



UNIVERSIDAD
DE MÁLAGA



E.T.S. DE INGENIERÍA DE
TELECOMUNICACIÓN
UNIVERSIDAD DE MÁLAGA

Programa de Doctorado en Ingeniería de Telecomunicación

Tesis Doctoral

Contribuciones al estudio y análisis de sistemas vestibles de detección de caídas

Escuela Técnica Superior de Ingeniería de Telecomunicación

Universidad de Málaga

2021



Autor: José Antonio Santoyo Ramón

Director: Eduardo Casilari Pérez



UNIVERSIDAD
DE MÁLAGA

AUTOR: José Antonio Santoyo Ramón

 <https://orcid.org/0000-0003-4252-1077>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña JOSÉ ANTONIO SANTOYO RAMÓN

Estudiante del programa de doctorado DE INGENIERÍA DE TELECOMUNICACIÓN de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: CONTRIBUCIONES AL ESTUDIO Y ANÁLISIS DE SISTEMAS VESTIBLES DE DETECCIÓN DE CAÍDAS

Realizada bajo la tutorización de EDUARDO CASILARI PÉREZ y dirección de EDUARDO CASILARI PÉREZ (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 12 de DICIEMBRE de 2021

 Fdo.: JOSÉ ANTONIO SANTOYO RAMÓN Doctorando/a	 Fdo.: EDUARDO CASILARI PÉREZ Tutor/a
 Fdo.: EDUARDO CASILARI PÉREZ Director/es de tesis	



UNIVERSIDAD
DE MÁLAGA



Escuela de Doctorado



EFQM AENOR



Edificio Pabellón de Gobierno. Campus El Ejido.
29071
Tel.: 952 13 10 28 / 952 13 14 61 / 952 13 71 10
E-mail: doctorado@uma.es

Agradecimientos

Se suele decir que las cosas no se demuestran con palabras sino con hechos, eso es algo que me habéis confirmado durante todos estos años mediante vuestro apoyo, gracias de corazón a todas las personas que me han acompañado. Siempre habéis estado ahí, en los momentos buenos y sobre todo cuando más os necesitaba: en los instantes en los que se camina a la deriva, en aquellos momentos en los que creéis en mí incluso cuándo ni yo mismo lo hago, en las situaciones en las que caigo y vuestra mano me levanta. Hemos vencido juntos a todos los obstáculos, a los imprevistos, a la ansiedad e incluso a una pandemia mundial, y lo mejor es que seguimos luchando, riendo y encima somos más fuertes; y no, no hablo solo de los que siguen a mi lado, incluyo también a las personas que ya no están, pero siempre formarán parte de mí. Gracias.

Quiero mostrar mi agradecimiento a mi director y tutor de tesis don Eduardo Casilari Pérez, el principal “culpable” de que este trabajo se haya elaborado. Durante todos estos años, de principio a fin, me ha apoyado y demostrado su profesionalidad, generosidad, ilusión y ganas de trabajar. Gracias por la infinita paciencia que has tenido conmigo, todos tus consejos y por haberme animado en todo momento a avanzar y finalizar esta tesis. También quiero hacer mención especial a don José Manuel Cano García, gracias por tu colaboración y ayuda siempre que lo he necesitado. Jamás os podré expresar toda mi gratitud y admiración.

Gracias a mi familia por siempre estar ahí y confiar en mí, por darme ánimos y cariño en todo momento, por ser mi apoyo. Gracias por recordarme dónde está esa fuerza y esa calma que a veces pierdo y olvido, sin vosotros esto no habría sido posible.

Gracias también a todos mis amigos y amigas, por creer en mí, por ayudarme a desconectar, por las risas, por las charlas filosóficas y también por las absurdas, por hacerme sentir acompañado en todo momento. Este trabajo también tiene un poco de cada uno de vosotros.

También agradecer de nuevo a mi familia y amigos por haber participado en mis “locos” experimentos repitiendo el mismo movimiento infinidad de veces e incluso tirándooos porque necesitaba caídas. UMAFall existe gracias a vosotros.

Por último, quiero agradecer al Departamento de Tecnología y a la Universidad de Málaga por la oportunidad de realizar este trabajo y haber podido participar en él. Así mismo, a los diferentes organismos que han financiado parte del proyecto del que ha formado parte este trabajo: los Fondos FEDER (con beca UMA18-FEDERJA-022), la Junta de Andalucía (beca PAIDI P18-RT-1652) y la Universidad de Málaga, Campus de Excelencia Internacional Andalucía Tech.



UNIVERSIDAD
DE MÁLAGA

*Dedicado a mi familia
y a cada una de las personas
que forman parte de mí*



UNIVERSIDAD
DE MÁLAGA

Resumen

Las caídas son la segunda causa mundial de muerte por lesiones accidentales o no intencionales según la Organización Mundial de la Salud. En este contexto, cabe destacar que las personas mayores de 65 años tienen un mayor riesgo de sufrir este tipo de accidentes. Estos datos, unidos al envejecimiento de la población, convierten las caídas en uno de los mayores problemas de salud pública actuales, por tanto, resulta de vital importancia encontrar soluciones que ayuden a paliar los graves problemas originados por estas. El avance de los dispositivos “vestibles” y su incorporación en nuestro día a día convierte a esta tecnología en una solución interesante para aplicarla al ámbito de estudio de las caídas. En el presente trabajo se realiza un estudio y análisis de sistemas vestibles de detección de caídas, describiendo las diferentes contribuciones científicas aportadas a este campo de manera temática y cronológica. En concreto, se expone en primer lugar la creación de un repositorio público, que incluye movimientos de la vida cotidiana y caídas, con el objetivo de proveer a la comunidad científica de una herramienta con la que poder analizar sistemáticamente diferentes aspectos y técnicas relacionados con la detección. Posteriormente, debido a la aparición de diferentes repositorios en la literatura, se revisan estas bases de datos y se investigan los problemas originados por la heterogeneidad existente y la inexistencia de una normalización a la hora de crear este tipo de *datasets*. Consecutivamente, se analizan y evalúan diferentes procedimientos de detección de caídas de menor a mayor complejidad, tales como algoritmos de detección de caídas basados en umbrales, algoritmos de aprendizaje automático, tanto de aprendizaje supervisado (considerando las estrategias más comunes de *machine learning*) como detectores de anomalías, así como técnicas de aprendizaje profundo como las redes neuronales

convolucionales, demostrando que existen arquitecturas de clasificación que proporcionan un gran rendimiento a la hora de discernir entre caídas y movimientos cotidianos. Asimismo, se estudia el efecto que provoca la posición y número de sensores utilizados, los estadísticos utilizados como características de entrada de los algoritmos u otros factores como los rasgos físicos de los sujetos que participan en los repositorios, la frecuencia de muestreo o la tipología de los movimientos que se utilizan a la hora de entrenar y testear los algoritmos, así como las prestaciones obtenidas al entrenar y testear con bases de datos cruzadas, esto es, con trazas de movimientos provenientes de repositorios diferentes.

Abstract

Falls are the world's second principal cause of death from accidental or unintentional injuries according to the World Health Organization. In this context, it should be noted that people over 65 have a greater risk of suffering this type of accident. These data, together with the aging of the population, make falls one of the greatest current public health problems, therefore, it is fundamentally to find solutions that help relieve the serious problems caused by them. The advancement of wearable devices and their incorporation into our daily lives turns this technology into an interesting solution to apply to the field of research of falls. In the present work a study and analysis of wearable fall detection systems is carried out, describing the different scientific contributions made to this field in a thematic and chronological way. Specifically, firstly, the creation of a public repository is exposed, which includes movements of daily life and falls, with the goal of providing the scientific community with a tool with which to systematically analyse different aspects and techniques related to the detection. Subsequently, due to the appearance of different repositories in the literature, these databases are reviewed, and the problems caused by the existing heterogeneity and the lack of standardization when creating this type of datasets are investigated. Consecutively, different fall detection procedures are examined and evaluated from lower to higher complexity, such as threshold-based fall detection algorithms, machine learning algorithms, both supervised learning (considering the most common machine learning strategies) and detectors of anomalies, as well as deep learning techniques such as convolutional neural networks, showing

that there are classification architectures that provide great performance when it comes to discerning between falls and activities of daily living. Likewise, the effect caused by the position and number of sensors used, the statistics used as input characteristics of the algorithms or other factors such as the physical characteristics of the subjects participating in the repositories, the sampling frequency or the typology of data is studied or the movements that are used when the algorithms are training and testing, in addition to the performance obtained when performing a cross-datasets validation process, that is, training and testing with traces of movements from different repositories.

Contenido

Capítulo 1. Introducción	1
1.1. Estado del arte	3
1.2. Objetivos y contribuciones principales de la tesis	9
Capítulo 2. Desarrollo de un repositorio multisensor	11
2.1. Introducción.....	11
2.2. Descripción del banco de pruebas.....	12
2.3. Descripción de los movimientos presentes en UMAFall.....	13
2.4. Análisis básico de las características de las trazas de los movimientos	14
2.5. Conclusiones	14
Capítulo 3. Evaluación de la posición óptima del sensor y de la utilidad de combinar sensores	15
3.1. Análisis mediante algoritmos basados en umbral.....	15
3.1.1. Introducción	15
3.1.2. Metodología.....	16
3.1.3. Conclusiones	19
3.2. Estudio mediante algoritmos <i>machine learning</i> y análisis ANOVA.....	21
3.2.1. Introducción.....	21
3.2.2. Metodología.....	21
3.2.3. Resultados y conclusiones.....	25
Capítulo 4. Comparación y análisis de repositorios.....	31
4.1. Revisión de <i>datasets</i> existentes y propuesta de clasificación de ADL	31
4.1.1. Introducción.....	31
4.1.2. Metodología de búsqueda e información básica de los repositorios existentes	32
4.1.3. Análisis de los sujetos experimentales y escenarios utilizados	33

4.1.4. Número, duración y tipología de los movimientos	34
4.1.5. Número, características y posición de los sensores.....	35
4.1.6. Análisis de las características básicas de la acelerometría en los repositorios..	36
4.2. Estudio del problema la heterogeneidad de los repositorios.....	37
4.2.1. Introducción	37
4.2.2. Revisión y selección de repositorios públicos.....	37
4.2.3. Selección de características	38
4.2.4. Comparación y discusión de los resultados	39
4.2.5. Conclusiones	41
4.3. Análisis del algoritmo <i>machine learning</i> SVM con varias bases de datos y análisis ANOVA.....	42
4.3.1. Introducción	42
4.3.2. Descripción del sistema experimental	42
4.3.3. Conclusiones	44
Capítulo 5. Análisis de la influencia de las características de usuarios	47
5.1. Análisis de la caracterización estadística de los movimientos respecto a su tipología y las características de los usuarios.....	47
5.1.1. Introducción	47
5.1.2. Metodología.....	48
5.1.3. Resultados y conclusiones.....	49
5.2. Estudio del impacto de las características de los usuarios en las prestaciones de sistemas de detección de caídas basados en algoritmos de aprendizaje supervisado.....	51
5.2.1. Introducción	51
5.2.2. Definición de entorno de evaluación.....	51
5.2.3. Resultados.....	53
Capítulo 6. Evaluación de clasificadores de clase única.....	57

6.1. Introducción.....	57
6.2. Metodología.....	58
6.2.1. Algoritmos de detección de anomalías	58
6.2.2. Selección de características	59
6.2.3. Selección de repositorios	60
6.2.4. Sistema de prueba y evaluación.....	60
6.2.5. Métricas de calidad.....	61
6.3. Resultados.....	62
6.3.1. Prueba equilibrada	62
6.3.2. Estudio del impacto de las categorías de movimientos existentes en el set de entrenamiento respecto el rendimiento del sistema	63
6.3.3. Estudio del beneficio de <i>ensemble learning</i>	65

Capítulo 7. Evaluación de sistemas de aprendizaje profundo.

Análisis de la frecuencia de muestreo67

7.1. Introducción.....	67
7.2. Metodología.....	68
7.2.1. Selección de repositorios.....	68
7.2.2. Estudio del comportamiento del sistema respecto la frecuencia de muestreo..	68
7.2.3. Configuración de la red neuronal convolucional	69
7.2.4. Métrica de calidad.....	70
7.3. Resultados.....	70

Capítulo 8. Conclusiones y trabajo futuro.....73

8.1. Conclusiones	73
8.2. Amenazas de validez y trabajo futuro.....	78

Apéndice A. Publicaciones93

A.1. Desarrollo de un repositorio multisensor	95
A.2. Análisis mediante algoritmos basados en umbrales.....	105

A.3. Estudio mediante algoritmos <i>machine learning</i> y análisis ANOVA.....	125
A.4. Revisión de <i>datasets</i> existentes y propuesta de clasificación de ADL	157
A.5. Estudio del problema la heterogeneidad de los repositorios	187
A.6. Análisis del algoritmo <i>machine learning</i> SVM con varias bases de datos y análisis ANOVA.....	225
A.7. Análisis de la caracterización estadística de los movimientos respecto a su tipología y las características de los usuarios.....	241
A.8. Estudio del impacto de las características de los usuarios en las prestaciones de sistemas de detección de caídas basados en algoritmos de aprendizaje supervisados.....	253
A.9. Evaluación de clasificadores de clase única.....	269
A.10. Evaluación de sistemas de aprendizaje profundo. Análisis de la frecuencia de muestreo	291
A.11. Evaluación del testeo cruzado de repositorios en sistemas vestibles de detección de caídas.....	311
Apéndice B. Análisis de las características que describen los movimientos	323
Apéndice C. Evaluación de utilización de técnicas de <i>Ensamble Learning</i>.....	363
Apéndice D. Evaluación de sistemas de aprendizaje profundo y elección de hiperparámetros – Redes Convolucionales	369
Apéndice E. Testeo cruzado de repositorios utilizando CNN	383
Apéndice F. Publicaciones derivadas de esta tesis	395
F.1. Artículos publicados	395
F.2. Artículos enviados o pendientes de envío	396
F.3. Ponencias en congresos	397

Lista de Acrónimos

ADL	<i>Activities of Daily Living</i>
ANN	<i>Artificial Neural Network</i>
ANOVA	<i>Analysis Of Variance</i>
AUC	<i>Area under the curve</i>
BAN	<i>Body Area Network</i>
CEP	<i>Complex Event Processing</i>
CSV	<i>Comma Separated Values</i>
doi	<i>Digital Object Identifier</i>
DT	<i>Decision Trees</i>
ECG	Electrocardiograma
E.T.S.I.T.	Escuela Técnica Superior de Ingeniería de Telecomunicación
FCD	<i>F-test correlation difference</i>
FCQ	<i>F-test correlation quotient</i>
FDS	<i>Fall Detection Systems</i>
FN	<i>False Negative</i>
FOF	<i>Fear Of Fall</i>
FP	<i>False Positive</i>
GMM	<i>Gaussian Mixture Model</i>
GPS	<i>Global Positioning System</i>
HAR	<i>Human Activity Recognition</i>
hctsa	<i>Highly comparative time-series analysis</i>
ICD	<i>International Classification of Diseases</i>
IMC	Índice de Masa Corporal
IMU	<i>Inertial Measurement Unit</i>

ISBN	<i>International Standard Book Number</i>
ISSN	<i>International Standard Serial Number</i>
JCR	<i>Journal Citation Reports</i>
KDD	<i>Kernel Desity Estimation</i>
kNN	<i>k-Nearest Neighbors</i>
LSD	<i>Least Significant Difference</i>
MID	<i>Mutual information difference</i>
MIQ	<i>Mutual information quotient</i>
mRMR	<i>Minimum Redundancy Maximum Relevance</i>
NB	<i>Naïve Bayes</i>
OC	<i>One-class classifier</i>
OMS	Organización Mundial de la Salud
PPNN	<i>Parzen Probabilistic Neural Network</i>
ReLU	<i>Rectified Linear Unit</i>
RNN	<i>Recurrent Neural Networks</i>
ROC	<i>Receiver Operating Characteristic</i>
SNR	Signal to Noise Ratio
STAC	Síndrome de Temor a Caerse
SVM	<i>Support Vector Machine</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
WHO	<i>World Health Organization</i>

Capítulo 1. Introducción

Actualmente, la población mundial (y en especial, en los países de desarrollo) vive un evidente proceso de envejecimiento: la esperanza de vida aumenta y la tasa de natalidad disminuye. Según algunos estudios, se pronostica que entre los años 2.015 y 2.050 la población en el mundo con una edad superior a 60 años pasará de ser 900 millones a 2.000 millones de personas, lo que en términos porcentuales supondrá transitar desde un 12% a un 22% del total de los habitantes de la Tierra [1]. Este cambio social está dando lugar a una serie de desafíos a los que se debe hacer frente para conseguir un aumento en la calidad de vida de los ciudadanos de mayor edad. El presente estudio se centra en uno de los mayores problemas de la salud pública a la que se enfrenta la sociedad: las caídas.

Las caídas, definidas por la Organización Mundial de la Salud (OMS) como “sucesos involuntarios que hacen perder el equilibrio y dar con el cuerpo en el suelo o en otra superficie firme que lo detenga” [2], se encuentran estudiadas y catalogadas en el documento “Clasificación Internacional de Enfermedades” (ICD), redactado por la propia OMS con el propósito de establecer una sistematización y codificación de enfermedades. Concretamente, en la versión actualmente en uso (ICD-10) las caídas se encuentran codificadas como W00-W19 dentro del apartado “Causas externas de morbilidad” [3,4].

Sufrir una caída puede ocasionar consecuencias leves como hematomas o contusiones, pero también puede derivar en situaciones más graves y peligrosas como fracturas, desórdenes psicológicos, o en el peor de los casos, la muerte

La OMS advierte que las caídas son la segunda causa mundial de muerte por lesiones accidentales o no intencionales [5]. Cada año se producen alrededor de 37.3 millones de caídas cuya gravedad requiere atención médica. Se estima que anualmente fallecen en todo el mundo unas 424,000 personas como consecuencia de caídas [5,6]. Un hecho

más que probado es que son las personas mayores de 65 años las más proclives a padecer este tipo de accidentes. Aproximadamente el 28-35% de personas de más de 65 años sufre alguna caída cada año, aumentando dicho porcentaje con la edad, pues se convierte en 32-42% para personas de más de 70 [7]. Asimismo, aparte de este grupo etario, existen colectivos de riesgo, que están expuestos a sufrir caídas durante su tiempo laboral o de ocio (ciclistas, montañeros, bomberos, antenistas, etc.) y para los que puede resultar de gran interés una tecnología que afronte este problema.

En este contexto, resulta de vital importancia la rapidez con que una persona accidentada tras una caída es socorrida, pues la posible demora en la asistencia está fuertemente correlacionada con las secuelas físicas provocadas por el accidente o incluso con la probabilidad de sobrevivir [8]. En el grupo de las personas mayores, esta circunstancia puede agravar el problema de las caídas ya que una elevada proporción de estos ciudadanos viven solos, de forma que, si ocurre un accidente, puede resultar difícil alertar a un cuidador (un familiar, personal sanitario, etc.) para que le brinde ayuda.

Además de las lesiones físicas, las caídas son una fuente importante de temor y pérdida de independencia en las personas mayores. De hecho, existe un síndrome específico para personas mayores denominado Síndrome de Temor a Caerse (STAC) o en inglés *Fear Of Fall* (FOF) *syndrome* que se define por el conjunto de síntomas que experimenta una persona (que normalmente ya ha experimentado un accidente) por el miedo a volver a caer [9]. Así, es frecuente que, tras sufrir una caída (incluso leve), un paciente pueda sufrir una disminución de su confianza y con ello una reducción de la movilidad en su día a día, provocando, con el transcurso del tiempo, una pérdida de su capacidad funcional, mayor dependencia de terceros, ingresos hospitalarios, ansiedad, neuroticismo (inestabilidad emocional) e incluso el fallecimiento [10].

Este miedo a las caídas repercute en la propia recurrencia de caídas; según estudios, como por ejemplo explica Masud et al. en [9], las personas que han sufrido una primera caída tienen una probabilidad mayor del 60% de volver a experimentar de nuevo una segunda caída durante el año inmediatamente posterior al percance.

Todos estos datos indican que es prioritario encontrar soluciones y llevarlas a cabo para paliar los graves problemas derivados de las caídas que afectan a la sociedad. En este entorno sociocultural, no podemos olvidar que la tecnología forma parte de la vida de los seres humanos y que un gran número de dispositivos nos acompañan en nuestro día a día facilitando gran cantidad de tareas, además de permitir la recopilación de infinidad

de información y la posibilidad de permanecer comunicados sin importar la distancia. Su utilización puede ser una solución interesante en el campo de estudio de las caídas.

1.1. Estado del arte

La utilización de la tecnología en este ámbito de estudio es un recurso interesante para tratar de reducir los peligros asociados a las caídas. Aunque estas soluciones no evitan las propias caídas, llevan a cabo una labor que aporta información muy importante: detectar de forma automática la caída de una persona, alertando del suceso ocurrido inmediatamente y agilizando el tiempo de respuesta para auxiliar a la persona afectada.

En la actualidad existen distintos sistemas que pretenden servir de ayuda para disminuir las consecuencias de las caídas, son conocidos como sistemas de detección de caídas o FDS (*Fall detection Systems*). En este sentido, esta tesis no entra a analizar aquellos sistemas focalizados en la propia prevención o análisis del riesgo de caídas (sobre ese tema, véase, por ejemplo la revisión de estudios presentada en [11]).

Los FDS, que se pueden contemplar como un caso muy específico de sistemas de reconocimiento de actividades o HAR (*Human Activity Recognition*) systems, se han venido clasificando tradicionalmente en dos grandes grupos, por un lado, se encuentran los sistemas sensibles al contexto (*context-aware system*) y por otro los sistemas vestibles (*wearable systems*) [12].

El primer grupo, los sistemas sensibles al contexto, se basan en la utilización de una red de sensores ambientales localizados en el entorno o lugar concreto donde se encontrará el paciente. En este estudio no se utilizará este tipo de sistema debido a sus desventajas, entre las que se incluyen: su alto coste de instalación y mantenimiento, su limitación en cuanto a la zona donde el usuario puede ser monitorizado, la alta probabilidad de interferencia de factores externos como luz, ruidos y la sensación por parte del paciente de vulnerabilidad de su privacidad a causa de la utilización de elementos como micrófonos o cámaras.

Por su parte, los sistemas vestibles aprovechan sensores que se integran con facilidad en la vestimenta o indumentaria del paciente, dichos sensores solo monitorizan normalmente magnitudes ligadas al movimiento del usuario como su aceleración y velocidad angular. Este tipo de sistemas presenta algunas ventajas sobre los anteriormente citados ya que eliminan la restricción en cuanto a la zona de

monitorización pues ahora los sensores acompañan siempre al paciente; además, resultan menos intrusivo y más económicos y robustos frente a alteraciones por factores alternos. Otra ventaja es que en gran parte de este tipo de sistemas se utiliza como base o nodo central un *smartphone*, elemento que en la actualidad la mayoría de las personas poseen. Por ejemplo, en 2014 alrededor del 80% de personas mayores de 60 años en Estados Unidos ya tenían y utilizaban diariamente el teléfono móvil mientras que el 50% poseía un *smartphone* [13]. Cada año estos porcentajes tienden a incrementarse. Según se predice en [14] a partir de los datos de [15,16] la población mayor de 60 años en los países occidentales utilizará móvil prácticamente en su totalidad en los próximos diez años.

Como consecuencia, un número no desdeñable de los estudios disponibles en la literatura utilizan el *smartphone* como único elemento del sistema detector de caídas, trabajando por tanto como sensor, nodo de procesamiento de datos y sistema de comunicación [14,17,18]. Sin embargo algunos trabajos [19–21] han demostrado que la posición del sensor que monitoriza los movimientos del sujeto es crucial en las prestaciones del sistema de detección de caídas. En ese sentido, la posición natural de un *smartphone* no resulta ser muy aconsejable; pues habitualmente (si no es transportado en un bolso) se encuentra ubicado en un bolsillo, junto al muslo, lugar que presenta un problema añadido, ya que su rendimiento dependerá de su holgura y por tanto de la libertad de movimiento del teléfono dentro del propio bolsillo. Para caracterizar la movilidad de una persona, se suele recomendar ajustar al cuerpo al dispositivo y ubicarlo en posiciones más cercanas al centro de gravedad del cuerpo humano, como el pecho o la cintura [22,23]; sin embargo, estos no son lugares necesariamente ergonómicos para colocar un *smartphone*, por lo que se sugiere emplear sensores de un tamaño reducido que se incorporen a la vestimenta del paciente. En este ámbito de actuación, poseen también un interés creciente aquellos sistemas FDS basados en dispositivos *wearables* comerciales, muy en particular los que aprovechan los relojes inteligentes o smartwatches, cada vez más extendidos entre la población [24]. De hecho, algunos relojes, como el Apple Watch 4, ya incorporan de fábrica sistemas de detección y alerta ante caídas (aunque no especifican los algoritmos empleados ni han reportado claramente su fiabilidad).

Hoy en día existe gran variedad de motas sensoras que pueden ser utilizadas para este cometido. Dichos dispositivos incorporan acelerómetros, giróscopos y magnetómetros,

además de incluir comunicación inalámbrica de corto alcance (Bluetooth, ZigBee, etc.) que puede aprovecharse para enviar la información a dispositivos con mayor capacidad de procesamiento como por ejemplo un *smartphone*. Esta combinación de dispositivos da lugar a redes de área corporal multisensor, conocidas en inglés como *Multisensor Body Area Networks* (BANs). Actualmente este tipo de sistemas está en auge, prueba de ello es el aumento de estudios en este ámbito de la literatura, habiéndose publicado diversos trabajos que evalúan las prestaciones de un sistema de clasificación de actividades de la vida diaria (ADL o *Activities of Daily Living*) y/o detección de caídas al utilizar varios sensores inerciales ubicados en diferentes partes del cuerpo [22,25–29]. Para este propósito es necesario evaluar sistemáticamente los algoritmos de clasificación, repitiendo diferentes pruebas, variando las técnicas de detección y la parametrización de los algoritmos. Es por ello por lo que, en la mayor parte de los trabajos existentes en la literatura, se crean repositorios (publicados o no) con las medidas recogidas por los sensores inerciales al realizar diferentes movimientos.

Conseguir monitorizar una caída real es una tarea difícil. Se estima que sería necesario monitorizar entre setenta mil y cien mil días la actividad de personas mayores de 65 años para capturar unas 100 caídas reales [30]. Por esta razón, el procedimiento habitual llevado a cabo en la literatura es monitorizar a una serie de sujetos, comúnmente personas jóvenes y en buen estado de salud, que realizan sistemáticamente y siguiendo un plan predefinido movimientos comunes y otros en los que simulan caerse. Estos movimientos suelen incluir tanto diferentes actividades típicas de la vida diaria como caminar, agacharse, sentarse o levantarse de una silla o una cama, como la emulación de distintos tipos de caídas (habitualmente sobre una superficie protectora con el fin de evitar posibles lesiones). A pesar de la dificultad comentada de grabar caídas reales sufridas por personas mayores, también existen estudios en los que se ha monitorizado el día a día de personas ancianas con el propósito de obtener la mayor cantidad de datos en escenarios y situaciones reales [31–33].

En gran parte de los estudios incluidos en la literatura, los autores no publican estos repositorios de movimientos, lo que impide que otros investigadores puedan utilizarlos para estudiar, analizar y validar otras técnicas de detección de caídas. Asimismo, se ha detectado que no existe una pauta a seguir para crear estos *datasets*, por lo que se manifiesta una gran heterogeneidad entre ellos, tanto por la tipología de movimientos

que incorporan, como por las características de los sensores con los que se han recogido los datos.

Por otro lado, en cuanto a las técnicas utilizadas en la literatura para la detección de caídas en FDS vestibles, pueden hallarse diferentes metodologías en este tipo de sistemas. Las técnicas más sencillas se fundamentan en la utilización de algoritmos basados en umbrales. Fundamentalmente estas estrategias asumen la ocurrencia de caídas si uno o varios parámetros derivados de las medidas de los sensores inerciales excede un determinado valor. Este tipo de algoritmos suelen presentar problemas a la hora de clasificar ADL con altos niveles de agitación, como pueden ser un movimiento deportivo tal como saltar o correr [34–37].

Otro procedimiento empleado en este ámbito es la utilización de técnicas de *machine learning* o aprendizaje automático. Este tipo de métodos reconoce patrones a partir de ser entrenados con datos, permitiendo crear modelos capaces de clasificar nueva información que se le entregue. Según diversos estudios, los resultados de esos clasificadores suelen superar claramente a los obtenidos por algoritmos basados en umbrales [38]. Dentro de este grupo de técnicas, se encuentran los algoritmos de aprendizaje supervisado, los cuales requieren ser entrenados con todos los tipos de datos que se desea clasificar, en el caso de este estudio ADL y caídas (ya que un FDS puede comprenderse como un sistema de clasificación binario). Hay numerosos estudios que utilizan este tipo de algoritmos para la detección de caídas (véanse por ejemplo las referencias en [27,39–41]). Otras técnicas de aprendizaje automático interesantes en el ámbito de la detección de caídas son los algoritmos de detección de anomalías o de clase única. A diferencia de lo que ocurre con los métodos de aprendizaje supervisado este tipo de algoritmos solo es entrenado con un tipo de dato, de forma que, una vez configurados, aquellos datos que diverjan sustancialmente de los usados durante aprendizaje son clasificados como anomalías. Esta clase de algoritmo puede presentar ventajas en el campo de la detección de caídas, ya que evitarían la necesidad de monitorizar caídas (fingidas o reales) para entrenar el sistema, pues solo necesitarían entrenar con movimientos de tipo ADL (mucho más fáciles de capturar, incluso entre personas mayores). En la literatura pueden encontrarse estudios que utilizan esta clase de algoritmos para detectar caídas [42,43].

Asimismo, también existen otros métodos de clasificación más complejos basados en aprendizaje profundo. Estas técnicas, al contrario de lo que ocurre con los algoritmos de

aprendizaje clásico, permiten extraer automáticamente las características que representan los datos de las propias señales obtenidas por los sensores inerciales. Existen también varios estudios en la literatura que hacen uso de este tipo de métodos, consiguiendo muy buenas prestaciones al discernir entre caídas y ADL [44–49].

En la literatura también se han empleado otras técnicas para la detección de caídas como el procesamiento de eventos complejos (Complex Event Processing – CEP), en el ámbito de los FDS conocido normalmente como “*sensor fusion*”. Estos se apoyan en sistemas que toman las decisiones basándose en datos de dispositivos de medida diversos como sensores de aceleración, giróscopos, sensores de monitorización de frecuencia cardiaca o electrocardiograma (ECG), presión sanguínea, temperatura del cuerpo, saturación de oxígeno en sangre, entre otros. Este tipo de sistemas analizan una gran cantidad de información variada que puede ser utilizada incluso para anticiparse a una caída o detectar posibles anomalías y problemas del paciente, además de mejorar la fiabilidad de los sistemas [50,51]. Este tipo de sistemas queda fuera del ámbito de estudio de esta tesis.

En el presente trabajo se lleva a cabo el estudio y análisis de sistemas vestibles de detección de caídas. La estructura de este documento viene fundamentada a través de un conjunto de trabajos ya publicados o enviados a revistas indexadas en el JCR (*Journal Citation Reports*) y, en algún caso, a un congreso internacional. La temática de estos trabajos se encuentra fuertemente relacionada, de manera que la unión de todos ellos, junto a otros estudios y resultados no publicados, entendemos que conforma un corpus de investigación coherente. Esta memoria, por ende, se presenta como un documento completo y cohesionado que describe a partir del mismo hilo conductor las diferentes contribuciones aportadas al ámbito de estudio de la detección de caídas con dispositivos vestibles. La configuración del documento es la siguiente:

En primer lugar, en la sección actual, se ha realizado una introducción donde se presenta la problemática existente, los motivos que impulsan a realizar el trabajo, así como un esbozo del estado de la cuestión y, por último, una descripción de la estructura del documento.

Posteriormente, se incluye una serie de secciones o capítulos que describen y resumen, en español, cada una de las publicaciones que conforman el núcleo central de este trabajo. Tras estos resúmenes, se incluye un apartado o sección final donde se reúnen las conclusiones finales de los resultados globales del presente trabajo. Siguiendo el

mismo orden de los resúmenes, en el Apéndice A se incluye al final una copia de estas publicaciones (obviamente en inglés). Del mismo modo, en apéndices subsiguientes se adjuntan algunos estudios y resultados complementarios (redactados también en español), relacionados con el propósito de este trabajo pero que no han sido publicados todavía.

Los capítulos se articulan y suceden temáticamente a partir de un discurso científico que por un lado responde a la lógica cronológica de las publicaciones (y de la propia labor doctoral), y por otro, va incrementando la complejidad de los algoritmos analizados. Así, en primer lugar, se expone la creación y desarrollo de un repositorio propio de movimientos y caídas con el objetivo de proporcionar a la comunidad científica una recopilación de datos para poder analizar sistemáticamente diferentes técnicas de detección de caídas. A partir del repositorio conseguido, el documento continúa con un capítulo que incluye una serie de estudios que pretenden analizar y evaluar la posición óptima de los sensores inerciales en el cuerpo humano, así como, la utilidad de combinar sensores ubicados en diferentes ubicaciones. Seguidamente, debido a la progresiva aparición de repositorios de movimientos en estos últimos cinco años, se realiza una serie de trabajos donde se revisan los diferentes *datasets*, investigando los problemas derivados de la alta heterogeneidad existente entre los mismos y constatando la ausencia de una normalización a la hora de crear este tipo de repositorios. Posteriormente se analizan diferentes técnicas de detección de caídas con algunos de los repositorios disponibles, haciendo uso de algoritmos basados en umbrales y algoritmos de aprendizaje automático, tanto de aprendizaje supervisado como de detección de anomalías, así como aprendizaje profundo, en concreto de redes neuronales convolucionales. Además, se analiza la influencia en las prestaciones de los sistemas de detección de caídas de las características utilizadas para describir los movimientos, así como los rasgos intrínsecos de los sujetos, la frecuencia de muestreo y la tipología de los movimientos que se utilizan en el proceso de entrenamiento y testeo.

Cabe destacar que, en casi la totalidad de la bibliografía relacionada, las evaluaciones y análisis se basan simplemente en el estudio de las diferentes medias de las magnitudes que se obtienen, sin comprobar su significación estadística, es decir, sin verificar que esas variaciones en las medias son realmente significativas entre sí. Existe algún estudio donde sí se comprueba la significación estadística de los resultados, como por ejemplo [52], en el que realiza un análisis valor-p para analizar si la muestra influye en la

valoración del método de detección. En este trabajo se comprueba la significación estadística a través de técnicas específicas como el análisis de varianza (ANOVA), lo cual constituye otra de las aportaciones relevantes.

1.2. Objetivos y contribuciones principales de la tesis

A continuación, se exponen los principales objetivos y contribuciones de esta tesis:

- Crear y desarrollar de un repositorio de movimientos y caídas a partir de un sistema de detección de caídas multisensor vestibular que pueda ser empleado por la comunidad científica para comparar las técnicas y algoritmos de detección de caídas.
- Estudiar el efecto de la posición de los sensores utilizados en la caracterización de movimientos en un sistema de detección de caídas vestibular, así como, las posibles ventajas de combinar los resultados de diferentes sensores utilizando un repositorio de movimientos y caídas propio y haciendo uso de diferentes tipos de algoritmos de detección de caídas basados en umbrales y de aprendizaje supervisado.
- Incorporación del análisis de varianza (ANOVA) como técnica para determinar la significación estadística existente entre los resultados de los diferentes experimentos (procedimiento que habitualmente no se considera en los estudios sobre FDS).
- Estudiar y analizar *datasets* existentes en la literatura, examinando distintos factores, como las características de los sujetos experimentales, escenarios utilizados, particularidades de los movimientos como la cantidad, duración o tipología, número, propiedades y posición de los sensores utilizados, así como, las características básicas de acelerometría de los diferentes repositorios.
- Demostrar la gran heterogeneidad existente entre los repositorios de movimientos y la ausencia de unas normas que definan el procedimiento que debe llevarse a cabo para definir los bancos de pruebas experimentales y justificar si es legítimo extrapolar las métricas de calidad obtenidas al utilizar una determinada técnica de detección de caídas usando un repositorio concreto a otros escenarios.
- Analizar el efecto de las características seleccionadas como entrada de los algoritmos *machine learning* en el rendimiento del sistema de clasificación.

- Examinar el comportamiento de los sistemas y técnicas de detección de caídas al utilizar como entrenamiento datos de un repositorio y testear con los movimientos de otro *dataset* (*cross-datasets*).
- Estudiar el efecto que tienen las características intrínsecas (edad, altura, peso, IMC, sexo) de los sujetos experimentales, que realizan los movimientos incluidos en los repositorios, sobre las prestaciones de un sistema de detección de caídas, así como sobre las propias propiedades de acelerometría de los movimientos.
- Estudiar la conducta de los clasificadores de clase única al ser aplicados en el campo de la detección de caídas., investigando los posibles beneficios de utilizar *ensemble learning* en la clasificación.
- Evaluar la aplicación de sistemas de aprendizaje profundo como redes neuronales convolucionales al ámbito de la detección de caídas.
- Estudiar el impacto de la frecuencia de muestreo en el comportamiento de un sistema de detección de caídas.
- Investigar el comportamiento de un sistema de detección de caídas basado en una red neuronal convolucional cuando se realiza un testeo cruzado de repositorios.

Capítulo 2. Desarrollo de un repositorio multisensor

Resumen del artículo [53]:

Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. In Proceedings of the 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) (Procedia Computer Science); Leuven, Belgium, July 24-26, 2017; Vol. 110, pp. 32–39.

Puede encontrar una copia de esta publicación en el apéndice A.1.

2.1. Introducción

A pesar del auge existente en la investigación de sistemas de detección de caídas, actualmente, no existe un procedimiento científico estandarizado que defina la metodología que se debe utilizar para evaluar este tipo de sistemas.

En concreto, uno de los factores más relevantes es la ausencia de un repositorio de movimientos de referencia universalmente aceptado por la comunidad científica que pueda ser empleado para comparar las técnicas y algoritmos de detección de caídas. En este sentido, la mayoría de los autores (al menos hasta hace poco tiempo) proponían en sus trabajos nuevas técnicas de detección de caídas y utilizaban sus propias muestras (posteriormente no publicadas) para evaluar el rendimiento de sus propuestas. Hasta fechas relativamente recientes en pocos casos se replicaban los experimentos utilizando muestras de otros estudios.

Las muestras generadas en las bases de datos empleadas en estudios sobre FDS tampoco seguían un patrón estandarizado para los experimentos, por tanto, en cada trabajo se utilizaban sensores para monitorizar a los sujetos con unas características determinadas, heurísticamente elegidas. Asimismo, la

localización de los sensores de movilidad (normalmente inerciales) era distinta en cada estudio, el número y tipología de los movimientos era diferente. Igualmente, los rasgos de los sujetos (altura, peso, edad, etc.) variaban arbitrariamente entre los estudios.

En consecuencia, cada estudio se validaba en unas condiciones diferentes y, por consiguiente, los resultados no podían ser comparados.

El objetivo de esta sección es presentar el repositorio de acceso público denominado UMAFall que fue diseñado para proporcionar a la comunidad científica un conjunto de movimientos y caídas simuladas que pueda ser utilizado para estudiar el rendimiento de sus técnicas y sistemas de detección de caídas. UMAFall fue una base de datos pionera, tratándose de uno de los primeros repositorios de movimientos que incluía trazas obtenidas a partir de la monitorización de varios sensores inerciales ubicados en diferentes posiciones del cuerpo de los sujetos participantes, así como, una amplia variedad de tipos de movimientos (básicos, deportivos, distintas clases de caídas, etc.), además de contener información sobre las características de los usuarios y los sensores. Actualmente puede considerarse como cierto referente en este ámbito de estudio. De acuerdo con tres de las principales bases bibliográficas el artículo donde se describe el repositorio (ver [53]), ha sido citado al menos 79 veces según Google Scholar [54], 60 veces basándose en las métricas de Scopus [55] y 48 en función de los datos de Web of Science [56]. Asimismo, existen otros trabajos en la literatura que referencian directamente el lugar web, Figshare, donde está alojado el repositorio. Las estadísticas de esta página indican que la web de base de datos ha sido visitada hasta ahora 12,736 veces y el repositorio descargado en 6,129 ocasiones [57].

2.2. Descripción del banco de pruebas

El sistema se compone de un *smartphone* interconectado mediante Bluetooth con cuatro motas sensoras ubicadas en diferentes partes del cuerpo de los sujetos experimentales. El conjunto presenta una tipología de estrella, en la cual, el *smartphone* actúa como nodo central. Tanto el *smartphone* como las motas sensoras capturan periódicamente las medidas de los sensores de movimiento. El *smartphone* dispone de un acelerómetro, las motas sensoras de un

acelerómetro, un giróscopo y un magnetómetro, todo ellos triaxiales. La información es recibida y guardada por el dispositivo móvil gracias a una aplicación diseñada con ese propósito. Además de la información recogida por los sensores, se adjunta información de cada experimento, así como las características de los dispositivos utilizados y de los sujetos que realizan las pruebas. El objetivo de este sistema es permitir recopilar trazas de movimientos para crear un repositorio, de manera que dichos datos puedan ser analizados posteriormente de manera sistemática y *offline*

De esta manera, este repositorio abre la posibilidad de nuevos campos de estudio en el ámbito de los FDS. Las muestras pueden ser investigadas sistemáticamente con diferentes fines, como el análisis del comportamiento del rendimiento del sistema detector según la ubicación de los sensores, de las características intrínsecas de los propios sensores o de los rasgos de los sujetos experimentales, entre otros.

2.3. Descripción de los movimientos presentes en UMAFall

Un conjunto de sujetos experimentales realiza sistemáticamente una serie de movimientos (ADL y caídas simuladas), efectuados en un entorno doméstico mientras portan el sistema descrito en la sección anterior. En una primera versión del repositorio, el rango de edad de las personas participantes se encuentra entre 18 y 55 años, siendo la media de edad de 26.9 años. La simulación de las caídas, efectuadas en un colchón amortiguador, solamente fue realizada por los sujetos menores de 50 años, por razones de seguridad.

Las motas sensoras se ubicaron en diferentes partes del cuerpo de los sujetos con bandas elásticas, concretamente, en la muñeca, pecho, cintura y tobillo. El *smartphone* se coloca en un bolsillo del pantalón (a la altura del muslo).

Los movimientos fueron guardados en ficheros en formato CSV (*Comma Separated Values*), de manera que posteriormente fueran fácilmente accesibles y procesables desde otros entornos (como Matlab).

2.4. Análisis básico de las características de las trazas de los movimientos

Como era de esperar, el patrón más relevante a la hora de identificar una caída mediante la información monitorizada por un acelerómetro es la existencia de un pico provocado por el impacto del sujeto contra el suelo. Por ese motivo, en este estudio se examina de manera básica la evolución del valor máximo y mínimo del módulo de la aceleración en diferentes tipos de ADL y caídas dependiendo de la ubicación del sensor. Los resultados ponen de manifiesto la dificultad existente a la hora de clasificar entre ADL y caídas aplicando simplemente una política basada en umbrales ya que se encuentran actividades como las deportivas que también dan lugar a picos de aceleración que pueden confundirse fácilmente con los provocados al sufrir una caída. Así mismo, ciertas ubicaciones como la muñeca o el tobillo parecen no ser recomendables para caracterizar los movimientos, ya que al encontrarse en las extremidades se producen movimientos más bruscos no relacionados con el resto del cuerpo que pueden confundirse con caídas.

2.5. Conclusiones

Este trabajo presenta UMAFall, un repositorio de movimientos (ADL y caídas) que al contrario de la mayoría de los repositorios existentes en la literatura para cada movimiento incluye las trazas de más de un sensor, en concreto se utilizan cinco dispositivos ubicados en diferentes partes del cuerpo de los sujetos, los cuales miden la aceleración, velocidad angular y la fuerza del campo magnético. Además, contiene información de las personas que realizan el movimiento y las características de los dispositivos utilizados. El objetivo es que pueda ser empleado para analizar sistemáticamente técnicas de detección de caídas, así como el comportamiento de estas dependiendo de diferentes factores como la ubicación de los sensores, rasgos de los sujetos, etc. Ha sido publicado y se encuentra abierto al público para su utilización en el repositorio *online* abierto Figshare [57]

Capítulo 3. Evaluación de la posición óptima del sensor y de la utilidad de combinar sensores

3.1. Análisis mediante algoritmos basados en umbral

Resumen del artículo [58]:

Casilari Pérez, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. Analysis of a smartphone-based architecture with multiple mobility sensors for fall detection. PLoS One 2016, 11, 1–17.

Puede encontrar una copia de esta publicación en el apéndice A.2.

3.1.1. Introducción

En los últimos años, el auge de los teléfonos inteligentes junto a su capacidad de comunicación y computación, así como la posibilidad de medir parámetros físicos gracias a los sensores que incorporan ha provocado que se desarrolle un gran número de investigaciones relacionadas con la detección de caídas, en las cuales las soluciones propuestas hacen uso de un *smartphone* como elemento principal. En estos sistemas, la ubicación de los sensores en el cuerpo de la persona que es monitorizada puede resultar ser un factor crucial en la capacidad del sistema para caracterizar los movimientos, por lo que el uso del *smartphone* como único elemento de medida puede presentar desventajas debido a que la posición más adecuada para la detección de caídas puede diferir de la posición natural de transporte de un teléfono móvil.

Por este motivo, este trabajo presenta un sistema híbrido que aprovecha las ventajas del *smartphone* y, a su vez, se interconecta con una serie de motas sensoras externas que pueden ser colocadas en diferentes partes del cuerpo.

La información obtenida con esta arquitectura (mediante un entorno de prueba con diferentes sujetos experimentales que llevan a cabo una serie de movimientos predefinidos) es analizada con detalle con el fin de evaluar el comportamiento de un sistema de detección de caídas basado en umbrales. El análisis busca encontrar el número de sensores y sus posiciones en el cuerpo de los sujetos que optimizan el rendimiento a la hora de clasificar los movimientos, así como, establecer cuál es el papel más conveniente del *smartphone* en este tipo de escenarios.

3.1.2. Metodología

3.1.2.1. Repositorio UMAFall

Para este trabajo se utilizan los datos recopilados en el repositorio UMAFall (descrito en la publicación anteriormente resumida [53]). El repositorio contiene movimientos efectuados por un grupo de sujetos experimentales que sistemáticamente ejecutaron una serie de ADL y caídas simuladas preestablecidas.

Como ya se comentó, el sistema desarrollado para crear el *dataset* se compone de un *smartphone*, que ejerce el rol de nodo central en una topología Bluetooth en estrella, al cual se interconectan cuatro motas, cada una dotada de una unidad de medición inercial o IMU (*Inertial Measurement Unit*). El *smartphone* se ubica en el bolsillo del pantalón y las motas en la muñeca, tobillo, pecho y cintura de los sujetos experimentales. Los dispositivos monitorizan las magnitudes medidas por el acelerómetro, giróscopo y magnetómetro, todos ellos triaxiales, proporcionados por la IMU.

3.1.2.2. Algoritmos de detección de caídas basados en umbrales

En este estudio se comparan cuatro algoritmos básicos de detección de caídas basados en umbrales. Con esta solución, partiendo de una determinada característica de las muestras, el algoritmo compara una o varias de las magnitudes inerciales medidas con valores umbrales predefinidos que, al ser sobrepasados (por encima o por debajo), podrán indicar al proceso de detección si se considera una caída o una ADL. En este estudio, todos los algoritmos considerados se sustentan en la información proporcionada por los

acelerómetros. Este tipo de algoritmos se fundamenta en el hecho de que, en una caída, normalmente se producen las siguientes fases: una caída libre inicial (durante la cual se desploma la medida del acelerómetro), el golpeo del cuerpo de la persona sobre el suelo o la superficie, que origina un fuerte aumento del módulo de la aceleración en un corto periodo de tiempo y, posteriormente, una fase (muy variable y difícil de simular con realismo en un laboratorio) en el que el cuerpo permanece herido o inconsciente sobre el suelo, con escasa movilidad.

La principal ventaja de este tipo de algoritmos (frente a los que son analizados por otros trabajos posteriores en esta tesis) radica en su bajo coste computacional y simplicidad, lo que puede facilitar su implementación en dispositivos vestibles de escasas capacidades de computación y memoria.

Los algoritmos considerados en este trabajo son:

- Monitorización de umbral básico: simplemente se comprueba si el módulo de la aceleración (SMV) supera un umbral máximo en algún momento para decidir si se produce o no la caída.
- *iFall* (algoritmo propuesto por Sposaro en [59]) Comprueba si el módulo de la aceleración pasa sucesivamente (en un tiempo menor de lo que marca cierta ventana de tiempo $- t_{WiFall} -$) por una bajada (caída libre) y una subida brusca (golpe). Por tanto, hace uso de un umbral inferior (SMV_u) y un umbral superior (SMV_l).
- *Fall Index* (algoritmo propuesto por Yoshida en [60]): Al igual que el algoritmo *iFall*, tiene en cuenta la existencia de una fase de disminución del módulo de la aceleración y posteriormente un pico provocado por un golpe, aunque, en lugar de utilizar dicho módulo como parámetro de observación, emplea el módulo de la diferencia de aceleración entre muestras consecutivas y, por ende, utiliza un único valor umbral.
- *PerFallD* (algoritmo propuesto por Dai en [61]): considera que existe caída si el módulo de la aceleración (SMV) y la aceleración en la dirección vertical absoluta (A_v), en un intervalo superan ciertos umbrales.

3.1.2.3. Evaluación del sistema

Los datos contenidos en el repositorio UMAFall se analizan sistemáticamente con los algoritmos anteriormente mencionados y se obtienen una serie de métricas de calidad. Para cada uno de los algoritmos se realiza un barrido del valor de los umbrales seleccionados:

- Para el algoritmo de umbral básico se realiza un barrido del único umbral utilizado en el proceso, el umbral máximo de la aceleración (SMV_{Th}).
- El algoritmo iFall posee dos umbrales, el umbral superior e inferior del módulo de la aceleración. En este caso, para calcular las métricas de calidad se realiza un barrido del umbral superior (SMV_u), mientras que se mantiene fijo el umbral inferior (SMV_l), en concreto se utiliza un valor de 0.65g, magnitud seleccionada tras diferentes pruebas iniciales con el propósito de maximizar las prestaciones del sistema.
- En cuanto al algoritmo Fall Index, como ya se comentó anteriormente, solamente hace uso de un único umbral, (FI), por tanto, las métricas de calidad se obtienen a partir de un barrido de los valores de éste.
- Por último, en el caso del algoritmo PerFallID, se dispone de dos umbrales diferentes, por un lado, el umbral superior del módulo de la aceleración (SMV_{ThP}) y por otro el umbral superior de la aceleración vertical absoluta (AV_{ThP}). En este caso, como ambas magnitudes se encuentran ligadas, la experiencia se realiza haciendo un barrido del umbral superior del módulo de la aceleración, el valor elegido para el umbral superior de la aceleración absoluta será un 75% del valor seleccionado para el otro umbral. Al igual que en el caso de iFall, este valor se selecciona tras una serie de pruebas iniciales con la intención de maximizar el rendimiento del algoritmo.

Las métricas de calidad utilizadas son tres características muy comunes en la evaluación de sistemas de clasificación binaria (esquema en el que se pueden englobar los FDS):

- Sensibilidad (Se): definida como la relación entre el número de caídas identificadas correctamente (“verdaderos positivos”) y el número de caídas total en el repositorio empleado.

- Especificidad (Sp): definida como la relación entre ADL identificados correctamente (“verdaderos negativos”) y el número total de ADL realmente ejecutadas.
- Área bajo la curva (AUC): Este parámetro se calcula a partir de la característica de operativa del receptor o curva ROC (*Receiver Operating Characteristic*), que se obtiene trazando Se frente a la tasa de falsos positivos ($1 - Sp$) para los diferentes valores del umbral de discriminación.

Como la sensibilidad y la especificidad se encuentran fuertemente ligadas (mediante una correlación inversa) y con el ánimo de definir una métrica de calidad global del sistema, se emplea la media geométrica de la sensibilidad y especificidad, que unifica en un solo valor la información aportada de ambos parámetros.

A la hora de analizar la importancia de la posición de los sensores corporales se hace uso de forma selectiva de los datos de los sensores de manera que se comprueban todas las combinaciones posibles (En el sistema se utilizan cinco sensores, por lo que se comprueban un total de $2^5 - 1 = 31$ combinaciones). Para cada combinación analizada y cada experimento probado, se identifica una caída solo si todos los sensores considerados clasifican individualmente la actividad como ocurrencia de caída, es decir, si la aplicación del algoritmo correspondiente a las señales captadas por todos los sensores contemplados produce un positivo.

3.1.3. Conclusiones

Tras el análisis llevado a cabo se concluye que:

- Un mayor número de sensores en el sistema no se traduce en una mejora del rendimiento del sistema.
- Los mejores resultados vienen dados por el sensor ubicado en la muñeca. Como se trata de una posición con alta probabilidad de sufrir movimientos bruscos, se deduce que combinarla con sensores como el pecho o la cintura puede ser favorable, ya que estos dos también proporcionan buenos resultados.

- La utilización del *smartphone* para la detección de caídas no parece ser aconsejable, ya que la ubicación natural de este tipo de dispositivo (un bolsillo), parece no ser la mejor opción para caracterizar movimientos. Sin embargo, la capacidad de comunicación y procesamiento, así como la posibilidad de informar sobre la posición mediante GPS (*Global Positioning System*), lo convierten en un dispositivo muy interesante para ser utilizado como nodo de interconexión con sensores corporales y para transmitir las correspondientes alarmas
- En cuanto a los algoritmos, *Fall Index* y umbral básico parecen superar las prestaciones de los otros algoritmos estudiados, *PerFallD* e *iFall*, por lo que la mayor complejidad de estos dos últimos no viene acompañada de una mejora significativa del rendimiento a la hora de clasificar.
- El tipo de movimientos contenidos en el repositorio de estudio repercute de forma significativa en el rendimiento del sistema al utilizar algoritmos de detección basados en umbrales. Los movimientos de tipo deportivo u otras actividades con movimientos agitados son fácilmente confundidos con caídas.
- Se concluye que los algoritmos de detección de caídas basados en umbrales no son lo suficientemente robustos para ser utilizados en un sistema de detección de caídas fiable, ya que para especificidades de al menos 95%, solamente se consiguen sensibilidades cercanas al 60% en aquellas pruebas donde existen ADL de tipología deportiva. Sin embargo, en escenarios donde se espera que no se produzcan movimientos de este tipo o similar, el rendimiento del sistema aumenta considerablemente alcanzando valores superiores al 90% de media geométrica de la sensibilidad y especificidad.

3.2. Estudio mediante algoritmos *machine learning* y análisis ANOVA

Resumen del artículo [62]:

Santoyo-Ramón, J.; Casilari Pérez, E.; Cano-García, J. Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection with Supervised Learning. Sensors 2018, 18, 1155.

Puede encontrar una copia de esta publicación en el apéndice A.3.

3.2.1. Introducción

Tal y como se mencionó en el apartado anterior, la implantación del teléfono móvil en la vida diaria de la sociedad, junto a su capacidad de comunicación, computación y medición de parámetros físicos lo ha convertido en el elemento base de diferentes estudios relacionados con la detección de caídas. Este hecho motiva el presente artículo, en el cual se estudia el mismo repositorio híbrido multisensor que en las secciones anteriores (UMAFall [53]), con la intención de analizar el comportamiento de las prestaciones de un sistema de detección de caídas cuando se combina el uso de sensores ubicados en diferentes posiciones en el cuerpo de los sujetos experimentales, esta vez utilizando algoritmos automáticos de aprendizaje supervisado. Asimismo, precedentemente se realiza un estudio sobre las “características” o “atributos” (*features*), derivados a partir de las medidas inerciales de los movimientos, que se utilizan como entrada para los algoritmos, con el propósito de determinar cuáles definen mejor los movimientos y resultan, por tanto, más adecuadas para optimizar las prestaciones del sistema de detección de caídas.

3.2.2. Metodología

3.2.2.1. Repositorio UMAFall

Al igual que en la sección anterior para el análisis sistemático del sistema detector de caídas se utiliza el repositorio de movimientos UMAFall, el cual fue creado para estos fines de estudio y se encuentra abierto al público en internet. Como ya se comentó en el resumen del artículo anterior, debido a que las trazas de los movimientos incluyen medidas de diferentes sensores posicionados en cinco ubicaciones distintas del cuerpo de los sujetos experimentales, el

repositorio resulta ideal para el estudio, ya que permite comprobar hasta $2^5 - 1 = 31$ combinaciones diferentes de sensores. Como particularidad, cabe reseñar que en este artículo se utiliza una versión actualizada del repositorio, en la cual se añadieron nuevos tipos de movimientos como aplaudir, subir los brazos o simular el proceso de contestar a una llamada con el móvil, así como la participación de 2 nuevos sujetos experimentales mayores de 65 años (hasta conformar un total de 19 voluntarios).

3.2.2.2. Algoritmos automáticos de aprendizaje supervisado

El presente estudio hace uso de algoritmos de aprendizaje supervisado. Este tipo de algoritmo precisa de una fase de entrenamiento supervisado antes de poder ser utilizado para clasificar muestras de prueba. Este entrenamiento se consigue proporcionando al algoritmo una serie de características (estadísticos que caracterizan los movimientos) que son extraídos de las muestras de entrenamiento (un subconjunto del repositorio empleado) junto con la decisión real (ADL o caída) que debería generar el detector cuando son alimentados con dichas entradas. Una vez entrenado, el algoritmo construye un modelo matemático que se puede utilizar para clasificar otros movimientos en la llamada fase de prueba (que se realiza con los movimientos del repositorio no utilizados en la fase de entrenamiento).

En este caso se utilizan cuatro de los algoritmos de aprendizaje supervisado más utilizados en la literatura: *Support Vector Machine* (SVM), *k-Nearest Neighbors* (kNN), *Naïve Bayes* (NB) y árboles de decisión o *Decision Trees* (DT). En este análisis se utilizan unos hiperparámetros fijos para cada uno de los algoritmos:

Support Vector Machine (SVM)	Kernel Lineal
k-Nearest Neighbors (kNN)	K=9 Tipo de distancia = Euclidean
Naïve Bayes (NB)	Distribución normal gaussiana
Decision Tree (DT)	Nº máximo de categorías = 10 Nº máximo de divisiones = 1

Tabla 3.1. Resumen de los algoritmos de aprendizaje supervisado utilizados junto a los hiperparámetros empleados.

3.2.2.3. Extracción de características

Uno de los aspectos claves en el campo de los algoritmos *machine learning* es la selección adecuada de las características de entrada utilizadas por los algoritmos. En este caso se proponen seis estadísticos computados a partir de las señales de acelerometría de los movimientos captadas por los sensores transportables. Todos ellos son calculados dentro de una ventana de tiempo fija (de 0.5 s) ubicada en la sección o segmento de señal de cada muestra (ADL o caída) que incluye la mayor diferencia de aceleración existente a lo largo de todo el movimiento.

Nombre	Descripción
μ_{SMV}	Media del módulo de la aceleración. Describe la media del nivel de movimiento o agitación del cuerpo.
$A_{\omega_{diff(max)}}$	Máxima variación de las componentes de la aceleración. Permite identificar la presencia de un cambio brusco en el módulo de la aceleración
σ_{SMV}	Desviación estándar del módulo de la aceleración. Describe la variabilidad del módulo de la aceleración.
μ_{θ}	Media del ángulo de rotación. Aporta información sobre los cambios de orientación del cuerpo del sujeto.
$\mu_{SMV(diff)}$	Media de la diferencia del módulo de la aceleración entre dos muestras consecutivas. Caracteriza la variación existente en el módulo de la aceleración entre dos muestras consecutivas.
μ_{Ap}	Media de la componente de la aceleración paralela al plano del suelo. Aporta información sobre los cambios de perpendicularidad del cuerpo del sujeto respecto al suelo.

Tabla 3.2. Características de entrada usadas para caracterizar los movimientos en el presente estudio.

Para el caso del estadístico de la media de la componente de la aceleración paralela al plano del suelo (μ_{Ap}) debe advertirse que se ha tenido en cuenta la disposición de los ejes de los sensores (ver Figura 3.1) en el *testbed*, en lo que se ha considerado la posición de referencia o inicial, esto es, el sujeto de pie en posición vertical y erguida, tal y como se muestra en la Figura 3.2.

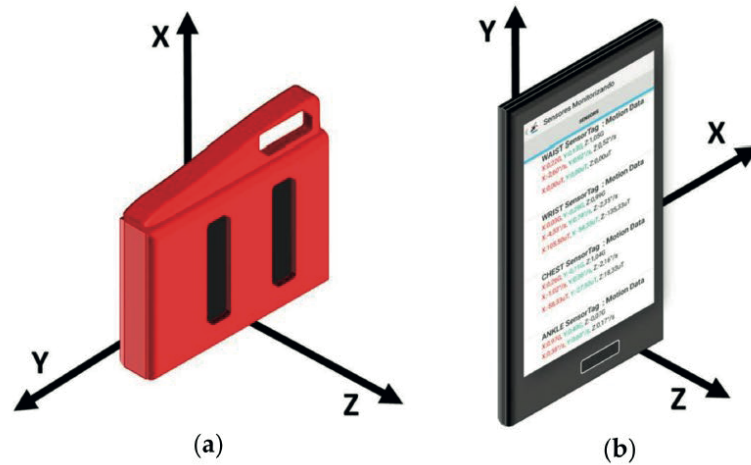


Figura 3.1. Representación del sistema de referencia espacial de los dispositivos de detección empleados en el repositorio UMAFall (los dispositivos están firmemente unidos al cuerpo de los sujetos para garantizar que el sistema de referencia no cambie durante los experimentos). (a) SensorTag (b) Smartphone.

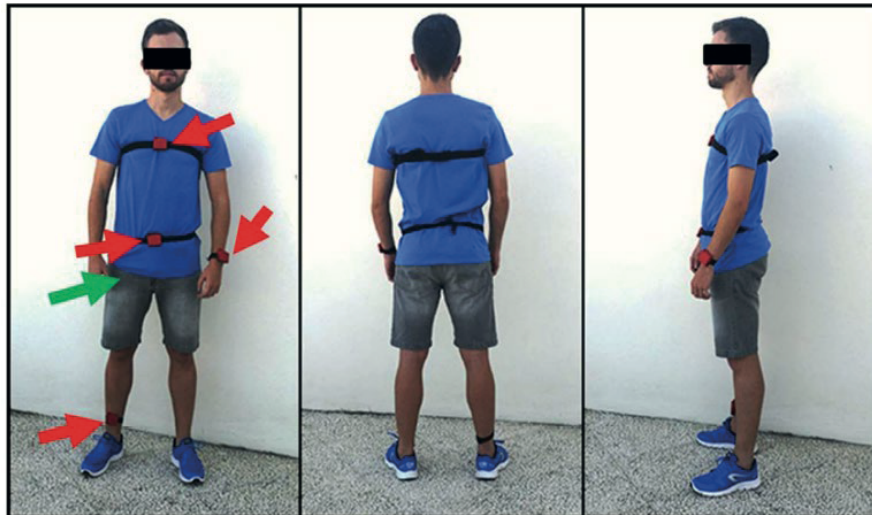


Figura 3.2. Ejemplo de la distribución de los sensores utilizados por un sujeto experimental en el repositorio UMAFall. La flecha verde indica la posición del smartphone. Las flechas rojas corresponden a los módulos SensorTag, que se unen al cuerpo del usuario mediante bandas elásticas.

3.2.2.4. Evaluación del sistema

Del mismo modo que en el artículo precedente se analizan sistemáticamente los movimientos presentes en el repositorio UMAFall consiguiéndose de este modo una serie de métricas de calidad (sensibilidad, especificidad y la media geométrica de la sensibilidad y la especificidad) que proporcionan información sobre el comportamiento del sistema. Aparte de comparar los resultados de los cuatro algoritmos *machine learning* seleccionados, el objetivo principal de este estudio es, por un lado, conocer el impacto en las prestaciones del sistema de detección de caídas respecto a las características de entrada seleccionadas para

los algoritmos, y por otro (como en el artículo anterior) averiguar cómo afecta el número y posición de los sensores al sistema.

Para el presente estudio se utiliza un diseño factorial 2^k , donde k es el número de posibles factores que pueden afectar en el proceso de detección (características de entrada o posiciones de los sensores).

Para determinar la significación estadística existente entre los resultados de los experimentos (algo que habitualmente no se considera en los estudios sobre FDS) se utilizan análisis de varianza (ANOVA), en concreto del tipo factorial. Una prueba ANOVA posibilita decidir si las medias de dos o más poblaciones diferentes difieren significativamente, de este modo, permite comprobar si diferentes tratamientos en un experimento provocan diferencias significativas en los resultados finales y, en consecuencia, evaluar la importancia de los diferentes factores que pueden alterar el funcionamiento del detector de caídas. Asimismo, el valor del error proporcionado por el ANOVA, es decir, el efecto provocado por factores diferentes a los estudiados [63] permite inferir la robustez del sistema, por un lado, ante la variabilidad introducida por el propio clasificador utilizado y por otro ante causas externas como por ejemplo la composición de los *datasets* estudiados o las características inherentes de los usuarios (altura, peso, edad, etc.).

3.2.3. Resultados y conclusiones

Tras la realización de las diferentes pruebas y el análisis de los resultados se llega a una serie de conclusiones que se resumen a continuación para cada estudio.

3.2.3.1. Impacto de la elección de características de entrada (features)

En primer lugar, se investiga qué combinación de características de entrada maximiza las prestaciones de los algoritmos. Para cada set de características las sensibilidades y especificidades globales se calculan como la media de las métricas resultantes de utilizar todas las combinaciones posibles de sensores, cada uno ubicado en un lugar diferente del cuerpo de los sujetos. Las conclusiones obtenidas son las siguientes:

- La utilización de un número mayor de características de entrada, y con ello, un aumento de la complejidad de los modelos de los sistemas de clasificación, no se corresponde con una mejora en el rendimiento del sistema de detección de caídas. De hecho, a veces, la utilización de un mayor número de características de entrada deteriora considerablemente las prestaciones. Esta situación se puede justificar en el hecho de que las *features* que proporcionan información menos relevante pueden introducir ruido perjudicial al sistema.
- La elección de una combinación de características de entrada óptima depende fuertemente del algoritmo utilizado. Sin embargo, se observan algunos patrones que se repiten:
 - Los estadísticos que con mayor frecuencia proporcionan mejores prestaciones y cuyo porcentaje de contribución en el resultado final es mayor son:
 - La desviación estándar del módulo de la aceleración (σ_{SMV}).
 - La media de la componente de la aceleración paralela al plano del suelo (μ_{Ap}).
 - La máxima variación de las componentes de la aceleración ($A_{w_{diff(max)}}$).
 - Por el contrario, se detecta que el estadístico $\mu_{SMV_{diff}}$ parece no aportar información suficientemente relevante al sistema de detección.
- Por otro lado, el ANOVA realizado muestra que existen ciertos factores diferentes a los estudiados que afectan en la variabilidad de los resultados. El algoritmo que parece ser más robusto frente a estos factores es SVM, por el contrario, kNN parece ser el más sensible.

Algoritmos	Intervalo de confianza del valor medio de $\sqrt{Se \cdot Sp}$
SVM	[73.5% - 78.5%]
k-NN	[79.9% - 85.0%]
Naïve Bayes	[67.1% - 74.5%]
Decision Tree	[80.1% - 84.9%]

Tabla 3.3. Intervalos de confianza obtenidos para cada uno de los algoritmos de aprendizaje supervisado en el análisis ANOVA para el valor medio de $\sqrt{Se \cdot Sp}$ utilizando todas las posibles combinaciones de sensores.

	μ_{SMV}	$A_{\omega_{diff(max)}}$	σ_{SMV}	μ_{θ}	$\mu_{SMV_{diff}}$	μ_{Ap}	Error
SVM	✓ 3.33%	✓ 42.31%	✓ 14.97%	– 0.70%	– 0.21%	✓ 25.81%	5.22%
KNN	✓ 0.95%	– 6.07%	✓ 21.33%	✓ 4.12%	– 1.33%	✓ 19.39%	21.25%
Naïve Bayes	– 0.98%	✓ 11.55%	✓ 7.47%	✓ 12.60%	– 2.14%	✓ 32.78%	13.48%
Decision Trees	✓ 3.71%	✓ 6.78%	✓ 41.18%	– 1.45%	– 1.16%	✓ 14.91%	12.11%

Tabla 3.4. Resumen de los resultados del análisis ANOVA. Se incluye el porcentaje de contribución a la media global del experimento de las características de entrada, así como el porcentaje de error obtenido. El símbolo (✓) indica si la característica correspondiente está incluida o no en la combinación que logra un mejor rendimiento estadísticamente significativo.

3.2.3.2. Impacto en la elección de posición y combinación de los sensores en la detección de caídas

En este segundo estudio se evalúa la importancia de la posición de los sensores de movimiento a la hora de diferenciar entre una caída y un ADL, así como si la combinación de varios sensores mejora las prestaciones del sistema. Este análisis se vuelve a realizar con cada uno de los algoritmos de aprendizaje supervisado. Las características de entrada que se tiene en cuenta en cada algoritmo son aquellas que presentaron unas mejores prestaciones en el apartado anterior.

Como en el estudio del apartado 3.1 con algoritmos basados en umbrales, cuando se emplea más de un sensor, a la hora de generar la clasificación binaria, se aplica el operador lógico 'AND', es decir, se supone que ha ocurrido una caída solo si es detectada simultáneamente por el algoritmo en todas las posiciones consideradas.

Las conclusiones extraídas se resumen a continuación:

- La localización de los sensores en el cuerpo de los sujetos tiene un efecto crucial en las prestaciones del sistema de detección de caídas.
- Las posiciones con los que se consiguen mejores resultados son el pecho y la cintura, ambas posiciones cercanas al centro de gravedad del cuerpo humano.
- El tobillo es la ubicación que parece caracterizar peor la movilidad de los sujetos, es la posición que, de forma notable, peores resultados proporciona.
- El bolsillo del pantalón no parece ser una posición recomendable para ubicar sensores de movimientos. Además de que los resultados son

peores que los obtenidos por otras posiciones como la cintura o el pecho, la variabilidad de las características del bolsillo (holgura, tamaño, posición, etc.) puede provocar un impacto negativo a la hora de clasificar. Por tanto, el comportamiento de esta posición obliga a reconsiderar el papel del *smartphone* como sensor.

- En el caso de los teléfonos inteligentes empleados en este estudio, la frecuencia de muestreo de su sensor de aceleración fue superior a la de las motas del sensor (200 Hz frente a 20 Hz), por lo que puede deducirse que una frecuencia de muestreo más alta no se correlaciona con un mejor desempeño. Este tema será tratado específicamente y con mayor detenimiento en un artículo posterior.
- La combinación simultánea de dos o más sensores en ningún caso aumenta el rendimiento del sistema, es decir, la posibilidad de incrementar la especificidad no compensa la pérdida de sensibilidad. Por tanto, el aumento de la complejidad de la red de sensores corporales, que afecta negativamente a su ergonomía y costo, no está ligado con una mayor eficiencia a la hora de clasificar.
- Una arquitectura de sensor único (con un acelerómetro unido al pecho) parece ser suficiente para maximizar la efectividad del sistema. El uso de un segundo sensor 'de respaldo' en la cintura no degrada el rendimiento mientras que podría ayudar a evitar falsos positivos en caso de cualquier anomalía que sufra el dispositivo en el pecho.
- Si se comparan los resultados obtenidos por el sensor de cintura (utilizado como sensor único) con los diferentes algoritmos, se puede concluir que las mejores prestaciones se consiguen con el algoritmo SVM. En este punto, no obstante, cabe decir que existe superposición en los intervalos de confianza obtenidos para los cuatro algoritmos estudiados. Para el caso de la SVM, el intervalo de confianza de la métrica de rendimiento se encuentra en el intervalo [95.6%-99.9%] con un valor medio de 97.75%, lo que implica valores medios mínimos para la especificidad y la sensibilidad superiores al 95,5%. Además, SVM es el algoritmo que produce el menor error (14.16%) para el análisis ANOVA. Este menor error reduce el posible impacto de otros posibles factores que no se han considerado (por ejemplo, diferencias en los patrones de movilidad o

características físicas del sujeto). En cualquier caso, el impacto de la variabilidad entre los rasgos de los sujetos será analizado en un estudio posterior.

- Las prestaciones obtenidas en posiciones como el pecho y la cintura demuestran que los cuatro algoritmos de aprendizaje supervisado utilizados son una opción interesante en el campo de la detección de caídas. En todos ellos se consiguen medias geométricas de la sensibilidad y especificidad superiores al 90%; valor muy superior al obtenido con los algoritmos basados en umbrales estudiados en el artículo anterior, siendo destacable el hecho de que, en este caso, los algoritmos sí clasifican correctamente movimientos más bruscos como los deportivos, tipología con la que existía problemas con los algoritmos basados en umbrales.

	Bolsillo	Pecho	Cintura	Muñeca	Tobillo	$\sqrt{Se \cdot Sp}$
SVM Error = 14,162%.	✓	–	–	–	–	[0,741-0,783]
	–	✓	–	–	–	[0,956-0,999]
	–	–	✓	–	–	[0,935-0,978]
	–	–	–	✓	–	[0,890-0,933]
	–	–	–	–	✓	[0,724-0,766]
	–	✓	✓	–	–	[0,918-0,961]
KNN Error = 22,857%	✓	–	–	–	–	[0,833-0,875]
	–	✓	–	–	–	[0,950-0,993]
	–	–	✓	–	–	[0,950-0,992]
	–	–	–	✓	–	[0,903-0,946]
	–	–	–	–	✓	[0,797-0,840]
	–	✓	✓	–	–	[0,926-0,969]
Naïve Bayes Error = 16,967%	✓	–	–	–	–	[0,792-0,853]
	–	✓	–	–	–	[0,922-0,981]
	–	–	✓	–	–	[0,903-0,963]
	–	–	–	✓	–	[0,863-0,899]
	–	–	–	–	✓	[0,614-0,675]
	–	✓	✓	–	–	[0,863-0,923]
Decision Trees Error = 40,620%	✓	–	–	–	–	[0,787-0,840]
	–	✓	–	–	–	[0,938-0,991]
	–	–	✓	–	–	[0,913-0,965]
	–	–	–	✓	–	[0,890-0,943]
	–	–	–	–	✓	[0,825-0,878]
	–	✓	✓	–	–	[0,896-0,949]

Tabla 3.5. Resumen de los resultados del estudio sobre la importancia de la ubicación del sensor en el cuerpo de un sujeto en un sistema detector de caídas. El símbolo (✓) indica que el sensor se tiene en cuenta a la hora de decidir; mientras el símbolo (–) identifica cuándo no se utiliza. Se han adjuntado las combinaciones donde trabaja cada sensor por separado y la combinación de sensores que mejor resultados alcanza con cada algoritmo. Se marcan en negrita aquellos sensores o combinación de sensores que consiguen medias geométricas de especificidad y sensibilidad mayores del 90%.

Capítulo 4. Comparación y análisis de repositorios

4.1. Revisión de *datasets* existentes y propuesta de clasificación de ADL

Resumen del artículo [64]:

Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.-M. Analysis of Public Datasets for Wearable Fall Detection Systems. Sensors 2017, 17, 1513

Puede encontrar una copia de esta publicación en el apéndice A.4.

4.1.1. Introducción

Tal y como se ha comentado reiteradamente en los capítulos anteriores, los repositorios de movimientos generados por sujetos experimentales constituyen la base más habitual para el estudio de sistemas y técnicas de detección de caídas, permitiendo la realización de evaluaciones y análisis sistemáticos de políticas de detección. Sin embargo, no existe un marco de trabajo estandarizado en este campo de estudio. Este hecho se refleja claramente (al menos en la fecha de publicación del artículo, en 2017) en la ausencia de un conjunto de datos de movimientos empleado masivamente por la comunidad investigadora para comparar las prestaciones de las diferentes propuestas en la literatura. De hecho, en la mayor parte de los estudios publicados hasta aquel momento, los autores validaban sus trabajos utilizando su propio banco de pruebas, definiendo ellos mismos los tipos de movimientos (ADL y caídas), y utilizando su propio grupo de sujetos experimentales. Asimismo, estos datos experimentales sobre los que se apoyaban los estudios no solían publicarse de forma abierta. Todas

estas circunstancias dificultan la reproducibilidad de las pruebas y la comparación cruzada de las técnicas utilizadas.

Sin embargo, el creciente interés científico por los sistemas FDS; propiciado por el auge de los dispositivos vestibles, el envejecimiento de la población y la problemática que conllevan las caídas en el ámbito de la salud; ha venido provocando la aparición de diferentes y sucesivos repositorios de movimientos públicos destinados a ofrecer un marco evaluativo que resuelva los problemas anteriormente mencionados.

En este artículo se revisaron y evaluaron una serie de bases de datos que contienen mediciones de ADL y caídas simuladas, teniendo en cuenta múltiples factores que intervienen en la definición de estos. El objetivo era poner de manifiesto la alta heterogeneidad existente entre los conjuntos de datos y la necesidad de normalizar y categorizar algunos aspectos para permitir que las evaluaciones de los sistemas de caídas sean comparables entre ellos.

4.1.2. Metodología de búsqueda e información básica de los repositorios existentes

La búsqueda de repositorios públicos se llevó a cabo a través de una exploración sistemática de las principales fuentes de recursos científicos (IEEE Explorer, Scopus, Ovid SP and Cochrane library) y complementada con administradores de referencias académicas como Mendeley o Google Scholar. La exploración se basó principalmente en búsquedas de cadenas de texto en el título, resumen y palabras clave de los artículos (se limita solamente a publicaciones en inglés). Los términos de búsqueda fueron “*fall detection*” combinado con “*dataset*”, “*repository*” o “*database*”.

En este estudio solamente se analizaron repositorios de movimientos creados a partir de sistemas vestibles que contengan caídas además de ADL. La búsqueda de este estudio se completó en abril de 2017.

El formato de las bases de datos encontradas es diverso, desde archivos CSV (texto plano separado por comas) a bases de datos SQL o archivos de Matlab. También existe diversidad a la hora de organizar los datos, en ocasiones toda la información es contenida en un único archivo, en otras los repositorios están

formados por una larga lista de archivos, cada uno correspondiente a un movimiento concreto ejecutado por un sujeto.

4.1.3. Análisis de los sujetos experimentales y escenarios utilizados

Todos los repositorios han sido generados por un grupo de sujetos experimentales que participan voluntariamente. Al analizar algunos rasgos y características de las personas participantes en cada base de datos se observa la heterogeneidad existente. Los parámetros estudiados son: número de sujetos, distribución por género, edad, peso y altura. A continuación, se resume algunas de las conclusiones alcanzadas:

- Existe un patrón que se repite en todos los repositorios: el grupo de voluntarios se compone principalmente de adultos jóvenes, con una media por debajo de los 30 años. En los casos donde participan personas mayores de 50 años, estos no simulan caídas y solamente se limitan a ejecutar ADL, para las que no existen riesgos de posibles lesiones. Asimismo, en todos los *datasets* se indica que los sujetos se encuentran en buen estado de salud y no comentan que ninguno de los participantes presente alguna patología que pueda afectar a sus movimientos.
- En 10 de los 12 repositorios el número de participantes hombres es superior al de mujeres.
- Los rangos de peso y altura existentes en cada uno de los repositorios son muy variados. Además, las desviaciones de altura y peso dentro del mismo repositorio también son significativas, en algunos casos superando desviaciones de 14 cm y 10 kg respectivamente.
- A pesar de que la mayoría de los repositorios incluyen información sobre los rasgos de los sujetos experimentales (sexo, altura, peso), estos no suelen emplearse en la evaluación de los sistemas de detección de caídas para discriminar muestras. En un estudio posterior, haremos uso de dichos datos para investigar la influencia de estos parámetros en el proceso de detección.

Por otro lado, otro aspecto analizado son las condiciones y el escenario donde los voluntarios desarrollan los movimientos.

- En la mayoría de los repositorios (10 de 12) los movimientos realizados están predefinidos y se siguen una serie de pautas.
- En la mayoría de los repositorios los movimientos se efectúan en escenarios “controlados” como gimnasios, laboratorios o entornos domésticos.
- En la mayoría de los casos las caídas se simulan sobre una superficie protectora (colchonetas o elementos similares).

4.1.4. Número, duración y tipología de los movimientos

Las conclusiones obtenidas en esta sección son las siguientes:

- La duración de las trazas de los movimientos es muy heterogénea entre repositorios y en algunos casos incluso entre movimientos de una misma base de datos.
- Así mismo, existe una gran diversidad en cuanto el número de muestras contenidas en cada repositorio, encontrando casos donde no se llega a la centena de muestras y otros donde se superan las 10.000 muestras.
- Por otro lado, también se observa una gran diferencia de tipos de ADL y caídas. En algunos casos los repositorios solamente contienen ADL cuyo grado de movilidad o agitación es bajo (caminar lento, sentarse o tumbarse) y otros incluyen movimientos donde se requiere un mayor esfuerzo físico como saltar o correr. Así, a la hora de evaluar las decisiones de un sistema de detección de caídas los resultados de un repositorio y otro no pueden ser comparables, ya que la clasificación de los movimientos del repositorio que contiene muestras de acciones de tipo ADL más bruscas o que impliquen movimientos más enérgicos será más compleja.
- Se propone categorizar las ADL en tres grandes grupos genéricos atendiendo a su grado de movilidad. De esta forma a la hora de evaluar un sistema de detección de caídas se puede caracterizar de una manera más concreta el escenario en el que ha sido probado. La clasificación propuesta sería la siguiente:
 - Movimientos básicos: Requieren un esfuerzo físico mínimo (sentarse, tumbarse, etc.)

- Movimientos estándar: actividades de la vida diaria que requieren cierto esfuerzo físico o una movilidad mayor (caminar, agacharse, subir escaleras).
- Movimientos deportivos: Aquellas actividades de aspecto más deportivo (saltar, correr, etc.) y una mayor exigencia física.
- Movimientos cercanos a una caída: Corresponde a tropiezos o pérdidas del equilibrio en los que no existe una caída porque el sujeto logra, mediante movimientos compensatorios, rehacer su posición sin caer. No obstante, son pocas las bases de datos que incluyen movimientos que simulan esta tipología

4.1.5. Número, características y posición de los sensores

Un sistema de detección de caídas vestible está formado por sensores que son transportados en el cuerpo para caracterizar los movimientos que el sujeto realiza. Sin embargo, no existe una pauta común en cuanto al número, características y posición de los sensores. Del estudio de los repositorios, en este ámbito, se concluye que:

- En los repositorios estudiados, la mayoría de los sistemas está compuesto por un único sensor, suele ser un acelerómetro, que es ubicado normalmente en una posición fija concreta. En algunos casos se incluye más de una posición de medición, así como, otros tipos de sensores (giróscopos o magnetómetros) que trabajan junto al acelerómetro.
- En 5 de los 12 repositorios se utiliza un sistema formado únicamente por un *smartphone* que trabaja como sensor.
- En cuanto a las características de los sensores (tasas de muestreo, rangos de medidas, resolución) utilizados también se detecta una gran variabilidad. Dichos datos no siempre están reportados directamente en la base de datos de movimientos, sin embargo, puede extraerse de la documentación o el modelo de los sensores.
- En algunos casos donde se utiliza un *smartphone* como sensor se encuentran rangos del acelerómetro muy bajos ($\pm 2 g$), valor suficiente para reconocer la orientación de la pantalla, pero posiblemente demasiado escaso para caracterizar movimientos y caídas.

- El intervalo de frecuencia de muestreo en el que trabajan los diferentes sensores también es variado. En artículos posteriores se estudia la influencia de este parámetro en el comportamiento del sistema de detección.

4.1.6. Análisis de las características básicas de la acelerometría en los repositorios

Por último, en este estudio se compara estadísticamente diferentes características de las muestras existentes en cada repositorio. Se considera únicamente las medidas tomadas por el acelerómetro triaxial (sensor común en todos los *datasets*) ubicado en la posición más repetida: la cintura o el muslo (en caso en el que no exista sensor en la cintura)

De nuevo, se comprueba una alta variabilidad entre los repositorios en los parámetros estudiados:

- Tal y como se indicaba anteriormente el rango de los sensores es un factor que puede limitar la caracterización de los movimientos, tal y como demuestran los resultados los valores de aceleración superan en gran medida el rango de $\pm 2 g$ que presentan algunos de los sensores utilizados. En el caso de las muestras obtenidas por estos sensores de rango bajo se observa que dichas medidas se saturan, ya que existe una alta concentración de medidas con dicho valor.
- Como ya se comprobó en capítulos anteriores, se detecta que existe una gran dificultad para discriminar entre ADL y caídas cuando se aplican simplemente decisiones basadas en umbrales de magnitudes básicos. Esta dificultad aumenta cuando el sensor posee un rango muy limitado.
- Por otro lado, y como cabía esperar, cuando las ADL son organizadas por tipologías de movimientos (movimientos básicos, movimientos estándar y movimientos deportivos) se comprueba que los movimientos básicos son más fácilmente reconocibles respecto las caídas y que los movimientos deportivos poseen valores para el máximo del módulo de la aceleración muy similar a las caídas, superando, en algunos casos, los valores obtenidos por las caídas.

- Estos resultados evidencian la necesidad de particularizar el análisis de cualquier sistema de detección de caídas en función de la naturaleza de las ADL con las que se ha validado el proceso de detección.

4.2. Estudio del problema la heterogeneidad de los repositorios

Resumen del artículo [65]:

Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems. J. Healthc. Eng. 2020, 2020, 1–36

Puede encontrar una copia de esta publicación en el apéndice A.5.

4.2.1. Introducción

Tras el estudio del artículo previamente resumido, ya se hizo patente la heterogeneidad existente entre los repositorios de movimientos que se encuentran publicados en la literatura. Como continuación del trabajo expuesto previamente se lleva a cabo una nueva investigación; el propósito es doble: actualizar la revisión de los *datasets* existentes (incluyendo las nuevas bases de datos que habían ido apareciendo desde 2017) y profundizar en la comparación entre los repositorios disponibles, para evidenciar, mediante un análisis más complejo, que incorpora nuevos estadísticos, la notable heterogeneidad entre los mismos.

4.2.2. Revisión y selección de repositorios públicos

Tal y como ya se comentó anteriormente para desarrollar arquitecturas de detección automática de caídas es necesario la existencia y utilización de repositorios de movimientos fiables que puedan ser utilizados para evaluar la capacidad del sistema en discernir entre ADL y caídas.

A diferencia de los primeros estudios sobre FDS, que generaban sus propios repositorios (que no eran posteriormente publicados), el uso de estos *datasets* públicos para la evaluación de los sistemas de detección de caídas se está convirtiendo cada vez más en una herramienta habitual en la literatura. No obstante, en la mayoría de los estudios solamente se utilizan uno o dos de estos

datasets. Como consecuencia, las conclusiones de la mayoría de los trabajos se basan en los resultados obtenidos cuando se utilizan un conjunto de muestras muy particular, por lo cual, dada la gran diversidad existente entre los repositorios, es posible cuestionar si las conclusiones obtenidas con un determinado conjunto de datos pueden extrapolarse a otros escenarios (distintos sujetos, diferentes movimientos, otros sensores, etc.).

Este artículo analiza sistemáticamente una serie de estadísticos, utilizados para caracterizar los movimientos, con la finalidad de comprender las diferencias presentes entre los repositorios destinados a la evaluación de FDS.

Para que la comparación sea lo más justa posible, tras revisar y tabular las características básicas de los *datasets* disponibles, se han seleccionado aquellos repositorios que poseen medidas de sensores inerciales en posiciones idénticas, en concreto, el estudio se centra en la posición de la cintura, ya que según varios trabajos es una de las posiciones más aconsejables para caracterizar la movilidad de las personas dado que se encuentra cerca del centro de gravedad del cuerpo humano y proporciona una mejor ergonomía a la hora de ubicar un sensor de manera fija [22,23]. Por otra parte, como el único factor común de todos los repositorios es que incluyen las medidas de un acelerómetro triaxial, solo se consideran las señales de acelerometría.

4.2.3. Selección de características

Las características o estadísticos utilizados en este estudio se basan principalmente en el hecho de que en el momento de producirse una caída el cuerpo experimenta varias fases inerciales (tal y como se mencionó en un capítulo anterior). En primer lugar, la persona accidentada sufre una caída libre y posteriormente un choque contra una superficie de impacto, lo que se corresponde con un descenso brusco de las componentes de aceleración seguido de un fuerte pico de aceleración respectivamente.

Para la obtención de las características, para cada movimiento ejecutado por los sujetos experimentales, a efectos de calcular los estadísticos de las medidas inerciales, se considera solamente una ventana de tiempo centrada en el punto donde se detecta la máxima variación de los componentes de aceleración. Los estadísticos analizados (que amplían en seis nuevas características las

empleadas en el artículo comentado en el apartado 3.2) se presentan a continuación.

Nombre	Descripción
μ_{SMV}	Media del módulo de la aceleración. Describe la media del nivel de movimiento o agitación del cuerpo.
$A_{\omega diff(max)}$	Máxima variación de las componentes de la aceleración. Permite identificar la presencia de un cambio brusco en el módulo de la aceleración
σ_{SMV}	Desviación estándar del módulo de la aceleración. Describe la variabilidad del módulo de la aceleración.
μ_{θ}	Media del ángulo de rotación. Aporta información sobre los cambios de orientación del cuerpo del sujeto.
$\mu_{SMV(diff)}$	Media de la diferencia del módulo de la aceleración entre dos muestras consecutivas. Caracteriza la variación existente en el módulo de la aceleración entre dos muestras consecutivas.
μ_{Ap}	Media de la componente de la aceleración paralela al plano del suelo. Aporta información sobre los cambios de perpendicularidad del cuerpo del sujeto respecto al suelo.
SMV_{max}	Pico máximo de aceleración. Caracteriza la violencia del impacto en el movimiento.
SMV_{min}	Valor mínimo de la aceleración. Representa la fase de caída libre.
γ_{SMV}	Asimetría del módulo de la aceleración. Describe la simetría de la distribución de la aceleración.
SMA	Área de magnitud de la señal. Estima la actividad física del movimiento.
E	Energía de la señal. Caracteriza los movimientos bruscos.
μ_R	Media de la función autocorrelación. Aporta información sobre movimientos inesperados.

Tabla 4.2.4.1. Características de entrada usadas para caracterizar los movimientos en el presente estudio. Todos los estadísticos se calcularon a partir de las señales de acelerometría medidas en un sensor en la cintura.

4.2.4. Comparación y discusión de los resultados

Para el estudio de las características se utilizan dos técnicas: por un lado, para captar visualmente la distribución de los estadísticos, se emplean diagramas de cajas o gráficos *boxplots* y por otro, tests de comparación múltiple, desarrollados a partir de los resultados obtenidos por análisis ANOVA al comparar los diferentes repositorios. Las pruebas se repiten con dos tamaños de ventana de observación diferentes (0.5 s y 1 s). Las conclusiones son las mismas para los dos tamaños de ventana de observación en todas las experiencias.

4.2.4.1. Comparación entre caídas y ADL

Los primeros resultados muestran un grado alto de variabilidad entre los estadísticos de las trazas de los diferentes repositorios (inter-variabilidad) e incluso entre las trazas de un mismo repositorio (intra-variabilidad).

4.2.4.2. Comparación de los diferentes tipos de ADL

La variabilidad estadística existente entre los ADL de un mismo repositorio (intra-variabilidad), resultante en la experiencia anterior, puede verse justificada por la presencia de un alto número de movimientos diferentes. Aunque todos ellos son etiquetados como ADL, el esfuerzo físico necesario para llevar a cabo su ejecución puede ser muy distinto.

Para minimizar los efectos de esta heterogeneidad en los ADL por este motivo, se propone volver a realizar el estudio, pero teniendo en cuenta la exigencia física de los movimientos. Tal y como ya se propuso en el artículo de la sección 4.1, se dividen los movimientos en varios grupos: movimientos básicos, movimientos estándar, movimientos deportivos y movimientos cercanos a una caída.

Los resultados de esta experiencia señalan que a pesar de la categorización de los movimientos sigue existiendo una alta variabilidad de los conjuntos de datos cuando se comparan los repositorios. Para los cuatro tipos de movimiento y para todas las métricas, la media de las características estadísticas de cada conjunto de datos es significativamente diferente de la calculada para, al menos, otros dos conjuntos de datos.

4.2.4.3. Comparación de un tipo específico de movimiento: Caminar

Para tratar de reducir aún más la posible heterogeneidad presente entre las muestras comparadas, se reduce el estudio a un tipo concreto de movimiento: caminar. Se ha elegido esta ADL por tratarse de un movimiento cotidiano en escenarios reales y porque se encuentra disponible en la mayoría de los repositorios.

El análisis vuelve a confirmar que existe una gran variabilidad estadística entre los movimientos de las diferentes bases de datos, incluso siendo de la misma

clase. Solamente en 3 de las 12 características, σ_{SMV} , $A_{w_{diff_{max}}}$ y SMV_{max} la media no puede considerarse significativamente diferentes de las calculadas para las otras bases de datos.

4.2.4.4. Resultados con las medidas del sensor ubicado en la muñeca

Con el fin de verificar que los resultados anteriores se repiten con las medidas de otros sensores ubicados en otras partes del cuerpo de los sujetos, se realiza un análisis con los datos del sensor de la muñeca, posición muy utilizada en muchos estudios gracias a su ergonomía, a pesar de que al encontrarse en una extremidad dificulta la caracterización del cuerpo completo de los sujetos (al menos, a efectos de detectar posibles caídas).

Se repiten las mismas pruebas realizadas anteriormente, pero a partir de los datos obtenidos con el sensor ubicado en la muñeca. Tal y como se esperaba los resultados presentan incluso una mayor heterogeneidad que los obtenidos con el sensor de la cintura.

4.2.5. Conclusiones

Tras el análisis de los resultados de este estudio se muestra que existe una diferencia estadísticamente significativa de las medias para casi todos los conjuntos de datos y características cuando se comparan los diferentes repositorios de movimientos.

Esta variabilidad de los datos podría estar justificada por la compleja interacción de un amplio conjunto de factores: la tipología y el número de actividades incluidas en el diseño del *dataset*, la metodología aplicada a la hora de ejecutar los movimientos predefinidos, los rasgos intrínsecos de los sujetos experimentales participantes, así como las características técnicas (posible sesgos o errores en las medidas, etc.) y la ergonomía de los sensores.

Como conclusión final, el estudio pone de manifiesto la ausencia de un normas o reglas que definan el procedimiento que debe llevarse a cabo para definir los bancos de pruebas experimentales en los que se generan los conjuntos de datos. Así, los trabajos de la literatura que evalúan las prestaciones de sistemas de detección de caídas sin tener en cuenta diferentes repositorios, están ignorando si su sistema funciona con el mismo rendimiento cuando se utilizan

otros conjuntos de datos y, por tanto, si es legítimo extrapolar dichas métricas de calidad a otros escenarios, para compararlas con otras técnicas de detección de caídas que han sido evaluadas con otros movimientos.

4.3. Análisis del algoritmo *machine learning* SVM con varias bases de datos y análisis ANOVA

Resumen del artículo [66]:

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Study of the Detection of Falls Using the SVM Algorithm, Different Datasets of Movements and ANOVA. In Proceedings of the IEEE IWBBIO: International Work-Conference on Bioinformatics and Biomedical Engineering; Springer: Granada, Spain, May 8-10, 2019; Vol. 11465 LNBI, pp. 415–42.

Puede encontrar una copia de esta publicación en el apéndice A.6.

4.3.1. Introducción

En los resúmenes presentados en las secciones anteriores se ha comentado la heterogeneidad de los repositorios de movimientos disponibles públicamente en la literatura, así mismo, se ha estudiado la influencia de las características de entrada de los algoritmos *machine learning* en las prestaciones del sistema de detección de caídas. Este estudio, en primer lugar, vuelve a analizar el efecto que produce la utilización de determinadas características de entrada (*features*) y se explora la posibilidad de extrapolar el aprendizaje conseguido con cierta base de datos cuando se testea con otra. Una de las mayores contribuciones del trabajo se centra en emplear el análisis ANOVA en la evaluación de las prestaciones de la *features*.

4.3.2. Descripción del sistema experimental

Se utiliza la misma metodología que la utilizada en estudios anteriores: un sistema de detección de caídas es evaluado sistemáticamente a partir de los datos contenidos en repositorios de movimientos que incluyen actividades de la vida diaria y caídas. En este caso, el sistema utiliza como algoritmo de detección el algoritmo de aprendizaje supervisado SVM. Los datos estudiados solamente provienen de sensores ubicados en la cintura de los sujetos.

Tal y como se ha comentado en apartados anteriores, los algoritmos de aprendizaje supervisado requieren de un proceso de aprendizaje a partir de una serie de muestras de entrenamiento junto a su respectiva “etiqueta” o decisión clasificatoria a tomar (ADL o caída). Posteriormente el algoritmo crea un modelo matemático que puede ser empleado para clasificar muestras diferentes a las de entrenamiento que se analizan como conjunto de test.

Como se mencionó y como corresponde en un método de *machine learning* supervisado, los datos de entrada o *features* que operan como entradas en este tipo de algoritmos son estadísticos que caracterizan los movimientos que se quieren clasificar. En este estudio, en concreto, se utilizan las características descritas en la sección 3.2 y en la tabla que sigue. La forma de calcular dichos valores adopta el mismo procedimiento que se utilizaba en dicha sección: En las trazas de acelerometría del movimiento se localiza la ventana de tamaño $t_w = 0.5$ s donde las componentes de la aceleración sufren una mayor variación y se calculan los estadísticos de los datos contenidos en dicha ventana.

Nombre	Descripción
μ_{SMV}	Media del módulo de la aceleración. Describe la media del nivel de movimiento o agitación del cuerpo.
$A_{\omega_{diff(max)}}$	Máxima variación de las componentes de la aceleración. Permite identificar la presencia de un cambio brusco en el módulo de la aceleración
σ_{SMV}	Desviación estándar del módulo de la aceleración. Describe la variabilidad del módulo de la aceleración.
μ_{θ}	Media del ángulo de rotación. Aporta información sobre los cambios de orientación del cuerpo del sujeto.
$\mu_{SMV(diff)}$	Media de la diferencia del módulo de la aceleración entre dos muestras consecutivas. Caracteriza la variación existente en el módulo de la aceleración entre dos muestras consecutivas.
μ_{Ap}	Media de la componente de la aceleración paralela al plano del suelo. Aporta información sobre los cambios de perpendicularidad del cuerpo del sujeto respecto al suelo.

Tabla 4.2. Características de entrada usadas para caracterizar los movimientos en el presente estudio.

Como en estudios anteriores, volverá a utilizarse la media geométrica de la sensibilidad (Se) y especificidad (Sp) del sistema como métrica global para caracterizar la calidad del sistema de detección. De esta forma se trata de unificar en un único parámetro ambas métricas, las cuales, como se dijo, presentan una correlación negativa entre ellas.

En este trabajo se selecciona el algoritmo SVM como núcleo de clasificación del FDS. Este algoritmo es uno de los métodos de aprendizaje supervisado más utilizados en el campo de los sistemas de detección de caídas y estudios anteriores han demostrado que el rendimiento de este algoritmo suele dar lugar a mejores prestaciones que otras técnicas de clasificación de aprendizaje supervisado y de técnicas basadas en umbrales [28,38].

4.3.3. Conclusiones

4.3.3.1. Análisis del efecto de influencia de las características de entrada en las prestaciones del sistema

En este estudio se evalúa en primer lugar cuáles de las características de entrada utilizadas permiten que el algoritmo SVM logre el mejor rendimiento al discernir caídas de ADL. Para ello, se lleva a cabo un diseño factorial 2^k , donde k corresponde al número de factores posibles (estadística de entrada seleccionada) que pueden influir en el proceso de detección. A su vez se lleva a cabo un análisis de tipo ANOVA para inspeccionar la representatividad estadística de los diferentes resultados. La prueba ANOVA permite determinar si diferentes tratamientos en un experimento provocan diferencias significativas en los resultados finales y, en consecuencia, evaluar la importancia de los diferentes factores que pueden alterar el funcionamiento del detector de caídas.

A partir del ANOVA factorial se obtiene el efecto relativo que produce la utilización de cada característica, así como, la combinación de dichas características a la media global de la métrica de calidad del sistema cuando es utilizado con cada repositorio.

Los resultados concluyen que la característica $A_{\omega_{diff(max)}}$, la cual aporta información sobre cambios bruscos de aceleración, es el parámetro que presenta un mayor efecto positivo en el sistema de detección de caídas cuando es utilizado con cualquiera de los repositorios estudiados. Por otro lado, μ_{Ap} , que describe cambios en la perpendicularidad del sujeto, también parece influir notablemente en las prestaciones del proceso de clasificación. Asimismo, la utilización de ambas conjuntamente también aporta un efecto positivo en las prestaciones. Por último, μ_{SMV} , la media del módulo de la aceleración, aunque

no presenta un fuerte efecto directo en la métrica de calidad del sistema, sí parece influir positivamente cuando se combina con las métricas mencionadas anteriormente.

Estas conclusiones concuerdan con un estudio previo [67] que presentamos en el XXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2018), y con otro estudio no publicado y que ofrecemos íntegramente (en castellano) en el Apéndice B de esta memoria. En este trabajo, además, se analizaba un número mayor de características (incorporando a tal efecto de selección de *features* o ‘fenotipado’ de series temporales mediante la herramienta software de Matlab *hctsa*).

4.3.3.2. Evaluación de la eficacia de detección de un sistema entrenado con un repositorio particular cuando se prueba con las muestras de otro

Por otro lado, una vez conocidas las características de entrada más representativas, se hace uso de ellas y se evalúa la eficacia de detección de un sistema entrenado con una base de datos particular cuando se prueba con las muestras de otro repositorio. Las características de entrada utilizadas son

$$\mu_{SMV}, A_{\omega_{diff(max)}} \text{ y } \mu_{Ap}.$$

Los resultados demuestran que los datos utilizados como entrenamiento son un elemento que afecta significativamente en las prestaciones del proceso de detección de caídas. Los mejores resultados se consiguen cuando se utilizan muestras del mismo repositorio para entrenar y testear. Sin embargo, también se comprueba que existen ocasiones donde se consiguen muy buenos resultados a pesar de utilizar muestras de diferentes bases de datos, como es el caso en el que se entrena con el repositorio Erciyes (creado por Özdemir et al. [22]) y se testea con UMAFall, consiguiendo métricas de calidad superiores al 90%, valor que no es significativamente diferente a cuando se utilizan datos de UMAFall para entrenar y testear.

Por otro lado, también se obtiene el resultado contrario, la aparición de combinaciones que deterioran las prestaciones del sistema significativamente, como es el caso de entrenar con datos de la base de datos SisFall (creado por Sucerquia et al. [25]) y entrenar con cualquiera de los otros dos repositorios

analizados. Circunstancia que coincide con las conclusiones de obtenidas en otro trabajo previo [67].

Este empeoramiento de la eficacia de los clasificadores, detectado cuando se evalúan con la técnica de *datasets* cruzados (*cross-datasets*), presenta un escenario explorado por escasos estudios, que parece apuntar a cierto sobreaprendizaje de los métodos de clasificación, que conduce a errores altos cuando los espacios muestrales empleados para la evaluación se corresponden con entornos diferentes a los usados para generar las muestras de entrenamiento.

En el Apéndice E se incluye un trabajo aun no publicado donde se realiza otro testeo cruzado de repositorios más extenso y en el que se hace uso de técnicas de detección de caídas más complejas basadas en redes neuronales convolucionales. Este trabajo se vincula a la publicación adjunta en el Apéndice A. Publicaciones, concretamente en la sección A.11. Evaluación del testeo cruzado de repositorios en sistemas vestibles de detección de caídas, la cual será enviada al congreso internacional *PErvasive Technologies Related to Assistive Environments* (PETRA 2022).

Capítulo 5. Análisis de la influencia de las características de usuarios

5.1. Análisis de la caracterización estadística de los movimientos respecto a su tipología y las características de los usuarios

Resumen del artículo: [68]

Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.M. Analysis of a Public Repository for the Study of Automatic Fall Detection Algorithms. In Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18); ACM: Berlin, Germany, September 20-21, 2018; pp. 1–10.

Puede encontrar una copia de esta publicación en el apéndice A.7.

5.1.1. Introducción

En los capítulos anteriores ya se ha hecho patente el efecto que pueden provocar en las prestaciones de los sistemas de detección de caídas factores como la posición de los sensores o la tipología de los movimientos estudiados. En este estudio se analiza la influencia que tienen las características de los sujetos que realizan las actividades, así como los factores mencionados anteriormente en la caracterización estadística de los movimientos. El objetivo principal es determinar aquellos parámetros que ofrezcan una caracterización detallada de las muestras de movilidad y, de este modo, ofrezcan una discriminación más eficiente entre caídas y ADL.

5.1.2. Metodología

5.1.2.1. Repositorio utilizado

En este trabajo se utilizan las muestras del repositorio UMAFall [53], descrito en el Capítulo 2. Como ya se comentó, las trazas de este repositorio contienen información de la aceleración de cinco ubicaciones del cuerpo de los sujetos (muslo, pecho, cintura, muñeca y tobillo). Se recuerda que en el desarrollo de esta base de datos participaron 19 sujetos experimentales cuya edad se encontraba en el intervalo de 18 a 68 años, siendo la media 31.2 años. El *dataset* incluye doce tipos diferentes de ADL y tres tipos de caídas; todos estos movimientos (excepto las caídas, las cuales no fueron ejecutadas por los sujetos mayores de 50 años) fueron repetidos al menos tres veces por cada persona participante.

5.1.2.2. Estadísticos analizados

Para la caracterización de los movimientos se emplean los siguientes estadísticos:

- El módulo máximo de la aceleración (SMV_{max})
- El módulo mínimo de la aceleración (SMV_{min})
- La máxima desviación del módulo de la aceleración ($SMV_{diff(max)}$)
- El máximo valor medio del módulo de aceleración dentro de una ventana de tiempo ($\overline{SMV}_{max}$)
- El valor medio del ángulo de rotación máximo dentro de un segundo ($\bar{\theta}_{max}$)
- El valor máximo de la media del módulo de la aceleración de las componentes no perpendiculares al suelo ($\overline{SMV}_{NV_{max}}$)

5.1.2.3. Pruebas realizadas

En primer lugar, se hace uso del método gráfico conocido como diagramas de cajas. Este tipo de representación permite analizar de manera visual la distribución estadística de los valores de los parámetros estudiados (cuartiles, percentiles, valores medios y anómalos). Mediante esta técnica se estudian todos los movimientos de manera conjunta (para cada ubicación de los

sensores). Posteriormente, para confirmar la importancia de la tipología de los movimientos en la variabilidad del comportamiento de la aceleración, se analizan los movimientos agrupándolos según el grado de movilidad (este tipo de clasificación ya fue propuesto en el Capítulo 4).

Por otro lado, se investiga la repercusión de las características de los sujetos experimentales en los estadísticos seleccionados para caracterizar los movimientos. Este tipo de estudio fue pionero en el momento de su realización, ya que, aunque muchos repositorios contenían información de los sujetos, nunca se había examinado el efecto que generaban dichos rasgos en las propiedades de las trazas de movilidad. Para este tipo de experiencia se llevan a cabo una serie de análisis de varianza de una sola vía (*one-way* ANOVA); de este modo se estudia si existen diferencias estadísticamente significativas cuando la población experimental se divide en dos grupos atendiendo a su sexo, edad, altura o peso.

5.1.3. Resultados y conclusiones

Todos los resultados ilustran la dificultad existente para discriminar entre caídas y ADL si se considera simplemente una política basada en umbrales. Los diagramas de caja incluso demuestran que ciertas actividades físicas como saltar o correr provocan valores de la aceleración superiores a los producidos cuando tiene lugar una caída.

En el momento en el que los movimientos de la vida diaria son agrupados en tres conjuntos atendiendo a su grado de movilidad (movimientos básicos, estándar y deportivos), se observa que existen divergencias significativas entre los estadísticos de dichas particiones. Por ejemplo, las caídas podrían distinguirse fácilmente respecto la mayoría de ADL catalogadas como básicas y estándar estableciendo un valor umbral. Por el contrario, la discriminación basada en umbrales para las actividades deportivas sería ineficaz. Por lo tanto, se considera que la catalogación elemental propuesta puede ayudar a establecer un marco de evaluación a la hora de estudiar las prestaciones de un determinado sistema de detección de caídas. Así, si el sistema está destinado a ser utilizado por personas mayores con algún tipo de problema de movilidad, deberá parametrizarse y probarse para optimizar sus prestaciones a la hora de distinguir

una caída respecto de movimientos de la vida diaria básicos. Por el contrario, si el sistema se prevé que sea empleado por personas sanas con patrones de movilidad más complejos (ancianos más activos, ciclistas, montañistas, antenistas, bomberos, etc.) el sistema debe ser configurando, teniendo en cuenta movimientos de mayor intensidad y actividades deportivas.

Así mismo, esta experiencia también ha confirmado que ciertas ubicaciones (en especial, el tobillo) no son recomendables para la monitorización de la aceleración, ya que la información proporcionada no caracteriza fielmente la movilidad del cuerpo del sujeto. Otras posiciones más cercanas al centro de gravedad del cuerpo humano, como el pecho o la cintura no presenta estos problemas. Estos hechos concuerdan con los resultados de los capítulos previos.

Por último, en cuanto a los efectos producidos por las características de los sujetos experimentales, se encuentran diferencias estadísticamente significativas solamente en casos muy particulares. El más llamativo se evidencia con el estadístico SVM_{min} , donde parece que el sexo, la edad y la altura (que se correlaciona con el sexo) puede determinar, en cierta medida, el patrón de aceleración de los movimientos básicos. En consecuencia, no se debe descartar dicho efecto en el rendimiento del sistema. Por esta razón a la hora de crear las bases de datos se deben seleccionar los sujetos de la manera más heterogénea posible para que ningún sector de la población se encuentre subrepresentado.

Por otro lado, se argumenta que estas diferencias también podrían estar justificadas por la presencia de observaciones atípicas en el conjunto de datos (provocadas por un comportamiento muy particular de un determinado sujeto). Por tanto, se sugiere que este estudio debe ampliarse en trabajos posteriores considerando otros repositorios, tal y como se hace en la siguiente sección.

5.2. Estudio del impacto de las características de los usuarios en las prestaciones de sistemas de detección de caídas basados en algoritmos de aprendizaje supervisado

Resumen del artículo: [69]

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A study on the impact of the users' characteristics on the performance of wearable fall detection systems. Sci. Reports 2021 111 2021, 11, 1–14.

Puede encontrar una copia de esta publicación en el apéndice A.8.

5.2.1. Introducción

Tal y como se ha explicado y comprobado a lo largo de este trabajo, la utilización de algoritmos de aprendizaje supervisado en sistemas vestibles de detección de caídas, utilizando las señales captadas por los sensores inerciales, son una opción que aporta muy buenas prestaciones. Sin embargo, debido a las dificultades de entrenar y probar este tipo de detectores en escenarios realistas, los sistemas suelen probarse a partir de conjuntos predefinidos de movimientos (ADL y caídas emuladas) ejecutados por sujetos voluntarios. En gran parte de la literatura, se utilizan muestras de los mismos sujetos experimentales tanto para entrenar como para testear.

En el presente trabajo se investiga el efecto que se produce en la calidad de la clasificación del sistema de detección de caídas al entrenar diferentes algoritmos de aprendizaje supervisado con muestras ejecutadas por sujetos que tienen rasgos físicos (altura, peso, edad, índice de masa corporal y sexo) diferentes a las de los usuarios que realizan los movimientos en la fase de testeo.

5.2.2. Definición de entorno de evaluación

5.2.2.1. Selección de repositorios

En este estudio se vuelve a hacer uso de diferentes repositorios de movimientos (ADL y caídas) públicamente accesibles por Internet.

Para definir un marco de evaluación común, se seleccionan los datos recogidos por sensores ubicados en la cintura, posición que en diferentes artículos se aconseja debido a encontrarse cerca del centro de gravedad del cuerpo humano

y que en secciones anteriores se comprobó que conseguía buenos resultados. Además, se descartan aquellos repositorios que no contienen información sobre las características de los sujetos experimentales y aquellos con menos de 400 muestras de movimientos (por ofrecer escasas posibilidades a la hora de dividir los sujetos por características manteniendo un número representativo de muestras tanto para evaluación como para entrenamiento). Tras aplicar estos criterios, se seleccionan cinco repositorios (DOFDA [70], Erciyes [22], Sisfall [25], UMAFall [53] y UP-Fall [71]).

5.2.2.2. Selección de algoritmos de clasificación

En este estudio se seleccionan 4 algoritmos de aprendizaje supervisado muy utilizados en la literatura. Las pruebas con ellos (implementados en script de bibliotecas de *machine learning* de Matlab), además, se repetirán para diferentes posibilidades de hiperparametrización:

- *Support Vector Machine* (SVM), considerándose como *kernels* alternativos para transformar los datos el tipo lineal, cúbico, cuadrático y gaussiano.
- *k-Nearest Neighbors* (k-NN), utilizando cuatro formas diferentes para calcular la distancia (Euclidea, Minkowski, Chebychev y coseno) y tres valores alternativos (5, 10 y 50) del número de vecinos (k).
- Naïve Bayes (NB), se prueba teniendo en cuenta dos posibles procedimientos para aproximar la distribución de los datos de entrada: función Gaussiana y KDE (*Kernel Density Estimation*).
- *Decision Tree* (DT). Se contempla su evaluación mediante dos tipos diferentes de estructuras, por un lado, el tipo *coarse* en el que solo se permiten hasta 4 divisiones, y, por otro lado, el tipo *fine*, que limita las divisiones hasta un máximo de 100.

Por tanto, el análisis se realiza con un total de 30 variantes de algoritmos (cuatro para SVM, doce para KNN, dos para NB y dos para DT).

5.2.2.3. Selección de características de entrada

En este estudio se utilizan como características de entrada, todas las comentadas en el apartado 4.3 (aquí se denominan como de “elección propia”), las cuales fueron seleccionadas tras un análisis de la literatura ya que definen

propiedades dinámicas del cuerpo humano. Igualmente, los algoritmos se probaron con las 12 características que resultaron ser más representativas y fueron extraídas por el paquete de MATLAB *hctsa* (*Highly comparative time-series analysis*). Este software provee una herramienta que analiza miles de características de series de tiempo para agrupar y extraer las más significativas con el objeto de proporcionar una representación parsimoniosa (“de baja dimensión”) de los datos. Para un estudio más detallado de la extracción de estas características, véase el extenso análisis complementario presentado en el Apéndice B.

Los estadísticos de elección propia se calculan a partir de la información del acelerómetro comprendida en la ventana de tiempo $t_w = 2\text{ s}$ centrada alrededor del valor máximo de aceleración del movimiento de cada traza (es decir, en el intervalo donde, caso de corresponderse el movimiento con una caída, cabe situar el impacto contra el suelo).

Por otro lado, con independencia del set de características empleado, los datos de entrada son estandarizados usando una normalización de tipo *z-score*.

5.2.2.4. Métricas de calidad

Del mismo modo que en los estudios anteriores se utiliza como métrica de calidad global la media geométrica de la sensibilidad y la especificidad, $\sqrt{Se \cdot Sp}$. La sensibilidad (*Se*) define la proporción de caídas que son identificadas correctamente y la especificidad (*Sp*) la fracción de movimientos ADL clasificados adecuadamente.

5.2.3. Resultados

Se evalúa el impacto que tienen los rasgos físicos de los sujetos en la capacidad de clasificación de los algoritmos de aprendizaje supervisado. Los resultados se describen en los siguientes apartados.

5.2.3.1. Resultados de una distribución equilibrada o ideal de las muestras

En primer lugar, se evalúan las prestaciones de los diferentes algoritmos de aprendizaje supervisado y sus diferentes configuraciones con subconjuntos de entrenamiento y pruebas que contienen muestras de todos los tipos de

movimientos y sujetos. El objetivo es establecer una prueba de referencia para un caso ideal o equilibrado (en el que los sujetos que participan en el entrenamiento y la evaluación son los mismos).

En esta prueba se emplea validación cruzada, en concreto *5-fold*, siguiendo la regla típica 80-20, es decir, se utiliza el 80% de las muestras para entrenar y el resto para testear en cada iteración.

Tras un análisis de los resultados se detecta cierta predisposición a que el algoritmo SVM y sus diferentes configuraciones, y KNN en menor medida, den lugar a unas prestaciones superiores al resto de algoritmos. Del mismo modo, las 12 características de propia selección (las cuales tienen una clara interpretación física) producen un mejor comportamiento en el sistema que las estadísticas obtenidas mediante la prueba masiva de estadísticas (*hctsa*).

Por último, de este análisis masivo con cinco *datasets*, se concluye que los algoritmos de aprendizaje supervisado son capaces de dar lugar a unas prestaciones aceptables, en algunos casos superando medias geométricas de la sensibilidad y la especificidad superior al 97%.

5.2.3.2. Resultados de una distribución basada en rasgos de los sujetos

En este caso, se repite la metodología del análisis anterior pero esta vez los subconjuntos de datos se dividen en función de un rasgo personal específico (edad, altura, peso o índice de masa corporal). Así, las muestras generadas por cierto sujeto experimental solamente estarán incluidas en una de las particiones (entrenamiento o test). Para fines comparativos (y buscando analizar si las posibles divergencias de resultados no tienen relación con el criterio de división), también se lleva a cabo una prueba de los clasificadores en la que los usuarios se reparten de manera aleatoria entre los dos subconjuntos.

El objetivo de este estudio es determinar si existen diferencias entre los resultados de la prueba donde se utilizaba una distribución equilibrada (prueba de referencia obtenida en el apartado anterior) y el resto de las distribuciones. De los resultados se colige que:

- Se observa cierta tendencia a que entrenar el sistema con individuos de mayor altura y peso y testear con sujetos más bajos y de menor peso

genera mejores resultados que cuando se realiza el proceso contrario. Esta situación puede explicarse por el hecho de que los sujetos de mayor altura y peso provocan patrones de movilidad más reconocibles.

- El índice de masa corporal (como criterio divisorio de los sujetos) no parece influir significativamente en las prestaciones del sistema.
- Las pruebas donde se estudia la edad se ven limitadas por el hecho de que en la mayoría de los repositorios no existen sujetos experimentales edad superior a 50 o estos no simulan caídas. Sin embargo, con el repositorio Sisfall se puede realizar un análisis preliminar. Se observa que cuando se entrena con personas jóvenes y el test se efectúa con muestras obtenidas de los movimientos de personas mayores de esa edad se puede conseguir especificidades altas pero la sensibilidad es baja (30,67%). Este resultado podría evidenciar el hecho de que las caídas en las personas mayores siguen patrones de movilidad diferentes que las sufridas por personas más jóvenes. Estos resultados preliminares pueden cuestionar la utilidad de evaluar sistemas de detección de caídas con bases de datos generadas a partir de caídas medidas a partir de un grupo de voluntarios compuesto íntegramente por personas jóvenes.
- La división por sexo no parece afectar significativamente en la calidad de clasificación del sistema. El único patrón destacable es que, en cuatro de los repositorios utilizados, el entrenamiento con hombres parece dar lugar a métricas de calidad ligeramente más altas que las obtenidas cuando el clasificador es entrenado por muestras ejecutadas por mujeres. Este hecho puede estar ligado a la correlación que tiene el género con el peso y la altura.
- En resumen, las características físicas, exceptuando el criterio de la edad, parecen no ser un factor crucial que perturben significativamente las prestaciones de los detectores de caídas y por tanto no pueden considerarse un elemento clave a la hora de caracterizar los movimientos.

5.2.3.3. Resultados de una distribución basada en la naturaleza de los movimientos

Para finalizar este estudio, en este último experimento se evalúa el sistema de detección cuando se emplean distintos tipos de movimiento en la fase de

entrenamiento y testeo. Los movimientos se dividen en las subcategorías dependiendo de su grado de esfuerzo físico, que ya se propusieron en el apartado 4.1:

- Movimientos básicos: Acciones cotidianas que requieren un bajo esfuerzo físico (acostarse, sentarse, llamar por teléfono, etc.).
- Movimientos estándar: Acciones cotidianas que requieren de un grado mayor de movilidad (caminar, subir escaleras, recoger algo del suelo, etc.).
- Movimientos deportivos tales como correr, saltar, trotar.
- Movimientos cercanos a una caída: Acciones en las que el sujeto pierde el equilibrio y tropieza, pero no llega a caer.

A la hora de evaluar, una subcategoría es excluida de la partición de entrenamiento y es utilizada exclusivamente para testear. Tanto en el entrenamiento como en el testeo se incluyen también caídas, en una proporción 80/20.

Los resultados muestran que la tipología de los movimientos utilizados como entrenamiento tiene una gran influencia en las prestaciones del sistema (mucho mayor, en todo caso, que las características de los usuarios empleados en ambos conjuntos –entrenamiento y test–).

En el caso de los movimientos básicos, tienden a ser reconocidos correctamente por el clasificador, a pesar de no haberlo utilizado en la fase de entrenamiento, por lo que el rendimiento del sistema no varía significativamente.

Sin embargo, no entrenar con movimientos con un mayor grado de movilidad como los deportivos, provocan un drástico empeoramiento del rendimiento del sistema a la hora de clasificarlos, ya que en la clasificación son confundidos con caídas, dando lugar a una brusca degradación de la especificidad (en algunos casos baja a valores del 3%). Por tanto, estas conclusiones apuntan a la gran dificultad existente para discernir entre caídas y movimientos de alta intensidad cuando el sistema no es entrenado con este tipo de movimientos.

Capítulo 6. Evaluación de clasificadores de clase única

Resumen del artículo [72]:

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A Study of One-Class Classification Algorithms for Wearable Fall Sensors. Biosensors 2021, 11, 284.

Puede encontrar una copia de esta publicación en el apéndice A.9.

6.1. Introducción

Tal y como se ha procedido en los análisis de los apartados anteriores, la metodología que se suele llevar a cabo para evaluar un sistema de detección de caídas es estudiar la capacidad de clasificación de la arquitectura a partir de la utilización de repositorios de movimientos.

Uno de los hechos más discutido de este procedimiento es la validez de emplear, en este tipo de base de datos, caídas simuladas por voluntarios jóvenes en lugar de caídas reales de adultos mayores (el público principal de los FDS). Cabe destacar que queda fuera del objetivo de este estudio discutir la validez de este procedimiento. Sin embargo, la combinación de esta circunstancia y la dificultad inherente de entrenar el sistema con caídas reales (y más aún caídas de personas mayores) impulsa la realización de este trabajo con el propósito de estudiar una alternativa que no requiera del uso de este tipo de movimientos “fingidos”. En particular, el análisis se centra en evaluar la eficacia que presentan los algoritmos de detección de anomalías o *one-class* en el ámbito de la detección de caídas.

6.2. Metodología

6.2.1. Algoritmos de detección de anomalías

Los algoritmos de detección de anomalías o clasificadores OCC (*one-class classifiers*) conforman un tipo de algoritmo de aprendizaje supervisado que partiendo de un conjunto de datos de una determinada clase (única) es capaz de realizar una clasificación binaria cuando se le introducen nuevos datos. Los clasificadores OCC poseen la capacidad de distinguir entre datos del mismo tipo (básico o común, en este caso ADL) con el que ha sido entrenado respecto de otros diferentes, a los que detectará como anomalías o muestras *outliers*. Los OCC están especialmente indicados en situaciones donde se desea detectar elementos o patrones muy poco frecuentes a partir de una muestra que solo cuenta con elementos de un tipo común. Su principal ventaja es que su entrenamiento no exige contar con patrones de esa clase anómala (normalmente difíciles de obtener) y pueden quedar configurados a partir de un subconjunto de patrones que solo incluyen muestras de elementos o eventos comunes. Durante la fase de test o de funcionamiento, esas anomalías se detectarán a partir de la posible divergencia (normalmente fijada por cierto parámetro umbral) con respecto a los patrones de entrenamiento o de un modelo derivado de los mismos.

En este artículo se evalúan los cinco algoritmos de detección de anomalías (muy utilizados en el ámbito de las aplicaciones de *machine learning*) cuando se aplican a detectar caídas a partir de un conjunto de ADL y caídas de repositorios públicos. Estos algoritmos y sus correspondientes hiperparámetros son los que se indican en la Tabla 6.1.

Nombre del Algoritmo	Hiperparámetros	Valores
Autoencoder	Nº neuronas capa oculta	6, 10, 12, 15
	Función de transferencia del codificador y decodificador	Logistic Sigmoid $f(z) = \frac{1}{1 + e^{-z}}$
	Nº máximo de epoch	1000
	Coefficiente para el regulador de peso en la función coste	0.001
	Función de pérdida	Error cuadrático medio ajustado para entrenar un codificador automático disperso
Modelos de Mezcla Gaussiana (GMM)	Tipo de covarianza	Diagonal
	Nº de componentes	3, 5 y 7
Parzen Probabilistic Neural Networks (PPNN)	Función Ventana	$f(x) = e^{x-1}$
One-Class K-Nearest Neighbour (OC-KNN)	Tipo de distancia	Euclidean, Minkowski, Chebychev, Cosine
	Nº de vecinos	5, 10, 50
One-Class Support Vector Machine (OC-SVM)	Kernel	Linear, cubic, quadratic, medium gaussian

Tabla 6.1. Tabla resumen de los algoritmos one-class utilizados en el presente estudio junto con los hiperparámetros utilizados

6.2.2. Selección de características

Al igual que en los estudios previos, a la hora de caracterizar los movimientos con estadísticos que alimenten los OCC, se utilizan los dos sets de características descritos en el apartado 4.3: por un lado, doce características con claro significado físico (que denominamos “elección propia”), seleccionadas tras un análisis de la literatura, las cuales definen propiedades dinámicas del cuerpo humano; por otro lado, el set de características extraído mediante prueba masiva de estadísticas a través del fenotipado proporcionado por la herramienta *hctsa* de Matlab (*highly comparative time-series analysis*).

En el caso del conjunto de características de elección propia para cada muestra se calculan las características sobre una ventana $t_w = 2 s$, centrada en el valor

de máxima amplitud del módulo de aceleración del movimiento (SMV_{Max}), técnica ya empleada en algún trabajo anterior. Se utilizan las dos combinaciones de características de este tipo que resultaron ser más aconsejables en el estudio realizado en el apartado 4.3.1. Una se compone de las doce características (μ_{SMV} , $A_{\omega_{diff(max)}}$, σ_{SMV} , μ_{θ} , $\mu_{SMV(diff)}$, μ_{Ap} , SMV_{Max} , SMV_{Min} , γ_{SMV} , SMA , E y μ_R) y la otra de siete de ellas ($A_{\omega_{diff(max)}}$, σ_{SMV} , μ_{θ} , μ_{Ap} , SMV_{Max} , γ_{SMV} y E).

En el caso del set de características extraídas mediante la herramienta *hctsa*, se utilizan las 12 características que resultaron más representativas al dividir las mismas en 12 clústeres que agrupan aquellas características con comportamientos similares cuando se analizaba el repositorio SisFall (uno de los más empleados por la literatura).

6.2.3. Selección de repositorios

Al igual que en los capítulos preliminares, para evaluar el sistema se vuelve a utilizar diferentes repositorios de movimientos accesibles por Internet. En este caso se utilizan los siguientes nueve *datasets*: DLR [73], DOFDA [70], Erciyes [22], FallAID [29], IMUFD [38], KFall [74], SisFall [25], UMAFall [58] y UP-Fall [75]. Asimismo, se hace uso de los datos recogidos por el sensor ubicado en la cintura, por la misma razón que en las secciones anteriores, es una posición aconsejable para caracterizar el movimiento de una persona ya que se encuentra cercano al centro de gravedad.

6.2.4. Sistema de prueba y evaluación

Con el propósito de evaluar el comportamiento de los algoritmos de detección de anomalías a la hora de ser utilizados para la detección de caídas, en este estudio, se realizan tres pruebas.

Por un lado, se realiza una evaluación del sistema utilizando la técnica de validación cruzada *k-fold* (en concreto con $k = 5$) cuando se reparten las muestras de entrenamiento y estudio de manera equitativa (*fair*), esto es, se intenta que en cada partición exista un número similar de muestras del mismo tipo (deportivas, actividades básicas, etc.) y de todos los usuarios. Con esta prueba se pretende simular un escenario ideal, donde el sistema ha sido entrenado con muestras del mismo tipo que las que se emplearán en la fase de

test y de los mismos sujetos que posteriormente utilizan el sistema. Los resultados así obtenidos funcionarán como datos de referencia. En esta prueba se realiza un barrido del parámetro utilizado como umbral en cada algoritmo para detectar si una muestra es anomalía o no.

Por otro lado, se pretende estudiar la conducta que presenta el sistema al testear con actividades ADL cuya tipología difiere de los movimientos con los que ha sido entrenado. Para ello, se reparten los movimientos presentes en cada repositorio por categorías, dependiendo del esfuerzo físico que requiere su ejecución (de nuevo, distinguiendo entre movimientos básicos, estándar, deportivos y cercanos a caídas), tal y como ya se propuso en apartados anteriores. Posteriormente se testea cada categoría junto a todas las caídas entrenando con el resto de los movimientos de la vida diaria (ADL). El valor de los umbrales que se utilizan en esta experiencia son los que obtuvieron mejores resultados en la prueba de referencia.

Por último, se analizan los beneficios de utilizar *ensemble learning*, método que combina varios resultados de diferentes clasificadores para tomar la decisión final de clasificación.

6.2.5. Métricas de calidad

Como métrica de calidad global vuelve a utilizarse la media geométrica de la sensibilidad y especificidad del sistema, ya que unifica en un solo parámetro ambas métricas capaces, por su parte, de describir la capacidad del sistema para detectar caídas y evitar falsas alarmas, respectivamente.

Así mismo, en la prueba de referencia o equilibrada, al realizarse el barrido de umbrales es posible obtener también el área bajo la curva ROC, conocida como AUC (*Area Under Curve*) una mayor área bajo la curva indicará una mayor cercanía, al comportamiento de un detector ideal (que presenta una AUC igual a 1).

Por otro lado, para medir la estabilidad del sistema respecto al valor del umbral, se ha generado una métrica denominada *Threshold sensivity* que mide la pendiente existente en el punto de mayor media geométrica de la sensibilidad y la especificidad de la curva resultante de representar dicha métrica respecto el

valor del umbral. Un valor cercano a cero indicará una gran estabilidad. Este aspecto es normalmente obviado en la literatura sobre caídas (y en el empleo en general de los detectores de clase única), sin embargo, resulta de gran relevancia práctica para evaluar la robustez de estos clasificadores de clase única.

Los resultados obtenidos en esta métrica no están presentes en el artículo publicado (por razones de extensión) y se presentan en el repositorio que ha sido creado con el objetivo de ser utilizado como anexo de esta tesis [76].

6.3. Resultados

6.3.1. Prueba equilibrada

En primer lugar, cabe destacar que el sistema proporciona valores de la media geométrica de la sensibilidad y la especificidad superiores del 90% con todos los repositorios, e incluso supera en la mayoría de ellos el 95%, haciendo uso de alguno de los algoritmos e hiperparámetros analizados.

Por otro lado, examinando los resultados, se puede detectar un patrón común: las características de entrada que parece dar lugar a un mejor rendimiento del sistema son las pertenecientes a la combinación de características de elección propia, donde se han seleccionado las siete *features* que se determinaron más relevantes. Este hecho se pone de manifiesto principalmente con los algoritmos autoencoder, GMM y OC-KNN. Sin embargo, el comportamiento resulta más evidente con KNN, este hecho puede estar fuertemente relacionado con que este algoritmo es bastante sensible a las características irrelevantes, las cuales provocan un ruido importante a la hora de clasificar, por lo que al reducir las características que menos información aporta parece influir positivamente.

Asimismo, se observa que dependiendo del repositorio estudiado los algoritmos que consiguen un mejor rendimiento del sistema varían. No obstante, puede afirmarse que existe una tendencia a que el algoritmo OC-KNN sea el algoritmo que mejores métricas aporta (con 5 de los 9 repositorios estudiados), constatándose que además la utilización de distancia tipo coseno favorezca este rendimiento. Además, la desviación típica existente entre las diferentes pruebas *k-fold* con este tipo de algoritmo suele ser baja, cercana al 1%. El segundo

algoritmo OCC que, en media, mejor se comporta es el OC-SVM, el cual resulta ser el clasificador con mejores resultados para los tres repositorios restantes.

GMM, autoencoder y especialmente PPNN son los algoritmos que peores prestaciones consiguen: con medias geométricas de la sensibilidad y especificidad (en algún repositorio) por debajo de 0.90, además de una mayor inestabilidad y variabilidad, tanto en la desviación obtenida en el 5-Fold como entre *datasets*.

Estos resultados también ponen en evidencia la importancia de la elección de un valor umbral personalizado y adecuado dependiendo de las propiedades inherentes de las muestras. El valor del umbral con el que se obtienen los mejores rendimientos varía sustancialmente de unos repositorios a otros.

6.3.2. Estudio del impacto de las categorías de movimientos existentes en el set de entrenamiento respecto el rendimiento del sistema

Al igual que ocurría en la prueba anterior se observa que la combinación con las siete características del set de características de elección propia proporciona en la mayoría de los casos mejores resultados que las otras dos combinaciones estudiadas. En el caso de las características del tipo *hctsa* existe una gran degradación de los resultados, no consiguiendo obtener valores superiores al 85% en la mayoría de los casos.

También vuelve a repetirse que el algoritmo KNN parece ser el más eficiente en este experimento, es el algoritmo que en la mayoría de los casos arroja mejores rendimientos, en concreto cuando se utiliza como tipo de distancia el coseno.

En cuanto al hecho de entrenar con una serie de categorías de movimientos y testear otro tipo diferente se produce una degradación de la calidad de clasificación. Este hecho se intensifica en aquellos repositorios que incluyen un número mayor de movimientos diferentes como puede ser SisFall o UMAFall.

En cuanto a las diferencias de rendimiento provocadas como consecuencia de testear diferentes categorías de movimientos, se detectan varias conductas:

- El conjunto que presenta mayores problemas es aquel que agrupa los movimientos deportivos. Así, gran parte de las pruebas en las que se testea con este tipo de categoría (sin incluir muestras de la misma en los patrones

del entrenamiento) la métrica de calidad se encuentra entre 75% y valores incluso menores al 50%.

- Otra combinación que provoca que el sistema se degrade se da cuando los movimientos básicos son empleados solamente en la fase de test de los OCC y no se entrena con ellos. Al tratarse de los movimientos menos enérgicos, la ausencia de estos en el entrenamiento provoca que sean detectados como anomalías y con ello la aparición de un número mayor de falsas alarmas.
- Por el contrario, los movimientos incluidos en el tipo movimientos estándar parecen ser el tipo de muestras que menos influye en la efectividad del sistema. Este hecho parece deberse a que, al encontrarse en un punto intermedio de intensidad, entrenar con movimientos de menor intensidad (movimientos básicos) y de mayor intensidad (deportivos) da información suficiente al sistema para no considerarlos anomalías.
- En cuanto a las pruebas en los que se testea y no se entrena con movimientos clasificados como cercanos a caídas (reportados en el repositorio que ha sido creado como anexo de esta tesis [76] y no en el artículo), presentes en solo algunos de los repositorios (Erciyas, FallAIID e IMUFD), se observa una gran dificultad, en la mayoría de los casos por parte del sistema para diferenciarlos de las caídas, como consecuencia de que no existen otros movimientos similares en ninguno de los otros grupos con los que se entrena.
- En lo que se refiere al tipo de movimiento que incluye muestras obtenidas dentro de un vehículo en movimiento (véase el anexo [76]), solo presente en FallAIID, se observa que el sistema obtiene altos rendimientos, superando incluso al rendimiento obtenido en la prueba de referencia (5-Fold – Fair). Parece que la aceleración vinculada al propio vehículo no afecta de manera considerable en la clasificación y la utilización del resto de tipos de movimientos es suficiente para no considerarlo anomalías.

6.3.3. Estudio del beneficio de *ensemble learning*

Los métodos de *Ensemble learning* son estrategias de combinación de modelos de *machine learning* para mejorar capacidad de clasificación y como consecuencia sus prestaciones.

En esta prueba se evalúa si su utilización mejora los resultados obtenidos en la prueba equilibrado o referencia. En concreto, se calcula la decisión de clasificación cuando se aplica una votación por mayoría simple de tres clasificadores. Los tres clasificadores elegidos para cada repositorio serán aquellos algoritmos, junto con los hiperparámetros correspondientes, que proporcionaron mejores resultados (en algunos casos los mejores algoritmos incluyen dos instancias del mismos OCC, pero con hiperparámetros distintos). En el apéndice C se reportan resultados similares cuando se fuerza que la votación se produzca con tres OCC distintos (de los cinco probados) que arrojan mejores resultados, así mismo, en este apéndice se aplican las mismas técnicas y metodología con algoritmos de aprendizaje supervisado de más de una clase.

Las conclusiones son que el uso de esta técnica mejora el rendimiento del sistema en seis de los nueve conjuntos de datos analizados (en varios casos se alcanza un valor de $\sqrt{Se \cdot Sp}$ cercano a 0.99.) Mientras que con un solo repositorio (DLR) la aplicación de la técnica de votación reduce la efectividad del proceso de clasificación binaria.

Capítulo 7. Evaluación de sistemas de aprendizaje profundo. Análisis de la frecuencia de muestreo

Resumen del artículo [72]:

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A Study of the Influence of the Sensor Sampling Frequency on the Performance of Wearable Fall Detectors.

Pendiente de ser enviado a la revista Measurements.

Puede encontrar una copia de esta publicación en el apéndice A.10.

7.1. Introducción

Tal y como se ha expuesto a lo largo de este trabajo, en el ámbito de los sistemas vestibulares de detección de caídas existe una alta heterogeneidad en las características de los sensores utilizados para la monitorización de los movimientos de los sujetos. Este hecho puede suscitar interrogantes sobre su fiabilidad y comportamiento en distintos escenarios. Es por ello, por lo que el estudio del comportamiento del rendimiento del sistema respecto dichas características se convierte en un campo de interés científico.

En el caso de este trabajo, la característica en la que se ha enfocado es la frecuencia de muestreo. En la bibliografía el rango de frecuencias de muestreo utilizado en sistemas de detección de caídas es amplio, oscilando desde repositorios que se limitan a una frecuencia de muestreo de 5 Hz, como el *dataset* Graz UT OL [77], a otros, como FallAID [29], donde se emplean tasas superiores a 200 Hz. El objetivo de este estudio es analizar la conducta de un sistema de detección de caídas basado en aprendizaje profundo respecto a la frecuencia de muestreo utilizada por los acelerómetros.

7.2. Metodología

7.2.1. Selección de repositorios

Del mismo modo que en los capítulos precedentes, la evaluación del sistema se vuelve a llevar a cabo mediante la utilización de diferentes repositorios de movimientos accesibles por Internet.

En este caso se eligen los siguientes repositorios y cuyas características se resumen en la siguiente tabla:

Repositorio	Ref.	Número de acelerómetros	Posiciones de los acelerómetros	Frecuencia de muestreo (Hz)
<i>AnkFall</i>	[78]	1	Tobillo	50
<i>Cogent Labs</i>	[79]	2	Pecho, muslo	100
<i>DOFDA</i>	[70]	1	Cintura	33
<i>Erciyes</i>	[22]	6	Pecho, cabeza, tobillo derecho, muslo derecho, muñeca derecha, cintura	25
<i>FallAIID</i>	[29]	1	Cintura	238
<i>IMUFD</i>	[38]	7	Pecho, cabeza, tobillo izquierdo, muslo izquierdo, tobillo derecho, muslo derecho, cintura	128
<i>KFall</i>	[74]	1	Cintura	100
<i>MobiAct</i>	[80]	1	Bolsillo del pantalón	88
<i>SisFall</i>	[25]	1	Cintura	200
<i>tFall</i>	[39]	1	Bolso de mano y bolsillo	50
<i>TST</i>	[81]	2	Muñeca derecha, cintura	100
<i>UMAFall</i>	[53]	5	Tobillo*, pecho*, bolsillo del pantalón**, cintura*, muñeca*	200** y 20*
<i>UP-Fall</i>	[71]	5	Tobillo, cuello, muslo derecho, cintura, muñeca	18.48

Tabla 7.1. Resumen de la frecuencia de muestreo y ubicaciones de los acelerómetros correspondientes a los repositorios de movimientos utilizados en el presente estudio.

7.2.2. Estudio del comportamiento del sistema respecto la frecuencia de muestreo

Este estudio se basa en la evaluación de las prestaciones de un sistema de detección de caídas a partir de los datos de los repositorios comentados anteriormente. Dicho análisis se repite para cada repositorio utilizando diferentes frecuencias de muestreo. Con ese objetivo, se realiza un barrido de frecuencias, llevando a cabo un diezmado de los datos. Los resultados finales se agrupan atendiendo a las posiciones de los sensores.

Para confirmar la relevancia de los componentes de baja frecuencia en la caracterización de las actividades, se repite el experimento utilizando las muestras sin submuestrear pero filtradas previamente por un filtro paso bajo. Se repite la experiencia con varias frecuencias de corte (5, 10 y 20 Hz). Al implementar el filtro, la frecuencia de corte seleccionada debe ser menor que la mitad de la frecuencia de muestreo (f_s) para que siempre se cumpla el teorema de Nyquist.

Así mismo, se efectúa un estudio del espectro de potencia de la aceleración de los movimientos de cada repositorio.

7.2.3. Configuración de la red neuronal convolucional

En este capítulo, como núcleo de la detección de caídas se utilizan redes neuronales convolucionales o CNN (*Convolutional Neural Networks*). Estas constituyen un tipo de red neuronal artificial con aprendizaje supervisado cuyo funcionamiento es similar al córtex visual humano. La principal ventaja de las redes neuronales convolucionales respecto las redes neuronales artificiales o ANN (*Artificial Neural Networks*) clásicas es la reducción del número de parámetros y con ello su complejidad [82] y, sobre todo, su capacidad de tomar directamente las señales a clasificar y aprender autónomamente (mediante lo que se conoce como aprendizaje profundo o *deep learning*) las características que permiten identificar las clases en las que se dividen los patrones, evitando de este modo la selección arbitraria de características de entrada de otros métodos de *machine learning* supervisado.

La configuración de los hiperparámetros de la red convolucional ha sido seleccionada tras un análisis (o *grid search*) exhaustivo previo, no publicado y que presentamos íntegramente (en castellano) en el Apéndice D de esta memoria.

Para cada muestra se introducen como entrada de la CNN los datos, con la frecuencia de muestreo correspondiente (conseguida si es preciso mediante diezmado), de los tres ejes del acelerómetro contenidos en una ventana de tiempo de tres segundos alrededor del pico de máxima aceleración (es decir, 1.5 s antes y después del instante donde, para cada movimiento o muestra, se encuentra el valor máximo del módulo de la aceleración).

Cabe destacar, que a lo largo de la tesis se ha hecho uso de diferentes tamaños de ventanas temporales, el motivo de esta variabilidad se debe a que se comenzó probando ventanas más pequeñas (ya que se tomó como referencia la base inicial de UMAFall y la duración estimada de una caída típica), posteriormente se tendió a aumentar el tamaño de estas ventas por recomendaciones de la literatura y para evitar que algunas caídas más largas de otras bases de datos pudieran quedar sin caracterizar correctamente. No obstante, tras las pruebas realizadas la elección de esta ventana (dentro de valores en el rango empleado) no parece tener gran repercusión sobre los resultados o sobre las conclusiones alcanzadas. En ese sentido cabe pensar que para el cómputo de las características de entrada elegidas o de la clasificación que realizan los algoritmos de aprendizaje profundo como las redes convolucionales el tomar algunas muestras de más, alejadas del ‘epicentro’ de la caída (la fase de caída libre y el impacto) no resulta especialmente relevante.

7.2.4. Métrica de calidad

Al igual que en los estudios de los capítulos anteriores se utiliza la media geométrica de la sensibilidad y la especificidad como métrica para evaluar las prestaciones del sistema clasificador.

Para validar los modelos generados se hace uso de la técnica de validación cruzada *k-fold*, específicamente se realiza un análisis *5-fold*. Cada repositorio ha sido dividido en cinco particiones, en las cuales existe aproximadamente el mismo número de muestras de los diferentes movimientos y en las que se incluyen ejemplos de todos y cada uno de los sujetos participantes. De esta manera, se utiliza un 60% de las muestras para entrenar al sistema, un 20% para el proceso de validación (como técnica para evitar el *overfitting* o sobreentrenamiento) y un 20% para testear. Para conocer la variación existente entre los resultados también se estudia el comportamiento de la desviación estándar de la métrica de calidad.

7.3. Resultados

Los resultados derivados de este trabajo permiten obtener las siguientes conclusiones. En primer lugar, se observa una tendencia a que el rendimiento

del clasificador neuronal profundo comience a empeorar de forma drástica a partir de frecuencias de muestreo inferiores a 5 Hz. Por otro lado, un aumento de la frecuencia de muestreo a partir de 20 Hz no mejora significativamente la métrica de calidad del sistema. Por tanto, los resultados demuestran que la utilización de frecuencias de muestreo alrededor de 5 Hz es suficiente para proporcionar medias geométricas de la sensibilidad y la especificidad cercanos al 90% en la mayoría de los casos e incluso superiores del 95% para muchos de los repositorios analizados.

Además, estos datos nos permiten observar cómo determinadas ubicaciones del acelerómetro proporcionan mejores rendimientos del sistema de detección, así como una mayor estabilidad en los resultados. En ese sentido, destaca la posición del pecho y la cintura, los cuales consiguen resultados muy cercanos al 100% en gran parte de las pruebas. Estos resultados coinciden con las conclusiones de otros trabajos anteriores, como los correspondientes al Capítulo 3.

En cuanto a la desviación estándar (como indicativo de la variabilidad) de las prestaciones alcanzadas por el modelo neuronal, esta tiende a aumentar al utilizar frecuencias inferiores a 20 Hz. Añadidamente, los resultados parecen indicar que, en ciertos casos, utilizar frecuencias superiores a unos 60 Hz pueden influir negativamente en el sistema, hecho probablemente relacionado el aumento de la dimensión de la entrada de la red, lo que provoca que alcanzar resultados similares, los filtros convolucionales deberían ser mayores y más complejos.

Al igual que ocurría con la media geométrica de la sensibilidad y la especificidad, la utilización del acelerómetro en pecho y cintura da lugar a resultados más estables, destacando la cintura, cuya desviación es muy cercana a valores nulos en gran parte de los casos.

En cuanto a la experiencia en la que se realiza un filtrado a las señales que se introducen en la red neuronal mediante un filtro paso bajo, se observa que solamente en muy pocos casos y para la frecuencia de corte más baja (5 Hz), el proceso de filtrado provoca cierta degradación de la métrica de calidad. Por el contrario, para la mayoría de los repositorios y posiciones, la CNN se comporta mejor al filtrar los datos por un paso bajo respecto a la utilización de las

frecuencias originales. Este fenómeno podría explicarse por la eliminación del ruido de alta frecuencia que dificulta el proceso de aprendizaje de la CNN.

Por otra parte, el estudio de espectro de frecuencia demuestra que más del 90% de la potencia en las ADL y más del 75%-80% de potencia en las caídas se encuentran en componentes de 0 a 2.5 Hz. Convirtiéndose respectivamente en porcentajes de 95% y 90% si la banda de estudio sube hasta los 5 Hz. Este hecho justifica que los estudios con frecuencias de muestreo bajas obtengan rendimientos muy aceptables.

Estos resultados pueden traducirse en una mayor simplificación de los sistemas de detección de caídas y con ello en una disminución de energía consumida y coste, ya que los requerimientos técnicos de los acelerómetros en cuanto a frecuencia de muestreo son mínimos. Así mismo, la utilización de frecuencias de muestreo bajas también simplifica la carga computacional de los algoritmos de detección de caídas dado que el volumen de la señal a procesar (en este caso, el número de muestras dentro de la ventana de tiempo de observación) es menor.

Capítulo 8. Conclusiones y trabajo futuro

8.1. Conclusiones

Una vez detectado el grave problema que originan las caídas a la sociedad y observar cómo la situación empeora año tras año debido al envejecimiento de la población, se emprende la realización de la presente tesis. En este trabajo se ha elaborado el estudio y análisis de sistemas vestibles de detección automática de caídas, capaces de discriminar estos accidentes frente a movimientos convencionales o ADL (*Activities of Daily Living*). Las aportaciones en este ámbito han sido diversas, enlazándose progresiva y temáticamente mediante el mismo hilo conductor y de manera cronológica.

En un primer lugar se ha desarrollado un repositorio de movimientos conocido como UMAFall. El objetivo de su creación fue proporcionar a la comunidad científica un conjunto de datos para poder analizar sistemáticamente distintas técnicas y sistemas de detección de caídas. Esta base de datos resultó pionera en el momento de su creación, ya que, a diferencia de otros *datasets* existentes en aquel momento, incluía trazas de ADL y caídas obtenidas mediante la monitorización simultánea de varios sensores ubicados en diferentes partes del cuerpo de los sujetos participantes; así mismo, la tipología y variedad de los movimientos era amplia y a su vez adjuntaba, además, información de los rasgos y características intrínsecas de los sujetos y los sensores inerciales utilizados. El repositorio fue publicado y se encuentra accesible al público para su uso. Ha sido utilizado en diversas contribuciones científicas.

Posteriormente, haciendo uso del repositorio descrito anteriormente, se han realizado diferentes estudios con el propósito de conocer el efecto que provoca

en el proceso de clasificación factores como la ubicación de los sensores en el cuerpo de los sujetos o la combinación de varios sensores. Para estas pruebas se utilizan dos de los métodos más utilizados en este campo para la detección de caídas: algoritmos basados en umbrales y algoritmos de *machine learning* de aprendizaje supervisado. En ambos estudios las conclusiones son similares: la utilización de un mayor número de puntos de sensorización no apareja una mejora de las prestaciones del sistema. Al mismo tiempo se comprueba que existen ubicaciones del cuerpo humano donde los sensores consiguen caracterizar de mejor modo ADL y caídas, siendo destacables la cintura y el pecho, ambos puntos cercanos al centro de gravedad de una persona. Otra conclusión destacable es que la utilización del *smartphone*, cuya ubicación natural es el bolsillo del pantalón o un bolso, no es aconsejable para detectar caídas, práctica que estaba tomando gran relevancia debido al auge de los teléfonos inteligentes. Sin embargo, su capacidad de procesamiento y comunicación lo convierten en un dispositivo interesante para ser utilizado como nodo de interconexión con los sensores corporales.

En cuanto a las prestaciones de los algoritmos, se concluye que los algoritmos basados en umbrales presentan problemas a la hora de discernir entre caídas y movimientos con gran agitación como los movimientos deportivos, lo que no los hace suficientemente robustos para escenarios donde exista la posibilidad de que existan movimientos bruscos. Por otro lado, los algoritmos de aprendizaje supervisado sí presentan medias geométricas de la sensibilidad y especificidad aceptables, superiores a 0.90. Como novedad en este tipo de análisis incorporamos la comprobación de la significación estadística de los resultados mediante análisis de varianza (ANOVA), metodología poco utilizada hasta el momento en la bibliografía relacionada, donde basan las evaluaciones y análisis simplemente en el estudio de las diferentes medias de las magnitudes, sin verificar si las variaciones son realmente significativas entre sí.

Asimismo, también se realiza un estudio preliminar (el cual es ampliado posteriormente en el trabajo incluido en el apéndice B), sobre el impacto de la selección de los estadísticos utilizados como entrada de los algoritmos de aprendizaje supervisado, con el objetivo de encontrar cuáles de ellos caracterizan mejor los movimientos y, por ende, incrementan las prestaciones

del sistema. En primer lugar, se deduce que la utilización de un mayor número de características de entrada no se traduce necesariamente en una mejora de las prestaciones; al contrario, la utilización de características innecesarias puede dificultar el aprendizaje del sistema al introducir ruido en el modelo. Los estadísticos que parecen aportar más información a la hora de clasificar ADL y caídas son en primer lugar $A_{\omega_{diff(max)}}$; que proporciona información sobre cambios bruscos de aceleración; μ_{Ap} , describe cambios en la perpendicularidad del sujeto y, por último, μ_{SMV} , la media del módulo de la aceleración (véase el capítulo 6 para una definición formal de estas características). Además, el empleo combinado de estos estadísticos influye positivamente en la eficacia del clasificador. Igualmente, se comprueba que la utilización de los estadísticos relacionados con las propiedades dinámicas del cuerpo humano seleccionados en este estudio produce un mejor comportamiento del sistema que el utilizar una herramienta de elección de *features* por análisis masivo de estadísticos arbitrarios (en particular la biblioteca *hctsa*, desarrollada en Matlab).

Por otra parte, gracias a la paulatina aparición de repositorios de movimientos durante estos últimos cinco años ideados para la evaluación comparativa (*benchmarking*), se lleva a cabo una serie de trabajos donde se revisan estas bases de datos. En este ámbito, se detecta una amplia heterogeneidad: desde el número de sujetos participantes y sus rasgos físicos (edad, altura, peso, IMC), o la cantidad, tipología y duración de los movimientos que contiene, hasta las características o ubicación de los sensores utilizados; lo que verifica la inexistencia práctica de un proceso normalizador o, al menos, unas recomendaciones, a la hora de crear este tipo de repositorios. Por tanto, se concluye que los trabajos de la literatura que evalúan las prestaciones de sistemas de detección de caídas sin tener en cuenta diferentes repositorios, están ignorando si realmente su sistema funciona con el mismo rendimiento cuando se utilizan otros datos y, por tanto, si es legítimo extrapolar dichas métricas a otros escenarios.

Seguidamente se han analizado las prestaciones al utilizar diversos algoritmos de aprendizaje supervisado, así como, distintas combinaciones de hiperparámetros. Se vuelven a obtener prestaciones muy interesantes, en algunos casos superando medias geométricas de sensibilidad y especificidad

superiores a 0.97. Además, se detecta que el conocido algoritmo SVM y, en menor medida, el también popular KNN, presentan un rendimiento superior al resto de algoritmos probados.

También se analiza la influencia en las prestaciones del sistema de detección de caídas de los rasgos físicos de los sujetos participantes en los repositorios. El estudio acaba determinando que las características físicas, exceptuando el criterio de la edad, parecen no ser un factor crucial que perturbe significativamente las prestaciones del sistema. Por otro lado, se evalúa también la influencia de la naturaleza de los movimientos con los que se configuran los clasificadores, comprobándose que este es un factor mucho más importante que los rasgos de los sujetos. En ese sentido, se demuestra que no entrenar con movimientos con un grado alto de movilidad, como las actividades deportivas, provoca un drástico empeoramiento del rendimiento del sistema a la hora de intentar clasificarlos (pues son fácilmente confundidos con caídas).

Por otra parte, uno de los hechos más discutidos en la metodología utilizada frecuentemente en el ámbito de la detección de caídas es la validez de emplear en los análisis y evaluaciones de las técnicas de clasificación caídas simuladas por voluntarios jóvenes y sanos. Es por ello por lo que se estudia la utilización de clasificadores de clase única, los cuales solamente tienen que ser entrenados con ADL. Los resultados son bastante prometedores, superando en ocasiones el 0.95 de media geométrica de la sensibilidad y la especificidad, siendo destacado el algoritmo OC-KNN, seguido de OC-SVM. A la hora de analizar el impacto de las categorías de movimientos existentes en la fase de entrenamiento, vuelve a ocurrir un caso similar al comentado con los algoritmos de aprendizaje supervisado: el sistema presenta serios problemas cuando se intenta clasificar movimientos deportivos cuando no se ha entrenado con ellos, ya que son detectados como anomalías y por tanto erróneamente identificados con caídas. También ocurre algo parecido si no se entrena con movimientos de baja intensidad o movimientos básicos, estos son detectados como anomalías. En esta línea, un estudio preliminar (ampliado en el apéndice C) indica que la utilización de técnicas simples de *ensemble learning* (o aprendizaje por conjuntos) suele mejorar el rendimiento de los sistemas mediante el agregado

(por ejemplo, por votación –o *voting*–) de las decisiones de clasificadores individuales.

Por último, se analizan y evalúan modelos de detección de caídas basados en redes neuronales convolucionales. Estos métodos de clasificación más complejos permiten extraer automáticamente las características que caracterizan a los movimientos, sin necesidad de elegir y computar previamente un conjunto de estadísticos o *features*, con las que caracterizar el movimiento. Los resultados de la media geométrica de la sensibilidad y especificidad son superiores en la mayoría de los casos a 0.98, siendo por tanto mejores que los del resto de técnicas estudiadas. Empleando estas redes, se examina además la influencia que tiene la frecuencia de muestreo en la calidad de la clasificación, concluyendo que dicho factor no es crucial si el rango utilizado es superior a 5 Hz. Así, estos resultados pueden justificar una mayor simplificación de los sistemas de detección y con ello a una disminución de los costes, complejidad y consumo de los detectores vestibles, un aspecto importante si pensamos que estos normalmente se implementarán mediante dispositivos de capacidades (batería, potencia de cómputo, etc.) muy limitadas.

Para finalizar, en el apéndice E, se estudia el comportamiento de los sistemas de detección de caídas basados en redes neuronales convolucionales cuando se realiza un entrenamiento y testeo cruzado de repositorios. Los resultados demuestran que entrenar y testear con repositorios diferentes puede degradar de manera significativa las prestaciones del sistema de detección de caídas. Asimismo, se comprueba que la personalización del modelo, añadiendo nuevos movimientos a la hora de entrenar, puede traducirse en una gran mejora, principalmente si dichos movimientos son caídas. Sin embargo, debe procederse con cautela cuando los movimientos añadidos son solamente ADL que comparten similitudes con los movimientos etiquetados como caídas, ya que puede dificultar al modelo a la hora de clasificar y provocar una disminución de su rendimiento. Estos resultados parecen apuntar a cierto sobre-aprendizaje de los clasificadores en relación con la tipología de movimientos (ADL y caídas) utilizados, lo que pone de relieve la necesidad de evaluar cualquier clasificador con bases de datos muy amplias y, a ser posible, con movimientos reales de los usuarios objetivo (la gente mayor).

8.2. Amenazas de validez y trabajo futuro

A pesar de los grandes avances científicos que se han producido en los últimos años en el ámbito de la detección de caídas sigue existiendo un amplio abanico de tareas que investigar y resolver. Así mismo, el progreso obtenido hasta ahora genera nuevas ideas y posibles líneas de trabajo. A continuación, se presentan algunas las posibilidades de estudio que pueden ser objeto de interés.

En un primer lugar, a lo largo de esta tesis se ha demostrado la existencia de una gran heterogeneidad en los repositorios de caídas; tanto en la tipología de movimientos que componen las bases de datos, como las características de los sensores utilizados hasta en los rasgos físicos de los sujetos participantes; es por ello, que se propone la creación de una normativa que permita originar un marco de trabajo estandarizado y defina la metodología que debe llevarse a cabo a la hora de realizar este tipo de repositorios. De este modo, la evaluación de los sistemas de detección de caídas cuando utilizan este tipo de *datasets* podrán ser legítimamente comparables entre sí, al contrario de lo que ocurre actualmente.

Asimismo, por otro lado, sería de vital interés la creación de repositorios masivos con movimientos de la vida diaria y caídas reales de personas mayores. A pesar de que existen repositorios que ofrecen medidas inerciales de caídas reales de personas mayores como FARSEEING [83], hasta la fecha el número es muy escaso (y con muy pocas muestras disponibles). Los repositorios utilizados en este estudio, y el mayor porcentaje de los repositorios existentes hoy en día en la literatura monitorizan movimientos predefinidos y caídas simuladas, estas además son realizadas normalmente sobre una superficie protectora para evitar posibles lesiones. Asimismo, estos movimientos suelen realizarlos personas jóvenes y sanas. Todos estos factores pueden provocar diferencias relevantes entre las experiencias simuladas y los accidentes que suceden en un escenario real y con personas de avanzada edad, tal y como se discute en algunos trabajos como [33,84]. La evolución de los dispositivos vestibles permite la integración de sensores en la vestimenta y complementos sin que ocasione ningún tipo de carga o molestia, por lo que podría recopilarse una gran cantidad de información

de manera no intrusiva si se desarrollase algún proyecto interdisciplinar con centros sanitarios o residencias de ancianos.

Sin embargo, a pesar de lograr monitorizar largos periodos de tiempo, tal y como ya se comentó al comienzo de la tesis, lograr monitorizar una caída real es complicado; se calcula que sería necesario monitorizar entre setenta mil y cien mil días la actividad de personas mayores de 65 años para capturar unas 100 caídas reales [30]. Existe una alternativa que puede ser utilizada como solución intermedia entre la simulación de movimientos y caídas por personas jóvenes y sanas y la monitorización de personas mayores en entornos reales, se trata de la utilización de simuladores de edad, este tipo de sistema permite la posibilidad que personas jóvenes y sanas puedan experimentar limitaciones típicas de las personas ancianas, tanto físicas como cognitivas, como puede ser restricción del campo visual, sordera de alta frecuencia, rigidez articular, pérdida de la fuerza, reducción de la coordinación entre otras dificultades. En la literatura existen algunos trabajos [85–87] que ya han empleado este tipo de sistemas y demuestran cómo su utilización puede emular de una manera bastante fiel los efectos del envejecimiento y algunas patologías.

En cuanto a las técnicas de clasificación utilizadas para la detección de caídas también existen otras alternativas que no han sido probadas en este trabajo; por ejemplo, sería interesante el empleo de otros tipos de métodos de aprendizaje profundo como las redes neuronales recurrentes. Este tipo de red, ya empleada con éxito en algún trabajo sobre detectores de caídas, permite analizar series temporales. Al contrario que otras clases de red neuronal, las recurrentes poseen memoria, por tanto, la información del pasado es útil para la red y le otorga la capacidad de intentar predecir qué va a ocurrir posteriormente. De este modo, este tipo de técnica podría ser interesante para dar un paso más en la búsqueda de soluciones para este problema de salud mundial: no solo detectar la caída sino intentar anticiparse a ella. Relacionado con esta capacidad de anticipación de las caídas, también resulta atractivo el estudio de la detección de movimientos que pueden desembocar a una caída (mareos, desmayos, tropiezos, etc.) , tal y como propone Luna Perejón et al en [78] y la utilización de sistemas que protejan de dichas caídas como sistemas basados en *airbags* [88–91].

Referencias

1. World Health Organization (WHO) Fact Sheet No 404: Ageing and health Available online: <http://www.who.int/mediacentre/factsheets/fs404/en/> (accessed on Apr 30, 2017).
2. Yoshida, S. *A Global Report on Falls Prevention Epidemiology of Falls*; Switzerland, 2007;
3. World Health Organization *International Classification of Diseases, 11th version (ICD-11)*; 2018;
4. World Health Organization (WHO) Falls (Fact Sheets) Available online: <https://www.who.int/news-room/fact-sheets/detail/falls> (accessed on May 13, 2021).
5. WHO WHO | Falls - Fact sheet No 344 Available online: <http://www.who.int/mediacentre/factsheets/fs344/en/> (accessed on Apr 30, 2017).
6. EuroSafe *Injuries in the European Union: Summary of injury statistics for the years 2008-2010*; Amsterdam, 2013;
7. World Health Organization *WHO Global Report on Falls Prevention in Older Age.*; 2007;
8. Wild, D.; Nayak, U.S.; Isaacs, B. How dangerous are falls in old people at home? *Br. Med. J. (Clin. Res. Ed)*. **1981**, 282, 266–8, doi:10.1136/jech.35.3.200.
9. Masud, T.; Morris, R.O. Epidemiology of falls. *Age Ageing* **2001**, 30, 3–7, doi:10.1093/AGEING/30.SUPPL_4.3.

10. Olmos Zapata, P.; Abad Mateos, M.Á.; Pérez-Jara, J. Síndrome de temor a caerse en personas mayores de 65 años con mareos de repetición: estudio descriptivo. *Rev. Esp. Geriatr. Gerontol.* **2010**, *45*, doi:10.1016/j.regg.2010.02.005.
11. Rajagopalan, R.; Litvan, I.; Jung, T.-P. Fall Prediction and Prevention Systems: Recent Trends, Challenges, and Future Research Directions. *Sensors* **2017**, *17*, 2509, doi:10.3390/s17112509.
12. Igual, R.; Medrano, C.; Plaza, I. Challenges, issues and trends in fall detection systems. *Biomed. Eng. OnLine* **2013**, *12*, 1–24, doi:10.1186/1475-925X-12-66.
13. Board Of Governors Of The Federal Reserve System *Consumers and Mobile Financial Services 2015*; Washington, 2015;
14. Casilari, E.; Luque, R.; Morón, M.-J. Analysis of Android Device-Based Solutions for Fall Detection. *Sensors* **2015**, *15*, 17827–17894, doi:10.3390/s150817827.
15. Anderson, M. For vast majority of seniors who own one, a smartphone equals ‘freedom’ Available online: <http://www.pewresearch.org/fact-tank/2015/04/29/seniors-smartphones/> (accessed on Sep 3, 2021).
16. Smith, A. Older Adults and Technology Use. *Pew Res. Cent.* **2014**.
17. Zhang, T.; Wang, J.; Liu, P.; Hou, J. Fall Detection by Embedding an Accelerometer in Cellphone and Using KFD Algorithm. *J. Comput. Sci.* **2006**, *6*, 277–284.
18. Habib, M.A. shfak; Mohktar, M.S.; Kamaruzzaman, S.B. ahyah; Lim, K.S. eang; Pin, T.M. aw; Ibrahim, F.; Ashfak Habib, M.; Mohktar, M.S.; Bahyah Kamaruzzaman, S.; Seang Lim, K.; et al. Smartphone-based solutions for fall detection and prevention: Challenges and open issues. *Sensors* **2014**, *14*, 7181–7208, doi:10.3390/s140407181.
19. Kau, L.J.; Chen, C.S. A smart phone-based pocket fall accident detection, positioning, and rescue system. *IEEE J. Biomed. Heal. Informatics* **2015**, *19*, doi:10.1109/JBHI.2014.2328593.
20. Fang, S.-H.; Liang, Y.-C.; Chiu, K.-M. Developing a mobile phone-based

- fall detection system on Android platform. In Proceedings of the 2012 Computing, Communications and Applications Conference; IEEE: Hong Kong, China, January 11-13, 2012; pp. 143–146.
21. Ponce, H.; Martínez-Villaseñor, L.; Nuñez-Martínez, J. Sensor location analysis and minimal deployment for fall detection system. *IEEE Access* **2020**, *8*, doi:10.1109/ACCESS.2020.3022971.
 22. Özdemir, A.T. An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice. *Sensors* **2016**, doi:10.3390/s16081161.
 23. Dai, J.; Bai, X.; Yang, Z.; Shen, Z.; Xuan, D. Mobile phone-based pervasive fall detection. *Pers. Ubiquitous Comput.* **2010**, *14*, 633–643, doi:10.1007/S00779-010-0292-X.
 24. Wang, X.; Ellul, J.; Azzopardi, G. Elderly Fall Detection Systems: A Literature Survey. *Front. Robot. AI* **2020**, *0*, 71, doi:10.3389/FROBT.2020.00071.
 25. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A Fall and Movement Dataset. *Sensors* **2017**, *17*, 198, doi:10.3390/s17010198.
 26. Bao, L.; Intille, S.S. Activity Recognition from User-Annotated Acceleration Data. In *Lecture Notes in Computer Science*; Springer, Berlin, Heidelberg, 2004; Vol. 3001, pp. 1–17.
 27. Li, Z.; Tianchi, L.; Sijun, Z.; Zhiliang, Z.; Zhang, L.; Liu, T.; Zhu, S.; Zhu, Z. Human activity recognition based on triaxial accelerometer. In Proceedings of the Computing and Convergence Technology (ICCCT), 2012 7th International Conference on; Seoul, Korea (South), December 3-5, 2012; pp. 261–266.
 28. Aziz, O.; Klenk, J.; Schwickert, L.; Chiari, L.; Becker, C.; Park, E.J.; Mori, G.; Robinovitch, S.N. Validation of accuracy of SVM-based fall detection system using real-world fall and non-fall datasets. *PLoS One* **2017**, *12*, e0180318, doi:10.1371/journal.pone.0180318.
 29. Saleh, M.; Abbas, M.; Le Jeannes, R.B. FallAID: An Open Dataset of Human Falls and Activities of Daily Living for Classical and Deep Learning

- Applications. *IEEE Sens. J.* **2021**, *21*, doi:10.1109/JSEN.2020.3018335.
30. Becker, C.; Schwickert, L.; Mellone, S.; Bagalà, F.; Chiari, L.; Helbostad, J.L.; Zijlstra, W.; Aminian, K.; Bourke, A.; Todd, C.; et al. Proposal for a multiphase fall model based on real-world fall recordings with body-fixed sensors. *Zeitschrift für Gerontol. und Geriatr.* **2012**, *45*, 707–715, doi:10.1007/S00391-012-0403-6.
 31. Kangas, M.; Vikman, I.; Nyberg, L.; Korpelainen, R.; Lindblom, J.; Jämsä, T. Comparison of real-life accidental falls in older people with experimental falls in middle-aged test subjects. *Gait Posture* **2012**, *35*, 500–505, doi:10.1016/j.gaitpost.2011.11.016.
 32. de la Cal, E.; DaSilva, A.; Fáñez, M.; Villar, J.R.; Sedano, J.; Suárez, V. An Autonomous Fallers Monitoring Kit: Release 0.0. In *Intelligent Systems Design and Applications. ISDA 2019*; 2021; pp. 287–297 ISBN 978-3-030-49342-4.
 33. Bagalà, F.; Becker, C.; Cappello, A.; Chiari, L.; Aminian, K.; Hausdorff, J.M.; Zijlstra, W.; Klenk, J. Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls. *PLoS One* **2012**, *7*, e37062, doi:10.1371/journal.pone.0037062.
 34. Wang, F.T.; Chan, H.L.; Hsu, M.H.; Lin, C.K.; Chao, P.K.; Chang, Y.J. Threshold-based fall detection using a hybrid of tri-axial accelerometer and gyroscope. *Physiol. Meas.* **2018**, *39*, doi:10.1088/1361-6579/aae0eb.
 35. Bourke, A.K.; O'Brien, J. V.; Lyons, G.M. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait Posture* **2007**, *26*, doi:10.1016/j.gaitpost.2006.09.012.
 36. Lee, J.S.; Tseng, H.H. Development of an Enhanced Threshold-Based Fall Detection System Using Smartphones with Built-In Accelerometers. *IEEE Sens. J.* **2019**, *19*, 8293–8302, doi:10.1109/JSEN.2019.2918690.
 37. Kangas, M.; Konttila, A.; Lindgren, P.; Winblad, I.; Jämsä, T. Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait Posture* **2008**, *28*, 285–291, doi:10.1016/J.GAITPOST.2008.01.003.

38. Aziz, O.; Musngi, M.; Park, E.J.; Mori, G.; Robinovitch, S.N. A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials. *Med. Biol. Eng. Comput.* **2017**, *55*, 45–55, doi:10.1007/s11517-016-1504-y.
39. Medrano, C.; Igual, R.; Plaza, I.; Castro, M. Detecting falls as novelties in acceleration patterns acquired with smartphones. *PLoS One* **2014**, *9*, 94811, doi:10.1371/journal.pone.0094811.
40. Hakim, A.; Huq, M.S.; Shanta, S.; Ibrahim, B.S.K.K. Smartphone Based Data Mining for Fall Detection: Analysis and Design. *Procedia Comput. Sci.* **2017**, *105*, 46–51, doi:10.1016/J.PROCS.2017.01.188.
41. Liu, S.H.; Cheng, W.C. Fall detection with the support vector machine during scripted and continuous unscripted activities. *Sensors* **2012**, *12*, 12301–12316, doi:10.3390/s120912301.
42. Lisowska, A.; Wheeler, G.; Inza, V.C.; Poole, I. An Evaluation of Supervised, Novelty-Based and Hybrid Approaches to Fall Detection Using Silmee Accelerometer Data. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW); IEEE: Santiago, Chile. December 7-13, 2015; Vol. 2015-Febru, pp. 402–408.
43. Medrano, C.; Plaza, I.; Igual, R.; Sánchez, Á.; Castro, M. The effect of personalization on smartphone-based fall detectors. *Sensors* **2016**, *16*, doi:10.3390/s16010117.
44. Mauldin, T.R.; Canby, M.E.; Metsis, V.; Ngu, A.H.H.; Rivera, C.C. Smartfall: A smartwatch-based fall detection system using deep learning. *Sensors* **2018**, *18*, doi:10.3390/s18103363.
45. Luna-Perejón, F.; Domínguez-Morales, M.J.; Civit-Balcells, A. Wearable fall detector using recurrent neural networks. *Sensors* **2019**, *19*, doi:10.3390/s19224885.
46. Kraft, D.; Srinivasan, K.; Bieber, G. Deep Learning Based Fall Detection Algorithms for Embedded Systems, Smartwatches, and IoT Devices Using Accelerometers. *Technologies* **2020**, *8*, doi:10.3390/technologies8040072.

47. Casilari-Pérez, E.; García-Lagos, F. A comprehensive study on the use of artificial neural networks in wearable fall detection systems. *Expert Syst. Appl.* **2019**, *138*, doi:10.1016/j.eswa.2019.07.028.
48. Casilari, E.; Lora-rivera, R.; García-lagos, F. A study on the application of convolutional neural networks to fall detection evaluated with multiple public datasets. *Sensors* **2020**, *20*, doi:10.3390/s20051466.
49. Khojasteh, S.B.; Villar, J.R.; De La Cal, E.; González, V.M.; Chira, C. *Comparing model performances applied to fall detection*; Universidad de Madeira: Madeira, Portugal, July 9-12, 2018;
50. Blunda, L. La; Gutiérrez-Madroñal, L.; Wagner, M.F.; Medina-Bulo, I. A Wearable Fall Detection System Based on Body Area Networks. *IEEE Access* **2020**, *8*, doi:10.1109/ACCESS.2020.3032497.
51. Vaidehi, V.; Bhargavi, R.; Ganapathy, K.; Sweetlin Hemalatha, C. Multi-sensor based in-home health monitoring using Complex Event Processing. In *Proceedings of the International Conference on Recent Trends in Information Technology, ICRTIT 2012*; 2012.
52. Igual, R.; Medrano, C.; Plaza, I. A comparison of public datasets for acceleration-based fall detection. *Med. Eng. Phys.* **2015**, *37*, 870–878, doi:10.1016/j.medengphy.2015.06.009.
53. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. In *Proceedings of the 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) (Procedia Computer Science)*; Leuven, Belgium, July 24-26, 2017; Vol. 110, pp. 32–39.
54. Google Scholar Details of UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection Available online: https://scholar.google.com/citations?view_op=view_citation&hl=es&user=Rc0SBfQAAAAJ&citation_for_view=Rc0SBfQAAAAJ:u5HHmVD_uO8C (accessed on Sep 14, 2021).
55. Scopus Document details - UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection Available online:

- <https://www.scopus.com/record/display.uri?eid=2-s2.0-85028688489&origin=resultslist#metrics> (accessed on Sep 14, 2021).
56. Web Of Science Citations of UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection Available online: <https://www.webofscience.com/wos/alldb/citation-report/7b9e1ea6-80bf-449e-a38a-eda5944a79ac-084b9d58> (accessed on Sep 14, 2021).
 57. Casilari, E.; A.Santoyo-Ramón, J. UMAFall: Fall Detection Dataset (Universidad de Malaga) Available online: https://figshare.com/articles/UMA_ADL_FALL_Dataset_zip/4214283.
 58. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. Analysis of a smartphone-based architecture with multiple mobility sensors for fall detection. *PLoS One* **2016**, *11*, 1–17, doi:10.1371/journal.pone.0168069.
 59. Sposaro, F.; Tyson, G. iFall: An android application for fall monitoring and response. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; IEEE: Minneapolis, MN, USA, September 3-6, 2009; pp. 6119–6122.
 60. Yoshida, T.; Mizuno, F.; Hayasaka, T.; Tsubota, K.; Wada, S.; Yamaguchi, T. A wearable computer system for a detection and prevention of elderly users from falling. In Proceedings of the The 12th International Conference on Biomedical Engineering (ICBME); Springer: Singapore, December 7-10, 2005.
 61. Jiangpeng Dai; Xiaole Bai; Zhimin Yang; Zhaohui Shen; Dong Xuan PerFallD: A pervasive fall detection system using mobile phones. In Proceedings of the 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops); IEEE: Mannheim, Germany, March 29 - April 2, 2010; pp. 292–297.
 62. Santoyo-Ramón, J.; Casilari, E.; Cano-García, J. Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection with Supervised Learning. *Sensors* **2018**, *18*, 1155, doi:10.3390/s18041155.
 63. Montgomery, D.C. Design and Analysis of Experiments, 5th Edition. *Jhon*

- Wiley Sons, Inc* 2012.
64. Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.-M. Analysis of Public Datasets for Wearable Fall Detection Systems. *Sensors* **2017**, *17*, 1513, doi:10.3390/s17071513.
 65. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems. *J. Healthc. Eng.* **2020**, *2020*, 1–36, doi:10.1155/2020/6622285.
 66. Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Study of the Detection of Falls Using the SVM Algorithm, Different Datasets of Movements and ANOVA. In Proceedings of the IEEE IWBBIO: International Work-Conference on Bioinformatics and Biomedical Engineering; Springer: Granada, Spain, May 8-10, 2019; Vol. 11465 LNBI, pp. 415–428.
 67. Santoyo-Ramón, J.; Casilari, E.; Cano-García, J. Estudio de la detección de caídas utilizando el algoritmo SVM. In Proceedings of the Actas del XXXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2018); Granada, España, 2018; pp. 5–7.
 68. Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.M. Analysis of a Public Repository for the Study of Automatic Fall Detection Algorithms. In Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18); ACM: Berlin, Germany, September 20-21, 2018; pp. 1–10.
 69. Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A study on the impact of the users' characteristics on the performance of wearable fall detection systems. *Sci. Reports* **2021**, *11*, 1–14, doi:10.1038/s41598-021-02537-z.
 70. Cotechini, V.; Belli, A.; Palma, L.; Morettini, M.; Burattini, L.; Pierleoni, P. A dataset for the development and optimization of fall detection algorithms based on wearable sensors. *Data Br.* **2019**, doi:10.1016/j.dib.2019.103839.

71. Martínez-Villaseñor, L.; Ponce, H.; Espinosa-Loera, R.A. Multimodal Database for Human Activity Recognition and Fall Detection. *Proceedings* **2018**, doi:10.3390/proceedings2191237.
72. Santoyo-Ramón, J.A.; Casilari, E.; Cano-García, J.M. A Study of One-Class Classification Algorithms for Wearable Fall Sensors. *Biosensors* **2021**, *11*, 284, doi:10.3390/bios11080284.
73. Frank, K.; Vera Nadales, M.J.; Robertson, P.; Pfeifer, T. Bayesian recognition of motion related activities with inertial sensors. In Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Ubicomp '10; ACM Press: Copenhagen, Denmark, September 26-29, 2010; p. 445.
74. Human Factors and Ergonomics Lab- Korea Advanced Institute of Science and Technology KFall: A Comprehensive Motion Dataset to Detect Pre-impact Fall for the Elderly based on Wearable Inertial Sensors Available online: <https://sites.google.com/view/kfalldataset>.
75. Martínez-Villaseñor, L.; Ponce, H.; Brieva, J.; Moya-Albor, E.; Núñez-Martínez, J.; Peñafort-Asturiano, C. UP-Fall Detection Dataset: A Multimodal Approach. *Sensors* **2019**, *19*, 1988, doi:10.3390/s19091988.
76. Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Anexo: Contribuciones al estudio y análisis de sistemas-vestibulares-de detección de caídas Available online: https://github.com/jasr92/Anexo_Contribuciones-al-estudio-y-an-lisis-de-sistemas-vestibles-de-detecci-n-de-ca-das (accessed on Sep 17, 2021).
77. Wertner, A.; Czech, P.; Pammer-Schindler, V. An Open Labelled Dataset for Mobile Phone Sensing Based Fall Detection.; 2015.
78. Luna-Perejón, F.; Muñoz-Saavedra, L.; Civit-Masot, J.; Civit, A.; Domínguez-Morales, M. Ankfall—falls, falling risks and daily-life activities dataset with an ankle-placed accelerometer and training using recurrent neural networks. *Sensors* **2021**, *21*, doi:10.3390/s21051889.
79. Ojetola, O.; Gaura, E.; Brusey, J. Data set for fall events and daily activities from inertial sensors. In Proceedings of the 6th ACM Multimedia Systems

- Conference; ACM: Portland, Oregon. March 18-20, 2015; pp. 243–248.
80. Vavoulas, G.; Chatzaki, C.; Malliotakis, T.; Pediaditis, M.; Tsiknakis, M. The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones. In Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health; SCITEPRESS - Science and Technology Publications: Rome, Italy, April 21-22, 2016; pp. 143–151.
 81. Gasparrini, S.; Cippitelli, E.; Spinsante, S.; Gambi, E. A depth-based fall detection system using a Kinect® sensor. *Sensors* **2014**, *14*, 2756–2775, doi:10.3390/s140202756.
 82. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET); IEEE: Antalya, Turkey, August 21-23, 2017; Vol. 2018-Janua, pp. 1–6.
 83. Klenk, J.; Schwickert, L.; Palmerini, L.; Mellone, S.; Bourke, A.; Ihlen, E.A.F.; Kerse, N.; Hauer, K.; Pijnappels, M.; Synofzik, M.; et al. The FARSEEING real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls. *Eur. Rev. Aging Phys. Act.* **2016**, *13*, 1–7, doi:10.1186/S11556-016-0168-9/FIGURES/3.
 84. Klenk, J.; Becker, C.; Lieken, F.; Nicolai, S.; Maetzler, W.; Alt, W.; Zijlstra, W.; Hausdorff, J.M.; van Lummel, R.C.; Chiari, L.; et al. Comparison of acceleration signals of simulated and real-world backward falls. *Med. Eng. Phys.* **2011**, *33*, doi:10.1016/j.medengphy.2010.11.003.
 85. Groza, H.L.; Sebesi, S.B.; Mandru, D.S. Age Simulation Suits for Training, Research and Development. In *IFMBE Proceedings*; Cluj-Napoca, Romania, October 15, 2017; Vol. 59, pp. 77–80.
 86. Vieweg, J.; Schaefer, S. How an Age Simulation Suit affects Motor and Cognitive Performance and Self-perception in Younger Adults. *Exp. Aging Res.* **2020**, *46*, doi:10.1080/0361073X.2020.1766299.
 87. Timm, I.J.; Spaderna, H.; Rodermund, S.C.; Lohr, C.; Buettner, R.; Berndt,

- J.O. Designing a randomized trial with an age simulation suit—representing people with health impairments. *Healthcare* **2021**, *9*, doi:10.3390/healthcare9010027.
88. Tamura, T.; Yoshimura, T.; Sekine, M.; Uchida, M.; Tanaka, O. A wearable airbag to prevent fall injuries. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 910–914, doi:10.1109/TITB.2009.2033673.
89. Jung, H.; Koo, B.; Kim, J.; Kim, T.; Nam, Y.; Kim, Y. Enhanced algorithm for the detection of preimpact fall for wearable airbags. *Sensors* **2020**, *20*, doi:10.3390/s20051277.
90. Gelmini, S.; Strada, S.; Tanelli, M.; Savaresi, S.; Guzzon, A. Automatic detection of human's falls from heights for airbag deployment via inertial measurements. *Autom. Constr.* **2020**, *120*, doi:10.1016/j.autcon.2020.103358.
91. Shi, G.; Chan, C.S.; Li, W.J.; Leung, K.S.; Zou, Y.; Jin, Y. Mobile human airbag system for fall protection using mems sensors and embedded SVM classifier. *IEEE Sens. J.* **2009**, *9*, doi:10.1109/JSEN.2008.2012212.

Apéndice A. Publicaciones

Este apéndice incluye de manera ordenada una copia de cada una de las publicaciones asociadas a los resúmenes contenidos en los capítulos anteriores.

A.1. Desarrollo de un repositorio multisensor

Publicación

Casilari-Pérez, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. In Proceedings of the 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) (Procedia Computer Science); Leuven, Belgium, July 24-26, 2017; Vol. 110, pp. 32–39.

<https://doi.org/10.1016/j.procs.2017.06.110>

Nº páginas: 8

UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection¹

Abstract

The progress in the field of inertial sensor technology and the widespread popularity of personal electronics such as smartwatches or smartphones have prompted the research on wearable Fall Detection Systems (FDSs). In spite of the extensive literature on FDSs, an open issue is the definition of a common framework that allows a methodical and agreed evaluation of fall detection policies. In this regard, a key aspect is the lack of a public repository of movement datasets that can be employed by the researchers as a common reference to compare and assess their proposals.

This work describes UMAFall, a new dataset of movement traces acquired through the systematic emulation of a set of predefined ADLs (Activities of Daily Life) and falls. In opposition to other existing databases for FDSs, which only include the signals captured by one or two sensing points, the testbed deployed for the generation of UMAFall dataset incorporated five wearable sensing points, which were located on five different points of the body of the participants that developed the movements. As a consequence, the obtained data offer an interesting tool to investigate the importance of the sensor placement for the effectiveness of the detection decision in FDSs.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.2. Análisis mediante algoritmos basados en umbrales

Publicación

Casilari Pérez, E.; Santoyo-Ramón, J.A.; Cano-García, J.M.
Analysis of a smartphone-based architecture with multiple mobility
sensors for fall detection. PLoS One 2016, 11, 1–17

<https://doi.org/10.1371/journal.pone.0168069>

JCR: 2.806 (Ranking 15 sobre 64 en categoría
Multidisciplinary Sciences) (Q1)

Nº páginas: 17

Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection¹

Abstract

During the last years, many research efforts have been devoted to the definition of Fall Detection Systems (FDSs) that benefit from the inherent computing, communication and sensing capabilities of smartphones. However, employing a smartphone as the unique sensor in a FDS application entails several disadvantages as long as an accurate characterization of the patient's mobility may force to transport this personal device on an unnatural position. This paper presents a smartphone-based architecture for the automatic detection of falls. The system incorporates a set of small sensing nodes that can communicate with the smartphone to help in the fall detection decision. The deployed architecture is systematically evaluated in a testbed with experimental users in order to determine the number and positions of the sensors that optimize the effectiveness of the FDS, as well as to assess the most convenient role of the smartphone in the architecture.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.3. Estudio mediante algoritmos *machine learning* y análisis ANOVA

Publicación

Santoyo-Ramón, J.; Casilari Pérez, E.; Cano-García, J. Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection with Supervised Learning. Sensors 2018, 18, 1155

<https://doi.org/10.3390/s18041155>

*JCR: 3.031 (Ranking 15 sobre 61 en categoría
Instruments & Instrumentation) (Q1)*

Nº páginas: 29

Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection with Supervised Learning¹

Abstract

This paper describes a wearable Fall Detection System (FDS) based on a body-area network consisting of four nodes provided with inertial sensors and Bluetooth wireless interfaces. The signals captured by the nodes are sent to a smartphone which simultaneously acts as another sensing point. In contrast to many FDSs proposed by the literature (which only consider a single sensor), the multisensory nature of the prototype is utilized to investigate the impact of the number and the positions of the sensors on the effectiveness of the production of the fall detection decision. In particular, the study assesses the capability of four popular machine learning algorithms to discriminate the dynamics of the Activities of Daily Living (ADLs) and falls generated by a set of experimental subjects, when the combined use of the sensors located on different parts of the body is considered. Prior to this, the election of the statistics that optimize the characterization of the acceleration signals and the efficacy of the FDS is also investigated. As another important methodological novelty in this field, the statistical significance of all the results (an aspect which is usually neglected by other works) is validated by an analysis of variance (ANOVA).

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.4. Revisión de *datasets* existentes y propuesta de clasificación de ADL

Publicación

Casilari-Pérez, E.; Santoyo-Ramón, J.-A.; Cano-García, J.-M.
Analysis of Public Datasets for Wearable Fall Detection Systems.
Sensors 2017, 17, 1513

<https://doi.org/10.3390/s17071513>

*JCR: 2.475 (Ranking 16 sobre 61 en categoría
Instruments & Instrumentation) (Q2)*

Nº de páginas: 28

Analysis of Public Datasets for Wearable Fall Detection Systems¹

Abstract

Due to the boom of wireless handheld devices such as smartwatches and smartphones, wearable Fall Detection Systems (FDSs) have become a major focus of attention among the research community during the last years. The effectiveness of a wearable FDS must be contrasted against a wide variety of measurements obtained from inertial sensors during the occurrence of falls and Activities of Daily Living (ADLs). In this regard, the access to public databases constitutes the basis for an open and systematic assessment of fall detection techniques. This paper reviews and appraises twelve existing available data repositories containing measurements of ADLs and emulated falls envisaged for the evaluation of fall detection algorithms in wearable FDSs. The analysis of the found datasets is performed in a comprehensive way, taking into account the multiple factors involved in the definition of the testbeds deployed for the generation of the mobility samples. The study of the traces brings to light the lack of a common experimental benchmarking procedure and, consequently, the large heterogeneity of the datasets from a number of perspectives (length and number of samples, typology of the emulated falls and ADLs, characteristics of the test subjects, features and positions of the sensors, etc.). Concerning this, the statistical analysis of the samples reveals the impact of the sensor range on the reliability of the traces. In addition, the study evidences the importance of the selection of the ADLs and the need of categorizing the ADLs depending on the intensity of the movements in order to evaluate the capability of a certain detection algorithm to discriminate falls from ADLs.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.5. Estudio del problema la heterogeneidad de los repositorios

Publicación

Casilari-Pérez, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems. J. Healthc. Eng. 2020, 2020, 1–36

<https://doi.org/10.1155/2020/6622285>

JCR: 2.682 (Ranking 53 sobre 108 en categoría Health Care Sciences & Services) (Q2)

Nº de páginas: 36

On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems¹

Abstract

Due to the serious impact of falls on the autonomy and health of older people, the investigation of wearable alerting systems for the automatic detection of falls has gained considerable scientific interest in the field of body telemonitoring with wireless sensors. Because of the difficulties of systematically validating these systems in a real application scenario, Fall Detection Systems (FDSs) are typically evaluated by studying their response to datasets containing inertial sensor measurements captured during the execution of labelled nonfall and fall movements. In this context, during the last decade, numerous publicly accessible databases have been released aiming at offering a common benchmarking tool for the validation of the new proposals on FDSs. 'is work offers a comparative and updated analysis of these existing repositories. For this purpose, the samples contained in the datasets are characterized by different statistics that model diverse aspects of the mobility of the human body in the time interval where the greatest change in the acceleration module is identified. By using one-way analysis of variance (ANOVA) on the series of these features, the comparison shows the significant differences detected between the datasets, even when comparing activities that require a similar degree of physical effort. 'is heterogeneity, which may result from the great variability of the sensors, experimental users, and testbeds employed to generate the datasets, is relevant because it casts doubt on the validity of the conclusions of many studies on FDSs, since most of the proposals in the literature are only evaluated using a single database.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.6. Análisis del algoritmo *machine learning* SVM con varias bases de datos y análisis ANOVA

Publicación

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Study of the Detection of Falls Using the SVM Algorithm, Different Datasets of Movements and ANOVA. In Proceedings of the IEEE IWBBIO: International Work-Conference on Bioinformatics and Biomedical Engineering; Springer: Granada, Spain, May 8-10, 2019; Vol. 11465 LNBI, pp. 415–42.

https://doi.org/10.1007/978-3-030-17938-0_37

Nº de páginas: 14

Study of the Detection of Falls Using the SVM Algorithm, Different Datasets of Movements and ANOVA¹

Abstract

Falls are becoming a major public health problem, which is intensified by the aging of the population. Falls are one of the main causes of death among the elderly and in population groups that develop risk activities. In this sense, technologies can provide solutions to improve this situation. In this work we have analyzed different repositories of movements and falls designed to test decision algorithms in automatic fall detection systems. The objectives of the study are: firstly, to clarify what are the characteristics of the most significant accelerometry signals to identify a fall and secondly, to analyze the possibility of extrapolating the learning achieved with a certain database when tested with another one. As a novelty with respect to other works in the literature, the statistical significance of the results has been systematically evaluated by the analysis of variance (ANOVA).

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.7. Análisis de la caracterización estadística de los movimientos respecto a su tipología y las características de los usuarios

Publicación

Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.M. Analysis of a Public Repository for the Study of Automatic Fall Detection Algorithms. In Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18); ACM: Berlin, Germany, September 20-21, 2018; pp. 1–10

<https://doi.org/10.1145%2F3266157.3266211>

Nº de páginas: 10

Analysis of a Public Repository for the Study of Automatic Fall Detection Algorithms¹

Abstract

The use of publicly available repositories containing movement traces of real or experimental subjects is a key aspect to define an evaluation framework that allows a systematic assessment of wearable fall detection systems. This paper presents a detailed analysis of a public dataset of traces which employed five sensing points to characterize the user's mobility during the execution of ADLs (Activities of Daily Living) and emulated falls. The analysis is aimed at analysing two main factors: the importance of the election of the position of the sensor and the possible impact of the user's personal features on the statistical characterization of the movements. Results reveal the importance of the nature of the ADL for the effectiveness of the discrimination of the falls.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.8. Estudio del impacto de las características de los usuarios en las prestaciones de sistemas de detección de caídas basados en algoritmos de aprendizaje supervisados

Publicación

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A study on the impact of the users' characteristics on the performance of wearable fall detection systems. Sci. Reports 2021 111 2021, 11, 1–14.

<https://doi.org/10.1038/s41598-021-02537-z>

JCR: 4.380 (2020) (Ranking 17 sobre 72 en categoría Multidisciplinary Sciences) (Q1)

Nº de páginas: 14

A study on the impact of the users' characteristics on the performance of wearable fall detection systems¹

Abstract

Wearable Fall Detection Systems (FDSs) have gained much research interest during last decade. In this regard, Machine Learning (ML) classifiers have shown great efficiency in discriminating falls and conventional movements or Activities of Daily Living (ADLs) based on the analysis of the signals captured by transportable inertial sensors. Due to the intrinsic difficulties of training and testing this type of detectors in realistic scenarios and with their target audience (older adults), FDSs are normally benchmarked against a predefined set of ADLs and emulated falls executed by volunteers in a controlled environment. In most studies, however, samples from the same experimental subjects are used to both train and evaluate the FDSs. In this work, we investigate the performance of ML-based FDS systems when the test subjects have physical characteristics (weight, height, body mass index, age, gender) different from those of the users considered for the test phase. The results seem to point out that certain divergences (weight, height) of the users of both subsets (training and test) may hamper the effectiveness of the classifiers (a reduction of up to 20% in sensitivity and of up to 5% in specificity is reported). However, it is shown that the typology of the activities included in these subgroups has much greater relevance for the discrimination capability of the classifiers (with specificity losses of up to 95% if the activity types for training and testing strongly diverge).

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.9. Evaluación de clasificadores de clase única

Publicación

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A Study of One-Class Classification Algorithms for Wearable Fall Sensors. Biosensors 2021, 11, 284

<https://doi.org/10.3390/bios11080284>

*JCR: 5.519 (2020) (Ranking 8 sobre 64 en categoría
Instruments & Instrumentation) (Q1)*

Nº de páginas: 20

A Study of One-Class Classification Algorithms for Wearable Fall Sensors¹

Abstract

In recent years, the popularity of wearable devices has fostered the investigation of automatic fall detection systems based on the analysis of the signals captured by transportable inertial sensors. Due to the complexity and variety of human movements, the detection algorithms that offer the best performance when discriminating falls from conventional Activities of Daily Living (ADLs) are those built on machine learning and deep learning mechanisms. In this regard, supervised machine learning binary classification methods have been massively employed by the related literature. However, the learning phase of these algorithms requires mobility patterns caused by falls, which are very difficult to obtain in realistic application scenarios. An interesting alternative is offered by One-Class Classifiers (OCCs), which can be exclusively trained and configured with movement traces of a single type (ADLs). In this paper, a systematic study of the performance of various typical OCCs (for diverse sets of input features and hyperparameters) is performed when applied to nine public repositories of falls and ADLs. The results show the potentials of these classifiers, which are capable of achieving performance metrics very similar to those of supervised algorithms (with values for the specificity and the sensitivity higher than 95%). However, the study warns of the need to have a wide variety of types of ADLs when training OCCs, since activities with a high degree of mobility can significantly increase the frequency of false alarms (ADLs identified as falls) if not considered in the data subsets used for training.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.10. Evaluación de sistemas de aprendizaje profundo. Análisis de la frecuencia de muestreo

Publicación

Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A Study of the Influence of the Sensor Sampling Frequency on the Performance of Wearable Fall Detectors.

Pendiente de ser enviado a la revista Measurement.

Nº de páginas: 17

A Study of the Influence of the Sensor Sampling Frequency on the Performance of Wearable Fall Detectors¹

Abstract

Last decade has witnessed a major research interest on wearable fall detection systems. Sampling rate in these detectors strongly affects the power consumption and required complexity of the employed wearables. This study investigates the effect of the sampling frequency on the efficacy of the detection process. For this purpose, we train a convolutional neural network to directly discriminate falls from conventional activities based on the raw acceleration signals captured by a transportable sensor. Then, we analyze the changes in the performance of this classifier when the sampling rate is progressively reduced. In contrast with previous studies, the detector is tested against a wide set of public repositories of benchmarking traces. The quality metrics achieved for the different frequencies and the analysis of the spectrum of the signals reveal that a sampling rate of 20 Hz can be enough to maximize the effectiveness of a fall detector.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

A.11. Evaluación del testeo cruzado de repositorios en sistemas vestibles de detección de caídas

Publicación

A Cross-dataset Evaluation of Wearable Fall Detection Systems, submitted to PErvasive Technologies Related to Assistive Environments (PETRA) conference, June 29–July 01, 2022, Corfu, Greece.
Pendiente de ser enviado a PETRA conference.

Nº de páginas: 10

A Cross-dataset Evaluation of Wearable Fall Detection Systems¹

Abstract

The popularity of wearables and their seamless integration into our daily lives have transformed these devices into an appealing resource to deploy automatic fall detection systems. During last years, a massive literature on new methods and algorithms for these wearable detectors has been produced. However, in most cases these algorithms are tested against one single (or at best two) datasets containing signals captured from falls and conventional movements. This work evaluates the behavior of a fall detection system based on a convolutional neural network when different public repositories of movements are alternatively used for training and testing the model. After a systematic cross-dataset evaluation involving four well-known datasets, we show the difficulty of extrapolating the results achieved by a certain classifier for a particular database when another dataset is considered. Results seem to indicate that classification methods tend to overlearn the particular conditions (typology of movements, characteristics of the employed sensor, experimental subjects) under which the training samples were generated.

¹ La publicación íntegra ha sido sustituida por su resumen en esta edición de la tesis.

Apéndice B. Análisis de las características que describen los movimientos

B.1. Introducción

Una de las fases más importantes a la hora de utilizar técnicas Machine Learning es la elección de las características (*features*) que se utilizarán como magnitudes de entrada en los algoritmos. En este estudio, dichos valores deben aportar información representativa que permita a los algoritmos discernir si el movimiento estudiado es una caída o un movimiento de la vida diaria (ADL). En este análisis se emplean las características descritas en el capítulo 4.

Tal y como se expone en la bibliografía [1–3] y se reafirmará en este trabajo, la utilización de un número mayor de *features* de entrada no se traduce en un aumento del rendimiento (especificidad y sensibilidad) del algoritmo, sino que puede provocar diferentes problemas:

- Características adicionales que actúan como ruido para el modelo del algoritmo *Machine Learning*.
- Mayor carga computacional del algoritmo, utilización de recursos innecesarios.
- Mayor tiempo necesario para el entrenamiento.

Es por ello, que debe implementarse un proceso de elección de características, logrando la reducción de *features*, simplificando de este modo el modelo y mejorando la calidad de los datos. El enfoque habitual para este proceso es la utilización de una serie de métodos que pueden dividirse de la siguiente manera:

- *Feature Selection* (Selección de Características): Es un proceso que proporciona la información necesaria para determinar la elección de un subconjunto de M características (*features*) del conjunto original formado por N *features*, donde ($M \leq N$), basándose en un cierto criterio [1,3]. Este proceso filtra características irrelevantes o redundantes.
- *Feature Extraction* (Extracción de Características): Se trata de un proceso que da lugar a un conjunto de K nuevas características (*features*) a través del procesado de las N características originales, donde ($K < N$). Este proceso eliminará características innecesarias y combinará aquellas que proporcionen información redundante [1,3].
- *Feature Construction* (Construcción de Características): Este proceso descubre información no tenida en cuenta sobre la relación existente entre características y proporciona nuevas características. En este caso no se produce una reducción de características sino que se consiguen P nuevas [1,3].

En esta investigación se hace uso de una técnica del tipo *Feature Selection*. De forma posterior, el estudio se combina con un análisis de varianza Factorial (N-way ANOVA) del rendimiento de los algoritmos con cada una de las combinaciones de características de entrada, de manera que se comprobará la eficacia de estos métodos.

B.2. Selección de características (*Feature Selection*)

Para el presente estudio se hará uso de diferentes algoritmos. A continuación, se describirá brevemente el funcionamiento de cada uno de ellos.

B.2.1. Minimum Redundancy Maximum Relevance (mRMR) algorithm

Minimum Redundancy Maximum Relevance (mRMR) es un algoritmo de selección de características que ayuda a escoger las características más relevantes de un conjunto de datos etiquetados. Este algoritmo permite eliminar características que no aportan información significativa y minimizar información redundante previamente a utilizar un algoritmo *Machine Learning* [4,5].

B.2.1.1. Conceptos básicos

El algoritmo se basa en la información proporcionada por la medida denominada información mutua, mide el nivel de similitud entre dos variables. En el caso del presente estudio, las variables estudiadas son continuas, por lo que dicho valor se calcula a partir de test-F de Fisher a cada característica (g_i) de un total de m características y de K clases denotadas por $h = h_1, h_2 \dots, h_K$.

- En el caso de este estudio h está compuesto por $K = 2$ clases, esto es, la clase caída y la clase ADL.
- Por otro lado, existen $m = 12$ características, las cuales describen un movimiento y posteriormente como input de entrada en los algoritmos *Machine Learning*. Dichas características se denotan con g_i , donde $i \in \{1, \dots, m\}$.

El test-F se encuentra descrito matemáticamente en la ecuación (1) [4,6].

$$F(g_i, h) = \left[\sum_{k=1}^K \frac{n_k(\bar{g}_k - \bar{g})}{(K-1)} \right] / \sigma^2 \quad (1)$$

donde:

- $\bar{g}$ es la media del valor de g_i en todas las muestras estudiadas.
- $\bar{g}_k$ es la media de g_i para la clase k .
- n_k es el número de muestras que forman parte de la clase k .
- σ^2 es la varianza agrupada o varianza general definida como (2):

$$\sigma^2 = \left[\sum_{k=1}^K (n_k - 1) \cdot \sigma_k^2 \right] / (n - K) \quad (2)$$

donde:

- K es el número total de clases estudiadas.
- n_k es el número de muestras que forman parte de la clase k .
- n es el número total de muestras, esto es $\sum_{k=1}^K n_k$.
- σ_k^2 es la varianza de la clase k .

El valor de máxima relevancia para un grupo de características S se denota como:

$$\max V_F, \quad V_F = \frac{1}{m} \sum_{g_i \in S} F(g_i, h) \quad (3)$$

donde $|S|$ expresa el número de características que subgrupo de características que se está estudiando. Este valor permitirá conocer qué características describen mejor el conjunto de datos [4].

Por otro lado, el algoritmo también tiene en cuenta la condición de mínima redundancia puede ser definida a partir del coeficiente de correlación de Pearson, $c(i, j)$, como:

$$\min W_c, \quad W_c = \frac{1}{m^2} \sum_{(g_i, g_j) \in S} |c(g_i, g_j)| \quad (4)$$

Teniendo en cuenta este valor se pretende minimizar el número de características que aportan información similar [4].

La selección de características a partir del algoritmo mRMR se obtiene a través de la optimización de las condiciones de las ecuaciones (3) y (4) simultáneamente. Para optimizar ambas condiciones, es necesario combinarlas en una única función. En la bibliografía [4,5] se proponen dos combinaciones para variables continuas, *F-test correlation difference* (FCD) y *F-test correlation quotient* (FCQ), equivalentes. simultáneamente, a MID (*Mutual information difference*) y MIQ (*Mutual information quotient*) para variables discretas:

$$FCD = \max(V_F - W_c) \quad (5)$$

$$FCQ = \max(V_F / W_c) \quad (6)$$

En este estudio se utilizará la función *fscmrmmr* incorporada en la versión R2019B de Matlab. En esta función, el criterio utilizado para optimizar las condiciones descritas es *F-test correlation quotient* (FCQ).

B.2.1.2. Funcionamiento del algoritmo

El algoritmo mRMR sigue los siguientes pasos iterativos para obtener un ranking de características que minimizan la redundancia y maximizan la relevancia de la información del *dataset* [4,5].

Se comienza con un set de características no seleccionadas, S^c , que incluye todas las características estudiadas, y un set S vacío, en el que se irán incorporando las características seleccionadas, las cuales se eliminarán de S^c . Se tendrá en cuenta el orden de inclusión en S para crear el ranking.

1. Se selecciona la característica, incluida en S^c , con un mayor valor de relevancia ($\max V_F$). Se selecciona dicha característica y se incluye en el grupo de características S . En este primer paso, sólo contendrá esta primera característica.
2. Se buscan las características que presentan relevancia mayor a 0 y una redundancia nula de la información contenida en S .
 - a. Si ninguna de las características presentes en S^c presenta una relevancia mayor de cero y redundancia nula, se deberá pasar al paso 4.
 - b. En caso contrario, se seleccionará aquella característica con mayor relevancia y se incluirá en S .
3. Se deberá repetir el paso 2 hasta que las características incluidas en S^c presenten una redundancia diferente de cero.
4. Se seleccionará aquella característica con mayor valor FCQ con relevancia no nula, del set de características aun no seleccionadas (S^c) y se incluirá en S .
5. Se repite el paso 4 hasta que las características incluidas en S^c tengan una relevancia nula.
6. Se añaden a S las características con relevancia nula en orden aleatorio.

B.3. Resultados

B.3.1. Utilización del algoritmo mRMR

Se lleva a cabo la utilización del algoritmo mRMR para cada uno de los repositorios de movimientos estudiados. En este caso se escogen aquellos que poseen medidas de un acelerómetro ubicado en la cadera.

La metodología utilizada para cada repositorio es la siguiente (descrita gráficamente en la Figura B.1):

1. Se calculan las características de cada una de las muestras por las que se compone el repositorio (Método descrito en el capítulo 4)
2. Se obtiene la media y la desviación típica de cada una de las características
3. Se normalizan las características obtenidas
4. Se hace uso del algoritmo mRMR aportando las características normalizadas de las muestras junto con la descripción del tipo de movimiento al que corresponde cada una de ellas (caída o ADL).
5. Se obtiene un ranking donde se ordenan cada una de las características de manera que se maximice la información y se minimiza la redundancia aportada (Tabla B.1). Así mismo se obtiene una lista denominada *F-test correlation quotient*, la cual aporta información sobre la información mutua aportada por cada una de las características estudiadas (Tabla B.2).

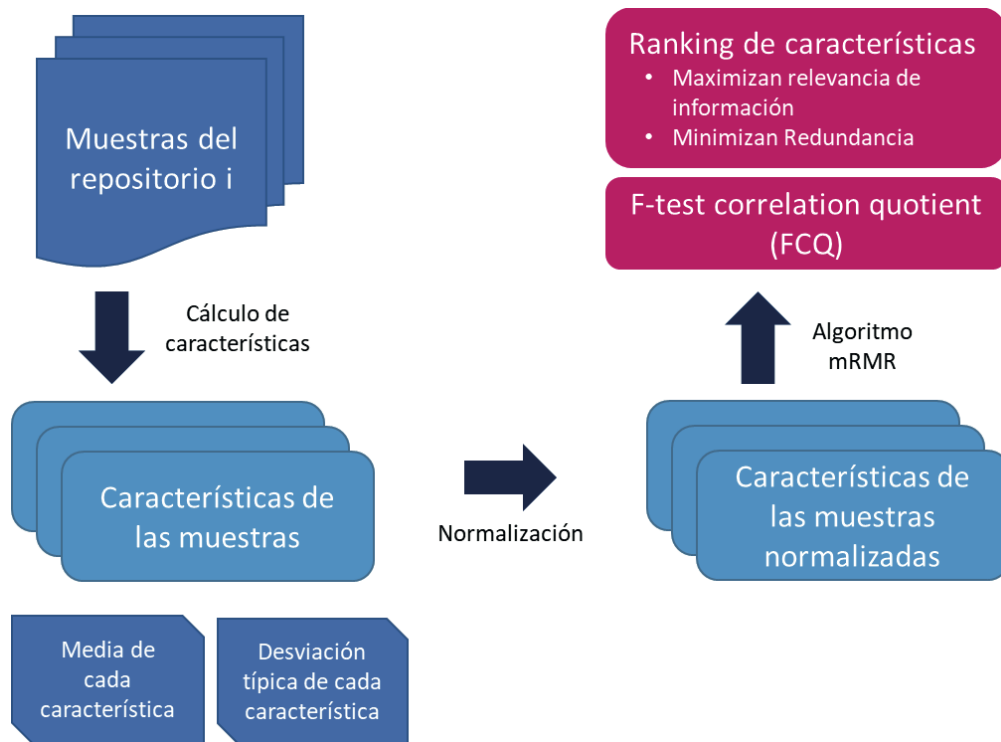


Figura B.1 – Metodología utilizada para el estudio del algoritmo mRMR

	Posición en el ranking de mRMR											
	A	B	C	D	E	F	G	H	I	J	K	L
	μ_{SMV}	$A_{\omega_{diff(max)}}$	σ_{SMV}	μ_{θ}	$\mu_{SMV(diff)}$	μ_{Ap}	SMV_{Max}	SMV_{Min}	γ_{SMV}	SMA	E	μ_R
DLR [7]	6	10	7	5	2	4	11	12	3	9	1	8
DOFDA [8]	7	1	11	6	2	4	8	3	5	10	9	12
Erciyes [9]	10	6	9	8	2	3	1	5	4	12	7	11
IMUFD [10]	5	9	1	6	4	3	7	11	2	8	10	12
SisFall [11]	12	1	11	8	2	3	10	6	4	5	9	7
Up-Fall [12]	10	8	1	4	9	11	7	6	3	12	5	2
UMAFall [13]	11	1	9	6	2	3	8	5	4	10	7	12

Tabla B.1. Ranking obtenido a partir de mRMR al analizar las características utilizadas para describir los movimientos, posteriormente serán empleadas como *inputs* de entrada del clasificador. El color verde más intenso significa una mejor clasificación en el ranking, por el contrario, el color rojo representa una peor posición.

	Predictor importance score (FCQ) mRMR											
	A	B	C	D	E	F	G	H	I	J	K	L
	μ_{SMV}	$A_{\omega_{diff(max)}}$	σ_{SMV}	μ_{θ}	$\mu_{SMV(diff)}$	μ_{Ap}	SMV_{Max}	SMV_{Min}	γ_{SMV}	SMA	E	μ_R
DLR [7]	1.15E-03	6.32E-04	9.68E-04	1.25E-03	1.28E-02	4.39E-03	6.03E-04	4.73E-04	1.07E-02	7.74E-04	4.68E-02	9.40E-04
DOFDA [8]	1.37E-02	3.76E-01	9.30E-03	2.08E-02	6.65E-02	2.92E-02	1.26E-02	5.60E-02	2.64E-02	1.08E-02	1.13E-02	4.14E-04
Erciyes [9]	6.85E-03	8.66E-03	7.33E-03	8.01E-03	1.98E-02	1.67E-02	5.56E-01	1.25E-02	1.25E-02	2.71E-03	8.54E-03	5.43E-03
IMUFD [10]	2.34E-01	2.00E-01	4.84E-01	2.18E-01	3.63E-01	4.08E-01	2.10E-01	1.44E-01	4.22E-01	2.08E-01	1.90E-01	7.99E-02
SisFall [11]	3.57E-03	3.72E-01	4.78E-03	6.52E-03	2.81E-02	2.43E-02	5.63E-03	7.55E-03	1.23E-02	7.73E-03	5.63E-03	6.63E-03
Up-Fall [12]	3.08E-03	8.11E-03	5.88E-01	1.10E-02	6.12E-03	0.00E+00	8.40E-03	1.08E-02	2.47E-02	0.00E+00	1.08E-02	3.02E-02
UMAFall [13]	8.61E-07	3.78E-01	1.50E-04	2.17E-04	9.35E-04	7.91E-04	1.73E-04	2.41E-04	4.14E-04	1.19E-04	1.84E-04	8.01E-05

Tabla B.2. FCQ obtenido a partir de mRMR al analizar las características utilizadas para describir los movimientos y que posteriormente serán empleadas como *inputs* de entrada del clasificador. El color verde indica un mayor FCQ, el rojo significa lo contrario.

B.3.1.1. Conclusiones

Los resultados obtenidos muestran que la heterogeneidad de los repositorios provoca que el algoritmo mRMR considere más relevantes distintas características dependiendo del caso.

Sin embargo, pueden observarse patrones comunes en cada una de las experiencias:

- $\mu_{SMV_{(diff)}}(E)$, $\mu_{Ap}(F)$ y $\gamma_{SMV}(I)$ se clasifican en la mayoría de los casos en los 5 mejores puestos del ranking.
- $\mu_{SMV}(A)$, $SMA(J)$ y $\mu_R(L)$ son las características que peores posiciones ocupan en el ranking y los que peores FCQ obtienen.
- $A_{\omega_{diff(max)}}(B)$ y $\sigma_{SMV}(C)$ son las características que han obtenido el primer puesto un mayor número de veces. Sin embargo, parece ocurrir que cuando una de ellas es seleccionada como la más relevante, la otra pierde importancia, ubicándola en los últimos puestos de la clasificación. Hecho que se observa claramente en los valores de FCQ. Por lo que podría entenderse como que ambas características están fuertemente relacionadas y el uso de una es redundante con la otra.

B.3.2. Comprobación ANOVA

Con el fin de evaluar el proceso de selección de características de entrada para los algoritmos de aprendizaje supervisado se realiza un análisis de varianza factorial (ANOVA de N-Way) del tipo 2^k . Dicho análisis se lleva a cabo a partir de los datos obtenidos al clasificar las diferentes muestras de los repositorios estudiados utilizando las diferentes combinaciones de características de entrada.

- En este estudio se utiliza un total de $k = 12$ características de entrada, que corresponden a los factores de estudio del ANOVA.
- Cada factor tiene dos niveles:
 - Activado, es decir, se utiliza como input de entrada del sistema de clasificación.

- Desactivado, no se tiene en cuenta a la hora de clasificar.
- Se realizan 6 réplicas completas del experimento, investigando de este modo todas las combinaciones posibles de los factores.
- Durante todo el estudio los diseños experimentales son completamente aleatorizados y se satisfacen los supuestos de normalidad.
- El *kernel* utilizado para el clasificador SVM es del tipo lineal, predeterminado en Matlab para el aprendizaje de dos clases, como es el caso de este estudio.

Con el ANOVA factorial se logrará comprobar si la utilización de una característica provoca un efecto significativo en la sensibilidad y especificidad del sistema y nos aportará información sobre la importancia de dicho efecto. Para ello, la métrica de calidad que estudiamos es la media geométrica de la sensibilidad y la especificidad.

$$\text{métrica de calidad} = \sqrt{Se \cdot Sp} \quad (7)$$

Donde *Se* y *Sp* corresponde a la sensibilidad y especificidad del sistema al clasificar las muestras correspondientes respectivamente.

La metodología seguida para cada uno de los repositorios de movimientos es la siguiente:

- Una parte del repositorio ha sido utilizada como datos de entrenamiento para los algoritmos de aprendizaje supervisado. Este *subset* de entrenamiento contendrá movimientos pertenecientes a todos los tipos contenidos en el repositorio completo. La selección de los movimientos se realiza de forma aleatoria.
- El resto de los datos disponibles se utiliza como muestras de testeo. Se han separado en diferentes bloques o replicas, concretamente en 6, de forma aleatoria y equitativamente según el tipo de movimiento; de esta forma cada grupo se analiza de manera independiente obteniendo la especificidad (*Sp*) y sensibilidad (*Se*) de las diferentes combinaciones características de manera que sea posible realizar posteriormente un

análisis ANOVA factorial. En la tabla Tabla B.3 se resume la composición de cada una de las particiones de los repositorios.

- Todas las características de las muestras, tanto las utilizadas en entrenamiento como las utilizadas en el testeo, son normalizadas tomando como referencia la media y varianza de las características de todas las muestras del repositorio.

Repositorio	Dataset Entrenamiento	Dataset estudio I	Dataset estudio II	Dataset estudio III	Dataset estudio IV	Dataset estudio V	Dataset estudio VI	TOTAL
DLR	158 29.10%	63 11.60%	62 11.42%	65 11.97%	66 12.15%	65 11.97%	64 11.79%	543 100.00%
DOFDA	186 51.24%	30 8.26%	30 8.26%	31 8.54%	30 8.26%	28 7.71%	28 7.71%	363 100.00%
Erciyes	216 6.56%	512 15.56%	514 15.62%	514 15.62%	511 15.53%	512 15.56%	512 15.56%	3291 100.00%
IMUFD	210 35.59%	63 10.68%	62 10.51%	63 10.68%	64 10.85%	64 10.85%	64 10.85%	590 100.00%
SisFall	204 4.55%	713 15.91%	711 15.86%	712 15.89%	712 15.89%	715 15.95%	715 15.95%	4482 100.00%
Up-Fall	209 41.97%	50 10.04%	46 9.24%	47 9.44%	49 9.84%	48 9.64%	49 9.84%	498 100.00%
UMAFall	184 31.62%	65 11.17%	67 11.51%	66 11.34%	66 11.34%	67 11.51%	67 11.51%	582 100.00%

Tabla B.3. Composición de los *datasets* creados a partir de cada repositorio, se indica el número total de muestras y el porcentaje del total al que corresponde.

B.3.2.1. Tamaño del efecto η^2

El tamaño del efecto de un ANOVA es el valor que permite medir qué proporción de la varianza en el resultado es consecuencia de la influencia de cada factor de estudio, es decir, cuánto afecta un factor o interacción de factores en el resultado final.

$$\eta_r^2 = \frac{SS_r}{SS_{Total}} \quad (8)$$

Donde:

- SS_r es la suma de cuadrados de la combinación de factores (r) que se quiere estudiar
- SS_{Total} es la suma de cuadrados total, el cual incluye todos los grupos de factores estudiados y el error.

Para evaluar la magnitud de los efectos obtenidos se suele adoptar como criterios los propuestos por Cohen en [14]. Dichos niveles de clasificación son:

Tamaño de Efecto	Valoración
$0.01 \leq \eta_r^2 \leq 0.09$	Efecto bajo
$0.09 \leq \eta_r^2 \leq 0.25$	Efecto medio
$0.25 \leq \eta_r^2$	Efecto alto

Tabla B.4. Criterios de valoración del tamaño de efecto de un estudio según el trabajo de Cohen [14].

Para calcular la suma de cuadrados de un factor o combinación se deberá conocer su efecto. El efecto de un factor o combinación se define como la variación en el resultado final producido por un cambio en el nivel del factor o combinación. [15]

$$Efecto = \frac{\sum_{i=1}^{2^k} c_i \bar{y}_i}{2^{k-1}} = \frac{Contraste}{n \cdot 2^{k-1}} \quad (9)$$

$$SS_r = \frac{(Contraste_r)^2}{n \cdot 2^k} = n \cdot 2^{k-2} efecto^2 \quad (10)$$

Donde:

- n denota el número de réplicas
- k denota el número total de factores estudiados
- c_i corresponde a una constante que tendrá valor +1 o -1 dependiendo de la interacción que se está estudiando y la combinación a la que precede, este valor puede obtenerse fácilmente creando una matriz de signos de interacciones.
- $\bar{y}_i$ corresponde al valor medio de la métrica estudiada (media geométrica de la sensibilidad y especificidad del sistema) para las n replicas utilizando la combinación de factores i -ésima.

Ejemplo con un estudio de 3 factores ($k = 3$):

- Existirán $2^k = 2^3 = 8$ combinaciones de factores.

Si se denominan a los factores como A, B y C, las combinaciones serán:

1. A
2. B
3. C
4. AB
5. AC
6. BC
7. ABC
8. (-) Todos los factores inactivos

En el caso del cálculo de SS_A :

$$SS_A = n \cdot \frac{(\sum_{i=1}^{2^3} c_i \bar{y}_i)^2}{2^3} = \frac{n}{8} \cdot (\bar{y}_A + \bar{y}_{AB} + \bar{y}_{AC} + \bar{y}_{ABC} - \bar{y}_B + \bar{y}_C - \bar{y}_{BC} - \bar{y}_-)$$

En el caso del cálculo de SS_{AB} :

$$SS_{AB} = n \cdot \frac{(\sum_{i=1}^{2^3} c_i \bar{y}_i)^2}{2^3} = \frac{n}{8} \cdot (\bar{y}_{AB} + \bar{y}_{ABC} + \bar{y}_- + \bar{y}_C - \bar{y}_A - \bar{y}_B - \bar{y}_{AC} - \bar{y}_{BC})$$

Seguidamente, en las tablas 5-7, se expone el tamaño de efecto en ANOVA de los efectos principales, las interacciones de primer y segundo orden respectivamente.

	Tamaño de efecto en ANOVA – Efectos Principales												ERROR
	A	B	C	D	E	F	G	H	I	J	K	L	
	μ_{SMV}	$A_{\omega_{diff(max)}}$	σ_{SMV}	μ_{θ}	$\mu_{SMV_{(diff)}}$	μ_{Ap}	SMV_{Max}	SMV_{Min}	γ_{SMV}	SMA	E	μ_R	
DLR	1.69%	2.97%	0.21%	0.03%	0.18%	0.66%	2.53%	0.70%	3.81%	0.17%	3.93%	0.06%	68.88%
DOFDA	0.79%	2.72%	0.32%	0.67%	0.24%	0.08%	0.35%	0.05%	0.17%	0.33%	0.43%	0.00%	61.53%
Erciyes	0.31%	8.24%	3.04%	5.47%	0.01%	0.71%	4.36%	1.06%	3.11%	0.09%	0.00%	0.00%	32.48%
IMUFD	0.28%	0.26%	0.91%	0.16%	0.01%	0.83%	0.05%	0.02%	0.01%	0.03%	1.33%	0.00%	77.43%
SisFall	0.68%	5.84%	2.19%	0.28%	0.00%	31.52%	2.50%	0.49%	5.71%	2.52%	2.87%	0.09%	16.74%
Up-Fall	0.26%	10.89%	3.07%	3.41%	0.01%	0.03%	4.79%	3.32%	3.81%	0.00%	1.86%	0.21%	24.61%
UMAFall	0.18%	9.85%	1.60%	0.95%	0.00%	7.37%	2.28%	3.58%	3.16%	0.05%	4.17%	0.00%	23.35%

Tabla B.5. Tamaño del efecto de los factores principales del ANOVA realizado a cada repositorio estudiado. Se hace uso de los datos del sensor ubicado en la cintura y el algoritmo de aprendizaje supervisado SVM

Tamaño de efecto en ANOVA – Interacciones de primer orden												
	B*C	B*F	B*G	B*H	B*I	C*G	C*H	C*K	G*H	G*I	F*I	F*J
DLR												
DOFDA												
Erciyes	2.86%		3.11%	1.47%	1.93%	2.75%	1.30%		1.35%	2.20%		
IMUFD								1.26%				
SisFall	1.21%	1.19%	2.45%			1.03%				2.30%		1.53%
Up-Fall	1.91%		4.31%	2.11%	2.19%	2.18%	1.51%		1.68%	1.89%		
UMAFall		3.89%	1.80%	1.37%	3.13%					1.48%	2.29%	

Tabla B.6. Tamaño del efecto de las interacciones de primer orden con una valoración mínima de efecto bajo del ANOVA realizado a cada repositorio estudiado.

Tamaño de efecto en ANOVA – Interacciones de segundo orden									
	B*C*G	B*C*H	B*C*I	B*F*G	B*F*I	B*G*I	B*G*H	C*G*H	C*G*I
DLR									
DOFDA									
Erciyes	2.67%	1.42%	1.17%			1.54%	1.38%	1.30%	1.43%
IMUFD									
SisFall	1.03%					1.98%			
Up-Fall	1.99%	1.63%				2.03%	1.78%	1.66%	
UMAFall				1.18%	1.84%	1.91%			

Tabla B.7. Tamaño del efecto de las interacciones de primer orden con una valoración mínima de efecto bajo del ANOVA realizado a cada repositorio estudiado.

B.3.2.2. Conclusiones

En primer lugar, observamos que en el análisis ANOVA el error, esto es, el efecto provocado por factores diferentes a la combinación de las características utilizadas es bastante alto. Esto se debe por un lado a la variabilidad introducida por el clasificador (SVM), ya sea causada por su propia parametrización o por su incapacidad intrínseca para clasificar de la misma forma muestras que, aunque son del mismo tipo, al analizarlas son detectadas como movimientos diferentes. Además, existen otros parámetros en los repositorios que pueden provocar desviaciones en los resultados como por ejemplo la existencia de desigualdades entre los *subsets* de testeo provocados por las características (peso, edad, altura, etc.) de los voluntarios que realizan los movimientos o por diferencias entre los movimientos simulados incluidos en cada *dataset*. Sin embargo, el análisis puede utilizarse como referencia para intuir diferentes patrones similares en los diferentes repositorios.

- $A_{\omega_{diff(max)}}$ (B) parece ser la característica que aporta un mayor efecto en los resultados finales. Teniendo en cuenta los criterios propuestos por Cohen [14], el efecto de este factor se califica 2 veces como medio y 4 veces como bajo, por lo tanto, solo en uno de los repositorios el efecto es despreciable.
- Los factores SMV_{Max} (G), E (K) y γ_{SMV} (I) también parecen cobrar importancia, consiguiendo cada uno de ellos un efecto bajo en 5 de los 7 repositorios.
- Las interacciones entre los factores comentados en los dos puntos anteriores son también destacables, obteniendo en gran parte de los casos una valoración de efecto bajo. Por lo que además del efecto aportado por ellos de manera individual, su combinación favorece de manera sinérgica el comportamiento del clasificador.
- Las características σ_{SMV} (C), μ_{θ} (D) y SMV_{Min} (H) también pueden ser interesantes, aunque no parecen ser relevantes en algunos de los repositorios estudiados, también obtienen una calificación de efecto bajo y consiguen interacciones de efecto bajo al combinarse con otras características.

- $\mu_{SMV}(A)$, $\mu_{SMV(diff)}(E)$, $SMA(J)$ y $\mu_{SMV(diff)}(L)$ no parecen aportar ningún efecto al resultado final. Ambas características parecen no ser significativas. Estos resultados coinciden con los extraídos del análisis del algoritmo mRMR, exceptuando el caso de $\mu_{SMV(diff)}(E)$, el cual difiere totalmente, ya que era una de las características con mejores resultados en el otro análisis.
- $\mu_{Ap}(F)$, parecen aportar positivamente al efecto en algunos casos concretos, consiguiendo en uno de los estudios un efecto clasificado como alto.

B.3.3. Resultados de la clasificación de los movimientos utilizando SVM

En este apartado se recopilan algunos de los valores de la media geométrica de la sensibilidad y la especificidad resultante de aplicar diferentes combinaciones de características utilizadas como parámetros de entrada de los algoritmos de aprendizaje supervisado, se han seleccionado aquellas combinaciones que según los análisis anteriores podrían ser más interesantes.

Para seleccionar las combinaciones de características que estudiaremos con el fin de tratar de obtener los mejores resultados del sistema se seguirán los siguientes criterios.

B.3.3.1. Combinaciones seleccionadas a partir de mRMR

Por un lado, se analizarán como *inputs* del sistema las combinaciones de las características que hayan obtenido un mejor puesto en el ranking de mRMR para cada repositorio. Por tanto, para una combinación de dimensión N, esto es, una combinación compuesta por N características, se utilizarán las primeras N características de los rankings (Tabla B.1). Estas combinaciones, se remarcarán con un borde de color morado en las tablas de resultados que se mostrarán más adelante (Tablas B.8 -B.16).

Así mismo, se ha creado el siguiente ranking según la información extraída del análisis global. El orden en la clasificación corresponderá con la prioridad que tendrá esa característica para ser añadida a la combinación de inputs de entrada. Este ranking se utilizará a la hora de fusionar los resultados del análisis mRMR y ANOVA.

1. $\mu_{SMV(diff)}$ (E) – Posteriormente se comprobará que esta característica no afecta significativamente.
2. γ_{SMV} (I)
3. μ_{Ap} (F)
4. $A_{\omega_{diff(max)}}$ (B)
5. σ_{SMV} (C),
6. SMV_{Min} (H), μ_{θ} (D)

Las características μ_{SMV} (A), SMV_{Max} (G), SMA (J), E (K) y $\mu_{SMV(diff)}$ (L) se desechan.

B.3.3.2. Combinaciones seleccionadas a partir de ANOVA

A partir de la información extraída de los ANOVA realizados para cada repositorio se ha creado el siguiente ranking. Se ha tenido en cuenta el tamaño de efecto obtenido para cada característica y las tendencias observadas con los diferentes repositorios. El orden en la clasificación corresponderá con la prioridad que tendrá esa característica para ser añadida a la combinación de inputs de entrada al igual que se ha explicado en el apartado anterior. Estas combinaciones, se remarcarán con un borde de color rosa en las tablas de resultados que se mostrarán a continuación (Tablas B.8 -B.16).

1. $A_{\omega_{diff(max)}}$ (B)
2. γ_{SMV} (I), SMV_{Max} (G) y E (K)
3. μ_{Ap} (F)
4. σ_{SMV} (C), SMV_{Min} (H), μ_{θ} (D)

Las características μ_{SMV} (A), $\mu_{SMV(diff)}$ (E), SMA (J) y $\mu_{SMV(diff)}$ (L) se desechan.

B.3.3.3. Combinaciones seleccionadas a partir de ANOVA y mRMR

También se estudiarán las combinaciones resultantes de fusionar ambos estudios. Estas combinaciones contendrán las características que sean comunes como relevantes en ambos estudios. Para ello se hará uso de los rankings globales que se han mostrado anteriormente.

Estas combinaciones, se remarcarán con un borde de color azul en las tablas de resultados que se mostrarán a continuación (Tablas B.8 -B.16).

NOTA: Los malos resultados obtenidos en el análisis del repositorio DLR se deben a que éste contiene un número muy bajo de movimientos del tipo caída, por lo que se consiguen especificidades muy altas, pero una sensibilidad nula en la mayoría de los casos. Lo que provoca que la media geométrica de la sensibilidad y especificidad tenga valores nulos o muy bajos.

		DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
A	μ_{SMV}	0.00%	84.97%	77.44%	89.08%	34.41%	0.00%	0.00%
B	$A_{\omega_{diff(max)}}$	0.00%	98.98%	94.18%	90.10%	82.88%	87.39%	59.57%
C	σ_{SMV}	0.00%	96.54%	91.85%	93.30%	77.90%	87.87%	0.00%
D	μ_{θ}	0.00%	84.62%	78.66%	86.71%	53.55%	75.22%	0.00%
E	$\mu_{SMV(diff)}$	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
F	μ_{Ap}	0.00%	0.00%	81.50%	83.36%	80.99%	0.00%	80.25%
G	SMV_{Max}	0.00%	97.57%	94.75%	86.25%	79.77%	86.31%	33.89%
H	SMV_{Min}	0.00%	62.19%	88.38%	63.64%	75.22%	84.52%	0.00%
I	γ_{SMV}	0.00%	88.10%	89.85%	78.52%	86.08%	85.09%	76.73%
J	SMA	0.00%	73.28%	59.01%	84.93%	70.79%	0.00%	0.00%
K	E	0.00%	97.55%	0.00%	93.50%	60.62%	0.00%	0.00%
L	μ_R	0.00%	84.97%	77.44%	89.08%	34.41%	0.00%	0.00%

Tabla B.8. Media aritmética de la media geométrica de la sensibilidad y especificidad obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 1 sola característica. Se remarca con color rosa el borde de los resultados de las combinaciones que según el análisis ANOVA hemos concluido que podría tener un mejor efecto en los resultados y en morado la combinación que podría tener mejores resultados según los resultados de mRMR

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
EK	0.00%	97.93%	0.00%	92.94%	62.37%	0.00%	0.00%
BE	0.00%	99.31%	93.92%	89.51%	83.46%	87.39%	72.18%
EG	0.00%	97.57%	94.35%	86.25%	81.19%	86.31%	47.11%
CI	0.00%	97.89%	94.44%	93.38%	86.75%	85.39%	77.98%
CL	0.00%	96.54%	91.85%	93.30%	77.93%	87.57%	0.00%
BI	0.00%	99.31%	96.48%	96.48%	87.30%	86.19%	79.02%
BG	0.00%	97.89%	95.23%	95.23%	79.98%	86.48%	64.40%
BK	16.67%	98.98%	94.18%	94.18%	88.46%	96.50%	96.07%

Tabla B.9. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 2 características. Se remarca con color rosa el borde de los resultados de las combinaciones que según el análisis ANOVA hemos concluido que podría tener un mejor efecto en los resultados y en morado la combinación que podría tener mejores resultados según los resultados de mRMR

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
EIK	0.00%	97.57%	90.16%	92.81%	86.36%	85.70%	78.65%
BEH	0.00%	98.24%	93.53%	89.88%	83.38%	88.35%	65.22%
EFG	0.00%	97.26%	96.14%	89.21%	88.66%	85.11%	77.72%
CFI	0.00%	97.89%	95.85%	92.56%	92.65%	87.85%	86.24%
BEF	0.00%	99.65%	94.62%	91.20%	89.30%	86.78%	81.15%
CIL	0.00%	97.89%	94.33%	93.38%	86.02%	86.52%	76.84%
BGI	0.00%	97.89%	96.55%	96.55%	87.77%	94.03%	90.05%
BGK	16.67%	97.59%	95.27%	95.27%	88.39%	96.50%	94.82%
BIK	49.87%	98.98%	96.48%	96.48%	89.44%	97.18%	93.87%
BFI	0.00%	99.65%	96.11%	92.00%	92.70%	87.58%	86.58%

Tabla B.10. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 3 características. Se remarca con color rosa el borde de los resultados de las combinaciones que según el análisis ANOVA hemos concluido que podría tener un mejor efecto en los resultados, en morado la combinación que podría tener mejores resultados según los resultados de mRMR y en azul la combinación obtenida al combinar ambos resultados.

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
EFIK	0.00%	97.57%	91.98%	94.01%	92.37%	86.02%	86.80%
BEFH	0.00%	98.24%	94.01%	91.54%	90.73%	85.82%	86.93%
EFGI	0.00%	97.57%	96.00%	91.31%	92.90%	88.60%	86.24%
CEFI	0.00%	97.89%	95.70%	92.57%	92.76%	88.21%	86.24%
BEFI	0.00%	99.65%	96.11%	92.00%	92.69%	87.94%	86.41%
CDIL	0.00%	97.57%	96.93%	92.98%	85.99%	90.23%	79.25%
BGIK	49.87%	97.59%	96.55%	96.55%	89.11%	97.18%	94.06%
BCFI	0.00%	97.89%	97.89%	93.08%	93.19%	94.70%	89.71%

Tabla B.11. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 4 características. Se remarca con color rosa el borde de los resultados de las combinaciones que según el análisis ANOVA hemos concluido que podría tener un mejor efecto en los resultados, en morado la combinación que podría tener mejores resultados según los resultados de mRMR y en azul la combinación obtenida al combinar ambos resultados.

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
DEFIK	0.00%	97.57%	92.26%	93.49%	92.72%	85.33%	88.27%
BEFHI	0.00%	97.89%	96.30%	92.21%	93.36%	89.78%	89.27%
EFGHI	0.00%	96.13%	96.14%	90.92%	93.68%	92.86%	93.76%
ACEFI	0.00%	97.89%	95.75%	92.21%	94.76%	90.20%	88.55%
BEFIJ	49.74%	99.65%	96.52%	92.00%	93.26%	93.63%	88.59%
CDIKL	49.74%	97.57%	96.92%	92.78%	86.77%	97.40%	88.06%
BFGIK	49.87%	97.89%	96.30%	96.30%	94.43%	97.18%	95.17%
BCDFI	0.00%	97.89%	97.89%	93.26%	94.02%	96.69%	89.83%
BCFHI	16.54%	97.89%	97.89%	93.14%	93.97%	97.71%	95.14%

Tabla B.12. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 5 características. Se remarca con color morado el borde de los resultados de las combinaciones que mRMR indicaba como mejores para cada repositorio.

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
BCFGIK	49.87%	97.89%	97.89%	92.16%	96.32%	97.20%	95.33%
BDFGIK	49.87%	97.89%	97.89%	92.70%	94.56%	97.76%	95.86%
BFGHIK	50.00%	97.89%	97.89%	92.34%	95.37%	98.57%	96.11%
BCDFHI	33.20%	97.89%	97.89%	93.14%	94.17%	98.54%	94.96%

Tabla B.13. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 6 características. Se remarca con color morado el borde de los resultados de las combinaciones que mRMR indicaba como mejores para cada repositorio.

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
BCDFGIK	49.87%	97.89%	98.13%	92.56%	96.36%	98.16%	95.31%
BCFGHIK	50.00%	97.89%	96.34%	92.37%	95.91%	98.20%	95.77%
BDFGHIK	50.00%	97.89%	97.73%	92.55%	95.26%	99.65%	95.94%

Tabla B.14. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 7 características. Se remarca con color morado el borde de los resultados de las combinaciones que mRMR indicaba como mejores para cada repositorio.

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
BCDFGHIK	49.87%	97.89%	97.72%	92.55%	95.89%	99.29%	95.78%

Tabla B.15. Algunas de las medias aritméticas de la media geométrica de la sensibilidad y especificidad ($\sqrt{Sp \cdot Se}$) obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM 8 características. Se remarca con color morado el borde de los resultados de las combinaciones que mRMR indicaba como mejores para cada repositorio.

	DLR	DOFDA	Erciyes	IMUFD	SisFall	Up-Fall	UMAFall
ABCDEFGHIIJKL	66.67%	99.65%	92.12%	92.12%	95.93%	98.55%	95.46%

Tabla B.16. Media aritmética de la media geométrica de la sensibilidad y especificidad obtenida por los bloques de experimentos realizados utilizando como input del algoritmo SVM las 12 características estudiadas.

B.3.4. La característica $\mu_{SMV(diff)}$ (E)

Como se puede observar en la Tabla B.12 y ya se comentó en anteriores apartados, mRMR cataloga a la característica $\mu_{SMV(diff)}$ (E) como una de las más importantes. Sin embargo, el análisis ANOVA nos indica que el efecto que produce en el resultado es insignificante. A continuación, compararemos esas combinaciones respecto a aquellas que no incluyen dicha característica.

	Combinación recomendada por mRMR	Resultado $\sqrt{Sp \cdot Se}$	Combinación recomendada por mRMR sin E	Resultado $\sqrt{Sp \cdot Se}$
DLR	EIK	0	IK	0
DOFDA	BEF	99.65%	BF	99.65%
ERCIYES	EFI	91.85%	FI	92.12%
IMUFD	CFI	92.56%	CFI	92.56%
SisFall	BEF	89.30%	BF	89.07%
UP-Fall	CIL	86.52%	CIL	86.52%
UMAFall	BEF	81.15%	BF	81.19%

Tabla B.17. Comparación de los resultados obtenidos con las combinaciones de 3 factores recomendadas por mRMR respecto a dichas combinaciones sin la utilización de la característica $\mu_{SMV(diff)}$ (E).

	Combinación recomendada por mRMR	Resultado $\sqrt{Sp \cdot Se}$	Combinación recomendada por mRMR sin E	Resultado $\sqrt{Sp \cdot Se}$
DLR	EFIK	0	FIK	0
DOFDA	BEFH	98.24%	BFH	98.24%
ERCIYES	EFGI	96.00%	FGI	95.93%
IMUFD	CEFI	92.57%	CFI	92.56%
SisFall	BEFI	92.69%	BFI	92.70%
UP-Fall	CDIL	90.23%	CDIL	90.23%
UMAFall	BEFI	86.41 %	FGHI	86.58 %

Tabla B.18. Comparación de los resultados obtenidos con las combinaciones de 4 factores recomendadas por mRMR respecto a dichas combinaciones sin la utilización de la característica $\mu_{SMV(diff)}$ (E).

	Combinación recomendada por mRMR	Resultado $\sqrt{Sp \cdot Se}$	Combinación recomendada por mRMR sin E	Resultado $\sqrt{Sp \cdot Se}$
DLR	DEFIK	0	DFIK	0
DOFDA	BEFHI	97.89%	BFHI	98.24%
ERCIYES	EFGHI	96.14%	FGHI	96.14%
IMUFD	ACEFI	92.21%	ACFI	92.56%
SisFall	BEFIJ	93.26%	BFIJ	93.12%
UP-Fall	CDIKL	97.40%	CDIKL	97.40%
UMAFall	BEFHI	89.27%	BFHI	88.88%

Tabla B.19. Comparación de los resultados obtenidos con las combinaciones de 5 factores recomendadas por mRMR respecto a dichas combinaciones sin la utilización de la característica $\mu_{SMV(diff)}$ (E).

Como se observa en las tablas B.17-B.19 la exclusión de la característica $\mu_{SMV(diff)}$ (E) no repercute de forma significativa en la media de $(\sqrt{Sp \cdot Se})$, tal y como indica el análisis ANOVA, el cual señalaba a dicha característica como no significativa.

B.3.5. Selección de las combinaciones con mejores resultados

Observando las tablas Tablas B.8 -B.16 se puede intuir cuáles son las combinaciones de características de entrada del sistema que permiten discernir entre caídas y ADL con una mayor sensibilidad y especificidad. Sin embargo, estos valores no nos permiten a simple vista demostrar las diferencias estadísticas existentes, por ello, se ha realizado un análisis ANOVA para cada uno de los repositorios, haciendo uso de los valores, ya calculados en el apartado B.3.2, de la media geométrica de la sensibilidad y especificidad obtenida para las diferentes combinaciones de entrada para cada uno de los 6 *subsets* de testeo. Los resultados se encuentran gráficamente representados mediante una prueba de múltiple comparación en la Figura B.2.

Los resultados del algoritmo mRMR nos permiten seleccionar, con un coste computacional mucho menor que un ANOVA factorial, algunas características que según podemos comprobar dan lugar a resultados bastante aceptables. No

obstante, las características y orden de importancia que el algoritmo recomienda para cada repositorio varían drásticamente. Este hecho provoca que, en determinadas ocasiones, al utilizar específicas combinaciones recomendadas, el sistema funcione con altas prestaciones al analizar cierto repositorio, pero su rendimiento baje drásticamente al probar con otro. Debido a que el objetivo de este estudio es conseguir un modelo de detección de caídas robusto y flexible, se opta por desechar dichas combinaciones.

Para esta experiencia se han seleccionado las combinaciones derivadas del ANOVA realizado en la sección B.3.2 y también las composiciones resultantes de fusionar la información de esta sección mencionada junto con los resultados del algoritmo mRMR (sección B.3.1). En ambas ocasiones, se analizan las combinaciones que tienen al menos 3 características, ya que son los casos donde podemos observar que en la mayoría de los repositorios se obtiene una métrica de la media geométrica de la sensibilidad y la especificidad superior o cercana al 90%.

Para diferenciar de una manera más sencilla cada combinación, se ha asignado un identificador a cada una de ellas, dicha información se recoge en la Tabla B.20.

ID	Combinaciones
1	BFI
2	BGI
3	BGK
4	BIK
5	BGIK
6	BCFI
7	BFGIK
8	BCDFI
9	BCFHI
10	BCFGIK
11	BDFGIK
12	BFGHIK
13	BCDFHI
14	BCDFGIK
15	BCFGHIK
16	BDFGHIK
17	BCDFGHIK
18	ABCDEFGHIIJKL

Tabla B.20. Combinaciones seleccionadas como posibles mejores junto al identificador asignado.

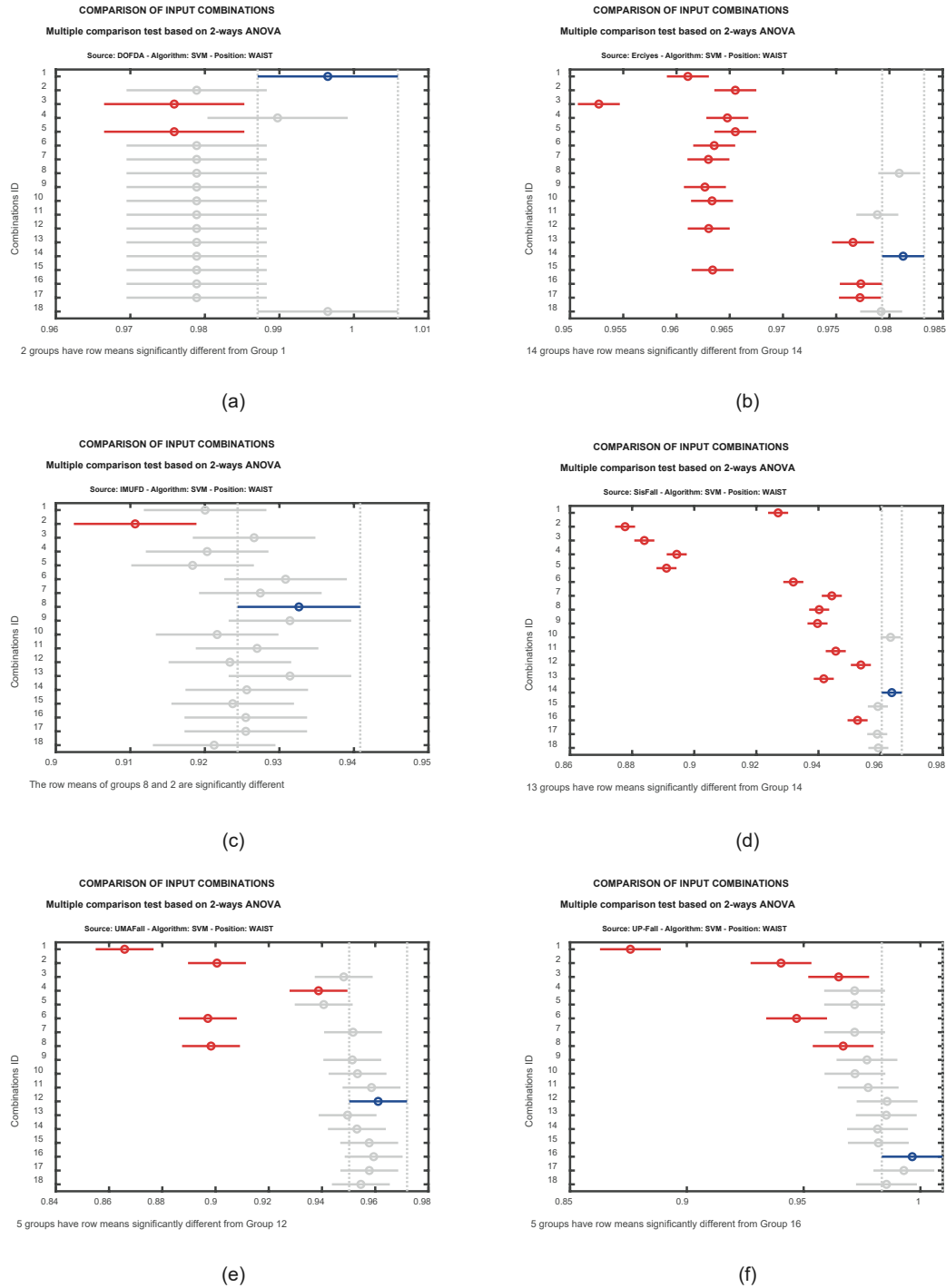


Figura B.2. Representación gráfica del test de comparación múltiple realizado con los resultados del ANOVA 2-way para comprobar la significancia estadística existente entre los resultados obtenidos por el sistema al utilizar diferentes características de entrada y el algoritmo *Support Vector Machine* (SVM). Se ha utilizado un $\alpha = 0.05$ y como tipo de valor crítico LSD (Fisher's Least Significant Difference) para cada repositorio.

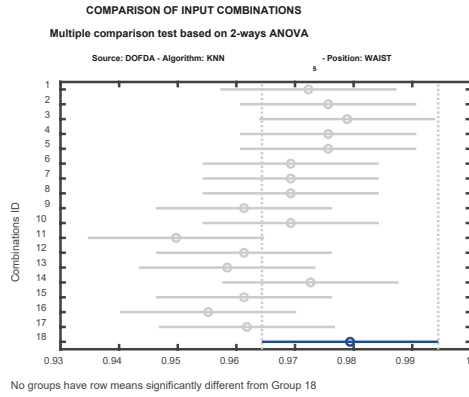
Tal y como se demuestra en las gráficas de la Figura B.2, las combinaciones formadas por 5 o más de las características seleccionadas como más representativas en gran parte de las experiencias no son significativamente diferentes las unas de las otras. El caso peor ocurre al analizar los repositorios Erciyes y SisFall, donde la media de la combinación que consigue mejor

rendimiento difiere significativamente de un número mayor de combinaciones. En esta experiencia destaca la combinación cuya ID es 18 (la combinación que incluye todas las características estudiadas) ya que nunca difiere significativamente de la combinación que aporta mayor rendimiento, pero de una manera aún más relevante la combinación con ID 14, ya que incluyendo solamente 7 de las 12 combinaciones estudiadas, es junto a la 18 la única que se mantiene como no significativamente diferente a la que mejor rendimiento proporciona en cada caso.

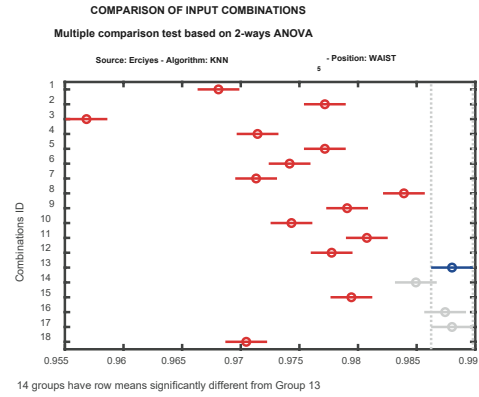
B.3.6. Utilización del algoritmo de aprendizaje supervisado k-NN

En este apartado se utilizan los mismos *subsets* de entrenamiento y testeo que han sido empleados hasta ahora, así como, las combinaciones de las características de entrada que se utilizaron en el apartado anterior, al considerarlas como supuestamente más recomendables. Sin embargo, en lugar de utilizar el algoritmo de aprendizaje supervisado *Support Vector Machine* (SVM) se hará uso de otro algoritmo de aprendizaje supervisado, en concreto de *k-Nearest Neighbours* (k-NN). El objetivo es analizar el comportamiento que tiene este algoritmo al utilizar las características seleccionadas para ser utilizadas con SVM.

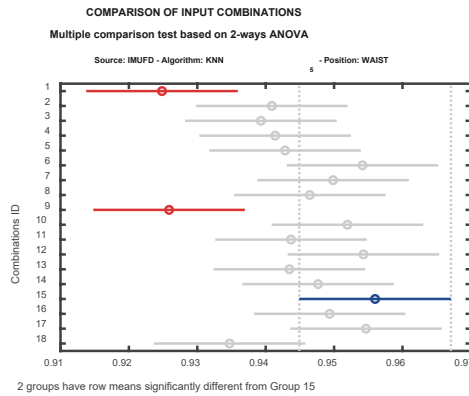
B.3.6.1. k-NN con k = 5



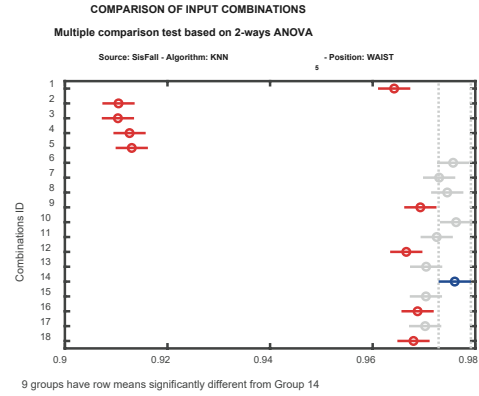
(a)



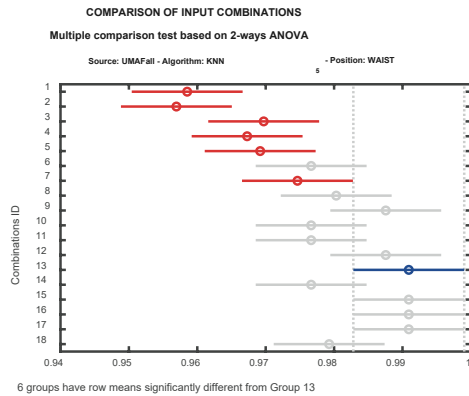
(b)



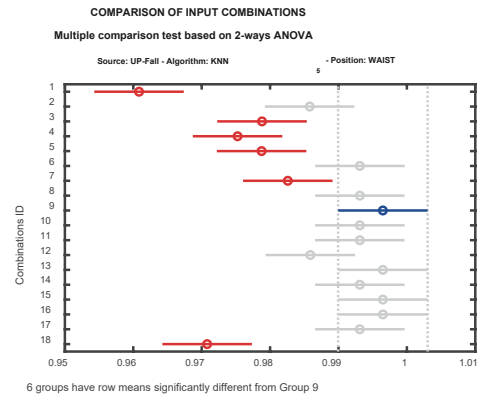
(c)



(d)



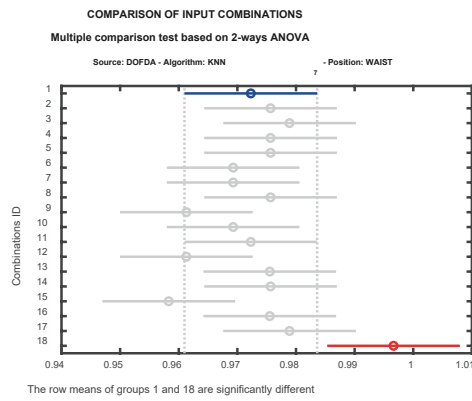
(e)



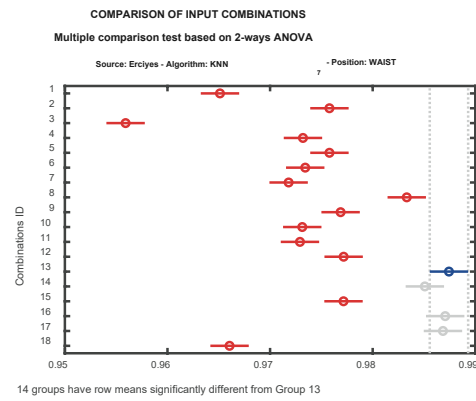
(f)

Figura B.3. Representación gráfica del test de comparación múltiple realizado con los resultados del ANOVA 2-way para comprobar la significancia estadística existente entre los resultados obtenidos por el sistema al utilizar diferentes características de entrada y el algoritmo *k*-Nearest Neighbours (*k*-NN) donde $k=7$. Se ha utilizado un $\alpha = 0.05$ y como tipo de valor crítico LSD (*Fisher's Least Significant Difference*) para cada repositorio.

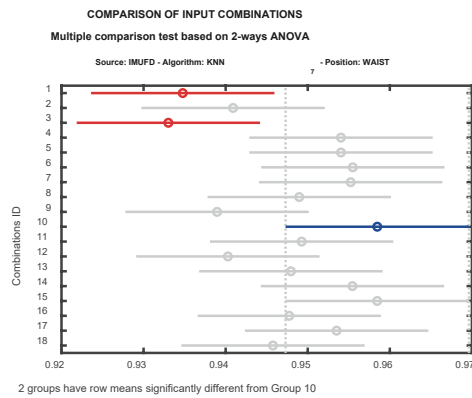
B.3.6.2. k-NN con $k = 7$



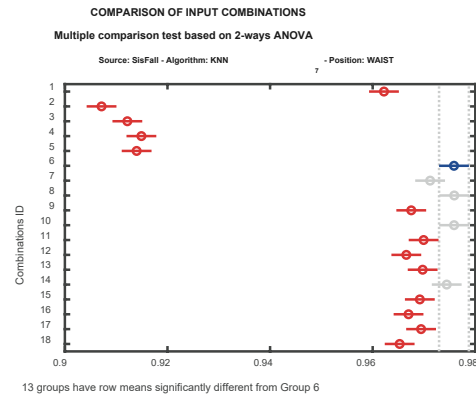
(a)



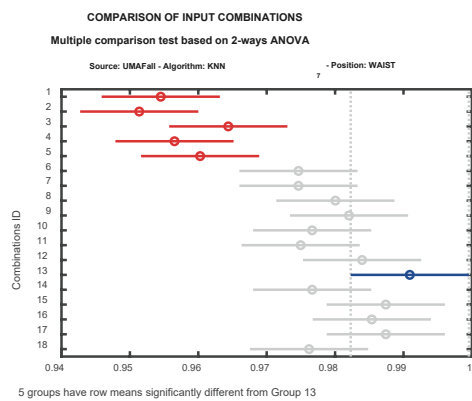
(b)



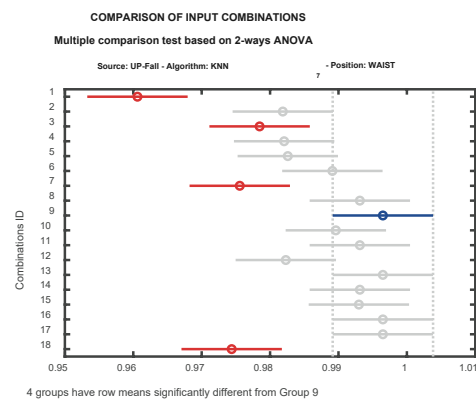
(c)



(d)



(e)



(f)

Figura B.4. Representación gráfica del test de comparación múltiple realizado con los resultados del ANOVA 2-way para comprobar la significancia estadística existente entre los resultados obtenidos por el sistema al utilizar diferentes características de entrada y el algoritmo *k*-Nearest Neighbours (*k*-NN) donde $k=7$. Se ha utilizado un $\alpha = 0.05$ y como tipo de valor crítico LSD (*Fisher's Least Significant Difference*) para cada repositorio.

B.4. Conclusiones

B.4.1. Efecto de las características de entrada

En primer lugar, tanto el análisis realizado con mRMR como el ANOVA muestran que la heterogeneidad de las características intrínsecas de cada repositorio [16] afectan a la hora de clasificar muestras. Los análisis exponen que los estadísticos de entrada para caracterizar los movimientos afectan de diferente forma según el repositorio estudiado. Sin embargo, también demuestran patrones que se repiten en los diferentes estudios:

- Existen características que parecen no aportar información relevante al sistema a la hora de clasificar. Tanto ANOVA como mRMR coinciden en que μ_{SMV} (A), SMA (J) y μ_R (L) no son características aconsejables.
- Por otro lado, las características $A_{\omega_{diff(max)}}$ (B), μ_{Ap} (F) y γ_{SMV} (I), son consideradas en ambos análisis como características relevantes a la hora de clasificar. Asimismo, ANOVA revela que la interacción de dichas características también aporta valor al resultado final.
- ANOVA demuestra que SMV_{Max} (G) afecta significativamente a la hora de clasificar y que, además, su interacción con las características indicadas en el punto anterior afecta positivamente y de forma significativa en el rendimiento del sistema.

B.4.2. mRMR y ANOVA

Por otro lado, se ha comprobado que el algoritmo mRMR puede aportar información bastante interesante a la hora de seleccionar características de entrada en un sistema de detección con un coste computacional bajo. Sin embargo, según las pruebas realizadas, los resultados que aporta pueden resultar demasiado específicos, lo que hace que el sistema se comporte adecuadamente con datos de una población determinada pero el rendimiento baje drásticamente con otra.

El ANOVA factorial es un proceso con un coste computacional mucho más alto, se necesita una cantidad de datos superior y además es necesario entrenar y

testear los datos con las combinaciones que se quieren entrenar. Pero nos ha proporcionado información más amplia y general del proceso.

B.4.3. Combinaciones recomendadas

En este caso, se ha seleccionado la combinación etiquetada como ID=14 como la combinación más recomendable para ser utilizada en el sistema de detección de caídas. Las razones que han motivado a esta elección han sido que se trata de la combinación que con una menor dimensión (compuesta por 7 características) ha obtenido los mejores resultados cuando se ha utilizado por el sistema (valores de la media geométrica de la especificidad y sensibilidad mayores del 95% en la mayor parte de las experiencias). La combinación con ID=14 (BCDFGIK), hace uso de las características $B = A_{\omega_{diff(max)}}$, $C = \sigma_{SMV}$, $D = \mu_{\theta}$, $F = \mu_{Ap}$, $G = SMV_{Max}$, $I = \gamma_{SMV}$, y $K = E$.

Otra conclusión relevante derivada del estudio es que la utilización de un mayor número de características de entrada no conlleva a una mejora en el rendimiento del sistema, a veces puede provocar “ruido” en el sistema y empeorar la clasificación.

B.4.4. Utilización de las características de entrada obtenidas con el algoritmo *k-Nearest Neighbours*

Por último, la utilización del algoritmo *k-Nearest Neighbours* ha mostrado que en el caso de la detección de caída el uso de las características de entrada recomendadas para el algoritmo SVM da lugar a un comportamiento similar del sistema, obteniendo métricas de la sensibilidad y especificidad por encima del 95%. Sin embargo, se ha observado que la combinación que incluye las 12 características (ID 18) en este algoritmo a veces sufre una disminución de rendimiento, el problema puede venir del hecho que este algoritmo es bastante sensible a las características irrelevantes, provocando un ruido importante a la hora de clasificar [17,18]. Por otro lado, la combinación etiquetada con el ID 14, sigue siendo recomendable con todos los repositorios.

B.5. Highly comparative time-series analysis (hctsa)

B.5.1. Introducción

Highly comparative time-series analysis (hctsa) es una herramienta software que permite realizar un análisis altamente comparativo de series de tiempo utilizando Matlab. Este software es capaz de extraer miles de características de un *dataset* de series de tiempo. Además incorpora una serie de herramientas para visualizar y analizar dichas características [19,20].

En este apartado se hará uso de este *toolbox* con el objetivo de estudiar si las características utilizadas en nuestro sistema (las cuales fueron elegidas teniendo en cuenta su sentido físico respecto al estudio de movimientos y caídas) pueden ser sustituidas por otras que permitan un mejor rendimiento del sistema.

B.5.2. Metodología

En primer lugar, se llevará a cabo el cálculo de características de los repositorios utilizando hctsa. (Este estudio se efectuará con los repositorios SisFall y UMAFall).

Este software es capaz de extraer más de 7700 características, sin embargo, no todas las operaciones son compatibles con todos los tipos de series temporales por lo que el siguiente paso implica filtrar las operaciones o/y series de tiempo que produjeron errores o salidas especiales. En el caso de este estudio, el número de características obtenidas es de alrededor 3600. Seguidamente, se realiza normalizan los datos mediante una transformada sigmoideal robusta para valores atípicos [19,20].

Posteriormente, se obtiene el rendimiento resultante de la clasificación de los datos al utilizar como input del sistema cada característica. Para ello hctsa realiza un k-fold con k=10, utilizando el algoritmo *Support Vector Machine* con *kernel* lineal.

Por último, se realiza un análisis que permite visualizar la correlación (Figura B.5) existente entre las características que han dado lugar a un mejor rendimiento. Dando lugar a un mapa de calor dividido en diferentes clústeres que agrupan aquellas características con comportamientos similares.

Debido a que durante todo este estudio se han utilizado 12 características, también se ha decidido utilizar 12 características extraídas con hctsa. Para ello, se ha ajustado el estudio hasta obtener 12 clústeres. Se seleccionará una característica de cada uno de ellos, en concreto, aquella que HTCSA marca como más representativa, la más cercana al centro del clúster. Las características seleccionadas provienen del estudio realizado con el repositorio SisFall, al ser considerado uno de los repositorios más completos.

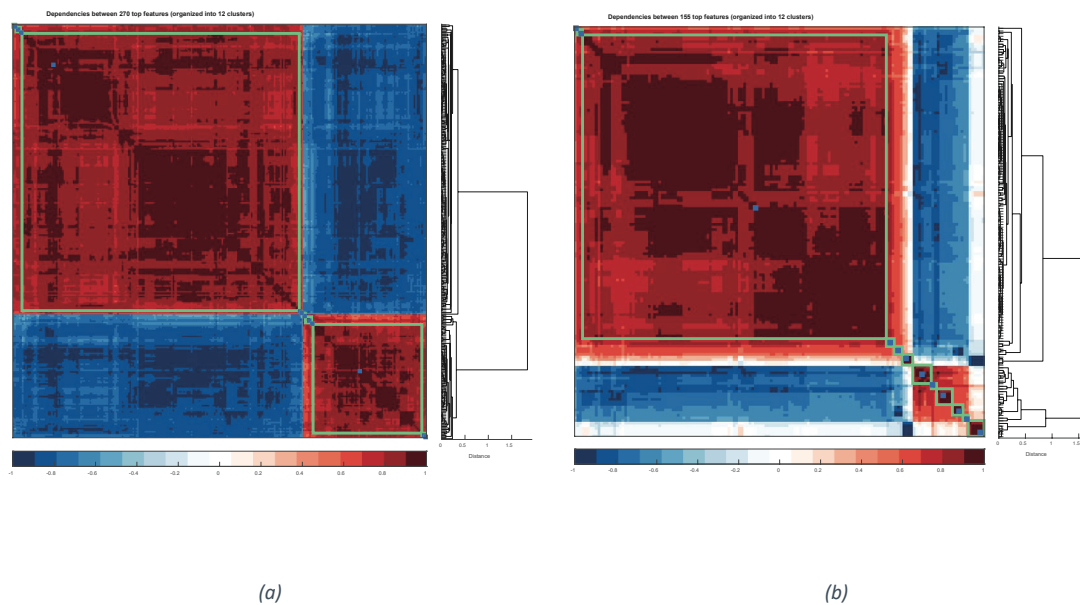


Figura B.5. Mapa de calor que representa la correlación existente entre las diferentes características extraídas para cada repositorio, clasificándolas en diferentes clústeres.

A continuación, se describen someramente las características seleccionadas para este estudio.

ID	Nombre	Descripción
A*	<i>SY_LocalGlobal_kurtosis_p5</i>	Curtosis del resultado de comparar esta misma estadística de una serie de tiempo respecto a la correspondiente de la primera mitad de dicha serie
B*	<i>SY_SlidingWindow_s_s2_2</i>	Parámetro vinculado a la estacionaridad de la desviación típica existente en una serie temporal a través de una ventana deslizante
C*	<i>AC_nl_11</i>	Autocorrelación no lineal de una serie de tiempo
D*	<i>DN_Withinp_20</i>	Estadístico ligado a la proporción de puntos de datos dentro de 2 desviaciones estándar de la media
E*	<i>NL_TSTL_ReturnTime_005_1_005_n1_1_3_statrtym</i>	Métrica relacionada con el análisis del histograma de tiempos de retorno de la serie temporal
F*	<i>NW_VisibilityGraph_norm_gaussnlogL</i>	Estadístico relacionado con el análisis gráfico de visibilidad de la serie temporal
G*	<i>DN_RemovePoints_absfar_01_saturate_kurtosisrat</i>	Estadístico ligado al efecto causado en las propiedades de la serie temporal a medida que se eliminan puntos
H*	<i>NW_VisibilityGraph_norm_olu90</i>	Estadístico relacionado con el análisis gráfico de visibilidad de la serie temporal
I*	<i>EN_mse_1</i>	Entropía multiescala de una serie temporal
J*	<i>NL_TSTL_ReturnTime_005_1_005_n1_1_3_hcgdist</i>	Métrica relacionada con el análisis del histograma de tiempos de retorno de la serie temporal
K*	<i>SY_SlidingWindow_s_sampen5_10</i>	Parámetro vinculado a la estacionaridad de la desviación típica existente en una serie temporal a través de una ventana deslizante
L*	<i>CP_ML_StepDetect</i>	Análisis de pasos discretos en una serie temporal

Tabla B.21. Estadísticos obtenidos a través de hctsa que han sido seleccionados para este estudio. Cada uno de ellos pertenece a clústeres diferentes cuando son ordenados por la correlación existente entre ellos.

B.6. Resultados

Para comprobar el rendimiento del sistema de clasificación al utilizar las características recomendadas por hctsa se utilizarán como inputs de entrada las características de forma individual y las combinaciones resultantes al utilizar simultáneamente. En las siguientes tablas se recopilan los resultados alcanzados.

B.6.1. Métricas obtenidas tras la clasificación

B.6.1.1. SisFall

ID	Combinaciones	Media geométrica de la <i>se</i> y la <i>sp</i>					Métricas k-fold	
		Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Media	STD
1*	A*	85.15%	83.85%	84.66%	83.11%	82.88%	83.93%	0.0098
2*	B*	84.98%	85.09%	84.86%	84.85%	84.44%	84.84%	0.0024
3*	C*	51.43%	48.24%	48.24%	51.13%	49.63%	49.74%	0.0152
4*	D*	86.89%	86.46%	87.05%	86.79%	87.53%	86.94%	0.0040
5*	E*	77.77%	78.28%	79.86%	77.60%	77.50%	78.20%	0.0097
6*	F*	87.13%	86.74%	86.48%	87.11%	86.53%	86.80%	0.0031
7*	G*	85.63%	84.75%	84.60%	84.16%	84.88%	84.80%	0.0054
8*	H*	91.77%	91.21%	91.08%	90.68%	92.51%	91.45%	0.0071
9*	I*	84.02%	84.54%	84.35%	84.96%	84.10%	84.39%	0.0038
10*	J*	88.31%	88.84%	90.43%	89.20%	89.39%	89.23%	0.0079
11*	K*	78.84%	79.96%	79.74%	80.64%	79.00%	79.64%	0.0073
12*	L*	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.0000
13*	H*J*	94.69%	95.15%	94.85%	94.65%	95.10%	94.89%	0.0023
14*	H*J*L	94.60%	95.15%	94.85%	94.71%	95.10%	94.88%	0.0024
15*	H*J*L*E*	94.51%	95.15%	94.85%	94.42%	95.10%	94.81%	0.0033
16*	H*J*L*E*D*	94.99%	95.96%	95.64%	95.23%	96.20%	95.60%	0.0050
17*	H*J*L*E*D*G*	95.05%	96.10%	96.53%	95.88%	95.77%	95.87%	0.0054
18*	H*J*L*E*D*G*C*	95.57%	96.34%	96.62%	95.75%	95.45%	95.95%	0.0051
19*	H*J*L*E*D*G*C*K*	96.01%	96.39%	96.86%	96.40%	96.34%	96.40%	0.0030
20*	H*J*L*E*D*G*C*K*I*	96.29%	96.39%	96.86%	96.40%	96.25%	96.44%	0.0024
21*	H*J*L*E*D*G*C*K*I*B*	96.06%	96.39%	96.86%	96.25%	96.63%	96.44%	0.0031
22*	H*J*L*E*D*G*C*K*I*B*A*	96.10%	96.39%	96.39%	96.36%	96.54%	96.35%	0.0016
23*	A*B*C*D*E*F*G*H*I*J*K*L*	96.24%	96.16%	96.53%	96.12%	96.30%	96.27%	0.0016

Tabla B.22. Métricas obtenidas tras la clasificación de las muestras del repositorio SisFall siguiendo la metodología k-fold utilizando como características de entrada las recomendadas por el toolbox htcsa.

		Media geométrica de la <i>se</i> y la <i>sp</i>					Métricas k-fold	
Combinaciones		Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Media	STD
14	BCDFGIK	96.88%	96.28%	96.36%	96.23%	97.02%	96.55%	0.0037
18	ABCDEFGHIJKL	97.21%	96.70%	97.21%	97.02%	97.30%	97.09%	0.0024

Tabla B.23. Métricas obtenidas tras la clasificación de las muestras del repositorio SisFall siguiendo la metodología k-fold utilizando las combinaciones de características que se recomendaron en el apartado B.4.3. Donde $A = \mu_{SMV}$, $B = A_{\omega_{diff(max)}}$, $C = \sigma_{SMV}$, $D = \mu_{\theta}$, $E = \mu_{SMV(diff)}$, $F = \mu_{Ap}$, $G = SMV_{Max}$, $H = SMV_{Min}$, $I = \gamma_{SMV}$, $J = SMA$, $K = E$, $L = \mu_R$.

B.6.1.2. UMAFall

ID	Combinaciones	Media geométrica de la <i>se</i> y la <i>sp</i>					Métricas k-fold	
		Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Media	STD
1*	A*	86.71%	86.42%	84.19%	84.94%	88.43%	86.14%	0.0165
2*	B*	93.51%	94.20%	96.13%	93.49%	96.74%	94.82%	0.0152
3*	C*	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.0000
4*	D*	83.64%	83.98%	78.91%	77.38%	71.99%	79.18%	0.0495
5*	E*	63.64%	65.70%	61.21%	76.13%	62.91%	65.92%	0.0593
6*	F*	89.41%	94.20%	94.02%	96.11%	92.84%	93.32%	0.0248
7*	G*	78.20%	79.64%	81.53%	81.89%	83.08%	80.87%	0.0194
8*	H*	83.64%	79.03%	78.91%	77.38%	70.52%	77.89%	0.0474
9*	I*	92.16%	88.81%	86.16%	87.99%	94.14%	89.85%	0.0324
10*	J*	74.27%	72.71%	73.61%	81.55%	71.56%	74.74%	0.0394
11*	K*	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.0000
12*	L*	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.0000
13*	H*J*	79.85%	85.12%	83.34%	90.78%	82.96%	84.41%	0.0404
14*	H*J*L	78.20%	80.60%	84.52%	90.01%	84.17%	83.50%	0.0448
15*	H*J*L*E*	79.05%	79.03%	84.52%	90.01%	84.17%	83.36%	0.0457
16*	H*J*L*E*D*	89.41%	91.40%	91.22%	94.77%	84.77%	90.31%	0.0365
17*	H*J*L*E*D*G*	90.80%	92.00%	92.02%	96.02%	88.78%	91.92%	0.0264
18*	H*J*L*E*D*G*C*	92.78%	90.19%	89.40%	96.02%	88.78%	91.43%	0.0298
19*	H*J*L*E*D*G*C*K*	92.78%	91.55%	90.01%	96.02%	90.77%	92.22%	0.0236
20*	H*J*L*E*D*G*C*K*I*	95.36%	93.51%	94.64%	95.40%	92.84%	94.35%	0.0114
21*	H*J*L*E*D*G*C*K*I*B*	96.71%	94.84%	97.26%	95.40%	97.40%	96.32%	0.0115
22*	H*J*L*E*D*G*C*K*I*B*A*	96.71%	94.84%	98.01%	95.40%	96.08%	96.21%	0.0123
23*	A*B*C*D*E*F*G*H*I*J*K*L*	95.36%	94.84%	98.64%	96.11%	96.76%	96.34%	0.0148

Tabla B.24. Métricas obtenidas tras la clasificación de las muestras del repositorio UMAFall siguiendo la metodología k-fold utilizando como características de entrada las recomendadas por el toolbox htcsa.

		Media geométrica de la <i>se</i> y la <i>sp</i>					Métricas k-fold	
Combinaciones		Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Media	STD
14	BCDFGIK	97.42%	96.14%	98.08%	96.64%	96.76%	97.01%	0.0075
18	ABCDEFGHIJKL	99.37%	98.08%	95.25%	98.64%	98.06%	97.88%	0.0156

Tabla B.25. Métricas obtenidas tras la clasificación de las muestras del repositorio SisFall siguiendo la metodología k-fold utilizando las combinaciones de características que se recomendaron en el apartado B.4.3. Donde $A = \mu_{SMV}$, $B = A_{\omega_{diff(max)}}$, $C = \sigma_{SMV}$, $D = \mu_{\theta}$, $E = \mu_{SMV(diff)}$, $F = \mu_{Ap}$, $G = SMV_{Max}$, $H = SMV_{Min}$, $I = \gamma_{SMV}$, $J = SMA$, $K = E$, $L = \mu_R$.

B.6.2. ANOVA de las combinaciones resultantes

Se realiza un ANOVA de los resultados de las tablas anteriores y se representan gráficamente en la Figura B.6 mediante un test de comparación múltiple utilizando un nivel de significación de 0.05 y como tipo de valor crítico LSD (*Fisher's Least Significant Difference*).

Las 23 primeras combinaciones de características corresponden a las originadas a partir del análisis con hctsa, mostradas en las tablas 22 y 24. Las dos últimas a las seleccionadas como aconsejables, utilizadas en las experiencias anteriores a esta sección y descritas en B.4.3.

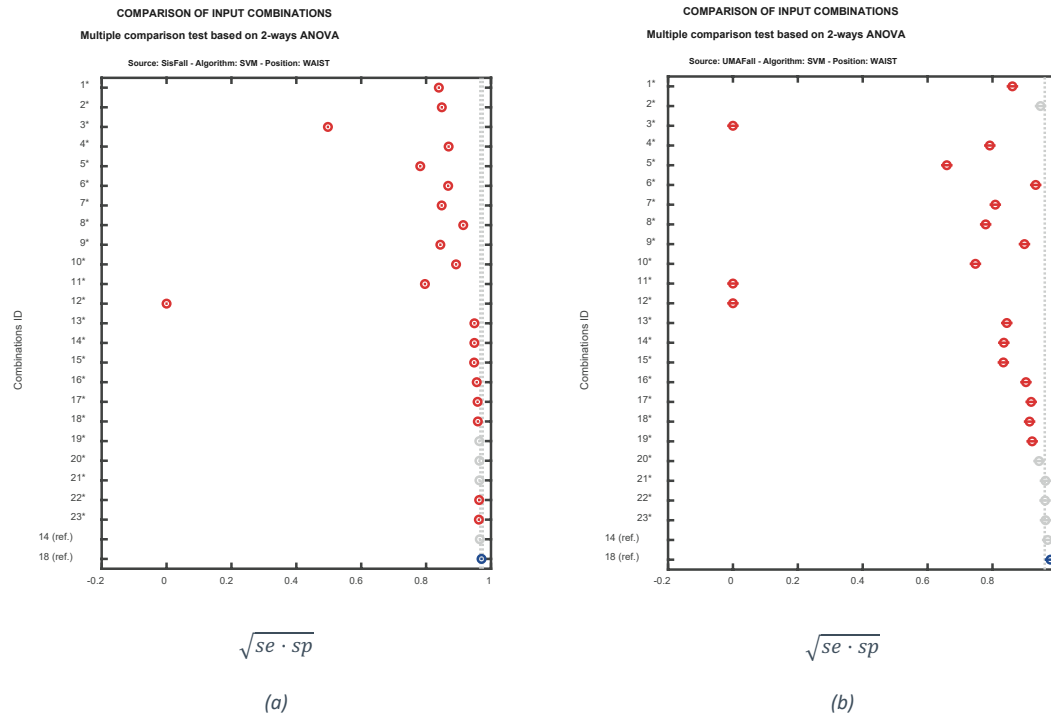


Figura B.6. Representación gráfica del test de comparación múltiple de los resultados del ANOVA realizado para comprobar la significancia estadística existente entre los resultados del sistema al hacer uso de diferentes combinaciones de características de input de entrada, utilizando un $\alpha = 0.05$ y como tipo de valor crítico LSD (Fisher's Least Significant Difference). El eje Y, combinations ID, corresponde a la combinación de inputs de entrada utilizada. El eje X corresponde al valor de la media geométrica de la sensibilidad y especificidad del sistema. Las 23 primeras combinaciones están realizadas utilizando los features recomendados por hctsa y las dos últimas son las combinaciones que se determinaron como mejores en el apartado B.4.3.

B.6.3. Conclusiones

En primer lugar, observamos que las pruebas etiquetadas como de referencia (éstas utilizan como inputs las características recomendadas en la sección previa) dan lugar al mejor rendimiento del sistema, nunca siendo significativamente inferior al resultante de ninguna otra combinación compuesta por características originadas a partir de hctsa.

Es cierto, que existen combinaciones que consiguen alcanzar rendimientos similares a las combinaciones de referencia, pero en la mayoría de los casos las operaciones para obtener el estadístico tienen una mayor carga computacional.

Por otro lado, llama la atención los resultados de las características etiquetadas como 3*, 11* y 12*, los cuales producen un rendimiento con una media geométrica de la sensibilidad y la especificidad de valor nulo, cuando en el estudio htcsa obtenía valores cercanos al 90%. La utilización de estadísticos aleatorios sin atender al sentido físico de las muestras puede provocar problemas de este tipo, ya que el sistema encuentra patrones ligadas a la propia naturaleza de las muestras en lugar del movimiento en sí.

Debido a este mismo hecho, también existe el peligro de que una característica que trabaja muy bien con un repositorio no sea compatible con otro, como ocurre con las características 3* y 11*, donde obtiene un rendimiento superior del 80% en SisFall pero 0% en UMAFall. Pueden existir patrones que se repiten en cada repositorio a la hora de grabar los movimientos (Ejemplo: longitud de la muestra menor para un determinado tipo de movimiento) que el sistema lo detecte como significativo para clasificar pero que en realidad no esté relacionado con el movimiento en sí.

Por tanto, tras estos resultados se llega a la conclusión de que las características seleccionadas al comienzo de este trabajo presentan ciertas ventajas respecto a las originadas con HTCSA:

- Tienen un sentido físico más evidente relacionado con las caídas y ADL.
- Pueden utilizarse con diferentes repositorios ya que están estrechamente ligados a la naturaleza de los movimientos.
- Se obtienen medias de la especificidad y sensibilidad igual o superiores a las pruebas con HTCSA.

Referencias

1. Motoda, H.; Liu, H. Feature selection, extraction and construction. *Commun. IICM* **2002**.
2. MTW; Liu, H.; Motoda, H. Feature Extraction Construction and Selection: A Data Mining Perspective. *J. Am. Stat. Assoc.* **1999**, doi:10.2307/2669967.
3. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, doi:10.1016/s0004-3702(97)00063-5.
4. Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. In Proceedings of the Proceedings of the 2003 IEEE Bioinformatics Conference, CSB 2003; 2003.
5. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, doi:10.1109/TPAMI.2005.159.
6. Ding, C.H.Q. Analysis of gene expression profiles: Class discovery and leaf ordering. In Proceedings of the Proceedings of the Annual International Conference on Computational Molecular Biology, RECOMB; 2002.
7. Frank, K.; Vera Nadales, M.J.; Robertson, P.; Pfeifer, T. Bayesian recognition of motion related activities with inertial sensors. In Proceedings of the UbiComp'10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing; 2010.
8. Cotechini, V.; Belli, A.; Palma, L.; Morettini, M.; Burattini, L.; Pierleoni, P. A dataset for the development and optimization of fall detection algorithms based on wearable sensors. *Data Br.* **2019**, doi:10.1016/j.dib.2019.103839.
9. Özdemir, A.T. An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice. *Sensors (Switzerland)* **2016**, doi:10.3390/s16081161.
10. Aziz, O.; Musngi, M.; Park, E.J.; Mori, G.; Robinovitch, S.N. A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials. *Med. Biol. Eng. Comput.* **2017**, doi:10.1007/s11517-016-1504-y.
11. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A fall and movement dataset. *Sensors (Switzerland)* **2017**, *17*, doi:10.3390/s17010198.
12. Martínez-Villaseñor, L.; Ponce, H.; Espinosa-Loera, R.A. Multimodal Database for Human Activity Recognition and Fall Detection. *Proceedings* **2018**, doi:10.3390/proceedings2191237.
13. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. *Procedia Comput. Sci.* **2017**, *110*, 32–39, doi:10.1016/j.procs.2017.06.110.

14. Cohen, J. Quantitative Methods in Psychology: A Power Primer. *Psychol. Bull.* **1992**.
15. Stangl, D.K. Encyclopedia of Statistics in Behavioral Science. *J. Am. Stat. Assoc.* **2008**, doi:10.1198/jasa.2008.s230.
16. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems. *J. Healthc. Eng.* **2020**, 2020, 1–36, doi:10.1155/2020/6622285.
17. Langley, P.; Iba, W. Average-Case Analysis of a Nearest Neighbor Algorithm. *Proc. Thirteen. Int. Jt. Conf. Artificial Intell.* **1993**.
18. Zhang, S. Cost-sensitive KNN classification. *Neurocomputing* **2020**, doi:10.1016/j.neucom.2018.11.101.
19. Fulcher, B.D.; Jones, N.S. hctsa: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell Syst.* **2017**, doi:10.1016/j.cels.2017.10.001.
20. Fulcher, B.D.; Little, M.A.; Jones, N.S. Highly comparative time-series analysis: The empirical structure of time series and their methods. *J. R. Soc. Interface* **2013**, doi:10.1098/rsif.2013.0048.

Apéndice C. Evaluación de utilización de técnicas de *Ensamble Learning*

C.1. Introducción

En el Capítulo 6, ya se comentó el interés de utilizar métodos de *ensemble learning*, este trabajo amplía el estudio realizado sobre este tema.

Este tipo de técnicas son estrategias de combinación de modelos de *machine learning* para mejorar capacidad de clasificación y como consecuencia sus prestaciones [1].

A continuación, se describen diferentes pruebas llevadas a cabo con la intención de averiguar si el uso de este tipo de procedimientos mejora la capacidad del sistema de detección de caídas en clasificar correctamente ADL y caídas. En estas pruebas se utilizan por un lado algoritmos de aprendizaje supervisado y por otros algoritmos de clase única.

C.2. Metodología

Al igual que en pruebas de apartados anteriores, se analiza el rendimiento del sistema de detección de caídas al utilizar diferentes repositorios de movimientos existentes en la literatura. En concreto, para estas pruebas se hace uso de los repositorios DLR [2], DOFDA [3], Erciyas [4], FallAIID [5], IMUFD [6], KFall [7], SisFall [8], UMAFall [9] y UP-Fall [10].

Sin embargo, en este caso la decisión final del sistema vendrá dada a partir de aplicar un método *ensemble learning*. En concreto, en este estudio se emplea la técnica conocida como *voting*. Es decir, la decisión final de clasificación se

obtiene aplicando una votación por mayoría simple de varios modelos de clasificación diferentes. Específicamente, en este estudio se utilizan 3 modelos.

En esta ocasión, los repositorios se dividen en 4 particiones, de manera que en cada una de ellas existe un número similar de muestras de cada tipo de movimiento y trazas de todos los sujetos experimentales.

Las pruebas realizadas pueden dividirse en 2 clases.

- Pruebas tipo A: Se utilizan 3 algoritmos diferentes (o con diferentes hiperparámetros). En este caso, una de las particiones se utiliza para el testeo del sistema y las restantes para el proceso de entrenamiento. Los 3 algoritmos se entrenan con los mismos datos, por tanto, el 75% de las muestras es utilizada para entrenar y el 25 % para testear.
- Prueba tipo B: Se utiliza 1 solo algoritmo con unos hiperparámetros concretos, el cual se entrena 3 veces con cada una de las particiones seleccionadas para entrenamiento para obtener 3 modelos diferentes, la partición restante es utilizada para testear el sistema. Por tanto, se utiliza un 25% de las muestras para entrenar y conseguir cada modelo y otro 25% para testear.

Para mejorar la precisión de los resultados, cada prueba se repite 5 veces y el resultado final es el resultante de aplicar la media a las métricas de calidad.

El objetivo de estos estudios es evaluar si la utilización de técnicas del tipo *ensemble learning* mejora las prestaciones de un sistema de detección de caídas.

C.2.1. Pruebas llevadas a cabo

Para mejorar la legibilidad del documento las hojas de cálculo con los resultados se han alojado en un repositorio creado con el propósito de servir de anexo de este trabajo [11].

C.2.1.1. Pruebas tipo A

Para este tipo de pruebas se realizan 3 experimentos diferentes.

- Se seleccionan los 3 algoritmos de cada tipo que proporcionaron mejores prestaciones en la prueba equilibrada de las experiencias de los capítulos 5 y 6, utilizando como características de entrada la combinación compuesta por 7 características de elección propia con la que se consiguieron mejores resultados en el apéndice B.
- Se seleccionan los 3 algoritmos (con independencia del tipo) junto con los hiperparámetros que proporcionaron mejores prestaciones en la prueba equilibrada de las experiencias de los capítulos 5 y 6, utilizando como características de entrada la combinación compuesta por 7 características de elección propia con la que se consiguieron mejores resultados en el apéndice B.
- Se seleccionan los 3 algoritmos (con independencia del tipo) junto con los hiperparámetros que proporcionaron mejores prestaciones en la prueba equilibrada de las experiencias de los apartados capítulos 5 y 6, sin embargo, esta vez utilizando como características de entrada cualquiera de las 3 combinaciones recomendadas en el apéndice B, esto es, la combinación de 7 características y la combinación que utiliza todas las características de elección propia y la combinación de 12 características extraídas a partir de HCTSA.

NOTA: En algunos casos algunas pruebas pueden coincidir.

C.3. Conclusiones

C.3.1. Pruebas tipo A

Los resultados demuestran que en el caso en el que se seleccionan 3 algoritmos de tipos diferentes, tanto para las pruebas realizadas con algoritmos de dos clases como de clase única, solamente en 2 de los 9 repositorios estudiados la métrica global (media geométrica de la sensibilidad y especificidad) mejora ligeramente respecto del algoritmo que proporcionaba mejores resultados trabajando individualmente. Del mismo modo, se demuestra que existe un ascenso de la especificidad en 5 de los 9 repositorios.

Algo similar ocurre cuando se permite utilizar cualquier tipo de algoritmo e hiperparámetros, con 3 de los 9 repositorios la métrica global mejora al utilizar algoritmos de clase única y 4 de 9 en el caso de los algoritmos de dos clases.

En el caso en el que se prueba con algoritmos de diferentes tipos, hiperparámetros, así como, distintos sets de características de entrada, se observa una mejora de los resultados en 6 y 7 repositorios de 9, cuando se utilizan algoritmos de clase única y de dos clases respectivamente. En este caso existe una mejora de la especificidad en la mayoría de los casos, combinada a veces también con una mejora de la sensibilidad.

Por tanto, la conclusión de estas primeras pruebas es que utilizar este tipo de *ensamble learning* para la detección de caídas puede ser positivo para las prestaciones del sistema, haciéndose más patente cuando se permite combinar algoritmos de cualquier tipo, incluso del mismo pero con hiperparámetros diferentes, y permitiendo la utilización de sets de características diferentes.

C.3.2. Prueba tipo B

En este tipo de prueba sólo se realiza un tipo de experiencia. Se selecciona el algoritmo y los hiperparámetros que mejores resultados produjeron en la prueba equilibrada de las experiencias de los capítulos 5 y 6, utilizando como características de entrada la combinación compuesta por 7 características de elección propia con la que se consiguieron mejores resultados en el Apéndice B y se entrena con 3 particiones de entrenamiento diferentes para obtener 3 modelos diferentes.

Los resultados de esta experiencia no mejoran en ningún caso los obtenidos al utilizar solamente un algoritmo entrenado con el 80% de las muestras de los repositorios. En algunos casos empeoran drásticamente el rendimiento del sistema, seguramente debido a que se utiliza un número menor de muestras para entrenar cada uno de los modelos, hecho que no se compensa con la posterior votación para elegir la decisión final. Por tanto, no se recomienda sacrificar un porcentaje de muestras de entrenamiento para entrenar varios modelos en lugar de utilizar todas las muestras de entrenamiento para entrenar un determinado algoritmo.

Referencias

1. Polikar, R. Ensemble Learning. In *Ensemble Machine Learning*; 2012.
2. Frank, K.; Vera Nadales, M.J.; Robertson, P.; Pfeifer, T. Bayesian recognition of motion related activities with inertial sensors. In *Proceedings of the UbiComp'10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing*; 2010.
3. Cotechini, V.; Belli, A.; Palma, L.; Morettini, M.; Burattini, L.; Pierleoni, P. A dataset for the development and optimization of fall detection algorithms based on wearable sensors. *Data Br.* **2019**, doi:10.1016/j.dib.2019.103839.
4. Özdemir, A.T. An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice. *Sensors (Switzerland)* **2016**, doi:10.3390/s16081161.
5. Saleh, M.; Abbas, M.; Le Jeannes, R.B. FallAID: An Open Dataset of Human Falls and Activities of Daily Living for Classical and Deep Learning Applications. *IEEE Sens. J.* **2021**, 21, doi:10.1109/JSEN.2020.3018335.
6. Aziz, O.; Musngi, M.; Park, E.J.; Mori, G.; Robinovitch, S.N. A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials. *Med. Biol. Eng. Comput.* **2017**, doi:10.1007/s11517-016-1504-y.
7. Human Factors and Ergonomics Lab- Korea Advanced Institute of Science and Technology KFall: A Comprehensive Motion Dataset to Detect Pre-impact Fall for the Elderly based on Wearable Inertial Sensors.
8. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A fall and movement dataset. *Sensors (Switzerland)* **2017**, 17, doi:10.3390/s17010198.
9. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. *Procedia Comput. Sci.* **2017**, 110, 32–39, doi:10.1016/j.procs.2017.06.110.
10. Martínez-Villaseñor, L.; Ponce, H.; Espinosa-Loera, R.A. Multimodal Database for Human Activity Recognition and Fall Detection. *Proceedings* **2018**, doi:10.3390/proceedings2191237.
11. Santoyo-Ramon, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Anexo: Contribuciones al estudio y análisis de sistemas-vestibles-de detección de caídas Available online: https://github.com/jasr92/Anexo_Contribuciones-al-estudio-y-an-lisis-de-sistemas-vestibles-de-detecci-n-de-ca-das (accessed on Sep 17, 2021).

Apéndice D. Evaluación de sistemas de aprendizaje profundo y elección de hiperparámetros – Redes Convolucionales

D.1. Introducción

Las redes neuronales convolucionales son un tipo de red neuronal artificial, en inglés *Artificial Neural Network* (ANN), con aprendizaje supervisado cuyo funcionamiento es similar al córtex visual humano. Contiene varias capas ocultas especializadas y jerarquizadas, de esta manera cada una de ellas es capaz de detectar diferentes características, conforme más profunda es la capa más compleja es la característica que reconoce. Este tipo de red es muy efectiva para la clasificación de imágenes y el reconocimiento de voz [1].

La principal ventaja de las redes neuronales convolucionales respecto las redes neuronales artificiales clásicas es la reducción del número de parámetros y con ello su complejidad. Este factor permite resolver tareas complejas que no son posibles con ANN clásicas. Además, las redes neuronales convolucionales son capaces de detectar patrones con independencia de la posición en la que se encuentre dentro de los datos.

D.2. Elementos de una CNN

D.2.1. Capa de entrada

Esta capa recibe los datos de entrada. En ella se define las dimensiones requeridas para los datos de entrada, el número de canales y normaliza los datos cada vez que son propagados hacia la siguiente capa.

D.2.2. Capa de convolución

Este tipo de capa es el componente fundamental en esta clase de red. Tal y como indica su propio nombre, su funcionamiento se basa en un operador matemático denominado convolución (*). Dadas dos matrices A y B , la convolución de ambas, $A * B$, se define como la matriz resultante al realizar el producto escalar entre ellas mientras se desplaza una sobre la otra.

Este tipo de capa evita conectar todos los datos de entrada con neuronas (*full-connection*), dando lugar a un proceso más eficiente y sencillo, en el cual se extraen características de regiones locales, eliminando datos superfluos.

En este tipo de operación se encuentran varios elementos y parámetros:

- **Datos de entrada:** Se trata de los datos que recibe la capa.
- **Filtro:** Matriz que permite extraer cierta característica de los datos de la entrada. Será la matriz que se desplaza en el proceso de convolución. A la hora de crear la capa debe definirse:
 - Tamaño del filtro
 - Paso de la convolución: Nº de muestras del dato que se recorre en cada avance del proceso de convolución.
 - Nº de filtros utilizados en cada capa convolucional: Cada filtro realiza una convolución con la matriz de entrada generando un mapa de características. Cada filtro extrae una característica de los datos de entrada y por tanto posee una serie de pesos y sesgos que se reajustan en el proceso de entrenamiento.

NOTA: Si los datos de entrada tienen varios canales, para cada filtro se crearán tantos *kernels* como canales, posteriormente se realizarán una serie de operaciones entre ellos y conformarán una única salida.

- **Padding:** Uno de los problemas de la convolución es la pérdida de información de los bordes de las matrices. Éste hecho puede ser solucionado mediante la utilización de *zero-padding*, es decir, añadiendo a la matriz de entrada filas y columnas de ceros en los bordes. Este mecanismo también permite controlar el tamaño del dato de salida del filtro y acelerar el proceso de aprendizaje, validación y testeo.
- **Paso (Stride):** Permite controlar el número de muestras que se desplaza el filtro en cada iteración de la convolución y por tanto el solapamiento existente en el proceso.

Como resultado se obtiene una matriz de salida cuyo tamaño depende de las siguientes ecuaciones:

$$Height_{out} = \frac{Height_{input} - Height_{Filter} + 2 * Height_{padding}}{Height_{Stride}} + 1 \quad (1)$$

$$Width_{out} = \frac{Width_{input} - Width_{Filter} + 2 * Width_{padding}}{Width_{Stride}} + 1 \quad (2)$$

$$NChannels_{out} = NFilters \quad (3)$$

D.2.3. Capa de normalización *Batch*

Esta capa normaliza las capas ocultas tras la convolución, aumentando la velocidad de entrenamiento y las prestaciones de la red. Realiza una normalización del tipo unidad tipificada a la salida de la capa anterior. Esto es, dado un valor x , se define el valor de x en unidades tipificadas, z , como:

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

Donde μ es la media de las muestras y σ la desviación típica.

D.2.4. Capa de activación

Esta capa se aplica para limitar o saturar la salida generada tras la convolución, lo que permite desechar valores superfluos. Existen diferentes funciones que pueden ser utilizadas en este tipo de capa, por ejemplo, la función sigmoide o la tangente hiperbólica; sin embargo, la función que actualmente suele usarse es la función conocida como ReLU (*Rectified Linear Unit*).

La función ReLU se denota como:

$$ReLU(x) = \max(0, x) \quad (5)$$

Presenta ciertas ventajas respecto otras funciones utilizadas en la capa de activación:

- Definición simple de la función y su gradiente
- Posee un gradiente constante para entradas positivas, lo que resuelve el problema existente en funciones como la sigmoide o el arcotangente hiperbólico a la hora de realizar propagación hacia atrás, donde la señal del gradiente comienza a desaparecer conforme más profunda es la red. Este hecho es provocado debido a que el gradiente de estas funciones es muy cercano a cero en la mayoría de las partes excepto en valores centrales [1].
- ReLU consigue una representación más dispersa que favorece el entrenamiento.

D.2.5. Capa de submuestreo

La labor de este tipo de capa es la de reducir la complejidad de los datos realizando un submuestreo de éstos. Reduce el tamaño de los datos de entrada, pero no afecta al número de canales o filtros. Este proceso provoca que la red se quede solamente con los datos más representativos [1,2].

En este estudio, la técnica utilizada en este tipo de capa es *Max-pooling*. Se basa en la división de los datos en ventanas de un tamaño determinado, m , en cada ventana se selecciona el valor más alto.

D.2.6. Capa totalmente conectada (*Fully Connected*)

Esta capa conecta cada una de las neuronas de la capa anterior con las neuronas de la siguiente capa. Realiza una tarea similar a las capas existentes en las redes neuronales tradicionales. Combina todos los mapas de características con el objetivo de identificar los patrones más importantes.

D.2.7. Capa de activación *Softmax*

Esta capa permite conocer la probabilidad con que la muestra de entrada pertenece a una clase u otra. En este estudio se encuentra ubicada justo después de la capa *Fully Connected*. Utiliza la función de activación *Softmax* o función exponencial normalizada, la cual comprime un vector k-dimensional de valores reales en un vector k-dimensional de valores reales en el rango [0,1]. Para $j = 1, \dots, K$.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (6)$$

D.2.8. Capa de clasificación

Esta capa recibe los datos de la capa de activación *Softmax* y asigna cada entrada a una de las clases en las que pueden clasificarse las muestras. Además, calcula la pérdida o error de entropía cruzada, valor utilizado en el entrenamiento para calcular y actualizar los parámetros de la red convolucional en cada iteración.

La función de pérdida viene dada por la siguiente fórmula:

$$loss = -\frac{1}{N} \sum_{m=1}^M \sum_{c=1}^C w_c \cdot t_{m,c} \cdot y_{m,c} \quad (7)$$

D.3. Metodología

D.3.1. Repositorios de movimientos

En este nuevo estudio se vuelve a seguir el mismo procedimiento realizado en las experiencias de los apartados preliminares: evaluar el sistema de detección de caídas a partir de un análisis sistemático de distintos repositorios disponibles en la literatura, en concreto se utilizarán 8 (DOFDA [3], Erciyas [4], FallAIID [5], IMUFD [6], KFall [7], SisFall [8], UMAFall [9] y UP-Fall [10]).

D.3.2. Métricas de calidad

De nuevo, la métrica global de calidad empleada para la evaluación es la media geométrica de la sensibilidad y especificidad del sistema, la cual permite valorar la precisión con que la arquitectura es capaz de discernir entre ADL y caídas.

D.3.3. Configuración de las redes neuronales

El estudio tiene dos objetivos, por un lado, evaluar las prestaciones que proporciona la utilización de una red convolucional en un sistema de detección de caídas, y por otro, analizar la influencia que poseen una serie hiperparámetros de la red convolucional en la métrica de calidad global a la hora de clasificar movimientos. Exactamente se han estudiado un total de 1758 combinaciones diferentes para cada una de las 8 bases de datos. En la Tabla D.1 se resumen los hiperparámetros analizados junto a los valores comprobados.

La configuración de las redes neuronales utilizadas en este estudio presenta la estructura que se muestra gráficamente en la Figura D.1.

Para cada muestra se introducen los datos de la aceleración en una ventana de tiempo alrededor del pico de máxima aceleración (en el caso de que se introduzcan los 3 ejes del acelerómetro existen 3 canales, en el caso en el que solamente se introduzca el SMV, un canal).

Nombre del hiperparámetro	Valores analizados
Tamaño de la ventana de estudio (t_w)	2 s, 3 s, 4 s
Nº de capas convolucionales	1, 3, 5
Nº de filtros	8, 16, 32, 64
Ancho de los filtros convolucionales	4, 8
Ancho de <i>Padding</i>	4
<i>Stride</i>	1x1
Método de optimización	<i>Stochastic gradient descent with momentum</i>
<i>L2 Regulation</i>	0.0001
Ancho de la ventana de <i>Max-Pooling</i>	2, 3
Tamaño del <i>miniBatch</i>	64, 256
Tasa de aprendizaje inicial	0.01
Paciencia de validación	5
Tipo de canal	3-axis, SMV
Tipo de normalización	Ninguno, <i>zscore</i> , <i>zero-center</i>

Tabla D.1. Resumen de los hiperparámetros de la red convolucional analizados en el presente estudio

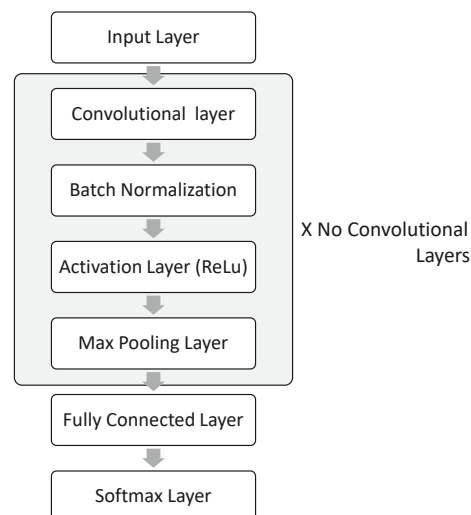


Figura D.1. Estructura de las redes convolucionales utilizadas en el presente estudio

D.3.4. Estudio de remuestreo

Para validar los modelos generados se hará uso de la técnica de validación cruzada k-Fold, específicamente se realiza un análisis 5-Fold. Cada repositorio ha sido dividido en 5 particiones, en los cuales existe aproximadamente el mismo número de muestras de los diferentes movimientos e incluye muestras de cada uno de los sujetos participantes. De esta manera, se utiliza un 60% de las muestras para entrenar al sistema, un 20% para el proceso de validación y un 20% para testear.

En cada experiencia se obtiene la media de la media geométrica de la especificidad y sensibilidad del sistema del análisis k-Fold junto con su desviación estándar.

D.3.5. Equipamiento y tiempo de cómputo

Las pruebas se han realizado utilizando la versión de Matlab R2020b y la *toolbox Deep Learning* versión 14.1. así como, *Parallel Computing Toolbox* versión 7.3.

Se ha dispuesto de dos ordenadores diferentes:

- Por un lado, el ordenador utilizado como principal, un MSI GL65 Leopard 10SEK, con un procesador Intel Core i7-10750H, 32 GB de RAM y una tarjeta gráfica compatible con CUDA, en concreto la NVIDIA GeForce RTX 2060.
- Por otro lado, un ordenador utilizado como apoyo, un ASUS con un procesador i7-4790, 24 GB de memoria RAM y una tarjeta gráfica Intel HD Graphics 4600 (No compatible con CUDA).

Combinando las dos *toolbox* especificadas anteriormente, teníamos la posibilidad de entrenar la red neuronal convolucional de dos maneras diferentes: utilizando la CPU y utilizando la GPU. Para utilizar la GPU es necesario que la tarjeta gráfica cumpla unas especificaciones concretas requeridas, en nuestro caso, solamente el primer ordenador era compatible [11].

La utilización de una GPU permite acelerar de manera considerable procesos computacionalmente intensivos como es el entrenamiento de las redes neuronales [12,13].

Por otra parte, para agilizar el proceso en ambos ordenadores se programa el script para que se traten en paralelo tres repositorios simultáneamente.

Haciendo una estimación aproximada del tiempo de cómputo de estas pruebas, se puede exponer que en el ordenador que dispone de la tarjeta gráfica compatible con CUDA ha tardado alrededor de cuatro días en procesar, para todos los repositorios utilizados en este estudio, las diferentes combinaciones de hiperparámetros utilizando un tipo de normalización de datos concreto. El tiempo de cómputo del ordenador cuya gráfica no es compatible con CUDA podía duplicarse para la misma tarea.

D.4. Resultados

En primer lugar, se analiza qué combinación de hiperparámetros consigue una mayor media geométrica de la sensibilidad y especificidad del sistema. Para ello, tal y como se ha comentado en la sección predecesora se analizan sistemáticamente los repositorios de movimientos elegidos para este estudio y se obtienen las métricas de calidad para cada una de las 1758 combinaciones de hiperparámetros. Posteriormente, se realiza la media de cada combinación y se selecciona aquella con la que se consigue una mayor métrica global de calidad.

Tras realizar las diferentes pruebas haciendo uso de distintos métodos de normalización se concluye que no existe una clara diferencia en las prestaciones del sistema cuando se utilizan unos u otros. Sin embargo, como se pretende que los modelos de predicción puedan ser utilizados entre poblaciones de movimientos diferentes, se elige la normalización tipo *zscore* o conocida como unidad tipificada. Debe puntualizarse, que la normalización se está realizando a partir de los datos de entrenamiento (como ocurriría en un escenario realista, en el que en principio se podría ir normalizando los datos en tiempo real a partir de variables ya medidas de antemano durante el *training*).

La combinación de hiperparámetros que tras el análisis proporciona una mejor clasificación de los repositorios utilizando como tipo de normalización *zscore* se resumen en la Tabla D.2. Con dichos parámetros, se obtiene una media de la media geométrica de la especificidad y sensibilidad de 0.9874, así como, una sensibilidad de 0.9934 y una especificidad de 0.9815.

En la Tabla D.3 se resumen las métricas de calidad obtenidas para cada repositorio al utilizar las muestras recomendadas en la tabla anterior. Tal y como se puede comprobar, en todos ellos la media geométrica de la sensibilidad y especificidad es superior a 0.95. Consiguiendo superar en 6 de los 8 repositorios el valor de 0.98, e incluso 0.995 en 4 de ellos. Así mismo, la desviación de esta métrica al realizar el análisis de remuestreo 5-fold es muy baja, no superando en la mayoría de los casos 0.01.

Puede encontrarse una versión extendida de los resultados en el repositorio que ha sido creado con el propósito de servir de anexo de esta tesis [14].

Nombre del hiperparámetro	Valores analizados
Tamaño de la ventana de estudio (t_w)	3 s
Nº de capas convolucionales	5
Nº de filtros	64
Ancho de los filtros convolucionales	4
Ancho de Padding	4
Stride	1x1
Método de optimización	Stochastic gradient descent with momentum
L2 Regulation	0.0001
Ancho de la ventana de Max-Pooling	2
Tamaño del miniBatch	64
Tasa de aprendizaje inicial	0.01
Paciencia de validación	5
Tipo de canal	3-axis
Tipo de normalización	zscore

Tabla D.2. Resumen de los hiperparámetros con los que se obtiene unas mayores prestaciones del sistema de detección de caída para los repositorios estudiados.

Repositorio	Se	Sp	$\sqrt{Se \cdot Sp}$	$\sigma_{\sqrt{Se \cdot Sp}}$
DOFDA	0.9962	1.0000	0.9981	0.0042
Erciyes	0.9984	0.9966	0.9975	0.0018
FallAIID	0.9244	0.9813	0.9524	0.0076
IMUFD	0.9521	0.9761	0.9638	0.0157
KFall	0.9991	0.9996	0.9994	0.0010
SisFall	0.9911	0.9996	0.9953	0.0034
UMAFall	0.9838	0.9974	0.9905	0.0115
UP-Fall	0.9859	0.9829	0.9842	0.0149

Tabla D.3. Métricas de calidad obtenidas al analizar los diferentes repositorios de movimientos utilizando los hiperparámetros seleccionados como recomendables y descritos en el Apéndice B

D.5. Conclusiones

En primer lugar, el análisis ha demostrado que la utilización de redes neuronales convolucionales en el ámbito de la detección de caídas es una opción que permite obtener muy buenas prestaciones. Si se realiza la media de la métrica de calidad global derivadas del uso de todos los repositorios seleccionados se obtiene un intervalo de valores comprendido entre 0.9470 y 0.9878 (valor obtenido para la combinación de hiperparámetros que peores y mejores resultados alcanza respectivamente).

Si se seleccionan los hiperparámetros que provocan un mejor rendimiento del sistema, cuando se normalizan los datos de entrada mediante *zscore*, tal y como se mencionó anteriormente, la mayoría de los repositorios, 5 de 8, obtienen valores superiores de 0.99 en sus medias geométricas de la sensibilidad y especificidad y el caso peor supera 0.95.

A su vez se pueden extraer otras conclusiones de este estudio, por ejemplo, se comprueba que la utilización de los 3 ejes del acelerómetro como parámetros de entrada de la red (utilizar 3 canales diferenciados), proporciona mejores prestaciones que alimentar al sistema solamente con el módulo de la aceleración. En el caso en que se utilizan los hiperparámetros recomendados, la diferencia de utilizar uno u otro tipo de canales afecta en un 0.7% del resultado final.

Así mismo, parece ser que la utilización de una ventana de estudio de 3 segundos favorece la caracterización de los movimientos comparado con una ventana de menor o mayor tamaño.

Por otro lado, la utilización de un número mayor de filtros en la capa de convolución tiende a mejorar las prestaciones del sistema, sin embargo, el porcentaje de mejora conseguido se hace más patente al pasar de 8 a 16 filtros que posteriormente al duplicar o cuadruplicar dicho número, por lo cual, a la hora de diseñar una red de este tipo, debe plantearse si es más conveniente conseguir esa ligera mejora o reducir la cantidad de filtros y con ello la carga computacional y complejidad de la red, traduciéndose a su vez en unos tiempos de espera en el proceso de entrenamiento y testeo menores.

En cuanto al ancho de la ventana de la capa de *max-pooling*, ésta puede provocar problemas si su valor es alto, ya que en los escenarios donde la frecuencia de muestreo es baja, el número de muestras introducidas en la red no será suficiente como para poder atravesar varias veces las capas de convolución (la capa de *max-pooling* va reduciendo el tamaño de los datos), impidiendo de este modo el correcto funcionamiento.

El resto de hiperparámetros no parece afectar de una manera clara y significativa al sistema. Su comportamiento depende de las combinaciones de los demás parámetros.

Referencias

1. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET); IEEE: Antalya, Turkey, August 21-23, 2017; Vol. 2018-Janua, pp. 1–6.
2. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning - An MIT Press book*; 2016;
3. Frank, K.; Vera Nadales, M.J.; Robertson, P.; Pfeifer, T. Bayesian recognition of motion related activities with inertial sensors. In Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Ubicomp '10; ACM Press: Copenhagen, Denmark, September 26-29, 2010; p. 445.
4. Özdemir, A.T. An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice. *Sensors* **2016**, doi:10.3390/s16081161.
5. Saleh, M.; Abbas, M.; Le Jeannes, R.B. FallAID: An Open Dataset of Human Falls and Activities of Daily Living for Classical and Deep Learning Applications. *IEEE Sens. J.* **2021**, *21*, doi:10.1109/JSEN.2020.3018335.
6. Aziz, O.; Musngi, M.; Park, E.J.; Mori, G.; Robinovitch, S.N. A comparison of accuracy of fall detection algorithms (threshold-based vs. machine learning) using waist-mounted tri-axial accelerometer signals from a comprehensive set of falls and non-fall trials. *Med. Biol. Eng. Comput.* **2017**, *55*, 45–55, doi:10.1007/s11517-016-1504-y.
7. Human Factors and Ergonomics Lab- Korea Advanced Intitute of Science and Technology KFall: A Comprehensive Motion Dataset to Detect Pre-impact Fall for the Elderly based on Wearable Inertial Sensors Available online: <https://sites.google.com/view/kfalldataset>.
8. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A Fall and Movement Dataset. *Sensors* **2017**, *17*, 198, doi:10.3390/s17010198.
9. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. In Proceedings of the 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) (Procedia Computer Science); Leuven, Belgium, July 24-26, 2017; Vol. 110, pp. 32–39.
10. Martínez-Villaseñor, L.; Ponce, H.; Brieva, J.; Moya-Albor, E.; Núñez-Martínez, J.; Peñafort-Asturiano, C. UP-Fall Detection Dataset: A Multimodal Approach. *Sensors* **2019**, *19*, 1988, doi:10.3390/s19091988.
11. MathWorks GPU Support by Release Available online: <https://uk.mathworks.com/help/parallel-computing/gpu-support-by-release.html> (accessed on Dec 8, 2021).
12. Vedaldi, A.; Lenc, K. MatConvNet Convolutional Neural Networks for MATLAB. *Proc. 23rd ACM Int. Conf. Multimed.*, doi:10.1145/2733373.
13. Mathworks MATLAB GPU Computing Support for NVIDIA CUDA Enabled

GPUs Available online: <https://uk.mathworks.com/solutions/gpu-computing.html> (accessed on Dec 8, 2021).

14. Santoyo-Ramon, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Anexo: Contribuciones al estudio y análisis de sistemas-vestibles-de detección de caídas Available online: https://github.com/jasr92/Anexo_Contribuciones-al-estudio-y-an-lisis-de-sistemas-vestibles-de-detecci-n-de-ca-das (accessed on Sep 17, 2021).

Apéndice E. Testeo cruzado de repositorios utilizando CNN

E.1. Introducción

Una de las problemáticas que se han abordado y puesto de manifiesto a lo largo de este trabajo es la gran heterogeneidad existente entre los repositorios de movimientos de la literatura, así como, la gran diversidad que también se encontrarían en los movimientos ejecutados en escenarios reales.

Por ello, el presente estudio trata de evaluar el comportamiento de un sistema de detección de caídas que utiliza una red neuronal convolucional o CNN (*Convolutional Neural Network*) para discernir entre ADL y caídas cuando se entrena con una población de movimientos y se testea con muestras provenientes de otro repositorio.

Este trabajo se vincula a la publicación adjunta en el apéndice A. Publicaciones en la sección A.11. Evaluación del testeo cruzado de repositorios en Sistemas Vestibles de Detección de Caídas, la cual será enviada al congreso internacional *PErvasive Technologies Related to Assistive Environments* (PETRA 2022).

E.2. Metodología

A continuación, se expone cada uno de los factores que deben tenerse en cuenta en la experiencia llevada a cabo en este trabajo.

E.2.1. Repositorios de movimientos

En primer lugar, tal y como se ha ido haciendo a lo largo de este trabajo, primero se seleccionan una serie de repositorios de movimientos con el fin de analizar sus muestras sistemáticamente.

En este caso se utilizan cuatro *datasets*: Erciyas [1], KFall [2], SisFall [3] y UMAFall [4]. Las razones para utilizar estas cuatro bases de datos son fundamentalmente dos: de un lado, son cuatro de los repositorios con mayor cantidad y variedad de tipologías de movimientos representados. Por otra parte, las frecuencias de muestreo empleadas para capturar las muestras son múltiplos o divisores exactos entre ellas, lo que facilitará el proceso de diezmado a la hora de entrenar y testear con unas y otras¹.

Para nuestro estudio, se analizan solamente las trazas recogidas por el sensor ubicado en la cintura de los sujetos, ya que es una posición común en todos los repositorios estudiados. Sin embargo, en el caso de UMAFall se añade también los datos recogidos por el sensor del *smartphone*, ubicado en el bolsillo del pantalón (muslo), ya que su frecuencia de muestreo es múltiplo de las frecuencias de muestreo de los otros repositorios y puede ser interesante analizar el comportamiento que presenta.

E.2.2. Métricas de calidad

Como en los estudios anteriores, se vuelve a utilizar como métrica global de calidad la media geométrica de la sensibilidad y la especificidad del sistema.

E.2.3. *Data Augmentation* o generación de datos artificiales

Data Augmentation es una técnica que permite crear nuevos datos de manera artificial a partir de alteraciones de los datos originales. El objetivo de esta práctica es permitir aumentar el número y la diversidad de los datos de

¹ En el caso de las muestras de Erciyas la frecuencia de muestreo es 25 Hz mientras que el sensor empleado en la cintura por UMAFall usaba una tasa de 20 Hz, razón por la que no se harán las pruebas cruzadas con estos dos repositorios.

entrenamiento [5], algo especialmente útil cuando el espacio muestral de entrenamiento y test es muy limitado o cuando algún tipo de dato está infrarrepresentado.

En el caso de este estudio, se utiliza esta técnica para simular la situación en la que se entrega al usuario a monitorizar un sistema de detección de caídas ya entrenado y se pretende que dicho sujeto que lo utiliza pueda añadir nuevas muestras personalizadas al modelo. Por tanto, en este trabajo se generan nuevos datos artificiales a partir de la selección de una serie de movimientos del *subset* o subconjunto de testeo. Esta selección se realiza de modo que se intenta que se incluyan movimientos de todos los tipos que están presentes en el repositorio empleado. En el proceso a estos datos se les añade cierto ruido blanco, mediante la función *awgn()* de Matlab, manteniendo una relación señal-ruido o SNR (*Signal to Noise Ratio*) de 20 dB entre la señal de la muestra original y el ruido agregado.

E.2.4. Configuración de la red convolucional utilizada

Para el proceso de clasificación se considera una red neuronal convolucional, en la cual se utiliza la configuración y estructura recomendada en el Apéndice D y resumida en la tabla E.1

Nombre del hiperparámetro	Valores utilizados
Tamaño de la ventana de estudio (t_w)	3 s
Nº de capas convolucionales	5
Nº de filtros	64
Dimensión de los filtros convolucionales	4 x 1
Ancho de <i>padding</i>	4
<i>Stride</i>	1x1
Método de optimización	<i>Stochastic gradient descent with momentum</i>
L2 Regulation	0.0001
Ancho de la ventana de <i>Max-Pooling</i>	2
Tamaño del <i>miniBatch</i>	64
Tasa de aprendizaje inicial	0.01

Paciencia de validación	5
Tipo de canal	Triaxial (un canal por cada dimensión del acelerómetro)
Tipo de normalización	z-score (o ninguna)

Tabla E.1. Resumen de los hiperparámetros de configuración de la CNN utilizados para este estudio.

E.2.5. Pruebas

El proceso de entrenamiento, validación y testeo es similar en todas las pruebas. Cada repositorio se divide en cinco particiones equilibradas, es decir, cinco fracciones o subconjuntos que incluyen un número similar de movimientos del mismo tipo y de todos los sujetos participantes.

En cada prueba, el repositorio seleccionado para entrenamiento utiliza cuatro de esas particiones (el 80% de las muestras) para el proceso de entrenamiento en tanto que el *subset* restante (20% de muestras) es utilizado en la validación. De este modo se entrena la red y se consigue el modelo del sistema de clasificación. Este proceso se repite cinco veces, variando la partición de validación, de manera que se obtienen cinco modelos de red CNN diferentes. Estos modelos ya configurados se testearán con repositorios distintos del utilizado.

Además de esta primera prueba, en la que el sistema se entrena utilizando solamente las muestras correspondientes al repositorio seleccionado para entrenamiento, se considera una segunda tipología de experimentos, en la se sigue el mismo procedimiento anterior, pero añadiendo al conjunto de entrenamiento, a su vez, una serie de muestras derivadas de aplicar una técnica de aumento de datos (*Data Augmentation*) o generación de datos artificiales a un 10% de las trazas del repositorio utilizado para el proceso de testeo. Este proceso simula el escenario en el que se entrega un sistema de detección de caídas previamente entrenado, que posteriormente puede además entrenarse con muestras personalizadas del sujeto que lo utiliza para personalizar o reajustar aún más el sistema a los movimientos del usuario final.

Dentro de este segundo tipo de prueba se analizan dos experiencias diferentes.

- En la primera experiencia se agregan sólo movimientos del tipo ADL generadas por *data augmentation* (en particular, se prueba

añadiendo un 10% del número de las muestras de ADL totales del repositorio original usado para el testeo).

- En la segunda experiencia se realiza el mismo proceso, pero añadiendo también las caídas (agregando un 10% del número de caídas totales del repositorio de test).

Como se ha apuntado, para el proceso de testeo se utilizan los cinco modelos de red obtenidos en la fase de entrenamiento y todas las muestras de un repositorio diferente al contemplado inicialmente para entrenar. A partir de ahí, se computan las métricas que permiten evaluar la calidad de la clasificación de cada modelo, de forma que la métrica de calidad global será la media de esas cinco métricas individuales.

E.2.6. Tratamiento de los datos para llevar a cabo el testeo cruzado

Los datos que se introducen en una red convolucional deben tener un tamaño predefinido. En los estudios realizados hasta ahora el tamaño venía determinado por de la ventana de observación seleccionada (t_w) y la frecuencia de muestreo de cada uno de los repositorios. Como cada repositorio puede tener una frecuencia de muestreo diferente, el número de muestras por las que están compuestas las trazas de los movimientos puede diferir entre repositorios.

A la hora de testear con movimientos de repositorios diferentes respecto a los que se utiliza para entrenar, se debe llevar a cabo un tratamiento de los datos con el objetivo de que los tamaños de las trazas coincidan y sean compatibles. Por ello, en esta experiencia, la manera de proceder es la siguiente: primero se analiza cuál de los repositorios utiliza una frecuencia de muestreo mayor. Una vez localizado, los movimientos de este repositorio se submuestran, es decir, se decrementa su frecuencia de muestreo por un factor entero M (proceso conocido como diezmo). De esta forma, puede trabajarse con ambos repositorios del mismo modo que se ha venido haciendo hasta el momento cuando se utilizaban muestras de un mismo *dataset*, ya que todos los datos trabajan con la misma frecuencia de muestreo, por consiguiente, el número de muestras contenidas en la ventana de estudio será igual y por tanto compatible.

A la hora de normalizar las entradas con las que se alimentan las redes convolucionales, se llevan a cabo tres tipos de estandarización diferentes:

1. Sin normalización: No se realiza ningún tipo de normalización a los datos de entrada.
2. Normalización *z-score* a los datos de entrenamiento y prueba de forma separada: Se realiza una normalización del tipo *z-score* (restando la media y dividiendo por la desviación típica) a los *subsets* de entrenamiento y testeo de manera individual. Para obtener los estadísticos necesarios para realizar la normalización se tienen en cuenta los valores de los tres ejes de los sensores de manera individual.
3. Normalización *z-score* a partir de los datos de entrenamiento: En este caso se vuelve a normalizar mediante la técnica *z-score* tanto al subconjunto de entrenamiento como al de testeo, sin embargo, esta vez, los estadísticos para realizar dicha normalización provienen solamente de los datos de entrenamiento (esta situación se correspondería con el escenario -más realista- en el que un sistema ya entrenado y operando como detector normaliza las medidas de los sensores a partir de valores pre-establecidos basados en las muestras empleadas durante el entrenamiento).

E.3. Resultados

Tras llevar a cabo las diferentes pruebas descritas anteriormente se obtienen las siguientes métricas de calidad resumidas en la Tabla E.1.

Repositorios empleados para entrenamiento y test	Tipo de normalización								
	Ninguna			z-score basado solo en los datos entrenamiento			z-score datos de entrenamiento y testeo de forma separada		
	Muestras añadidas al entrenamiento			Muestras añadidas al entrenamiento			Muestras añadidas al entrenamiento		
	Ninguna	Sólo ADL	ADL+Caídas	Ninguna	Sólo ADL	ADL+Caídas	Ninguna	Sólo ADL	ADL+Caídas
Erciyas (Entrenamiento)									
KFall (Test)	0.6547	0.2193	0.9984	0.7618	0.6078	0.9969	0.5141	0.2703	0.9978
SisFall (Test)	0.5523	0.4142	0.9908	0.8050	0.7226	0.9865	0.6537	0.3292	0.9893
UMAFall_Smartphone (Test)	0.8354	0.8720	0.9578	0.7262	0.9021	0.9189	0.7559	0.7215	0.9268
KFall (Entrenamiento)									
Erciyas (Test)	0.5790	0.3533	0.9933	0.5492	0.2209	0.9697	0.7222	0.5536	0.9916
SisFall (Test)	0.9791	0.9815	0.9928	0.9796	0.9817	0.9904	0.9830	0.9820	0.9916
UMAFall Cintura (Test)	0.5507	0.4349	0.9804	0.7344	0.5505	0.8523	0.6601	0.3204	0.9723
UMAFall_Smartphone (Test)	0.6845	0.6485	0.9582	0.6972	0.6756	0.9411	0.6725	0.5885	0.9376
SisFall (Entrenamiento)									
Erciyas (Test)	0.7139	0.4511	0.9937	0.4861	0.2697	0.9698	0.7241	0.4927	0.9918
KFall (Test)	0.9896	0.9910	0.9986	0.9933	0.9912	0.9989	0.9938	0.9953	0.9985
UMAFall Cintura (Test)	0.6583	0.4376	0.9670	0.6525	0.5311	0.8103	0.6086	0.4656	0.9609
UMAFall_Smartphone (Test)	0.7075	0.5955	0.9407	0.6799	0.6515	0.9493	0.7156	0.6228	0.9336
UMAFall Cintura (Entrenamiento)									
KFall (Test)	0.4118	0.2506	0.9976	0.3699	0.1281	0.9954	0.4278	0.2605	0.9968
SisFall (Test)	0.5185	0.3812	0.9907	0.4435	0.2848	0.9815	0.5492	0.3353	0.9888
UMAFall_Smartphone (Test)	0.8151	0.8886	0.9351	0.8330	0.8714	0.9459	0.8473	0.8615	0.9446
UMAFall_Smartphone (Entrenamiento)									
Erciyas (Test)	0.9300	0.9234	0.9937	0.7775	0.7651	0.9674	0.9467	0.9245	0.9934
KFall (Test)	0.9686	0.9470	0.9975	0.9619	0.9436	0.9968	0.9489	0.9063	0.9973
SisFall (Test)	0.9139	0.8870	0.9865	0.9068	0.8911	0.9831	0.9155	0.8510	0.9853
UMAFall Cintura (Test)	0.9694	0.9706	0.9897	0.9673	0.9677	0.9839	0.9199	0.9241	0.9768

Tabla E.2. Media geométrica de Se y Sp obtenida por el sistema de detección de caídas al utilizar para el entrenamiento y testeo diferentes repositorios. También se reúnen los resultados obtenidos al añadir en el proceso de entrenamiento muestras derivadas de aplicar una técnica de aumento de datos (Data Augmentation) al repositorio utilizado para el proceso de testeo, así como, diferentes técnicas de normalización.

E.4. Discusión de los resultados

En primer lugar, los resultados demuestran que entrenar y testear con repositorios diferentes puede degradar notablemente las prestaciones del sistema de detección de caídas, ya que, como se muestra en el Apéndice D (véase Tabla D.3.), el mismo sistema de detección basado en una red neuronal y utilizando los mismos hiperparámetros, conseguía medias geométricas de la especificidad y la sensibilidad cercanos a 0.99 para la mayoría de los casos, cuando se entrenaba y testeaba con muestras de provenientes del mismo repositorio. Sin embargo, en esta ocasión, en la que se utilizan movimientos de diferentes repositorios para el entrenamiento y el testeo, las prestaciones empeoran notablemente, llegando a bajar en ocasiones del 0.5, con independencia del sistema de normalización utilizado.

Se podría pensar que esta degradación de los resultados puede justificarse en pérdidas de información derivadas de disminuir la frecuencia de muestreo al tratar los datos (al aplicar el diezmado). Sin embargo, tal y como se demuestra en el Capítulo 7 la utilización de frecuencias de muestreo superiores a 5 Hz es suficiente para conseguir buenas prestaciones.

Cabe presuponer que el problema proviene del hecho que los algoritmos tienden a “sobre-aprender” los movimientos con los que se entrena, es por ello, que entrenar con repositorios que contienen una mayor variedad de muestras y tipos de movimientos puede provocar una generalización mejor.

Así mismo, se comprueba que, al añadir al entrenamiento muestras provenientes de llevar un proceso de aumento de datos a las muestras de testeo, el sistema mejora drásticamente sus prestaciones cuando los movimientos añadidos son caídas. Este hecho se explica a partir del mismo razonamiento anterior, al añadir una mayor variedad de movimientos, el sistema es capaz de reconocerlos de forma correcta posteriormente. Sin embargo, también existen ocasiones en las que al personalizar el entrenamiento con solamente ADL las prestaciones decaen de manera crítica, por ejemplo, al entrenar con SisFall o KFall y testear con Erciyos o UMAFall. Esta situación puede deberse a que los movimientos añadidos comparten similitudes con los movimientos etiquetados como caídas en el repositorio utilizado como entrenamiento (puede comprobarse en el estudio

realizado por Casilari et al en [6]), por lo que, a la hora de entrenar con estos nuevos datos, el sistema se encuentra con mayores dificultades para discernir entre ambas clases de movimientos.

Otro hecho llamativo observado es la variación que padece la prestación del clasificador con las muestras del repositorio UMAFall dependiendo de la posición de captura de los datos (sensor ubicado en la cintura o el sensor del *smartphone* localizado en el bolsillo del pantalón). A pesar de que ambos sensores han monitorizado los mismos movimientos y que, como ya se ha comentado, la frecuencia de muestreo (distinta en este caso, pues el *smartphone* utilizaba 200 Hz y el sensor en la cintura 20 Hz) no debería afectar a la hora de realizar la clasificación, de los resultados parece deducirse que entrenar con los datos obtenidos por el sensor del *smartphone* produce unas métricas de calidad mucho mejores que al hacerlo con el sensor ubicado en la cintura. Esta situación puede venir dada por determinados sesgos intrínsecos del sensor que provocan que los algoritmos interpreten peor los resultados. A partir de este hecho, se puede concluir que los conjuntos de datos públicos de evaluación comparativa dependen en gran medida de los sensores utilizados.

Por otro lado, la normalización de los datos mediante la técnica de unidad tipificada o *z-score* no parece mejorar el rendimiento a la hora de clasificar en todas las ocasiones, existen situaciones que incluso afecta negativamente. Esta situación puede venir motivada por el hecho de que los estadísticos utilizados en esta técnica (la media y la desviación típica) son muy sensibles a los valores anómalos (muy grandes o pequeños). En el caso de las trazas de los movimientos existen picos de aceleración muy abruptos, principalmente en las caídas y en otros movimientos con gran agitación, por tanto, normalizar mediante esta técnica provocaría una gran atenuación en estos valores picos y también en los valores valle, lo que finalmente repercutiría negativamente en el sistema a la hora de detectar patrones significativos para clasificar.

En resumen, se demuestra con este estudio la dificultad existente en el proceso de validación cruzada de repositorios (*cross dataset validation*), provocada principalmente, a nuestro entender, por la heterogeneidad de la tipología de los movimientos que componen los repositorios, así como, de las características de los sensores y por la tendencia al sobre-aprendizaje de los métodos de

clasificación. Del mismo modo, queda patente que añadir datos personalizados en el entrenamiento de los sistemas de detección de caídas puede suponer una mejora eficaz de sus prestaciones, siempre y cuando, dichos movimientos personalizados sean caídas. Estos resultados ponen de relieve la enorme importancia de emplear un número cuanto más amplio posible de muestras y repositorios a la hora de evaluar cualquier sistema de detección de caídas y la necesidad de testear los sistemas entrenados con los usuarios finales reales o, al menos, con un universo muestral distinto (en tipología de usuarios y movimientos) del empleado para el entrenamiento, con el objeto de determinar la robustez del clasificador y su auténtica viabilidad en un escenario de aplicación real.

Referencias

1. Özdemir, A.T. An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice. *Sensors (Switzerland)* **2016**, doi:10.3390/s16081161.
2. Human Factors and Ergonomics Lab- Korea Advanced Institute of Science and Technology KFall: A Comprehensive Motion Dataset to Detect Pre-impact Fall for the Elderly based on Wearable Inertial Sensors.
3. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A fall and movement dataset. *Sensors (Switzerland)* **2017**, *17*, doi:10.3390/s17010198.
4. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. *Procedia Comput. Sci.* **2017**, *110*, 32–39, doi:10.1016/j.procs.2017.06.110.
5. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. *2018 Int. Interdiscip. PhD Work. IIPhDW 2018* **2018**, 117–122, doi:10.1109/IIPHDW.2018.8388338.
6. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems. *J. Healthc. Eng.* **2020**, *2020*, 1–36, doi:10.1155/2020/6622285.

Apéndice F. Publicaciones derivadas de esta tesis

F.1. Artículos publicados

1. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. Analysis of a smartphone-based architecture with multiple mobility sensors for fall detection. *PLoS One* 2016, 11, 1–17.
 - **ISSN:** 1932-6203
 - **doi:** <http://dx.doi.org/10.1371/journal.pone.0168069>
 - **Factor de Impacto:** 2.806 (Ranking 15 sobre 64 en categoría Multidisciplinary Sciences) (Q1).
2. Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.-M. Analysis of Public Datasets for Wearable Fall Detection Systems. *Sensors* 2017, 17, 1513.
 - **ISSN:** 1424-8220
 - **doi:** <http://dx.doi.org/10.3390/s17071513>
 - **Factor de Impacto:** 2.475 (Ranking 16 sobre 61 en Categoría Instruments & Instrumentation) (Q2).
3. Santoyo-Ramón, J.; Casilari, E.; Cano-García, J. Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection with Supervised Learning. *Sensors* 2018, 18, 1155.
 - **ISSN:** 1424-8220
 - **doi:** <http://dx.doi.org/10.3390/s18041155>
 - **Factor de Impacto:** 2.806 (Ranking 15 sobre 64 en categoría Multidisciplinary Sciences) (Q1).

4. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. On the Heterogeneity of Existing Repositories of Movements Intended for the Evaluation of Fall Detection Systems. *J. Healthc. Eng.* 2020, 2020, 1–36.
 - **ISSN:** 2040-2295
 - **doi:** <http://dx.doi.org/10.1155/2020/6622285>
 - **Factor de Impacto:** 2.682 (Ranking 53 sobre 108 en categoría HEALTH CARE Sciences & Services) (Q2).
5. Santoyo-Ramón, J.A.; Casilari, E.; Cano-García, J.M. A Study of One-Class Classification Algorithms for Wearable Fall Sensors. *Biosensors* 2021, 11, 284.
 - **ISSN:** 2079-6374
 - **doi:** <http://dx.doi.org/10.3390/bios11080284>
 - **Factor de Impacto:** 5.519 (2020, último valor disponible) (Ranking 8 sobre 64 en categoría Instruments & Instrumentation) (Q1).
6. Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A study on the impact of the users' characteristics on the performance of wearable fall detection systems. *Sci. Reports* 2021 111 2021, 11, 1–14.
 - **ISSN:** 2045-2322
 - **doi:** <https://doi.org/10.1038/s41598-021-02537-z>
 - **Factor de Impacto:** 4.380 (2020, último valor disponible) (Ranking 17 sobre 72 en categoría Multidisciplinary Sciences) (Q1)

F.2. Artículos enviados o pendientes de envío

1. Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. A Study of the Influence of the Sensor Sampling Frequency on the Performance of Wearable Fall Detectors. *Pendiente de envío a Measurement*. 2021
2. Santoyo-Ramón, J.A.; Casilari, E.; Cano-García, J.M. Evaluation of the cross-datasets validation process of a FDS based on a CNN. *Pendiente de envío al congreso "PErvasive Technologies Related to Assistive Environments (PETRA 2022)", a celebrar en Corfú del 29 de junio al 1 de julio de 2022.*

F.3. Ponencias en congresos

1. Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection. In Proceedings of the 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) (Procedia Computer Science); Leuven, Belgium, July 24-26, 2017; Vol. 110, pp. 32–39.
 - **ISSN:** 1877-0509
 - **doi:** <http://dx.doi.org/10.1016/j.procs.2017.06.110>
2. Santoyo-Ramón, J.; Casilari, E.; Cano-García, J. Estudio de la detección de caídas utilizando el algoritmo SVM. In Proceedings of the Actas del XXXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2018); Granada, España, 2018; pp. 5–7.
 - **ISBN:** 978-84-17293-44-4
3. Casilari, E.; Santoyo-Ramón, J.-A.; Cano-García, J.M. Analysis of a Public Repository for the Study of Automatic Fall Detection Algorithms. In Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '18); ACM: Berlin, Germany, September 20-21, 2018; pp. 1–10.
 - **ISBN:** 978-14-50364-87-4
 - **doi:** <https://doi.org/10.1145%2F3266157.3266211>
4. Santoyo-Ramón, J.A.; Casilari-Pérez, E.; Cano-García, J.M. Study of the Detection of Falls Using the SVM Algorithm, Different Datasets of Movements and ANOVA. In Proceedings of the IEEE IWBBIO: International Work-Conference on Bioinformatics and Biomedical Engineering; Springer: Granada, Spain, May 8-10, 2019; Vol. 11465 LNBI, pp. 415–428.
 - **ISBN:** 978-30-30179-37-3
 - **doi:** https://doi.org/10.1007/978-3-030-17938-0_37

