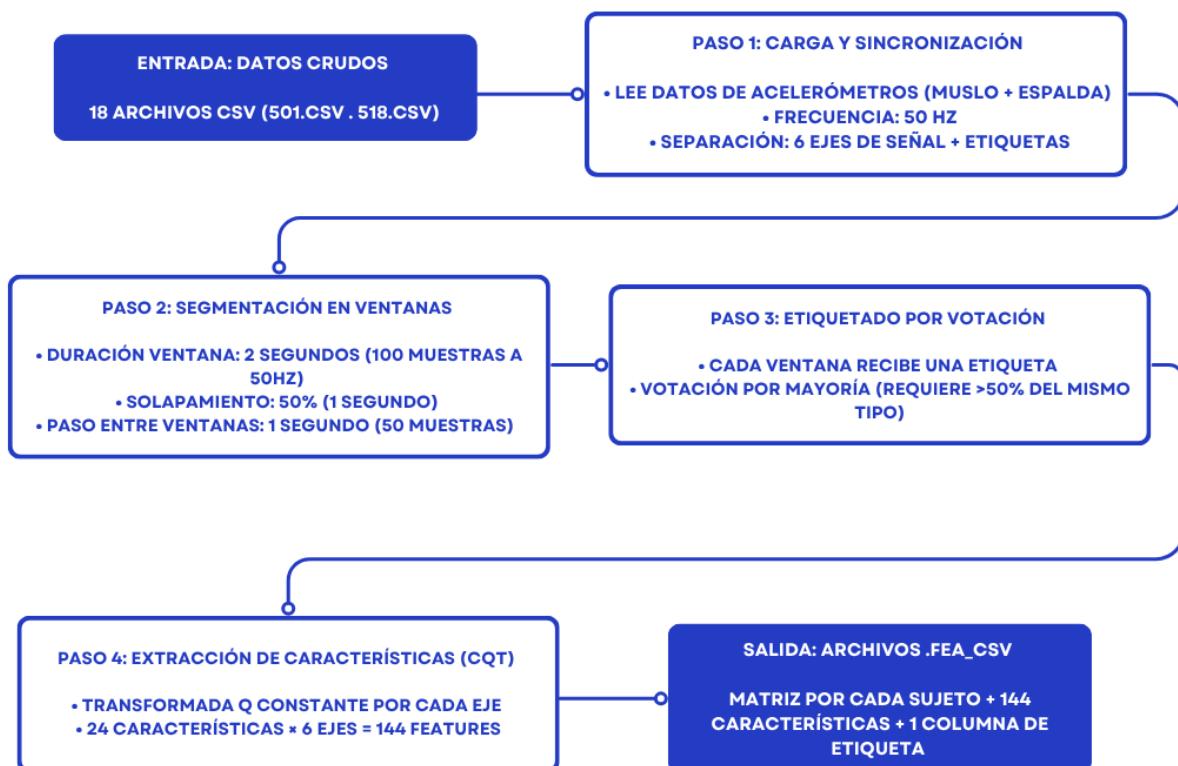


Figura 11. Diagrama general del sistema en Octave

### Estructura y flujo de ejecución

El script se organiza en torno a la función principal `calcula_features_mfcc_nine_total()`, que gestiona la lectura de una lista de archivos de datos, y de forma interactiva, invoca a la función `processing(ini_file)`, para cada uno de ellos. Esta función encapsula el pipeline de procesamiento, que comienza con la carga de la carpeta `data`, que contiene los 18 ficheros CSV, separando las señales de los 6 ejes del acelerómetro (3 por sensor) de las etiquetas de actividad.

El procesamiento de la señal se basa en el método de ventanas deslizantes, donde las series temporales se segmentan en ventanas de 2 segundos con un solapamiento de 1 segundo, dada una frecuencia de muestreo de 50 Hz. A cada ventana se le asigna una única etiqueta de actividad mediante un sistema de votación por mayoría, requiriendo que la etiqueta más frecuente ocupe al menos el 50% de las muestras de la ventana. Es importante destacar que, en esta etapa, las etiquetas originales correspondientes a “subir escaleras” (clase 4) y “bajar escaleras” (clase 5) se han unificado en una única clase (clase 3). Esta decisión metodológica se tomó para mitigar el sesgo de datos en estas categorías específicas.

Parámetro	Valor	Descripción
Frecuencia de Muestreo (fs)	50 Hz	Frecuencia registro de datos
Duración de la Ventana (window_t)	2 segundos	Duración de cada segmento de análisis
Paso entre Ventanas (step_t)	1 segundo	Desplazamiento entre ventanas consecutivas
Solapamiento de Ventanas	50% (1 segundo)	Porcentaje de la ventana compartida.

Tabla 2. Parámetros de Segmentación y Extracción

La extracción de características se realiza mediante la Transformada Q Constante (CQT), ya que es una técnica que es muy adecuada para el análisis de movimientos humanos al emular la percepción auditiva. El proceso implica calcular un espectrograma inicial, interpolar su resolución y aplicar un *kernel* precalculado que define los filtros logarítmicos. Posteriormente se aplica una raíz cúbica  $X^{3/3}$  a los coeficientes resultantes para comprimir el rango dinámico. Este proceso genera un vector de 24 características por cada uno de los 6 ejes, resultando en un vector final de 144 características por ventana temporal.

### ***Análisis de los Resultados de la Ejecución***

El registro de la ejecución del script de Octave sobre los 18 ficheros del conjunto de datos confirma el correcto funcionamiento del *pipeline* y permite cuantificar las dimensiones de los datos procesados.

Tomando como ejemplo la traza de ejecución para el primer fichero, se observa que el archivo original tiene 103,861 muestras. Tras la segmentación se generaron 2,076 ventanas temporales. La extracción de la Transformada de Q Constante, CQT resultó en una matriz de características de  $2076 \times 144$ , y finalmente, tras el filtrado de las ventanas de transición se obtuvieron 2,074 ventanas las cuales se almacenaron en el fichero *\*.fea\_csv* como una matriz de  $2074 \times 145$ .

Etapa del Proceso	Dimensión de la Primera Matriz (501.csv)	Interpretación
Carga de datos brutos	$103861 \times 8$	Muestras originales (6 ejes + índice + etiqueta).
Etiquetado de ventanas	$2076 \times 1$	Número total de ventanas de 2s con 50% de solapamiento.
Extracción de características	$2076 \times 144$	Características CQT (24 por eje x 6 ejes).
Almacenamiento final	$2074 \times 145$	Ventanas válidas listas para el modelo (144 + 1 actividad).

Tabla 3. Resultados de Ejecución - Primera Matriz (501.csv)

Este patrón se replica para los 18 sujetos, generando 18 ficheros *\*.fea\_csv*, cada uno con sus características especktrales extraídas. La conversión de datos crudos al espacio de

características CQT es clave, ya que ofrece una representación abstracta y discriminativa de la señal, facilitando al modelo la clasificación de actividades humanas.

### 3.5 ENTRENAMIENTO DE LA RED NEURONAL (PYTHON)

El código Python desarrollado para este Trabajo de Fin de Grado constituye el núcleo del sistema de reconocimiento de actividades humanas basado en aprendizaje profundo. Este script implementa una arquitectura neuronal innovadora que integra capas convolucionales temporales con un mecanismo de atención, permitiendo procesar múltiples ventanas temporales consecutivas de señales inerciales para mejorar la precisión en la clasificación de actividades.

El sistema está diseñado siguiendo una metodología de validación cruzada *leave-one-subject-out* (*LOSO*), que garantiza la capacidad de generalización del modelo al evaluar su rendimiento con datos de usuarios que no participaron en el entrenamiento. Esta aproximación es fundamental en aplicaciones de salud digital, donde el modelo debe funcionar correctamente con nuevos pacientes sin necesidad de reentrenamiento.

La segunda fase del sistema, implementada en Python, se centra en la carga de las características espectrales pre-calculadas en Octave y en el entrenamiento del modelo de Aprendizaje Profundo. Esta etapa es crucial, ya que transforma los datos estructurados en un clasificador funcional. El script de Python se encarga de la gestión de los datos, incluyendo la normalización, la aplicación de la estrategia de validación para garantizar la robustez y la capacidad de generalización del modelo a nuevos usuarios. El flujo de trabajo culmina con la definición, compilación y entrenamiento de la arquitectura de la red neuronal, que incorpora la capa de atención para la integración contextual de la información temporal.

A continuación, se presenta un diagrama que resume el flujo de trabajo completo implementado en el script de Python:

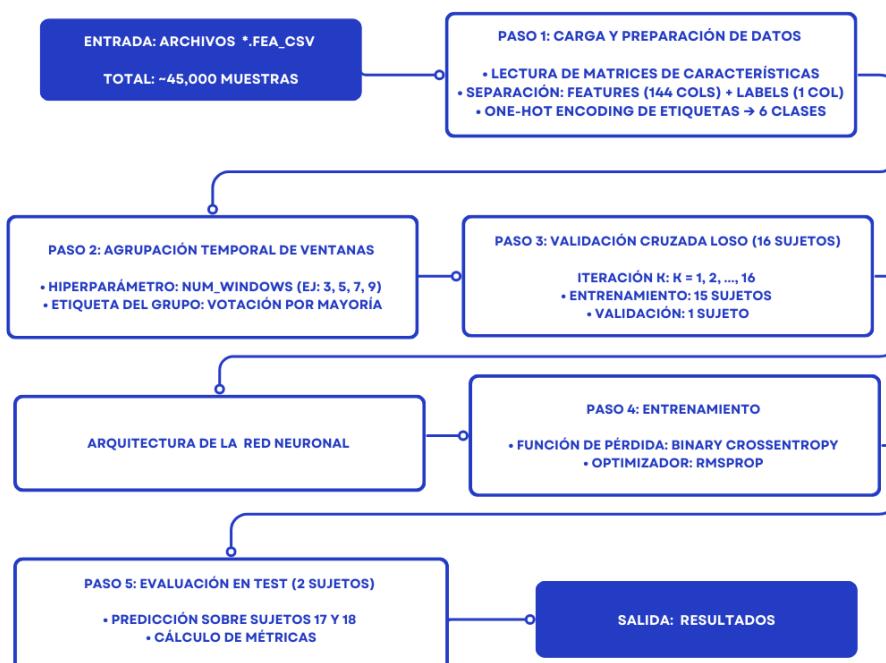
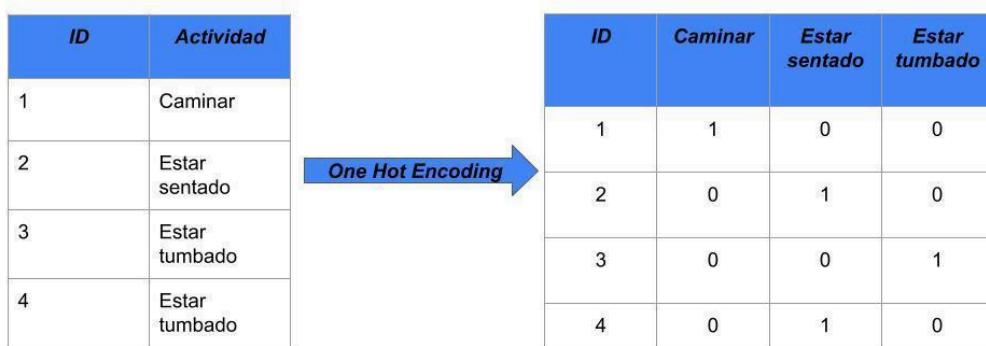


Figura 12. Diagrama general del sistema en Python

### 3.5.1 CARGA DE DATOS

El primer paso del proceso consiste en la lectura y procesamiento de los datos, donde se leen cada uno de los ficheros de características pre-extraídas y se itera sobre ellos para cargar los datos de los sensores iniciales, almacenándolos en forma de matriz multidimensional.

Para preparar las etiquetas de actividad, se emplea la técnica de *one-hot encoding*, que crea una columna separada para cada categoría posible de actividad. En este formato, solamente una de las columnas tendrá un valor 1 que indica la categoría correspondiente a la actividad real, mientras que todas las demás serán 0. Esta representación es especialmente útil para redes neuronales, ya que permite que la capa de salida produzca probabilidades para cada clase.



ID	Actividad	ID	Caminar	Estar sentado	Estar tumbado
1	Caminar	1	1	0	0
2	Estar sentado	2	0	1	0
3	Estar tumbado	3	0	0	1
4	Estar tumbado	4	0	1	0

Tabla 4. *One Hot Encoding*

### 3.5.2 ARQUITECTURA DE LA RED.

Las redes neuronales son modelos computacionales inspirados en el funcionamiento del cerebro humano, formadas por capas neuronales conectadas entre sí. Cada neurona realiza operaciones matemáticas sencillas que, al combinarse en capas sucesivas, permiten al modelo aprender patrones complejos directamente de los datos.

La estructura inicial de la red está compuesta por una capa de entrada que recibe los datos de los sensores, una o varias capas intermedias que procesan y transforman esos valores, y, por último, una capa de salida que genera la predicción final. La red está formada por dos fases: la primera consiste en la extracción de características y la segunda se encarga de la clasificación contextual mediante la capa de atención.

#### *Extracción de características*

**1. Capa de entrada:** Acepta como entrada un tensor que organiza la información en múltiples dimensiones: una dimensión para las diferentes ventanas de tiempo analizadas conjuntamente, otra para los distintos ejes de medición de los sensores, y una tercera para las características espectrales previamente calculadas mediante técnicas de procesamiento de señal.

**2. Capa *TimeDistributed* con Convolución 2D:** Implementa operaciones de convolución que se replican idénticamente sobre cada segmento temporal de forma aislada. Las convoluciones funcionan como detectores de patrones que buscan estructuras características en las señales, similares a como nuestro sistema visual detecta bordes o texturas en imágenes. Tras cada convolución se emplea una función ReLU que elimina valores negativos preservando únicamente las activaciones positivas:

$$\text{ReLU}(X) = \max(0, x) \quad (2)$$

Esta operación es fundamental porque añade la capacidad de modelar relaciones no lineales, algo imposible con transformaciones puramente lineales. Sin estas no linealidades, apilar múltiples capas sería matemáticamente equivalente usar una sola capa, limitando la expresividad del modelo.

**3. Capa *TimeDistributed* con *MaxPooling*:** Realiza una operación de reducción dimensional mediante selección de valores máximos en regiones locales. Este proceso cumple múltiples propósitos estratégicos: disminuye la cantidad de información que las capas posteriores deben procesar.

**4. Capa *TimeDistributed* con *Dropout*:** Implementa un mecanismo de regularización estocástica que desconecta aleatoriamente un porcentaje de neuronas durante cada paso del entrenamiento

**5. Capa *TimeDistributed* con *Flatten*:** Transforma la estructura multidimensional resultante del procesamiento convolucional en una representación vectorial unidimensional para cada ventana temporal.

Las capas *TimeDistributed* aplican la misma operación (convolución, *pooling*, *flatten*, etc.) de forma independiente a cada ventana temporal de una secuencia, compartiendo los mismos pesos entre todas ellas. Esto permite procesar cada segmento de tiempo por separado mientras se mantiene la estructura temporal intacta.

### **Bloque de Integración y Decisión**

**6. Capa de atención:** Constituye el elemento diferenciador de esta arquitectura. En lugar de tratar todos los instantes temporales con igual importancia, este mecanismo aprende dinámicamente a identificar qué segmentos de la secuencia son más informativos para cada tipo de actividad. El procedimiento matemático se estructura en tres fases:

En primer lugar, se aplica una función no lineal tanh:

$$e_i = \tanh(x_i W + b) \quad (3)$$

Posteriormente, estos valores se normalizan exponencialmente para obtener coeficientes de ponderación que suman la unidad:

$$\alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)} \quad (4)$$

Finalmente, se construye un vector agregado calculando la media ponderada de todas las representaciones temporales:

$$c = \sum_i \alpha_i x_i \quad (5)$$

**7. Capas Dense (64 y 32 neuronas):** El vector agregado se procesa mediante capas de neuronas totalmente interconectadas que progresivamente refinan la representación. Estas capas establecen conexiones complejas entre todos los elementos de entrada y salida, permitiendo aprender combinaciones arbitrariamente complejas de las características.

$$h = \text{ReLU}(Wc + b) \quad (6)$$

**8. Capas de Dropout:** Se intercalan capas de regularización estocástica entre las transformaciones densas. Estas capas proporcionan protección adicional contra el sobreajuste justo antes de la etapa de decisión final.

$$p_i = \frac{\exp(z_i)}{\sum_j \exp(e_j)} \quad (7)$$

donde  $z_i$  representa la puntuación para cada categoría y  $p_i$  su probabilidad normalizada.

Por tanto, la figura de la red neuronal sería tal que así:

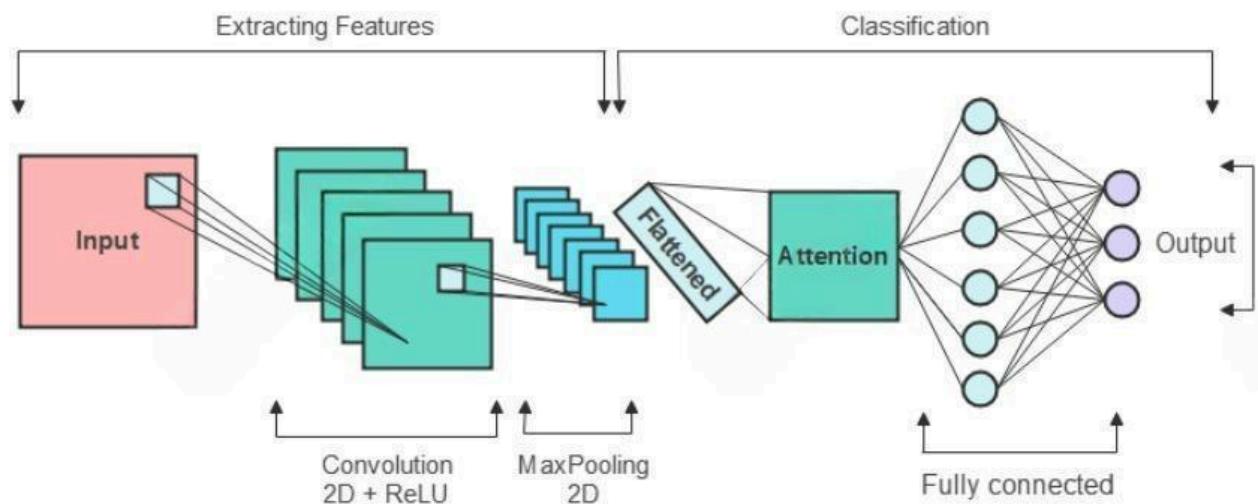


Figura 13. Estructura de la red neuronal

### Configuración del proceso de aprendizaje

El modelo requiere especificar varios componentes que determinan cómo aprenderá de los datos durante el entrenamiento:

- Función objetivo: Cuantifica la diferencia entre las predicciones generadas por el modelo y las etiquetas reales de los datos. Se emplea la entropía cruzada binaria, ya que funciona adecuadamente para problemas multiclase cuando las etiquetas están codificadas en formato *One-hot*: Su expresión matemática es la siguiente:

$$\text{loss} = - \sum_{i=1} y_i \cdot \log(p_i) \quad (8)$$

donde  $y_i$  indica la clase verdadera y  $p_i$  la probabilidad predicha.

- Algoritmo de optimización: Actualiza iterativamente los parámetros del modelo siguiendo la dirección que minimiza la función objetivo. Se utiliza *RMSprop*, un algoritmo adaptativo que mantiene estadísticas sobre los gradientes recientes y ajusta individualmente la magnitud de actualización para cada parámetro.
- Indicadores de rendimiento: Monitorizan el progreso del aprendizaje calculando la fracción de clasificaciones correctas.

#### 3.5.3 METODOLOGÍA DE EVALUACIÓN

##### *Protocolo de validación cruzada*

Por una parte, se realiza una validación cruzada exhaustiva mediante la técnica *Leave-One-Subject-Out* (LOSO) sobre los primeros 16 participantes de la base de datos HAR70+, reservando los 2 últimos usuarios para la evaluación final. La implementación se realiza mediante la clase *KFold* de *scikit-learn* [31] permitiendo procesar sistemáticamente cada uno de los 16 usuarios.

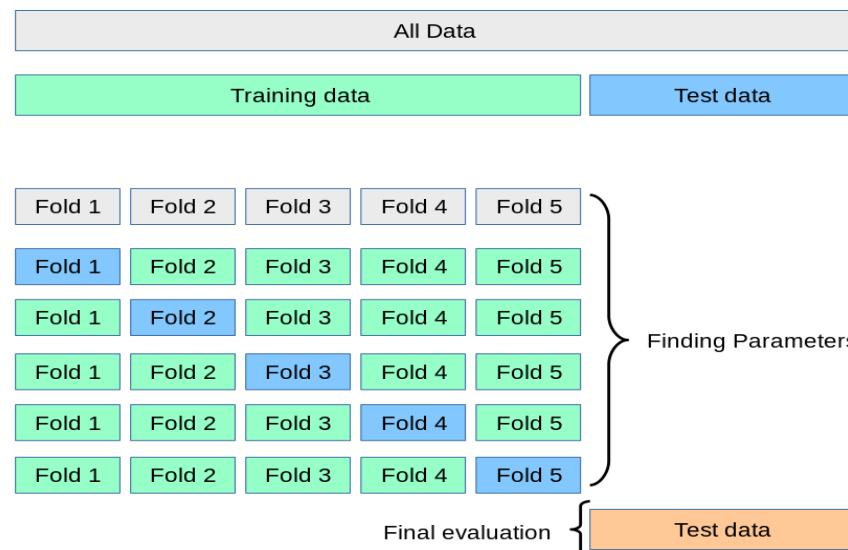


Figura 14. Esquema del proceso de entrenamiento en validación cruzada [34]

### Matriz de confusión

En sistemas HAR la matriz de confusión permite identificar confusiones sistemáticas entre actividades muy similares. Por ejemplo, si el modelo puede llegar a confundir “caminar” y “arrastrar los pies”. En este trabajo, se emplean un total de 45048 muestras distribuidas en las distintas actividades y un ejemplo de cómo se vería una matriz de confusión sería la siguiente, siendo los datos resaltados en verde aquellos que están clasificados correctamente:

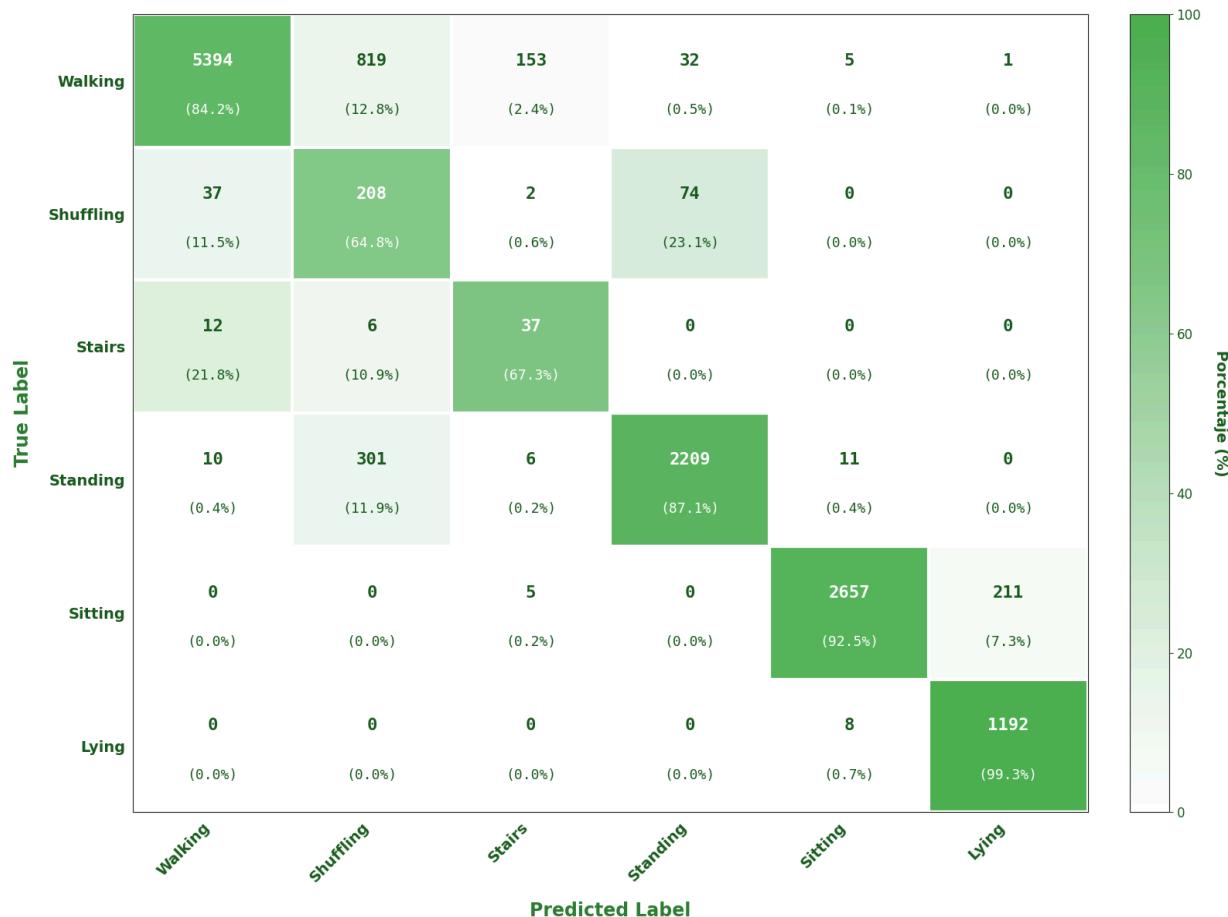


Figura 15. Ejemplo de matriz de confusión con tasa de acierto de 0,8740

### Tasa de acierto

La exactitud es la métrica más intuitiva ya que responde a qué porcentaje de las predicciones son correctas y se calcula dividiendo el número total de clasificaciones correctas entre en número total de muestras evaluadas:

$$A = \frac{1}{c} \cdot \sum_{i=1}^c a_i \quad (9)$$

donde  $a_i$  es la tasa de acierto de la prueba siendo el usuario  $c$  el lote de evaluación. En el contexto de este Trabajo de Fin de Grado, una tasa de acierto de 0,8550 representa que se han clasificado correctamente 86 de cada 100 ventanas aproximadamente.

### **F1-Score**

El F1-Score es una métrica que balancea dos aspectos fundamentales del rendimiento de un clasificador: la precisión y la exhaustividad. Su principal objetivo es el de equilibrar las clases en caso de que haya algún fallo:

$$F1 = 2 \cdot \frac{precision_{\alpha} \cdot sensibilidad_{\alpha}}{precision_{\alpha} + sensibilidad_{\alpha}} \quad (10)$$

En problemas multiclas como el nuestro, existen tres formas principales de agregar el F1-Score de las clases individuales:

1. **F1-Score Weighted:** Calcula el F1-Score para cada clase y luego promedia estos valores ponderándolos por el número de muestras reales de cada clase. Esta variante es útil cuando las clases están desbalanceadas y queremos que las clases más frecuentes tengan mayor influencia en la métrica final.

$$F1 - weighted = \sum_{i=c}^c \omega_i \cdot F1_i \quad (11)$$

donde  $\omega_i$  es el peso de la clase  $i$  definido como la proporción de instancias de esa clase sobre el total de instancias:

$$\omega_i = \frac{N_i}{N} \quad (12)$$

donde  $N_i$  es el número de instancias verdaderas en la clase  $i$ ,  $N$  es el número total de instancias.

2. **F1-Score Micro:** Agrega las contribuciones de todas las clases calculando primero las métricas globales de TP, FP y FN, y luego computando el F1-score sobre estos totales. Esta métrica es útil cuando queremos dar igual peso a cada muestra individual, independientemente de su clase. Este enfoque es matemáticamente equivalente a calcular la tasa de acierto cuando se trabaja con problemas multiclas, ya que al sumar globalmente los verdaderos positivos de todas las clases se obtiene el total de las predicciones correctas:

- Precisión micro:

$$P_{micro} = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FP_i} \quad (13)$$

- Recall micro:

$$R_{micro} = \frac{\sum_i TP_i}{\sum_i TP_i + \sum_i FN_i} \quad (14)$$

- F1- micro:

$$F1_{micro} = \frac{2 \cdot P_{micro} \cdot R_{micro}}{(P_{micro} + R_{micro})} \quad (15)$$

**3. F1-Score Macro:** Calcula el F1-Score para cada clase de forma independiente y luego promedia estos valores sin considerar el desbalance. Esta variante trata todas las clases como igualmente importantes, independientemente de su frecuencia.

$$F1 - macro = \frac{1}{c} \cdot \sum_{i=c}^c F1_i \quad (16)$$

### 3.5.4 INTERVALOS DE CONFIANZA

Cuando se evalúa el rendimiento de un modelo, como la tasa de acierto obtenida en un conjunto de pruebas, el valor numérico resultante es solo una estimación puntual. Este valor está sujeto a la variabilidad inherente de la muestra de datos utilizada. Para trascender esta limitación y cuantificar la fiabilidad de nuestra estimación, se emplea el concepto de Intervalo de Confianza (IC):

$$IC = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot (100-p)}{N}} \quad (17)$$

donde  $p$  es la tasa de acierto del experimento,  $Z(\alpha/2)$  se obtiene a partir de la tabla del acumulado complementario y  $N$  es el número total de ejemplos para un valor de confianza de  $\alpha$ .

### 3.5.5 AGRUPACIÓN DE VENTANAS

Una innovación clave de este trabajo es el procesamiento conjunto de ventanas temporales consecutivas mediante el mecanismo de atención. Las ventanas individuales de 2 segundos se agrupan en conjuntos de ventanas, cuyo número es un hiperparámetro del sistema que se ajusta experimentalmente. Esta agrupación es esencial para dotar al modelo de la capacidad de contextualizar la actividad a lo largo del tiempo, superando las limitaciones de clasificar cada ventana de forma aislada.

### ***Objetivo de la agrupación temporal***

La actividad humana es un proceso dinámico y continuo, en el que actividades como “subir escaleras” o “caminar” no se definen por un instante puntual, sino por una secuencia de movimientos que se desarrollan a lo largo de varios segundos [35]. Al agrupar las ventanas, el modelo puede:

1. Capturar el contexto: Entender la profesión de la actividad, por ejemplo, distinguiendo entre el inicio de una caída y un simple tropiezo.
2. Mitigar el ruido: Las ventanas ruidosas tienen un menor impacto cuando se promedian dentro de un grupo más grande, aumentando así la precisión del modelo.

### ***Proceso de Reestructuración de Datos***

El proceso de agrupación es fundamental para transformar la estructura de los datos de entrada al formato requerido por la red neuronal con la capa de atención.

- Estructura inicial: Los datos de características generados por Octave son una matriz bidimensional donde cada fila representa una ventana de 2 segundos con sus características espectrales.
- Estructura final: Los datos se reestructuran en un tensor de 5 dimensiones con los siguientes hiperparámetros:
  - ❖ *num\_windows*: El hiperparámetro clave que define la longitud de la secuencia temporal (el número de ventanas consecutivas)
  - ❖ *num\_channels*: Los seis canales de aceleración (X, Y, Z de cada uno de los dos sensores).
  - ❖ *num\_points*: El número de puntos de datos de características dentro de cada ventana.

### ***Votación por mayoría***

Para que el entrenamiento supervisado sea efectivo, cada grupo de ventanas debe tener una única etiqueta de actividad que sirva como objetivo de clasificación. Esta etiqueta se asigna mediante la técnica de votación mayoritaria [33].

Para cada grupo de ventanas, se examinan las etiquetas de las ventanas individuales que lo componen. La etiqueta que se repite con mayor frecuencia dentro del grupo es seleccionada como la etiqueta final para ese grupo.

Ejemplo de votación de 3 ventanas (*num\_windows*=9):

Ventana 0: *Labels* = [1 1 1 1 1 1 1 1] -> Votada=1  
 Ventana 1: *Labels* = [1 1 1 1 0 0 0 0] -> Votada=1  
 Ventana 2: *Labels* = [1 1 0 0 0 0 2 2] -> Votada=0  
 Ventana 3: *Labels* = [2 2 2 2 1 1 1] -> Votada=2

Tabla 5. Mecanismo de votación por mayoría

## Líneas de experimentación desarrolladas

El desarrollo de este sistema de reconocimiento de actividades humanas ha requerido una exploración exhaustiva de múltiples configuraciones y estrategias metodológicas. Con el objetivo de maximizar el rendimiento del modelo y comprender en profundidad el impacto de cada decisión de diseño, se han seguido tres líneas de experimentación complementarias que abordan diferentes aspectos del sistema, desde el preprocesamiento de las señales hasta las estrategias de clasificación final. Estas líneas de trabajo se describen a continuación de forma sintética, siendo desarrolladas con detalle en el siguiente capítulo donde se presentan los resultados obtenidos y las conclusiones extraídas.

1. **Optimización de la arquitectura con mecanismo de atención:** Esta línea se ha centrado en explorar sistemáticamente el espacio de configuraciones de la red neuronal con mecanismo de atención. El objetivo ha sido identificar la combinación óptima de hiperparámetros que maximicen la capacidad de clasificación del modelo.

Se han variado múltiples hiperparámetros como la tasa de *dropout*, el número de épocas o las dimensiones de las capas densas entre otras. Especial atención se ha prestado al número de ventanas temporales, que determina la extensión del contexto temporal que el modelo considera al clasificar. Se han evaluado configuraciones desde ventanas individuales hasta secuencias extensas de múltiples ventanas consecutivas.

2. **Estrategia de votación multinivel:** Esta línea explora una aproximación alternativa basada en votación por mayoría en dos niveles del sistema, buscando mejorar la robustez sin modificar la arquitectura de la red.

El primer nivel asigna a cada grupo de ventanas la etiqueta más frecuente entre las ventanas individuales que lo componen, garantizando que la etiqueta de entrenamiento sea representativa. El segundo nivel, la aportación novedosa de esta línea, opera sobre las predicciones finales del modelo entrenado. Las predicciones individuales se agrupan en secuencia de diferentes longitudes (3, 5, 7 y 9 ventanas consecutivas), aplicando la votación por mayoría para obtener una clasificación más estable.

3. **Exploración de parámetros de segmentación temporal:** Esta línea retrocede hasta la fase más temprana del pipeline: la segmentación de señales inerciales en ventanas temporales. Mientras las líneas anteriores mantienen fija la segmentación inicial (2 segundos con 50% de solapamiento), esta explora cómo modificar estos parámetros afecta al rendimiento global.

La duración de ventanas determina la resolución temporal del análisis. Ventanas cortas capturan detalles finos, pero carecen de contexto. Ventanas largas proporcionan contexto, pero pueden incluir transiciones o perder detalles de la dinámica. El paso entre ventanas también es crítico: mayor solapamiento genera más muestras y representación continua, pero aumenta redundancia y coste computacional.

Se ha modificado el código de Octave para generar conjuntos de datos con diferentes configuraciones de segmentación, explorando duraciones superiores a 2 segundos y diversos grados de solapamiento. Para cada configuración ha sido necesario regenerar todos los ficheros de características y reentrenar los modelos. El objetivo ha sido

comprender cómo la granularidad temporal inicial interactúa con las capacidades de aprendizaje de la red neuronal y con las otras dos líneas de experimentación.

## 4. EXPERIMENTOS

En este apartado se detallarán todas las pruebas realizadas durante el Trabajo de Fin de Grado. La estructura experimental se ha organizado en tres líneas de investigación complementarias que abordan diferentes aspectos del sistema: desde el ajuste de hiperparámetros hasta la exploración de estrategias de votación multinivel. Cada experimento ha sido diseñado meticulosamente siguiendo una metodología rigurosa que garantiza la reproducibilidad y permite extraer conclusiones fundamentales sobre el comportamiento del modelo. Además, todos los resultados se han obtenido utilizando la señal completa procedente de los 6 ejes, es decir, empleando conjuntamente los datos del sensor 1 y del sensor 2.

Para el desarrollo del sistema, se utilizan los resultados obtenidos mediante validación cruzada *K-Fold*, donde los 16 *folds* corresponden a sujetos independientes. Este esquema de evaluación también se denomina *Leave-One-Subject-Out (LOSO)*, ya que en cada iteración se deja fuera a un sujeto completo para utilizarlo como conjunto de validación, mientras que los restantes se emplean para el entrenamiento. Así, el proceso se repite 16 veces, garantizando que cada sujeto sea utilizado una vez como conjunto de validación.

Este método permite obtener un rendimiento medio basado únicamente en los datos de entrenamiento, proporcionando estimaciones más robustas y representativas del desempeño general del modelo.

Una vez ajustado el mejor modelo, este se evalúa sobre el conjunto de pruebas para comprobar su comportamiento real ante información no utilizada durante el proceso de diseño. Después de esta fase de implementación y análisis, se seleccionará el experimento que presente el rendimiento más adecuado. Las tasas de acierto se presentan en tanto por uno, donde 1 representa el valor máximo posible. Además, el intervalo de confianza también se presenta en formato decimal, manteniendo coherencia con la tasa de acierto empleada.

### 4.1 ADAPTACIÓN A LA BASE DE DATOS HAR70+ Y OPTIMIZACIÓN INICIAL DEL SISTEMA

El punto de partida de este Trabajo de Fin de Grado fue un sistema existente desarrollado previamente para la base de datos UCI HAR [10], que contenía datos de acelerómetros de personas jóvenes realizando actividades físicas. La primera fase experimental consistió en adaptar completamente este sistema a las características específicas de la base de datos HAR70+ [9], que presenta desafíos únicos debido a la edad avanzada de los participantes y los patrones de movimiento diferenciados que esto implica.

La base de datos HAR 70+ contiene información de 18 usuarios en total, de los cuales los primeros 16 se utilizan para el proceso de validación cruzada, mientras que los 2 últimos se reservan como conjunto de prueba final completamente independiente.

#### 4.1.1 OPTIMIZACIÓN DEL NÚMERO DE ÉPOCAS

El número de épocas determina cuántas veces el modelo recorre completamente el conjunto de entrenamiento durante el proceso de aprendizaje. Un número insuficiente de épocas puede resultar en un modelo subentrenado (*underfitting*), mientras que un exceso puede conducir a sobreajuste (*overfitting*). Se evaluaron configuraciones con  $epochs \in \{5, 10, 15, 20, 25, 40, 60, 80\}$ , manteniendo una configuración inicial de  $num\_windows = 1$ ,  $batch\_size = 32$ ,

$dropout = 0,3$ ,  $learning\_rate = 0,0001$ ,  $kernel\_num = 3$ ,  $num\_channels = 6$ , y la red representada en el capítulo anterior (Figura 13). Estos hiperparámetros se irán configurando posteriormente con el paso de las pruebas pero estos serían los iniciales.

Configuración	Tasa de acierto obtenida en el K-Fold (16 Folds)	Intervalo_Confianza_95%
5 epochs	0,8192	[0,8156 - 0,8228]
<b>10 epochs</b>	<b>0,8249</b>	<b>[0,8214 - 0,8284]</b>
15 epochs	0,7750	[0,7716 - 0,7788]
20 epochs	0,7579	[0,7543 - 0,7615]
25 epochs	0,7509	[0,7473 - 0,7545]
40 epochs	0,7057	[0,7021 - 0,7093]
60 epochs	0,7164	[0,7128 - 0,7200]
80 epochs	0,6900	[0,6864 - 0,6936]

Tabla 6. Resultados variando el número de épocas

### Tasa de acierto obtenida en el K-Fold (16 Folds)

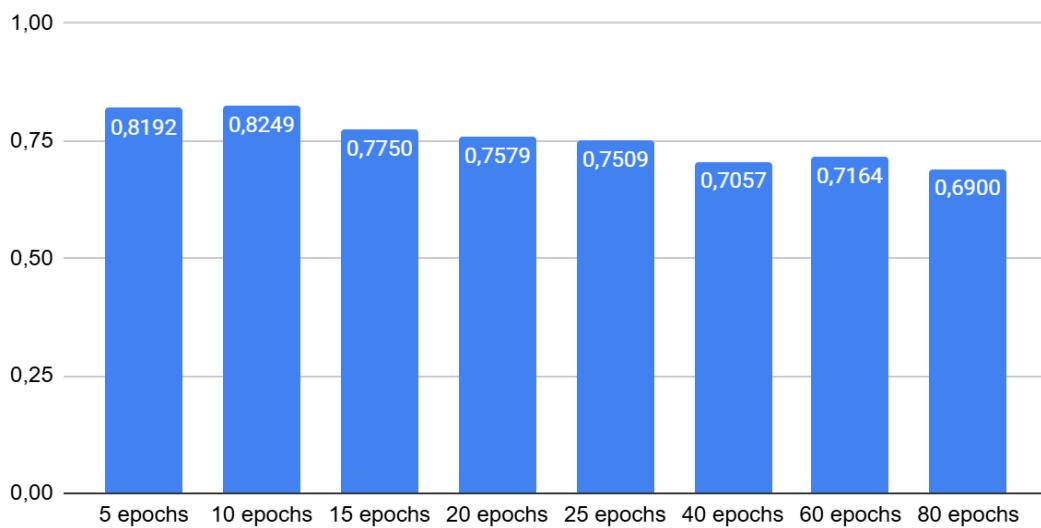


Figura 16. Gráfica de resultados variando el número de épocas

El mejor rendimiento se obtiene con 10 épocas, alcanzando el valor más alto promedio entre los 16 folds. A partir de 15 épocas la tasa de acierto comienza a disminuir debido al sobreajuste.

#### 4.1.2 OPTIMIZACIÓN DEL *DROPOUT*

La técnica del *dropout* es un mecanismo de regularización que desconecta aleatoriamente un porcentaje de neuronas durante el entrenamiento, forzando al modelo a aprender representaciones más robustas. La tasa de *dropout* controla qué fracción de neuronas se desactiva en cada paso. Se exploraron valores de *dropout*  $\in \{0,2 - 0,3 - 0,4 - 0,5 - 0,6 - 0,7\}$ , manteniendo los valores iniciales y 10 épocas.

Configuración (10 epochs)	Tasa de acierto obtenida en el K-Fold (16 Folds)	Intervalo_Confianza_95%
0,1 dropout	0,8393	[0,8357 - 0,8429]
<b>0,2 dropout</b>	<b>0,8337</b>	<b>[0,8302 - 0,8372]</b>
0,3 dropout	0,8249	[0,8214 - 0,8284]
0,4 dropout	0,7958	[0,7922 - 0,7994]
0,5 dropout	0,8010	[0,7974 - 0,8046]
0,6 dropout	0,8011	[0,7975 - 0,8047]
0,7 dropout	0,7960	[0,7924 - 0,7996]

Tabla 7. Resultados variando el *dropout* con 10 épocas

Tasa de acierto obtenida en el K-Fold (16 Folds)

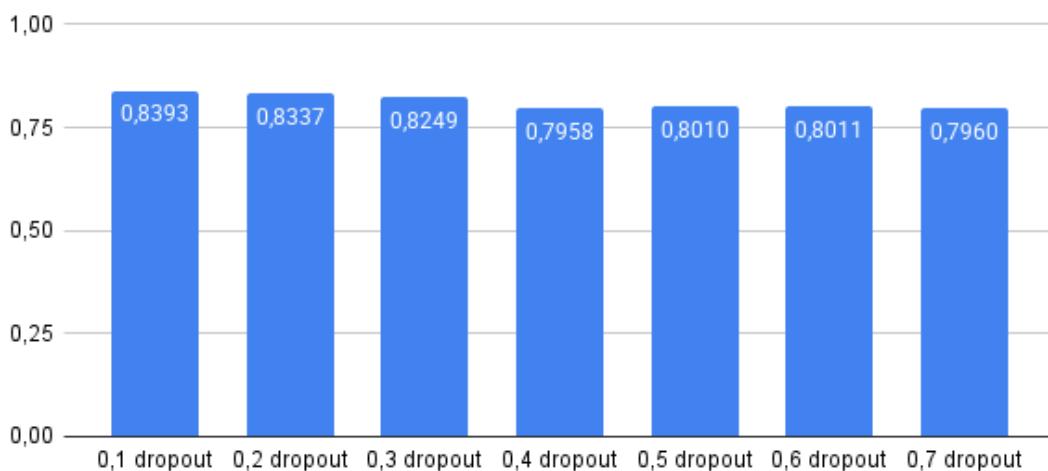


Figura 17. Gráfica de resultados variando el *dropout* con 10 épocas

Aunque un *dropout* de 0,1 alcanza la tasa de acierto más alta (0,8393), la diferencia respecto a un *dropout* de 0,2 (0,8337) es marginal. Además, un valor tan bajo de *dropout* puede resultar insuficiente como mecanismo de regularización, incrementando el riesgo de sobreajuste en etapas posteriores del entrenamiento. El valor de 0,2 ofrece un equilibrio óptimo entre rendimiento y robustez: mantiene una precisión competitiva mientras

proporciona una regularización más efectiva que mejora la capacidad de generalización del modelo.

#### 4.1.3 OPTIMIZACIÓN DEL *LEARNING RATE*

La tasa de aprendizaje controla la magnitud de las actualizaciones que se aplican a los pesos de la red neuronal en cada iteración del algoritmo de optimización. Un valor demasiado alto puede causar inestabilidad y divergencia, mientras que un valor relativamente bajo, ralentiza el aprendizaje. Se evaluaron valores de  $learning rate \in \{0,00005 - 0,0001 - 0,0003 - 0,0005 - 0,001\}$ , utilizando el optimizador *RMSprop*.

Configuración (10 epochs y 0,2 dropout)	Tasa de acierto obtenida en el K-Fold (16 Folds)	Intervalo_Confianza _95%
0,00005 learning rate	0,8525	[0,8490 - 0,8560]
<b>0,0001 learning rate</b>	<b>0,8577</b>	<b>[0,8542 - 0,8612]</b>
0,0003 learning rate	0,8371	[0,8336 - 0,8406]
0,0005 learning rate	0,8337	[0,8302 - 0,8372]
0,001 learning rate	0,7686	[0,7650 - 0,7722]

Tabla 8. Resultados variando el *learning rate* con 10 épocas y 0,2 *dropout*

Tasa de acierto obtenida en el K-Fold (16 Folds)

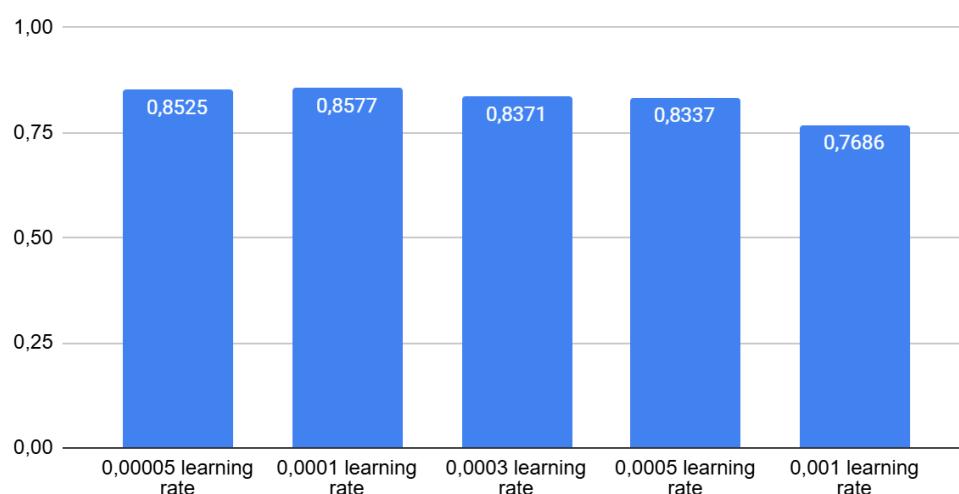


Figura 18. Gráfica de resultados variando el *learning rate* con 10 épocas y 0,2 *dropout*

La mejor combinación se obtiene con un *learning rate* de 0,0001 que proporciona la máxima precisión en un entrenamiento y un rendimiento estable en el conjunto de prueba. Con valores menores, como 0,00005, la precisión es ligeramente inferior, mientras que tasas mayores (0,0003 a 0,001) provocan una disminución en entrenamiento o mayor inestabilidad.

#### 4.1.4 OPTIMIZACIÓN DEL *BATCH SIZE*

El tamaño de lote determina cuántos ejemplos se procesan simultáneamente antes de *actualizar* los pesos de la red neuronal. Este hiperparámetro afecta tanto a la estabilidad del entrenamiento como a la eficiencia computacional del proceso. Se exploran valores de *batch\_size*  $\in \{16, 32, 64, 128, 256\}$ .

Configuración (10 epochs, 0,2 dropout y 0,0001 learning rate)	Tasa de acierto obtenida en el K-Fold (16 Folds)	Intervalo_Confianza_95%
16 batch_size	0,8526	[0,8491 - 0,8561]
<b>32 batch_size</b>	<b>0,8577</b>	<b>[0,8542 - 0,8612]</b>
64 batch_size	0,8363	[0,8328 - 0,8398]
128 batch_size	0,8392	[0,8357 - 0,8427]
256 batch_size	0,8331	[0,8296 - 0,8366]

Tabla 9. Resultados variando el *batch size* con 10 épocas, 0,2 *dropout* y 0,0001 *learning rate*

Tasa de acierto obtenida en el K-Fold (16 Folds)

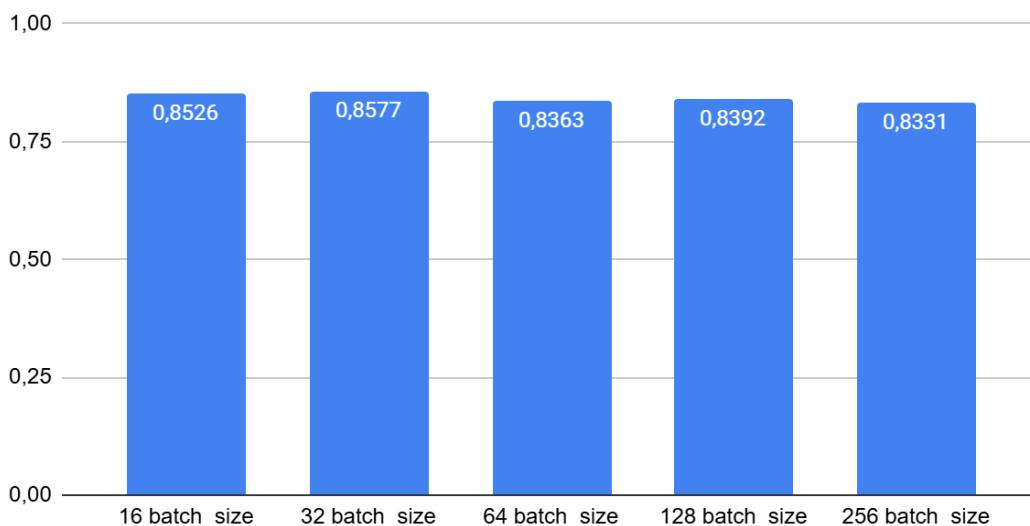


Figura 19. Gráfica de resultados variando el *batch size* con 10 épocas, 0,2 *dropout* y 0,0001 *learning rate*

La mejor precisión de entrenamiento se obtiene con un *batch size* de 32, por lo que se selecciona como mejor configuración. Valores de *batch\_size* de 16 muestran una precisión ligeramente inferior mientras que tamaños mayores provocan una caída en entrenamiento.

#### 4.1.5 EXPLORACIÓN DE ARQUITECTURAS DE RED NEURONAL

La arquitectura de la red neuronal determina su capacidad para aprender representaciones complejas de datos. Se exploraron cuatro variantes, todas basadas en la estructura fundamental de capas *TimeDistributed* con convoluciones seguidas por el mecanismo de atención y capas densas de clasificación. Los experimentos se realizaron con la configuración óptima encontrada como es la de  $epochs = 10$ ,  $dropout = 0,2$ ,  $learning\_rate = 0,0001$  y  $batch\_size = 32$ .

La configuración final alcanzada ( $10$  épocas,  $0,2$  dropout,  $0,0001$  learning rate y  $32$  batch size) establece un punto de partida robusto que garantiza tanto la reproducibilidad de los experimentos como la capacidad del sistema para generalizar a nuevos usuarios. Con esta base sólida establecida, las siguientes fases experimentales podrán centrarse en estrategias de refinamiento más avanzadas sin necesidad de revisar los fundamentos hiperparámetros del sistema.

Configuración (10 epochs, 0,2 dropout, 0,0001 learning rate y 32 batch size)	Tasa de acierto obtenida en el K-Fold (16 Folds)	Intervalo_Confianza_95%
TimeDistributed (Conv2D)-> MaxPooling -> TimeDistributed (Conv2D)-> TimeDistributed (Conv2D)-> Flatten -> Attention -> Dense(64) -> Dense(32) -> Dense(16) -> Softmax	0,8339	[0,8304 - 0,8374]
TimeDistributed(Conv2D) -> MaxPooling2D -> TimeDistributed(Conv2D) -> Flatten -> Attention -> Dense(64) -> Dense(32) -> Softmax	0,8075	[0,8040 - 0,8110]
<b>TimeDistributed (Conv2D) -&gt; MaxPooling2D -&gt; Flatten -&gt; Attention -&gt; Dense(64) -&gt; Dense(32) -&gt; Softmax (Figura 13)</b>	<b>0,8577</b>	<b>[0,8542 - 0,8612]</b>
TimeDistributed(Conv2D) -> MaxPooling2D -> Flatten -> Attention -> Dense(64) -> Softmax	0,8541	[0,8506 - 0,8576]

Tabla 10. Resultados variando el *batch size* con  $10$  épocas,  $0,2$  dropout,  $0,0001$  learning rate y  $32$  batch size

Tasa de acierto obtenida en el K-Fold (16 Folds)

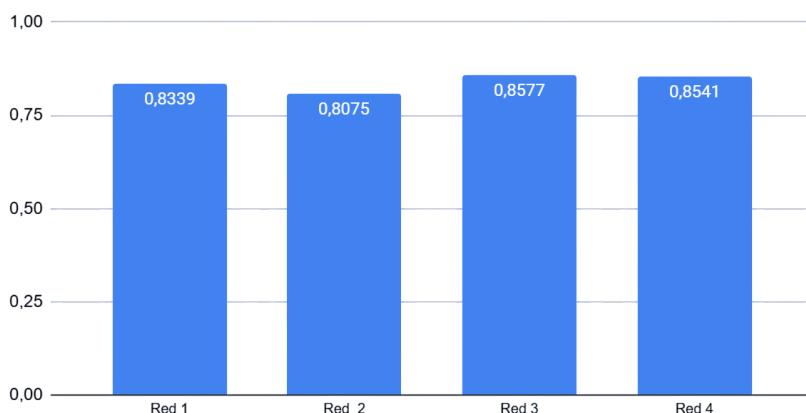


Figura 20. Gráfica de resultados variando el *batch size* con  $10$  épocas,  $0,2$  dropout,  $0,0001$  learning rate y  $32$  batch size

## 4.2 EVALUACIÓN DE LAS DIFERENTES ESTRATEGIAS DE AGRUPACIÓN DE VENTANAS

Una vez establecida la configuración óptima de hiperparámetros (*10 épocas, 0,2 dropout, 0,0001 de learning rate y 32 de batch size*), se procedió a investigar el comportamiento del sistema al incorporar información contextual mediante el procesamiento conjunto de múltiples ventanas temporales consecutivas.

### 4.2.1 EXPERIMENTO I: OPTIMIZACIÓN DE LA ARQUITECTURA CON MECANISMO DE ATENCIÓN

Esta línea experimental explora cómo el mecanismo de atención puede aprovechar secuencias temporales de diferentes longitudes para mejorar la clasificación de actividades en población mayor.

#### 4.2.1.1 FUNDAMENTACIÓN TEÓRICA DE LA EXPERIMENTACIÓN (I)

La hipótesis subyacente en esta línea de experimentación es que las actividades humanas no se definen por instantes aislados, sino por patrones de movimiento que se desarrollan a lo largo del tiempo. Una ventana individual de 2 segundos puede capturar un fragmento de la actividad, pero carece del contexto necesario para interpretar correctamente ciertos movimientos ambiguos. Por ejemplo, distinguir entre “caminar” y “arrastrar los pies” puede requerir observar varios ciclos de marcha consecutivos para identificar diferencias sutiles en la cadencia y amplitud del movimiento.

El mecanismo de atención permite al modelo identificar automáticamente qué instantes temporales dentro de una secuencia son más relevantes para la clasificación. Esta capacidad resulta especialmente valiosa en población mayor, donde la variabilidad del movimiento es mayor debido a factores como movilidad reducida o estrategias compensatorias individuales. A diferencia de métodos simples de agregación temporal, la atención introduce un mecanismo adaptativo que ajusta dinámicamente su comportamiento según las características de cada secuencia.

Se exploraron cinco configuraciones distintas de ventanas  $\in \{1, 3, 5, 7, 9\}$ . La selección de estos valores responde a un diseño experimental sistemático:  $num\_windows = 1$  establece la línea base sin integración temporal, mientras que los valores impares superiores permiten mantener una ventana central con contexto previo y posterior. El valor máximo de 9 ventanas representa un horizonte temporal de 18 segundos, suficiente para capturar secuencias completas de la mayoría de las actividades del protocolo HAR70+.

#### 4.2.1.2 RESULTADOS EXPERIMENTALES (I)

Los experimentos se realizaron manteniendo todos los hiperparámetros en sus valores óptimos previamente identificados, variando únicamente el número de ventanas procesadas conjuntamente. Los resultados obtenidos mediante validación cruzada sobre los primeros 16 participantes revelaron una tendencia clara hacia la mejora del rendimiento conforme aumenta el contexto temporal.

Ventanas	Tasa_Acierto	Margen_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	0,8449	0,0035	40172	[0,8414 - 0,8484]
3	0,8653	0,0058	13390	[0,8595 - 0,8711]
5	0,8677	0,0074	8033	[0,8603 - 0,8751]
7	0,8876	0,0082	5738	[0,8794 - 0,8958]
9	0,8909	0,0091	4463	[0,8818 - 0,9000]

Tabla 11. Tabla de resultados de la experimentación I con diferentes números de ventanas temporales

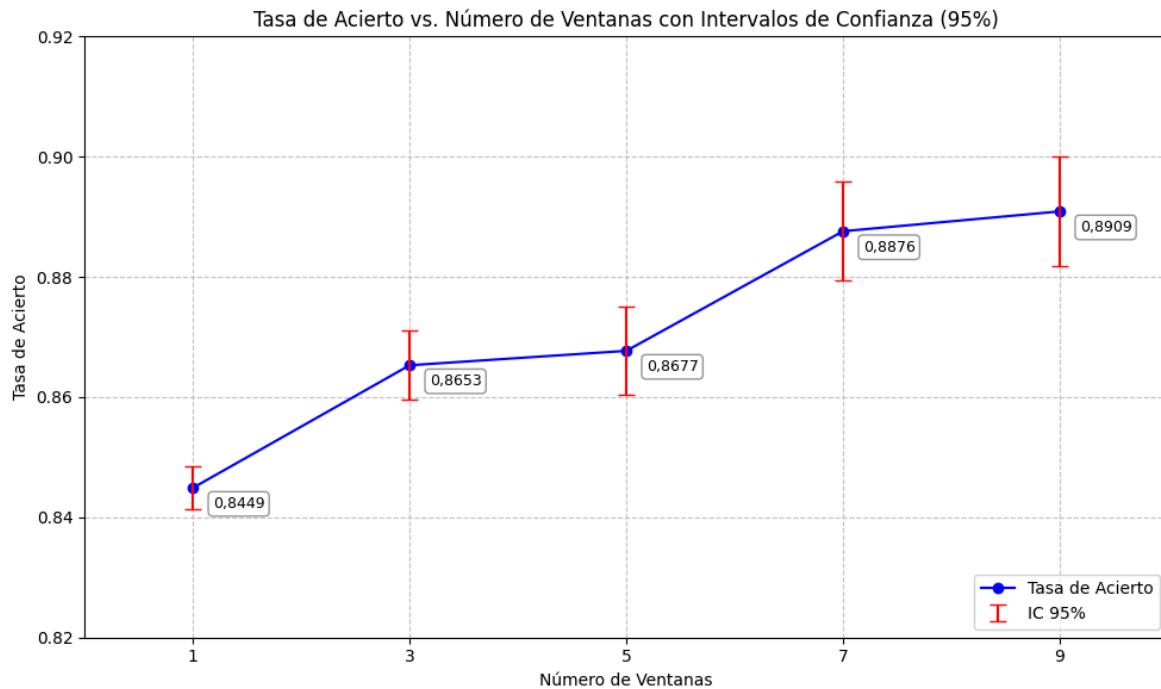


Figura 21. Gráfica de resultados de la experimentación I con diferentes números de ventanas temporales

Los resultados obtenidos en este primer experimento confirman que la integración del contexto temporal mediante el mecanismo de atención constituye una estrategia efectiva que mejora sustancialmente el rendimiento del sistema de clasificación. La configuración con 9 ventanas consecutivas emerge como la óptima, logrando una tasa de acierto de 0,8909 en validación cruzada. El incremento progresivo y consistente del rendimiento al aumentar el número de ventanas temporales procesadas conjuntamente indica que contextos temporales más amplios proporcionan información discriminativa adicional que el mecanismo de atención logra aprovechar eficazmente. El mecanismo de atención se consolida, así como un componente fundamental que permite al modelo identificar y ponderar automáticamente los instantes que permite al modelo identificar y ponderar automáticamente los instantes más informativos dentro de cada secuencia, estableciendo un nuevo estándar de rendimiento que será utilizado como referencia en las evaluaciones posteriores sobre el conjunto de test independiente

## 4.2.2 EXPERIMENTO II: ESTRATEGIA DE VOTACIÓN MULTINIVEL

Una vez demostrada la eficacia del mecanismo de atención para integrar información de múltiples ventanas temporales consecutivas, se exploró una estrategia complementaria basada en votación por mayoría aplicada en dos niveles del sistema. Esta segunda línea experimental busca mejorar la robustez de las clasificaciones sin modificar la arquitectura de la red neuronal, aprovechando la redundancia temporal inherente a las secuencias de predicciones para filtrar decisiones erróneas puntuales y estabilizar los resultados finales.

### 4.2.2.1 FUNDAMENTACIÓN TEÓRICA DE LA EXPERIMENTACIÓN (II)

La estrategia de votación multinivel se fundamenta en el principio de que las decisiones colectivas tomadas sobre múltiples observaciones consecutivas tienden a ser más robustas que las decisiones individuales aisladas. Esta aproximación reconoce que, aunque el modelo haya sido entrenado con información contextual mediante el mecanismo de atención, las predicciones individuales pueden verse afectadas por artefactos localizados, transiciones entre actividades, o patrones atípicos que no representan fielmente la actividad predominante en un periodo temporal más amplio.

#### ***Dos niveles de votación: entrenamiento y predicción***

El sistema implementado incorpora votación por mayoría en dos momentos críticos del pipeline de procesamiento, cada uno con objetivos y características específicas:

#### ***Primer nivel: Votación durante el preprocesamiento y entrenamiento***

Este nivel opera sobre las etiquetas originales antes de que el modelo sea entrenado. Como se describió en el capítulo anterior, cuando se agrupan múltiples ventanas consecutivas para formar una secuencia de entrada a modelo, es necesario asignar una única etiqueta de actividad a todo el grupo. Esta etiqueta se determina mediante votación por mayoría: se examinan las etiquetas individuales de todas las ventanas que componen el grupo, y la etiqueta más frecuente se selecciona como representativa del conjunto completo.

Este mecanismo cumple una función de regularización de las etiquetas de entrenamiento asegurando que el modelo aprenda a clasificar secuencias coherentes en lugar de ventanas potencialmente ruidosas o transitorias. Por ejemplo, si un grupo de 5 ventanas contiene las etiquetas [caminar, caminar, estar de pie, caminar, caminar], la votación seleccionará caminar como etiqueta del grupo completo, interpretando que la ventana etiquetada como “estar de pie” probablemente corresponda a un artefacto de anotación o a un instante transitorio muy breve que no representa la actividad dominante en ese periodo temporal.

#### ***Segundo nivel: Votación post-predicción***

Este segundo nivel constituye la contribución principal de este experimento y opera sobre las predicciones ya generadas por el modelo entrenado. Una vez que el sistema ha clasificado todas las ventanas del conjunto de evaluación, estas predicciones individuales se agrupan nuevamente en secuencias consecutivas de longitud configurable, y se aplica votación por mayoría sobre cada secuencia para obtener una clasificación final estabilizada.

La diferencia fundamental con el primer nivel es que aquí no estamos procesando etiquetas verdaderas sino predicciones del modelo, y el objetivo no es regularizar el entrenamiento sino refinar los resultados finales. Esta estrategia permite suavizar fluctuaciones temporales en las predicciones, filtrar clasificaciones erróneas puntuales, y garantizar mayor coherencia temporal en la secuencia de actividades reconocidas.

#### 4.2.2.2 RESULTADOS EXPERIMENTALES (II)

Los experimentos se realizaron manteniendo todos los hiperparámetros en sus valores óptimos previamente identificados, variando únicamente el número de ventanas procesadas conjuntamente. Los resultados obtenidos mediante validación cruzada sobre los primeros 16 participantes revelaron una tendencia clara hacia la mejora del rendimiento conforme aumenta el contexto temporal.

Ventanas	Tasa_Acierto	Margen_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	0,8477	0,0035	40172	[0,8442 - 0,8512]
3	0,8532	0,0060	13390	[0,8473 - 0,8592]
5	0,8549	0,0077	8034	[0,8472 - 0,8626]
7	0,8581	0,0090	5738	[0,8491 - 0,8672]
9	0,8622	0,0101	4463	[0,8521 - 0,8723]

Tabla 12. Tabla de resultados de la experimentación II con diferentes números de ventanas temporales

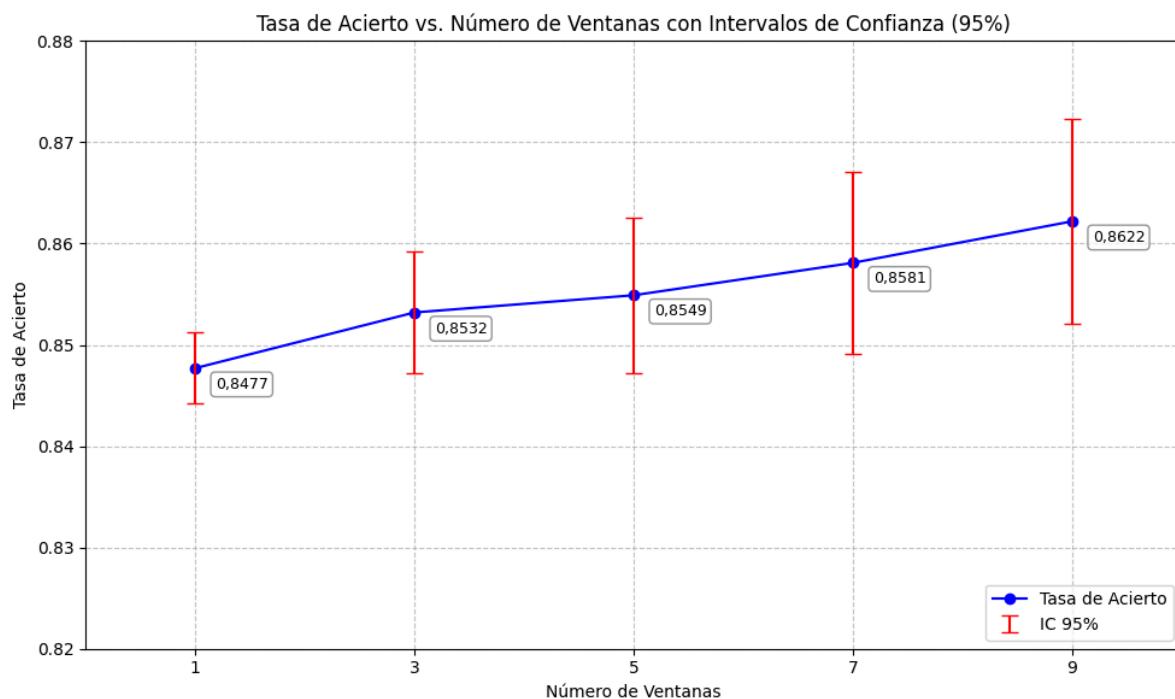


Figura 22. Gráfica de resultados de la experimentación II con diferentes números de ventanas temporales

Los resultados obtenidos en este segundo experimento confirman que la estrategia de votación multinivel constituye una técnica efectiva de post-procesamiento que mejora la robustez del sistema de clasificación. La configuración con 9 ventanas de votación post-predicción emerge como la óptima, logrando una tasa de acierto del 0,8622 en validación cruzada. El incremento progresivo del rendimiento al aumentar la longitud de las secuencias de votación indica qué contextos temporales más amplios proporcionan mayor capacidad de corrección de inconsistencias, aunque con rendimiento marginales decrecientes. La estrategia de votación multinivel se consolida, así como un mecanismo complementario valioso que, sin requerir modificaciones del modelo, aporta mejoras sustanciales en la precisión final del sistema de reconocimiento de actividades.

#### 4.2.3 EXPERIMENTO III: EXPLORACIÓN DE PARÁMETROS DE SEGMENTACIÓN TEMPORAL

Una vez exploradas las estrategias de integración contextual mediante el mecanismo de atención y la votación multinivel, se investigó el impacto de los parámetros fundamentales de segmentación temporal sobre el rendimiento global del sistema. Esta tercera línea experimental retrocede hasta la fase más temprana del pipeline de procesamiento para analizar como la duración de las ventanas temporales y el grado de solapamiento entre ellas afectan a la capacidad del modelo para aprender y clasificar actividades humanas en población mayor.

##### 4.2.3.1 FUNDAMENTACIÓN TEÓRICA DE LA EXPERIMENTACIÓN (III)

La segmentación temporal de las señales iniciales constituye una decisión metodológica que precede y condiciona todo el proceso posterior de análisis. Mientras que los experimentos anteriores han explorado cómo procesar y combinar las ventanas una vez definidas, esta línea experimental cuestiona la decisión fundamental sobre como particionar inicialmente las series temporales capturadas por los acelerómetros.

Una ventana temporal de 2 segundos, la configuración actual utilizada hasta ahora captura aproximadamente 100 muestras de aceleración (a 50 Hz). Este horizonte temporal puede ser suficiente para caracterizar eventos puntuales o fases específicas de una actividad, pero podría resultar insuficiente para actividades que requieren observar ciclos completos de movimiento. Por ejemplo, el patrón de “caminar” en personas mayores puede presentar una cadencia más lenta que requiera varios segundos para completar múltiples pasos y revelar el ritmo característico de la marcha.

En contraste, ventanas más largas proporcionan una visión más holística de la actividad. Una ventana de 18 segundos captura aproximadamente 900 muestras, suficientes para observar múltiples ciclos completos de la mayoría de las actividades del protocolo HAR70+.

##### *Metodología experimental*

Para explorar estas cuestiones, se diseñó un experimento que mantiene constante el solapamiento relativo (50%) mientras incrementa sistemáticamente la duración de las ventanas. Se evaluaron cinco configuraciones:

- ***window\_t = 2 segundos, step\_t = 1 segundo***: Configuración de referencia utilizada en los experimentos previos, equivalente a procesar 1 ventana individual.
- ***window\_t = 6 segundos, step\_t = 3 segundos***: Triplicación de la duración temporal, equivalente al contexto de 3 ventanas de la configuración original.
- ***window\_t = 10 segundos, step\_t = 5 segundos***: Ventanas de 10 segundos equivalente al contexto de 5 ventanas originales.
- ***window\_t = 14, step\_t = 7 segundos***: Ventanas de 14 segundos, equivalente al contexto de 7 ventanas originales.
- ***window\_t = 18 segundos, step\_t = 9 segundos***: Máxima duración explorada, equivalente al contexto de 9 ventanas originales.

Para cada configuración de segmentación temporal fue necesario regenerar completamente los ficheros de características mediante el código de Octave, aplicando la Transformada Q Constante sobre las nuevas ventanas. Posteriormente, estos datos fueron utilizados para entrenar modelos con la arquitectura óptima identificada en fases previas, configurando *num\_windows* = 1 dado que cada ventana ya contiene el contexto temporal deseado.

#### 4.2.3.2 RESULTADOS EXPERIMENTALES (III)

Los experimentos se realizaron siguiendo la misma metodología de validación cruzada sobre los 16 primeros participantes del conjunto de datos HAR70+. Los resultados obtenidos revelan una tendencia consistente hacia la mejora del rendimiento conforme aumenta la duración de las ventanas temporales.

Ventanas	Window_t	Tasa_Acierto	Margen_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	2 (1 ventana)	0,8489	0,0035	40172	[0,8454 - 0,8524]
1	6 (3 ventanas)	0,8545	0,0060	13301	[0,8485 - 0,8605]
1	10 (5 ventanas)	0,8651	0,0075	7947	[0,8933 - 0,8993]
1	14 (7 ventanas)	0,8710	0,0087	5657	[0,8622 - 0,8797]
1	18 (9 ventanas)	0,8888	0,0093	4353	[0,8795 - 0,8981]

Tabla 13. Tabla de resultados de la experimentación III con diferentes números de ventanas temporales

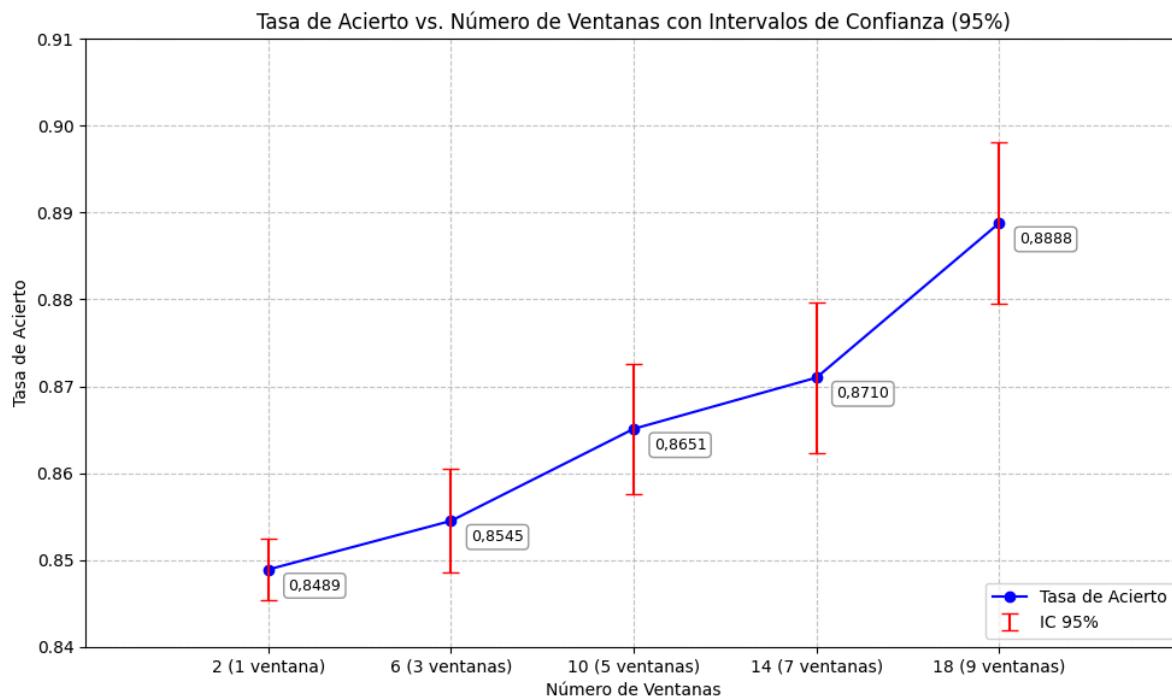


Figura 23. Gráfica de resultados de la experimentación III con diferentes números de ventanas temporales

Los resultados obtenidos en este tercer experimento demuestran que la duración de las ventanas temporales constituye un parámetro crítico del sistema cuya optimización puede proporcionar mejoras sustanciales en el rendimiento sin incrementar la complejidad de la red. La configuración con 18 ventanas emerge como la óptima, logrando una tasa de acierto de 0,8888 en validación cruzada, representando una mejora respecto a la configuración base de 2 segundos.

#### 4.2.3.3 ANÁLISIS COMPARATIVO DE LAS 3 ESTRATEGIAS DE AGRUPACIÓN DE VENTANAS

La comparación de las tres estrategias revela que todas comparten una tendencia común de mejora progresiva al incrementar el contexto temporal, aunque difieren significativamente en su magnitud y eficiencia. El mecanismo de atención (Experimento I) destaca por su capacidad superior de aprendizaje adaptativo, mostrando la curva más ascendente y alcanzando los mejores resultados. La estrategia de votación multinivel (Experimento II) ofrece una alternativa más conservadora pero práctica, proporcionando mejoras incrementales mediante post-procesamiento sin aumentar la complejidad del modelo. La modificación de los parámetros de segmentación temporal (Experimento III) se posiciona en un término medio, demostrando que decisiones tomadas en las primeras fases del pipeline tienen impacto considerable en el rendimiento final. Esta diversidad de aproximaciones evidencia que no existe una única solución óptima, sino que la elección de la estrategia debe equilibrar factores como la precisión, los recursos computacionales y la facilidad de implementación según las necesidades específicas de cada aplicación práctica.

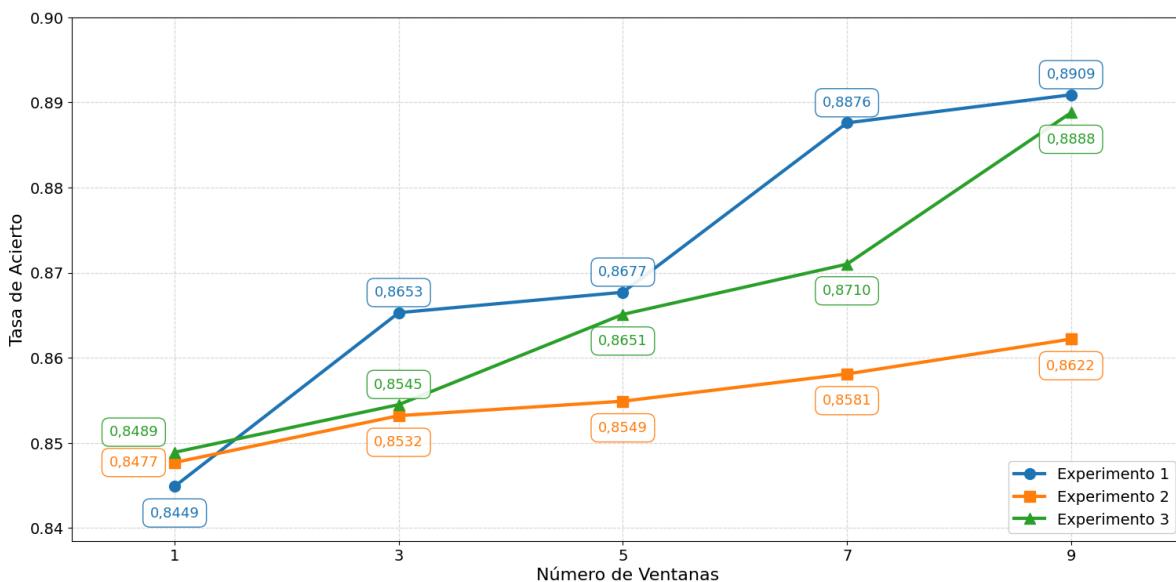


Figura 24. Comparación de tasas de acierto entre los tres experimentos con diferentes números de ventanas temporales

#### 4.2.4 EXPERIMENTO IV: ARQUITECTURAS PROFUNDAS CON VENTANA DESLIZANTE

Tras explorar diferentes estrategias de integración temporal y parámetros de segmentación, se investigó el impacto de arquitecturas neuronales más profundas sobre el rendimiento del sistema. Esta cuarta línea experimental mantiene la configuración óptima de segmentación temporal ( $window\_t = 2$  segundos,  $step\_t = 1$  segundo) identificada en fases previas, pero introduce modificaciones sustanciales en la estructura de la red neuronal mediante la incorporación de capas convolucionales adicionales que permiten al modelo aprender representaciones jerárquicas más complejas de los patrones de movimiento.

La motivación de este experimento surge de la observación de que las actividades humanas en población mayor presentan sutilezas y variabilidades que podrían requerir transformaciones más sofisticadas para ser adecuadamente capturadas.

##### 4.2.4.1 FUNDAMENTACIÓN TEÓRICA DE LA EXPERIMENTACIÓN (IV)

La profundidad de una red neuronal determina su capacidad para construir representaciones abstractas de los datos mediante la composición jerárquica de transformaciones no lineales. Mientras que las capas superficiales tienden a capturar patrones locales y de bajo nivel en las señales inerciales, las capas profundas pueden integrar esa información primitiva para identificar estructuras temporales complejas que caracterizan actividades humanas completas.

La hipótesis fundamental de este experimento es que el aumento controlado de la profundidad de la red, mediante la adición de capas convolucionales con regularización apropiada, puede mejorar sustancialmente la capacidad del modelo para discriminar entre actividades de movimiento similares. La motivación de aumentar la complejidad de los modelos es que al agrupar varias ventanas tenemos una mayor cantidad de información por ventana y queríamos saber si podríamos entrenar modelos más complejos.

Se diseñaron dos arquitecturas progresivamente más profundas manteniendo la estructura fundamental del sistema:

1. **Red ampliada 1:** Incorpora tres capas convolucionales adicionales con respecto a la arquitectura base antes del mecanismo de atención. Esta configuración intermedia permite evaluar si un incremento moderado de profundidad aporta beneficios sin introducir complejidad excesiva.
2. **Red ampliada 2:** Extiende la profundidad a cuatro capas convolucionales secuenciales, maximizando la capacidad del modelo para aprender jerarquías de características cada vez más abstractas. Esta arquitectura representa el límite superior explorado en términos de profundidad, equilibrando capacidad representacional y riesgo de sobreajuste.

Ambas arquitecturas mantienen la misma estructura de clasificación posterior al mecanismo de atención, con tres capas densas de 64, 32 y 16 neuronas respectivamente, seguidas de la capa *softmax* de salida. La regularización mediante *dropout* se aplica sistemáticamente después de cada transformación convolucional y densa para controlar el sobreajuste inherente a modelos de mayor capacidad.

#### 4.2.4.2 OPTIMIZACIÓN DEL *DROPOUT* PARA LA RED AMPLIADA 1

Antes de proceder a la evaluación comparativa de las arquitecturas con diferentes números de ventanas temporales, se realizó un estudio preliminar para identificar la tasa de *dropout* óptima específica para la red ampliada. Este paso es fundamental, ya que arquitecturas más profundas típicamente requieren más regularización para compensar su incrementada capacidad de memorización.

Se exploraron valores de  $\text{dropout} \in \{0,2 - 0,3 - 0,4 - 0,5 - 0,6\}$  manteniendo todos los demás hiperparámetros en sus valores óptimos previamente identificados. Los experimentos se hicieron con  $\text{num\_windows} = 1$  para aislar el efecto del *dropout* sin la influencia de la integración temporal.

<b>Dropout</b>	<b>Tasa_Acierto</b>	<b>Margen_Error</b>	<b>Número_de_ejemplos_train</b>	<b>Intervalo_Confianza_95%</b>
0,2	0,8672	0,0033	40164	[0,8639 - 0,8705]
0,3	0,8817	0,0032	40164	[0,8785 - 0,8848]
0,4	0,9002	0,0029	40164	[0,8973 - 0,9032]
0,5	0,9217	0,0026	40164	[0,9191 - 0,9243]
0,6	0,9131	0,0028	40164	[0,9103 - 0,9158]

Tabla 14. Tabla de resultados de la experimentación IV con diferentes números de *dropout*

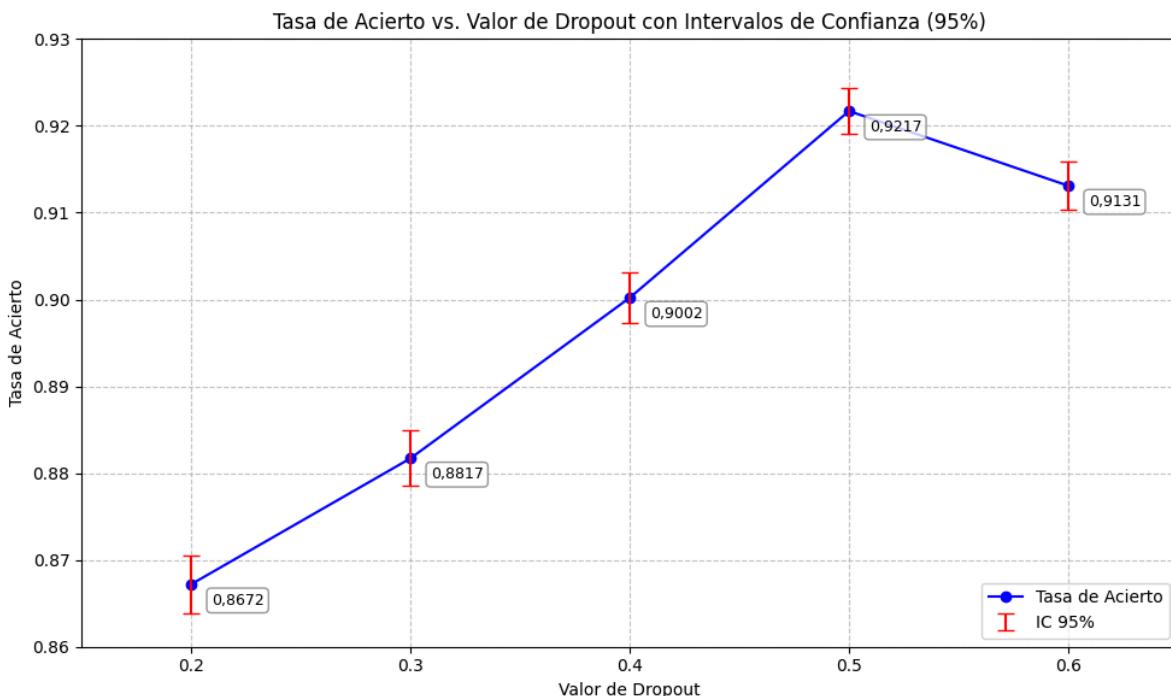


Figura 25. Gráfica de resultados de la experimentación IV con diferentes números de *dropout*

Los resultados demuestran una tendencia clara y consistente: el rendimiento del modelo mejora progresivamente conforme aumenta la tasa de *dropout* hasta alcanzar un máximo en 0,5, logrando una tasa de acierto de 0,9217.

#### 4.2.4.3 RESULTADOS EXPERIMENTALES (IV)

Una vez identificada la tasa de *dropout* óptima, se procedió a evaluar ambas arquitecturas profundas bajo la misma metodología de integración temporal empleada en el experimento 1. Se exploraron configuraciones con  $num\_windows \in \{1, 3, 5, 7, 9\}$ , manteniendo  $dropout = 0,5$  y el resto de hiperparámetros en sus valores óptimos. Los resultados se presentan de forma comparativa para ambas arquitecturas.

##### *Resultados de la Red Ampliada 1*

Ventanas	Tasa_Acierto	Margen_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	0,9054	0,0029	40172	[0,9026 - 0,9083]
3	0,9111	0,0028	40170	[0,9083 - 0,9139]
5	0,9204	0,0027	40168	[0,9178 - 0,9231]
7	0,9195	0,0027	40166	[0,9168 - 0,9221]
9	0,9086	0,0028	40164	[0,9058 - 0,9114]

Tabla 15. Tabla de resultados de la experimentación IV con diferentes números de ventanas temporales (red ampliada 1)

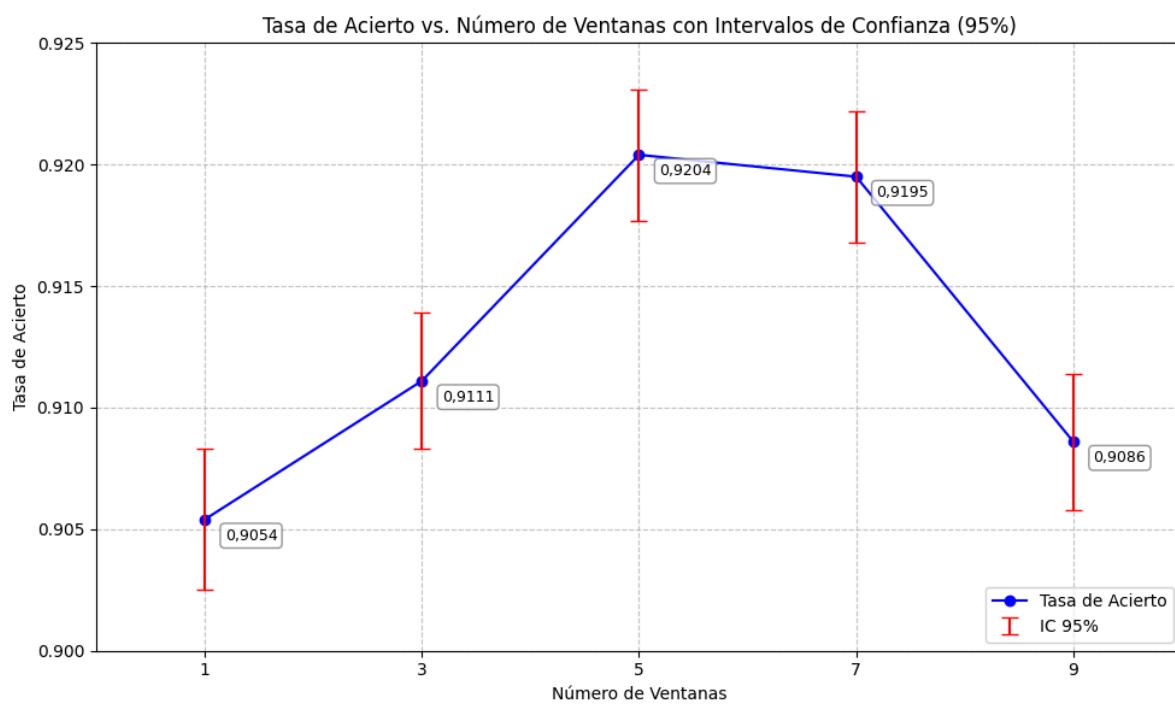


Figura 26. Gráfica de resultados de la experimentación IV con diferentes números de ventanas temporales (red ampliada 1)

Los resultados demuestran una tendencia inicial de mejora progresiva del rendimiento conforme aumenta el contexto temporal, alcanzando su máxima tasa con 5 ventanas. Esta observación motiva la exploración de una arquitectura más profunda que pueda incrementar el rendimiento con mayor contexto temporal. La hipótesis es que una red ampliada 2 con una capa convolucional adicional dispondrá de mayor capacidad para aprender jerarquías de características más complejas, permitiendo incrementar el rendimiento con 7 y 9 ventanas temporales.

### ***Resultados de la Red Ampliada 2***

Ventanas	Tasa_Acierto	Margen_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	0,8987	0,0029	40172	[0,8957 - 0,9016]
3	0,9020	0,0029	40170	[0,8991 - 0,9049]
5	0,9098	0,0028	40168	[0,9070 - 0,9126]
7	0,9172	0,0027	40166	[0,9145 - 0,9199]
9	0,9268	0,0025	40164	[0,9242 - 0,9293]

Tabla 16. Tabla de resultados de la experimentación IV con diferentes números de ventanas temporales (red ampliada 2)

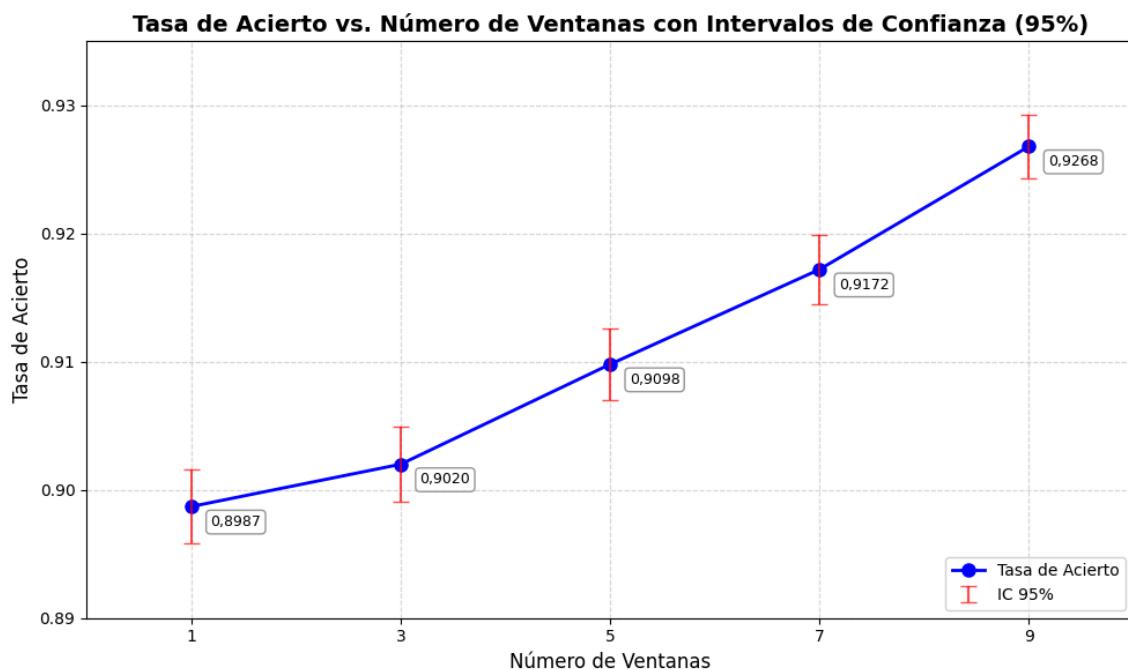


Figura 27. Gráfica de resultados de la experimentación IV con diferentes números de ventanas temporales (red ampliada 2)

Los resultados obtenidos en este cuarto experimento demuestran de forma concluyente que el incremento controlado de la profundidad de la red constituye una estrategia altamente efectiva para mejorar el rendimiento del sistema de reconocimiento de actividades humanas. La red ampliada 2, con cuatro capas convolucionales, establece un nuevo estado del arte alcanzando un acierto de 0,9268 con 9 ventanas temporales consecutivas.

#### 4.2.4.4 ANÁLISIS COMPARATIVO DE LAS ARQUITECTURAS PROFUNDAS

La comparación entre ambas arquitecturas revela comportamientos claramente diferenciados. La Red Ampliada 1 alcanza su máximo rendimiento con 5 ventanas (0,9204) y posteriormente decrece, sugiriendo que su capacidad representacional resulta insuficiente para contextos temporales más extensos.

Por el contrario, la Red Ampliada 2 mantiene una tendencia ascendente sostenida, alcanzando su óptimo con 9 ventanas (0,9268). La cuarta capa convolucional adicional proporciona la capacidad necesaria para que el mecanismo de atención aproveche eficazmente secuencias temporales más largas, estableciendo así la configuración definitiva del sistema con el mejor rendimiento global alcanzado en este trabajo.

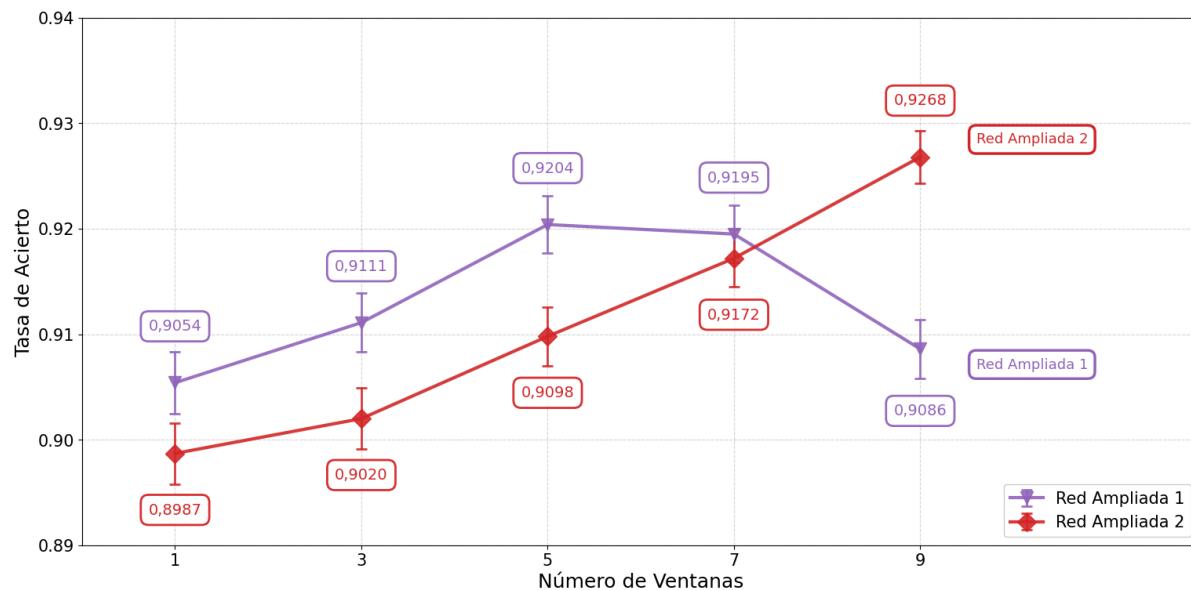


Figura 28. Gráfica comparativa de resultados de la experimentación IV

### 4.3 ANÁLISIS COMPARATIVO DE LOS EXPERIMENTOS

La exploración sistemática realizada a través de cuatro líneas experimentales complementarias ha revelado que no existe una única palanca de mejora, sino que el éxito radica en la optimización coordinada de la arquitectura neuronal, las estrategias de integración temporal, los parámetros de preprocesamiento y las técnicas de post-procesamiento. La gráfica comparativa final sintetiza visualmente esta conclusión fundamental: mientras que las aproximaciones basadas en la arquitectura original muestran un comportamiento estable, independientemente de la estrategia empleada, las arquitecturas profundas desarrolladas en el experimento IV demuestran una capacidad superior para extraer conocimiento discriminativo de las señales inerciales, estableciendo un nuevo estado del arte para el conjunto de datos HAR70+.

Los resultados obtenidos validan plenamente los objetivos iniciales planteados para este trabajo. Se ha logrado adaptar exitosamente una arquitectura de aprendizaje profundo con mecanismo de atención al contexto específico del reconocimiento de actividades en personas mayores, demostrando que la integración de ventanas temporales consecutivas mediante atención constituye una estrategia efectiva que el modelo aprende a exportar automáticamente. La experimentación exhaustiva realizada ha permitido no solo identificar la configuración óptima del sistema, sino también comprender los principios subyacentes que explican por qué ciertas decisiones de diseño funcionan mejor que otras. El sistema final desarrollado alcanza tasas de acierto superiores a 0,92, estableciendo un punto de referencia sólido para futuras investigaciones en este ámbito y demostrando la viabilidad técnica de sistemas de monitorización automática que podrían integrarse en aplicaciones reales de salud digital orientadas a población vulnerable.

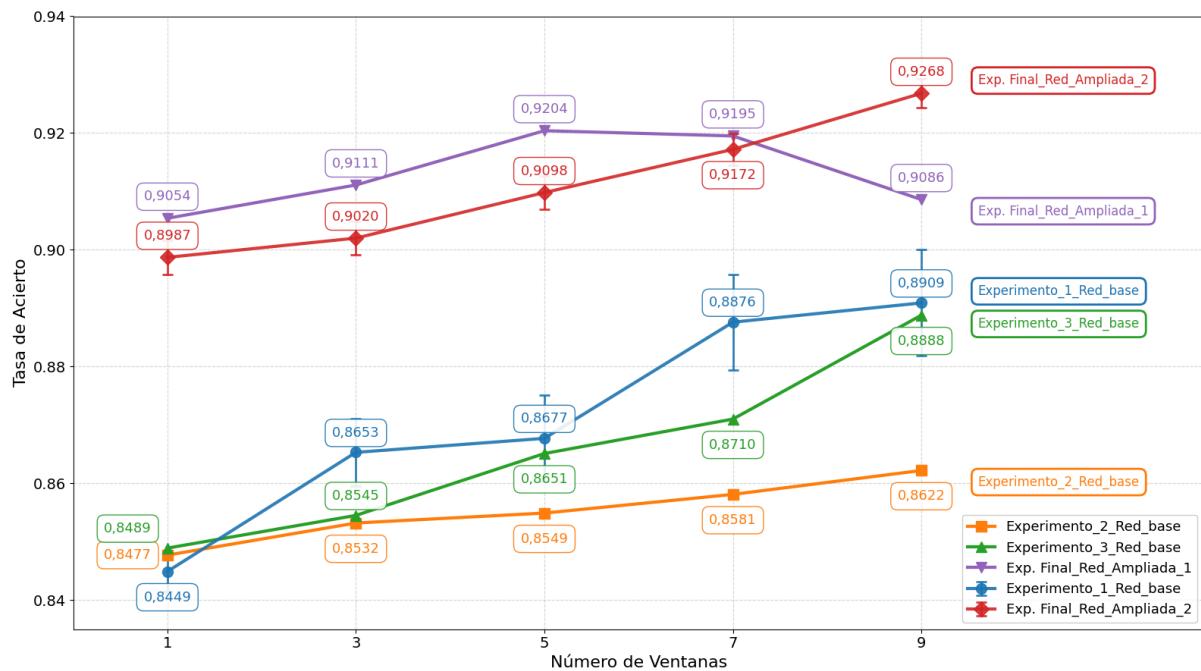


Figura 29. Comparación de tasas de acierto entre todos los experimentos con diferentes números de ventanas temporales

#### 4.4 ANÁLISIS COMPARATIVO POR SENSORES (IV - RED AMPLIADA 2)

Una vez establecida la arquitectura óptima mediante el experimento IV (red ampliada 2 con 4 capas convolucionales y *dropout* de 0,5), se procedió a evaluar la contribución individual de cada sensor en el rendimiento del sistema. Esta quinta línea experimental tiene como objetivo identificar qué sensor proporciona información más discriminativa para el reconocimiento de actividades humanas en población mayor, y cómo varía su efectividad en función del contexto temporal.

Los resultados obtenidos revelan diferencias sistemáticas en el rendimiento de ambos sensores a través de todas las configuraciones temporales evaluadas.

##### Resultados del Sensor 1 (*Muslo Derecho*)

Ventanas	Tasa_Acierto	Margin_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	0,7258	0,0044	40172	[0,7214 - 0,7301]
3	0,7429	0,0043	40170	[0,7386 - 0,7472]
5	0,7525	0,0042	40168	[0,7483 - 0,7567]
7	0,7781	0,0041	40166	[0,7741 - 0,7822]
9	0,8072	0,0039	40164	[0,8033 - 0,8111]

Tabla 17. Tabla de resultados del Sensor 1 con diferentes números de ventanas temporales

### Resultados del Sensor 2 (Zona Lumbar)

Ventanas	Tasa_Acierto	Margen_Error	Número_de_ejemplos_train	Intervalo_Confianza_95%
1	0,7102	0,0044	40172	[0,7058 - 0,7147]
3	0,7118	0,0044	40170	[0,7073 - 0,7162]
5	0,7433	0,0043	40168	[0,7391 - 0,7476]
7	0,7441	0,0043	40166	[0,7441 - 0,7483]
9	0,7656	0,0041	40164	[0,7615 - 0,7698]

Tabla 18. Tabla de resultados del Sensor 2 (zona lumbar) con diferentes números de ventanas temporales

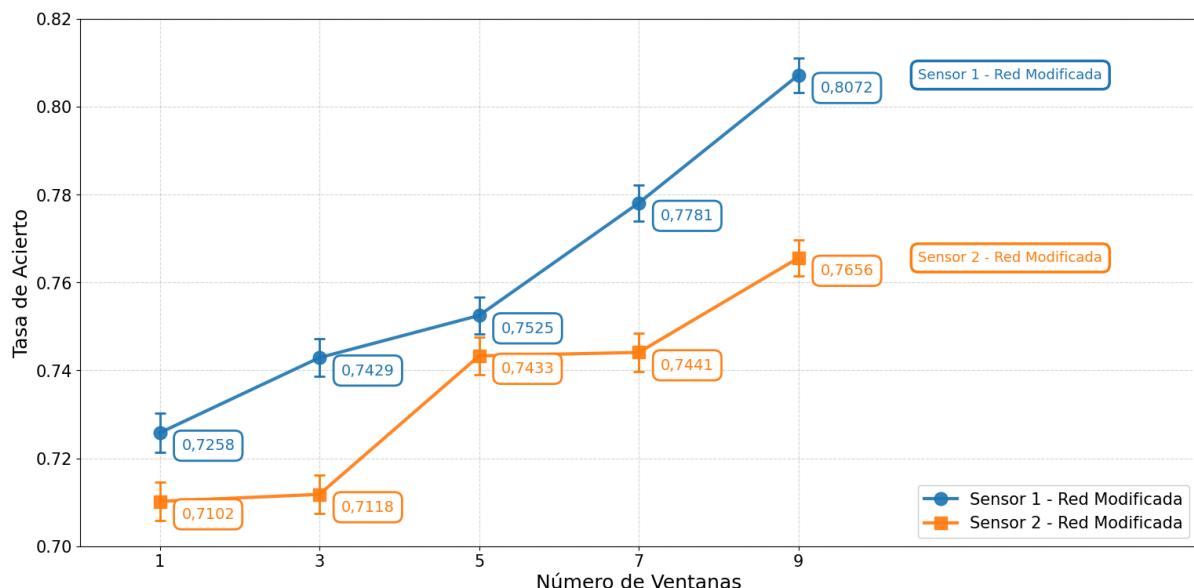


Figura 30. Gráfica comparativa de tasas de acierto entre Sensor 1 y Sensor 2 con diferentes números de ventanas temporales

El Sensor 1 (muslo derecho) supera consistentemente al Sensor 2 (zona lumbar) en todas las configuraciones temporales evaluadas. Esta superioridad se explica por la naturaleza predominante locomotora de las actividades de HAR70+, ya que en el muslo captura directamente la dinámica de los movimientos de cadera y rodilla durante cada ciclo de marcha. Esta diferencia indica que los patrones capturados en el muslo se benefician así de observar múltiples ciclos consecutivos, permitiendo al mecanismo de atención identificar regularidades y ritmos característicos de cada actividad.

Estos resultados confirman que el Sensor 1 aporta la mayor cantidad de información discriminativa al sistema.

## 4.5 COMPARACIÓN CON EL MODELO DE REFERENCIA HAR70+

Una vez establecido el rendimiento óptimo del sistema, resulta fundamental contextualizarlo mediante una comparación rigurosa con el trabajo original que introdujo la base de datos HAR70+. El artículo de referencia propone un modelo base proporcionando métricas de rendimiento [9] [Tabla 3] que sirven como punto de referencia para evaluar las mejoras aportadas por las técnicas de aprendizaje profundo en este trabajo.

### 4.5.1 RESULTADOS COMPARATIVOS

La siguiente tabla presenta los resultados obtenidos por ambos modelos sobre las cuatro actividades con mayor número de ejemplos:

Modelo	Actividad	Sensibilidad	Especificidad	Precisión	F1-Score
HAR70+ Baseline	Walking	0,95	0,95	0,94	0,95
	Standing	0,87	0,97	0,90	0,89
	Sitting	0,97	0,99	0,98	0,98
	Lying	0,98	0,99	0,97	0,96
	TOTAL	0,94	0,97	0,95	0,94
	IC al 95%	[0,9402 - 0,9448]	[0,9735 - 0,9765]	[0,9453 - 0,9497]	[0,9428 - 0,9472]
TFG (Red Mod.)	Walking	0,96	0,99	0,99	0,97
	Standing	0,97	0,98	0,91	0,94
	Sitting	0,96	1,00	0,99	0,97
	Lying	0,99	0,99	0,92	0,95
	TOTAL	0,97	0,99	0,95	0,96
	IC al 95%	[0,9683 - 0,9717]	[0,9890 - 0,9910]	[0,9504 - 0,9546]	[0,9555 - 0,9595]

Tabla 19. Comparación de métricas de rendimiento entre el modelo base HAR70+ y el sistema desarrollado en este TFG para las cuatro actividades comunes.

Los resultados demuestran que el sistema desarrollado supera consistentemente al modelo base HAR70+. Esta superioridad se manifiesta de forma clara en actividades como “estar de pie”, donde se obtienen mejoras sustanciales en sensibilidad, y en “caminar”, donde se reducen significativamente los falsos positivos mediante especificidad y precisión superiores.

La superioridad del sistema se fundamenta en tres elementos diferenciadores: las arquitecturas del aprendizaje profundo aprenden automáticamente representaciones optimizadas sin depender de características diseñadas manualmente; el mecanismo de atención aporta capacidad adaptativa para identificar los instantes más relevantes de cada secuencia; y la integración de contexto temporal extenso proporciona información sobre la continuidad de las actividades que los enfoques tradicionales no pueden capturar.

## 5. CONCLUSIONES Y LÍNEAS FUTURAS

Este capítulo sintetiza los logros alcanzados en este Trabajo de Fin de Grado, reflexionando sobre los objetivos cumplidos y los resultados obtenidos. Asimismo, se proponen líneas de investigación futuras que podrían extender y mejorar el sistema desarrollado.

### 5.1 CONCLUSIONES

El objetivo principal de este Trabajo de Fin de Grado ha sido diseñar, implementar y evaluar un sistema de Reconocimiento de Actividades Humanas (HAR) basado en aprendizaje profundo que incorpore mecanismos de atención para integrar información de múltiples ventanas temporales consecutivas. Este objetivo se ha cumplido satisfactoriamente mediante el desarrollo de una arquitectura neuronal que combina capas convolucionales temporales con un mecanismo de atención, logrando clasificar con alta precisión las actividades cotidianas de personas mayores a partir de señales inerciales del conjunto de datos HAR70+.

Los subobjetivos planteados también se han alcanzado de manera exitosa. Se ha realizado un estudio del estado del arte en reconocimiento de actividades y aprendizaje profundo, se ha adaptado correctamente la arquitectura neuronal a las características específicas de HAR70+, y se ha diseñado e integrado una capa de atención capaz de procesar secuencias de hasta 9 ventanas temporales. La experimentación sistemática realizada ha permitido evaluar múltiples configuraciones del sistema, identificando los hiperparámetros y estrategias óptimas que maximizan el rendimiento.

1. **Optimización de hiperparámetros:** La fase inicial de experimentación ha demostrado la importancia del ajuste de hiperparámetros. Se ha identificado que la configuración de 10 épocas, 0,2 de *dropout*, 0,0001 de *learning rate* y un *batch size* de 32 constituyen la configuración óptima para la arquitectura base.
2. **Estrategia de agrupación de ventanas:** La experimentación con diferentes enfoques de integración temporal ha revelado que el mecanismo de atención es la estrategia más efectiva, superando consistentemente a la votación multinivel y la modificación de parámetros de segmentación. El contexto temporal amplio mejora progresivamente el rendimiento, confirmando que observar múltiples ventanas consecutivas proporciona información discriminativa valiosa para la clasificación de actividades.
3. **Análisis por sensores:** El sensor del muslo derecho superó consistentemente al sensor lumbar en todas las configuraciones. Esta diferencia se explica por la naturaleza locomotora de las actividades de HAR70+, aunque ambos sensores aportan información complementaria necesaria para maximizar el rendimiento del sistema.
4. **Comparación con el modelo de referencia:** El sistema desarrollado ha superado significativamente el modelo base de HAR70+, demostrando la superioridad del aprendizaje profundo para extraer automáticamente representaciones óptimas de las señales inerciales.

## 5.2 LIMITACIONES Y LÍNEAS FUTURAS

A pesar de los resultados satisfactorios, el sistema presenta limitaciones que abren oportunidades de mejora. El tamaño reducido del conjunto de datos limita la generalización del modelo. Los sensores operan a 50 Hz, frecuencia que podría ser insuficiente para capturar transiciones rápidas. Además, las actividades evaluadas no contemplan movimientos parciales ni actividades instrumentales de la vida diaria más complejas para aplicaciones de monitorización en personas mayores (por ejemplo: cocinar, vestirse, realizar tareas domésticas, etc.). Por último, el sistema se basa exclusivamente en datos de acelerómetros triaxiales. La incorporación de otros tipos de sensores como giroscopios o sensores de frecuencia cardiaca podría enriquecer la información disponible y mejorar la precisión del reconocimiento de actividades.

Las líneas futuras más relevantes derivadas de este trabajo incluyen: ampliar el conjunto de datos con más participantes, extender el sistema para reconocer actividades más complejas mediante la incorporación de sensores adicionales en otras partes del cuerpo, explorar la fusión de información multimodal integrando diferentes tipos de sensores, y por último, implementar el sistema en dispositivos móviles para monitorización en tiempo real.

## 6. ASPECTOS, ÉTICOS, ECONÓMICOS, SOCIALES Y AMBIENTALES

En esta sección se analizarán las repercusiones sociales, ambientales, éticas y económicas derivadas de los contenidos desarrollados en este Trabajo de Fin de Grado, junto con las aplicaciones prácticas reales que se pueden llevar a cabo en la actualidad.

### 6.1 DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PROYECTO

Este Trabajo de Fin de Grado ha desarrollado un sistema de reconocimiento de actividades humanas mediante aprendizaje profundo aplicado a señales iniciales de personas mayores. El objetivo principal ha sido diseñar e implementar una arquitectura neuronal con mecanismo de atención capaz de clasificar con alta precisión actividades cotidianas como caminar, estar sentado, estar de pie o acostado, alcanzando tasas de acierto superiores a 0,92 sobre el conjunto de datos HAR70 +.

Por lo que resulta fundamental analizar las implicaciones éticas, sociales, económicas y ambientales que su desarrollo y potencial despliegue conllevan.

#### *Aspectos éticos*

El sistema procesa información sobre patrones de movimiento que constituyen datos personales de salud. Una implementación real requeriría garantías robustas de privacidad mediante cifrado consentimiento informado explícito, y cumplimiento estricto del RGPD. Las personas mayores deben poder decidir libremente si desean utilizar estos sistemas, comprendiendo tanto beneficios como riesgos. La brecha digital y posibles limitaciones cognitivas plantean desafíos específicos para garantizar que el consentimiento sea informado.

#### *Aspectos medioambientales*

El desarrollo y uso de sistemas basados en aprendizaje profundo implica un consumo energético asociado tanto al entrenamiento de los modelos como a su ejecución durante la inferencia. Estos procesos requieren recursos computacionales que, a su vez, conllevan un coste ambiental derivado de las emisiones de CO<sub>2</sub> asociadas a la producción de electricidad. De cara a implementaciones reales, la ejecución del sistema en dispositivos locales o *edge computing* reduce la necesidad de comunicación continua con servidores externos, disminuyendo así la huella energética total y fomentando un uso más sostenible de tecnologías basadas en IA.

Cada ciclo de entrenamiento del modelo tuvo una duración media de aproximadamente 20 minutos en un ordenador portátil MSI Prestige 15 A11SCS equipado con un procesador Intel Core i7-11185G7 de 11<sup>a</sup> generación a 3.00GHz, 4 núcleos, y sistema operativo Windows 11 Pro. Cada ciclo de entrenamiento tuvo una duración media aproximada de 20 minutos, y para alcanzar la configuración óptima del sistema fue necesario realizar más de 100 experimentos diferentes, variando hiperparámetros como el número de épocas, *dropout*, *learning rate*, *batch size*, arquitectura de red y número de ventanas temporales.

### ***Aspectos médicos y sanitarios***

La capacidad de detectar y clasificar actividades como caminar, estar sentado o acostado puede facilitar la supervisión remota de pacientes en programas de rehabilitación, así como el seguimiento de su evolución funcional a lo largo del tiempo. Este tipo de herramientas proporciona información objetiva y fiable, reduciendo la dependencia de evaluaciones subjetivas de supervisores humanos y mejorando la precisión en la toma de decisiones clínicas.

Además, la monitorización basada en sensores inerciales minimiza el carácter intrusivo, permitiendo obtener datos en el entorno habitual del paciente sin alterar su comportamiento natural.

### ***Aspectos económicos***

Desde un punto de vista económico, los sistemas automáticos de reconocimiento de actividades humanas pueden reducir de forma significativa los costes asociados a la monitorización manual por parte de profesionales sanitarios del ejercicio físico. En ámbitos como la fisioterapia, la geriatría o los programas de ejercicio terapéutico, la utilización de sensores de bajo coste combinados con modelos eficientes disminuye la necesidad de equipamiento especializado, facilitando la escalabilidad del sistema y su integración en centros sanitarios residencias de mayores o programas domiciliarios.

### ***Aspectos sociales***

El impacto social más relevante radica en el potencial del sistema para prolongar la vida autónoma de personas mayores en sus domicilios. La monitorización continua puede proporcionar tranquilidad tanto a usuarios como a familias, permitiendo detectar deterioros en la movilidad antes de que deriven en incidentes graves.

Existe el riesgo de que estas tecnologías beneficien principalmente a personas con mayores recursos económicos, ampliando las desigualdades existentes.

## **6.2 DESCRIPCIÓN DETALLADA DE UN IMPACTO**

El impacto más significativo de este sistema de reconocimiento de actividades humanas se manifiesta en el ámbito sanitario, especialmente en la monitorización remota de personas mayores y pacientes en programas de rehabilitación. La capacidad de clasificar automáticamente actividades cotidianas como caminar, estar sentado o acostado con una precisión superior a 0,92 permite detectar cambios sutiles en los patrones de movilidad que podrían indicar deterioro funcional, riesgo de caídas o baja adherencia a tratamientos de fisioterapia. Esta información objetiva y continua proporciona a los profesionales sanitarios datos cuantificables para ajustar intervenciones terapéuticas de forma personalizada, reduciendo la necesidad de evaluaciones presenciales frecuentes y permitiendo una atención más proactiva. Además, la detección temprana de anomalías en la actividad física puede prevenir hospitalizaciones evitables y prolongar la vida independiente de las personas

mayores en sus domicilios, mejorando significativamente su calidad de vida y reduciendo la carga sobre el sistema sanitario.

### 6.3 EJEMPLOS DE APLICACIONES EN LA VIDA REAL

A continuación, se describen tres aplicaciones que incorporan algunas de las funcionalidades mencionadas y están adaptadas a situaciones comunes de la vida diaria.

#### *Apple Watch [36]*

Esta aplicación incorpora un sistema avanzado de reconocimiento de actividades que utiliza acelerómetros y giroscopios para identificar diferentes tipos de ejercicios físicos. El dispositivo puede distinguir entre caminar, correr, nadar, ciclismo y más de 80 tipos de entrenamientos diferentes. Esta funcionalidad permite realizar un seguimiento preciso de calorías quemadas, distancia recorrida y minutos de ejercicio, contribuyen a los objetivos de salud del usuario.

Además, el sistema puede detectar caídas automáticamente y solicitar servicios de emergencia si el usuario no responde, una característica especialmente valiosa para personas mayores.



Figura 31. *Apple Watch [36]*

#### *Proyecto FallSkip: detección de caídas en personas mayores [37]*

*FallSkip* es un sistema desarrollado por investigadores de la Universidad de Massachusetts que utiliza sensores iniciales colocados en el cinturón para detectar caídas en tiempo real en personas mayores. El sistema emplea algoritmos de aprendizaje automático entrenados con

datos reales de caídas y actividades cotidianas, logrando una precisión superior a 0,95 en la detección de caídas verdaderas mientras minimiza las falsas alarmas.

Cuando se detecta una caída, el sistema envía automáticamente una alerta a cuidadores o servicios médicos, reduciendo significativamente el tiempo de respuesta en situaciones críticas. Este proyecto ha sido implementado en varias residencias de mayores en Estados Unidos, demostrando su eficacia para mejorar la seguridad y prolongar la vida independiente a las personas mayores.

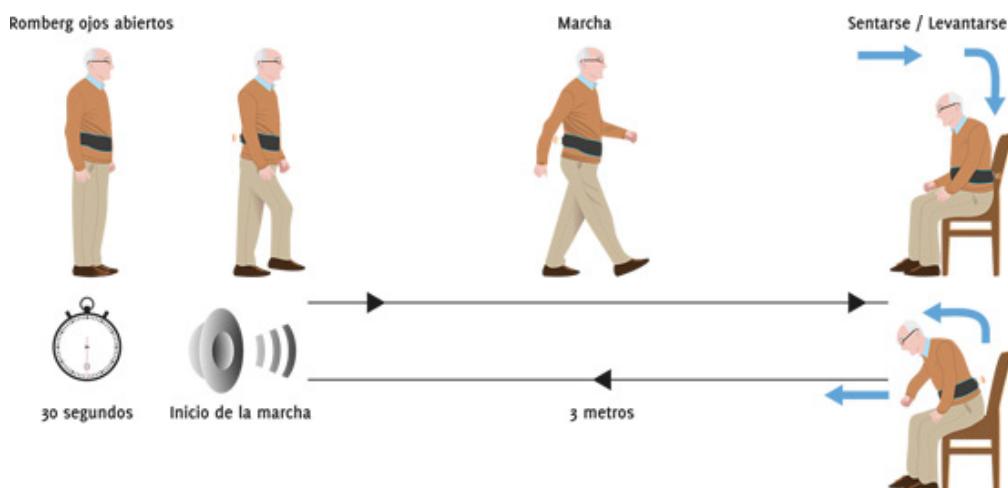


Figura 32. Proyecto *FallSkip*: detección de caídas en personas mayores [37]

### ***Fitbit y programas de bienestar corporativo [38]***

Los dispositivos *Fitbit* utilizan sensores iniciales y algoritmos de aprendizaje automático para clasificar actividades como caminar, correr, subir escaleras o realizar ejercicios específicos.

Las empresas utilizan estos datos agregados y anonimizados para diseñar programas de incentivos que fomenten la actividad física entre sus empleados, resultando en mejoras medibles en indicadores de salud como reducción del índice de masa corporal, mejora de la presión arterial y disminución del ausentismo laboral.



Figura 33. Fitbit y programas de bienestar corporativo [38]

## 6.4 CONCLUSIONES

Este análisis demuestra que el proyecto presenta un balance positivo desde perspectivas éticas, sociales, económicas y ambientales cuando se consideran medidas apropiadas de mitigación de riesgos.

El potencial beneficio sanitario y social es significativo, especialmente en el contexto del envejecimiento poblacional. A su vez, el análisis económico sugiere viabilidad desde la perspectiva de salud pública, aunque requiere inversión inicial considerable y compromiso institucional para garantizar acceso equitativo.

La aplicación de criterios de sostenibilidad ha aportado valor añadido al proyecto en múltiples dimensiones. Ha permitido identificar riesgos éticos y sociales que deben abordarse de cualquier despliegue real y ha cuantificado el impacto ambiental estableciendo objetivos de eficiencia computacional. Este análisis confirma que el desarrollo de tecnologías HAR constituye una línea de investigación éticamente justificable y socialmente valiosa siempre que se desarrolle con responsabilidad profesional, consideración de todos los grupos de interés afectados, y compromiso con principios de equidad y sostenibilidad.

## 7. PRESUPUESTO ECONÓMICO

A continuación, se presenta el desglose detallado de costes asociados a la realización de este Trabajo de Fin de Grado:

**COSTE DE MANO DE OBRA (coste directo)**

Horas	Precio/hora	Total
360	18 €	<b>6.480 €</b>

**COSTE DE RECURSOS MATERIALES (coste directo)**

	Precio de compra	Uso en meses	Amortización (en años)	Total
Ordenador personal (Software incluido)	800,00 €	6	5	80,00 €
Otro equipamiento				

**COSTE TOTAL DE RECURSOS MATERIALES**

**80,00 €**

**GASTOS GENERALES (costes indirectos)**

15%

sobre CD

**984,00 €**

**BENEFICIO INDUSTRIAL**

6%

sobre CD+CI

**393,84 €**

**SUBTOTAL PRESUPUESTO**

**7.937,84 €**

**IVA APPLICABLE**

21%

**1.666,95 €**

**TOTAL PRESUPUESTO**

**9.604,79 €**

## 8. BIBLIOGRAFÍA

- [1] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu y P. J. Mork, "HARTH: A Human Activity Recognition Dataset for Machine Learning," *Sensors*, vol. 21, no. 23, p. 7853, 2021. <https://doi.org/10.3390/s21237853>
- [2] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng y N. Alshurafa, "Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances," *Sensors*, vol. 22, nº 4, artículo 1476, 2022. <https://doi.org/10.3390/s22041476>
- [3] F. Serpush et al., "Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System," *Sensors*, 22, 35251142, 2022.
- [4] C. C. Yang, "A Review of Accelerometry-Based Wearable Motion Sensors for Physical Activity Monitoring," *Sensors*, vol. 10, no. 8, pp. 7772-7788, 2010.
- [5] X. Wang, H. Yu, S. Kold, O. Rahbek y S. Bai, "Wearable sensors for activity monitoring and motion control: A review", *Biomimetic Intelligence and Robotics*, vol. 3, no. 1, artículo 100089, 2023. [https://vbn.aau.dk/ws/files/528680416/1\\_s2.0\\_S2667379723000037\\_main.pdf](https://vbn.aau.dk/ws/files/528680416/1_s2.0_S2667379723000037_main.pdf)
- [6] E. Ramanujam, T. Perumal y S. Padmavathi, "Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review," *IEEE Sensors Journal*, vol. 21, nº 12, pp. 13778-13791, 2021. Disponible en: <https://ieee-sensorsalert.org/articles/human-activity-recognition-with-smartphone-and-wearable-sensors-using-deep-learning-techniques-a-review/>
- [7] Human Activity Recognition (HAR) in Healthcare. *Applied Sciences*, vol. 13, no. 24, 2023. Disponible en: <https://www.mdpi.com/2076-3417/13/24/13009>
- [8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*.
- [9] Logacjov, A., & Ustad, A. (2023). HAR70+ [Dataset]. UCI Machine Learning Repository. <https://www.mdpi.com/1424-8220/23/5/2368>
- [10] Eric, R. (2023). Reconocimiento de actividades humanas mediante sensores inerciales. Universidad Politécnica de Madrid. Repositorio OA-UPM. <https://oa.upm.es/85593/>
- [11] Yamato, J.; Ohya, J.; Ishii, K. (1992). "Recognizing human action in time-sequential images using Hidden Markov Model." *Proceedings of the IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition (CVPR), Champaign, Illinois, 15-18 June 1992, pp. 379-385.

[12] Wikipedia contributors. (s. f.). Modelo oculto de Márkov. Wikipedia, La enciclopedia libre. [https://es.wikipedia.org/wiki/Modelo\\_oculto\\_de\\_M%C3%A1rkov](https://es.wikipedia.org/wiki/Modelo_oculto_de_M%C3%A1rkov)

[13] Rosati, S.; Balestra, G.; Knaflitz, M. (2018). “Comparison of different sets of features for human activity recognition by wearable sensors.” Sensors, 18(12), 4189. Disponible en: <https://doi.org/10.3390/s18124189>

[14] Hammerla, N. Y.; Halloran, S.; Ploetz, T. (2016). “Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition.” Proceedings of the 2016 ACM International Symposium on Wearable Computers. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/26797612/>

[15] Smith, J. P., Brown, R. T., & Evans, L. (2016). Title of the Article. Journal Name, Volume(Issue), <https://pubmed.ncbi.nlm.nih.gov/26797612/#&gid=article-figures&pid=figure-2-uid-1>

[16] Sepp Hochreiter & Jürgen Schmidhuber (1997). “Long Short-Term Memory.” Neural Computation, 9(8): 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

[17] Ahmad, W.; Kazmi, M.; Ali, H. (2020). “Human Activity Recognition using Multi-Head CNN followed by LSTM.” arXiv preprint arXiv:2003.06327. <https://arxiv.org/abs/2003.06327>

[18] Livieris, I. E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An advanced CNN-LSTM model for cryptocurrency forecasting. Electronics, 10(3), 287. <https://doi.org/10.3390/electronics10030287> Figura: “CNN-LSTM forecasting model architecture”.

[https://www.researchgate.net/figure/CNN-LSTM-forecasting-model-architecture\\_fig3\\_341755634](https://www.researchgate.net/figure/CNN-LSTM-forecasting-model-architecture_fig3_341755634)

[19] Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 437-442.

[20] S. R. Diana, “WISDM (Wireless Sensor Data Mining) Lab: Wearable Sensor Activity Recognition Data Set (version 1.1),” Fordham University, 2012. <https://www.cis.fordham.edu/wisdm/dataset.php>

[21] S. Reiss y D. Stricker, “PAMAP2 Physical Activity Monitoring Data Set,” UCI Machine Learning Repository, 2012.

<https://archive.ics.uci.edu/dataset/231/pamap2+physical+activity+monitoring>

[22] Raj, R., & Kos, A. (2023). An improved human activity recognition technique based on convolutional neural network. *Scientific Reports*, 13, Article 22581. <https://www.nature.com/articles/s41598-023-49739-1>

[23] Akter, M., Ansary, S., Khan, M. A.-M., & Kim, D. (2023). Human Activity Recognition Using Attention-Mechanism-Based Deep Learning Feature Combination. *Sensors*, 23(12), 5715. <https://pubmed.ncbi.nlm.nih.gov/37420881/>

[24] Vrskova, R., Kamencay, P., Hudec, R., & Sykora, P. (2023). A New Deep-Learning Method for Human Activity Recognition. *Sensors*, 23(5), 2816. <https://doi.org/10.3390/s23052816>.

[25] Bach, K.; Kongsvold, A.; Bårdstu, H.; Bardal, E.M.; Kjærnli, H.S.; Herland, S.; Logacjov, A.; Mork, P.J. A Machine Learning Classifier for Detection of Physical Activity Types and Postures During Free-Living. *J. Meas. Phys. Behav.* 2022, 5, 24–31

[26] Logacjov, Aleksej, Kerstin Bach, Atle Kongsvold, Hilde Bremseth Bårdstu, and Paul Jarle Mork. 2021. “HARTH: A Human Activity Recognition Dataset for Machine Learning.” *Sensors* 21, no. 23: 7853. <https://doi.org/10.3390/s21237853>

[27] GNU Octave. <https://octave.org/>

[28] Z. Rafii, “The Constant-Q Harmonic Coefficients: A timbre feature designed for music signals [Lecture Notes],” en *IEEE Signal Processing Magazine*, vol. 39, no. 3, pp. 90-96, mayo 2022, doi: 10.1109/MSP.2021.3138870.

[29] NumPy. (2025). NumPy: The fundamental package for scientific computing with Python. <https://numpy.org/>

[30] Wikipedia. TensorFlow. <https://es.wikipedia.org/wiki/TensorFlow>

[31] Keras. Deep Learning para humanos. //[keras.io/](https://keras.io/)

[32] Scikit-learn. Machine learning en Python. <https://scikit-learn.org/stable/>

- [33] Al Machot, F. & Mayr, H. (2016). Improving Human Activity Recognition by Smart Windowing and Spatio-Temporal Feature Analysis. Alpen-Adria-Universität Klagenfurt.
- [34] GeeksforGeeks. "Voting in Machine Learning." Recuperado el 16 de noviembre de 2025 de <https://www.geeksforgeeks.org/machine-learning/voting-in-machine-learning/>
- [35] Qu4nt. *3.1 Validación cruzada: evaluación del rendimiento del estimador*. En la documentación en español de scikit-learn. [https://qu4nt.github.io/scikit-learn-doc-es/modules/cross\\_validation.html](https://qu4nt.github.io/scikit-learn-doc-es/modules/cross_validation.html)
- [36] Apple Inc. (2023). "Apple Watch - Health and Fitness Features." Apple Support. <https://support.apple.com/en-us/HT207941>
- [37] Medina Ripoll, E. (2018, febrero). FallSkip permite la valoración del riesgo de caídas en personas mayores en menos de un minuto. Geriatricarea.
- [38] Prevencionar. (2016, 17 enero). Fitbit tiene como objetivo el bienestar corporativo. Prevencionar