DISEÑO DE UN SISTEMA BASADO EN TECNOLOGÍAS MÓVILES, WEARABLES Y ANÁLISIS DE DATOS PARA PROMOVER EL ENVEJECIMIENTO ACTIVO EN MAYORES

Francisco Manuel García Moreno

Tesis Doctoral

Programa de Doctorado en Tecnologías de la Información y la Comunicación

Directoras:
María José Rodríguez Fórtiz
María Bermúdez Edo

UNIVERSIDAD DE GRANADA

Departamento de Lenguajes y Sistemas Informáticos

Editor: Universidad de Granada. Tesis Doctorales
Autor: Francisco Manuel García Moreno
ISBN: 978-84-1117-629-3
URI: https://hdl.handle.net/10481/79154

UNIVERSIDAD
DE GRANADA

DISEÑO DE UN SISTEMA BASADO EN TECNOLOGÍAS MÓVILES, WEARABLES Y ANÁLISIS DE DATOS PARA PROMOVER EL ENVEJECIMIENTO ACTIVO EN MAYORES

Tesis Doctoral

Programa de Doctorado en Tecnologías de la Información y la Comunicación

Francisco M. Garcia-Moreno

Directoras:

María José Rodríguez Fórtiz

María Bermúdez Edo

Departamento de Lenguajes y Sistemas Informáticos

Granada, noviembre 2022

A mis padres, por todo

A Emi, por hacer habitable cada página

A mis politos, por existir

AGRADECIMIENTOS

Si me dijeran que el lenguaje tiene límites, no dudaría de que estos párrafos torpes son el ejemplo perfecto. Solo ahora mis ojos —la distancia— me permiten la contemplación de todo este bosque, aún con un destino incierto, pero con unas raíces profundas que no sólo merecen un reconocimiento, sino el más sincero —sin adorno— de los agradecimientos, siempre insuficientes. Sin cada una de estas raíces el bosque no sería un bosque, más bien, sería un desierto. Y esta tesis: un oasis —una ilusión—, una terrible hoja en blanco.

Quiero dar mi profundo agradecimiento a todos los miembros del grupo de investigación «Modelling & Development of Advanced Software Systems» (MYDASS), de la Universidad de Granada (UGR), por su generosa acogida desde el primer momento y apoyo en mi trayectoria investigadora. Especialmente a la Dra. María Visitación Hurtado Torres, Mavi, y a Dr. José Luis Garrido Bullejos, por su confianza, apoyo y entusiasmo constantes; al Dr. Ángel Ruiz Zafra por su conversación y complicidad; a mi *compañera siamesa* (como diría la Dra. Carmen Villaverde, conste aquí también mi gratitud hacia a ella) la Dra. Estefanía Rodríguez García y gran enfermera; y a mi compañera de oficina y vivencias Fernanda Lugmaña, Fer.

Me gustaría dejar constancia el merecido reconocimiento al Dr. Payam Barnaghi, catedrático de Aprendizaje Automático (aplicado a la medicina) del Imperial College London de Reino Unido, por abrirme las puertas del Dementia Research Institute, por su calurosa acogida, disposición y apoyo incesante durante mi estancia en el Departamento Brain Sciences, en la etapa final de mi investigación. Del mismo modo, merece mi agradecimiento el doctorando Alex Capstick, por su confianza, compañerismo ejemplar y su ayuda desinteresada; así como al futuro doctor Juan David García Corchero, por hacerme la estancia más amena en Londres.

También quiero agradecer especialmente tanto a los trabajadores como a los usuarios de los distintos centros y residencias de mayores de Granada, que nos abrieron sus puertas y su corazón incondicionalmente: la Residencia El Pinar (La Zubia), donde realizamos la primera prueba piloto; la Asociación de Cerrillo de Maracena, donde iniciamos la fase experimental; y los centros cívicos del Zaidín y Maracena. Mis agradecimientos infinitos, insisto, a sus usuarios —en su mayoría mujeres— mayores de 65 años, por su predisposición a colaborar en el estudio, confianza, paciencia y conversaciones.

También agradezco a los miembros del grupo de investigación «Discriminación y estigma: aspectos psicosociales» de la UGR, especialmente, por su apoyo a una nueva línea de proyectos y estudios sobre emociones: a la Dra. Francisca Expósito Jiménez, catedrática de Psicología Social de la UGR; y a la investigadora —futura doctora—, mi amiga inseparable y confidente Marta Badenes Sastre, gracias por acompañarme en cada aventura personal y académica. De la misma manera, mi más profundo agradecimiento por su confianza plena, también a mi amiga e investigadora —inminente doctora— Ana Álvarez Muelas (Dpto. de Personalidad, Evaluación y Tratamiento Psicológico, de la UGR), compañeros de anhelos, proyectos y pasión académica.

Gracias por su confianza, conversación y apoyo incesante a mi trayectoria académica al Dr. Miguel Ángel Naranjo Gutiérrez, del Dpto. de Ciencias de la Computación e Inteligencia Artificial (CCIA) de la Universidad de Sevilla, con quien, además, comparto la afición a la escritura creativa. Asimismo, agradecer también su conservación y confianza a los doctores Carlos Cano Gutiérrez, Igor Zwir Nawrocki, Jesús Alcalá Fernández, Alberto Fernández Hilario y a la Dra. Coral del Val, del Dpto. CCIA de la UGR.

Gracias también a mis amigos y compañeros de fatigas, por hacer que el tiempo no pase y que esta investigación fuera un camino más interesante y ameno, más incluso en los tiempos de encierro en pandemia: David Granados, Dani, Moti, Richy, Fran Navarro, Coral y Lorena.

Agradezco, muy especialmente, su guía por estos círculos y su entrega total a mis directoras de tesis, María José Rodríguez Fórtiz y María Bermúdez Edo. A María José por su capacidad y entusiasmo para ayudar a la asistencia de personas con necesidades especiales. A María por su rigor, su sensibilidad y respeto. A ambas por su mano tendida desde el minuto cero, su confianza plena, su inquebrantable paciencia, su buen criterio y su amistad.

Gracias a toda mi familia y a los que ya no están entre nosotros, por confiar siempre en mí y brindarme su apoyo incondicional, sin pedirme nunca nada a cambio. A mis padres, Manuel y Jero, por las conversaciones, los abrazos y por haber sabido tolerar, pacientemente, con el cariño más sincero e infinito la pasión desmedida de un investigador entregado. Por todas las llamadas a deshoras que me ayudaron a desconectar —posponer esta obsesión para el día siguiente.

Con la misma torpeza e impotencia del lenguaje agradezco muy especialmente la existencia de Emi, mi fiel compañera en el viaje y en estas páginas. Quién no merece que al menos una vez en la vida le besen la cara y las manos. Gracias, Emi, por los besos, por tu presencia y aparecer en mi camino, por tu sonrisa, complicidad, sensibilidad y empatía y, como no, por no dejarme caer.

Y a todos mis *politos* —los que estuvieron, los que están y los que estén por venir—, por existir y darme el amor más incondicional.

RESUMEN

En las evaluaciones tradicionales del estado de salud de las personas mayores, como la fragilidad o la dependencia, subyace un componente subjetivo basado en las respuestas de los encuestados, unos costes asociados de recursos para los sistemas sociosanitarios y una falta de gestión de la prevención realista, pues no se detectan problemas de salud de forma temprana. Por ello, es necesario el desarrollo e implementación de nuevos sistemas tecnológicos que sean capaces de monitorizar y evaluar estos estados de salud, de forma objetiva, ubicua y transparente para sus usuarios, contribuyendo a la disminución de los costes derivados y promoviendo la prevención.

En este trabajo se presenta el diseño de un sistema *mobile-Health*, lo que significa utilizar diversos dispositivos móviles como *wearables*, para recopilar datos fisiológicos de las personas mayores, procesarlos y analizarlos de forma objetiva. El trabajo está motivado por la superación de las limitaciones de las evaluaciones tradicionales de los estados de fragilidad y dependencia en las personas mayores. En este sentido, se propone la arquitectura del sistema dentro del paradigma del Internet de las cosas (IoT), *m-IoTHealth*, donde los distintos dispositivos móviles están interconectados. En concreto el sistema se basa en el estilo arquitectónico de *microservicios*, donde cada uno de los componentes del sistema está especializado en una tarea concreta. Más concretamente, se proponen los componentes necesarios para recopilar y federar las diferentes fuentes de datos heterogéneas, procesarlas y analizarlas. Además, en este trabajo se generan modelos de aprendizaje automático para evaluar la fragilidad y la dependencia. Asimismo, se propone un modelo de detección de la intención motora como una primera aproximación en el estudio de datos cognitivos.

Los resultados de esta tesis aportan una solución tecnológica integral para abordar las evaluaciones de fragilidad y dependencia en personas mayores, la cual no se había abordado antes de esta manera en el estado del arte. Esta solución incluye el diseño de alto nivel de una arquitectura basada en *microservicios* para organizar el sistema en componentes altamente especializados en las diferentes tareas que se esperan de un *m-Health*: recopilación, procesamiento y análisis de los datos de los pacientes. En este sentido, hasta donde sabemos, es la primera vez que se hace uso de dispositivos móviles para recopilar datos fisiológicos en un entorno real —durante el desempeño de una actividad compleja (multidimensional: física, social y cognitiva) como es ir a *hacer la compra*—, procesar esos datos y generar modelos de aprendizaje automático, objetivamente precisos en la evaluación de los estados de fragilidad y dependencia en mayores. Asimismo, gracias al diseño propuesto basado en *microservicios* especializados, la solución es flexible y permite la reutilización e incorporación de sus distintos componentes para generar nuevos modelos de evaluación e integrarlos en el sistema, como el caso de nuestro modelo de detección de la intención motora.

Para la generación de estos modelos, se utilizaron los dispositivos *wearables* Samsung Gear S3 y Empatica E4, sus sensores de acelerómetro, giroscopio, ritmo cardíaco, temperatura, actividad electrodérmica de la piel (EDA) y la diadema Muse 2, que mide señales de electroencefalograma (EEG). Los datos recopilados de estos dispositivos se preprocesaron con diferentes técnicas como la segmentación en ventanas y el alineamiento de señales. Asimismo, se exploraron diferentes algoritmos de aprendizaje automático como k-Nearest Neighbors, Random Forest, Naïve Bayes y redes neuronales para series temporales (LSTM). Con todo ello, se optimizó la generación de los mejores modelos realizando una búsqueda de los mejores *hiperparámetros* y la selección de características para reducir la

dimensionalidad y encontrar las características más relevantes para la evaluación de fragilidad, dependencia e intención motora.

Esta propuesta tiene el potencial, en general, de promover la prevención del deterioro del estado de salud y, en particular, de fragilidad y dependencia, suponiendo una evaluación y monitorización objetiva y una reducción de costes sociosanitarios.

ABSTRACT

Traditional assessments of the elderly's health status, such as frailty or dependence, rely on subjective responses to questionnaires. These questionaries have associated resource costs for healthcare systems and entail a lack of realistic prevention management. Hence, it is necessary to develop new technological systems capable of evaluating these health conditions, in an objective, ubiquitous and transparent way for the users, contributing to reducing the derived costs and promoting prevention.

This work presents a mobile-Health system that collects and analyzes physiological data from the elders using various mobile devices. The system architecture follows the Internet of Things (IoT) paradigm, m-IoTHealth, interconnecting different mobile devices. Specifically, the system adopts the microservices architectural style, in which each component specializes in a single, fine-grained, task. These tasks consist of collecting, federating, and analyzing the heterogeneous data sources. To assess our system, we generated machine-learning models to evaluate the fragility and dependency status. We also proposed a motor imagery detection model as the first approach to study cognitive data.

The results of this thesis provide a comprehensive technological solution to assess frailty and dependency of elderly people. This solution includes the high-level design of a microservices-based architecture with highly specialized components for the different tasks expected from an m-Health. To the best of our knowledge, this is the first time a system collects and processes physiological data generating machine-learning models during the performance of a complex activity (multidimensional: physical, social, and cognitive) such as shopping. The results are objective and accurate in the assessment of the frailty and dependency status of older adults. Thanks to the proposed design based on specialized microservices, the solution is flexible, and allows the reuse and integration of its different components to generate new assessment models, as in the case of our motor imagery detection model

The generation of these models involved the use of different sensors embedded in the Samsung Gear S3, Empatica E4, and Muse 2 wearables, such as the accelerometer, gyroscope, heart rate, temperature, electrodermal skin activity (EDA), and electroencephalogram (EEG) signals. We use different preprocessing techniques, such as window segmentation and signal alignment, and explore several machine learning algorithms such as k-Nearest Neighbors, Random Forest, Naïve Bayes, and neural networks for time series (LSTM). To optimize the models, we search for hyperparameters that reduce dimensionality and identify relevant features.

This proposal could promote the prevention of the deterioration of the health status by the early detection of decay. In particular, it could prevent frailty and dependence, reducing, at the same time, social and healthcare costs.

ÍNDICE DE CONTENIDOS

LISTA DE FIGURAS

I PhD Dissertation

1. Introducción

Un sistema *mobile-health* (*m-Health*) consiste en utilizar dispositivos móviles o inalámbricos, equipados con distintos sensores, para recolectar datos de los pacientes (físicos, mentales, etc.), almacenarlos y enviarlos a servidores conectados a Internet para su posterior procesamiento, análisis y extracción de conocimiento [1], [2]. Un *m-Health* es capaz de monitorizar a los pacientes a distancia de forma ubicua e incluso, en muchas ocasiones, sustituir procedimientos tradicionales costosos [2]. Generalmente, la arquitectura de un *m-Health* [3] (Figura 1) se compone de dos partes diferenciadas: la *body area network* (BAN) y el sistema *back-end* (BESys). Para ejecutar la monitorización de los pacientes, los sensores se organizan en la red de comunicación BAN y ésta, a su vez, envía los datos recogidos hacia otros dispositivos cercanos —unidades base móviles (MBU)— que actúan de puerta de enlace (*gateway*). De esta forma, la *gateway* se encarga de conectar la BAN con una red exterior, donde se encuentra el sistema *back-end*, BESys, esperando a recibir los datos sensoriales para su posterior tratamiento y análisis. El BESys está formado por uno o varios servidores *back-end* y aplicaciones complementarias para procesar las bioseñales y otros datos recibidos por los servidores. Todos estos datos se podrían utilizar en los sistemas sanitarios para monitorizar diferentes dimensiones de la salud humana, como pueden ser la dimensión física y la dimensión cognitiva o mental, entre otras. La adopción de los sistemas *m-Health* y el uso de dispositivos inalámbricos en los sistemas sanitarios contribuirá a una reducción de costes socioeconómicos [4].

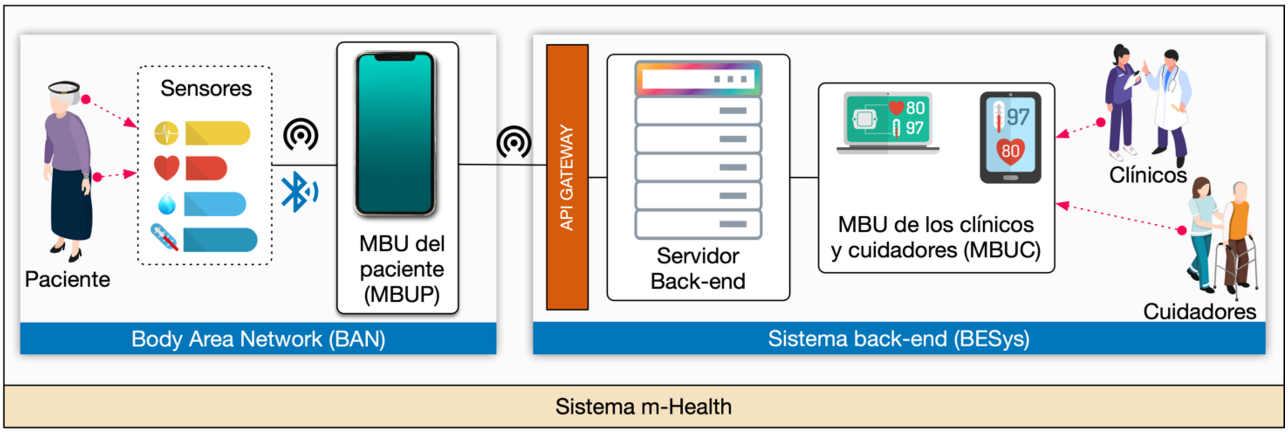

Figura 1. Escenario m-Health genérico

La BAN puede incluir dispositivos inalámbricos como teléfonos móviles o los que llamamos *vestibles*. Más conocidos como *wearables*, estos dispositivos pueden ofrecer en su interior distintos tipos de sensores capaces de medir señales de diversa naturaleza. Aunque los más extendidos captan señales físicas —a partir de sensores como el acelerómetro, el giroscopio, la sudoración de la piel, el ritmo cardíaco, oxígeno en sangre, etc.—, ya existen también *wearables* capaces de medir señales mentales —electroencefalografía (EEG)—. Por un lado (obviando el smartphone, ampliamente conocido), algunos ejemplos de dispositivos *wearables* que miden señales de dimensión física son: relojes inteligentes (Samsung Gear, Apple Watch, Fitbit Versa, Mobvoi TicWatch, etc.), pulseras (Empatica E4, Xiaomi Smart Band, etc.), bandas de pecho (Aidlab, BioHarness, etc.), bandas de brazo (BodyMedia SenseWear), bandas de pierna o cintura (DataLog), anillos (por ejemplo, Oura Ring, Moodmetric…). De todos ellos, los más completos suelen ser los relojes y algunas pulseras específicas (como la Empatica E4), ya que un mismo dispositivo puede disponer de distintos sensores (más de 3 sensores de salud e incluir geolocalización y sensores ambientales como de presión o humedad). Sin embargo, las bandas de pecho, brazo, cintura y pierna son más limitadas, contando con un número reducido de sensores reales (entre 1 y 3), y si incluyen más sensores, no son reales, se derivan a partir de datos calculados de otros sensores, llamándose sensores virtuales. Por otro lado, encontramos distintas bandas de cabeza —en inglés, *headsets*— que miden señales mentales EEG (además de algunas físicas) e, incluso recientemente, se han comercializado unos los auriculares, Neurable[1], que son capaces de tal hazaña. Al igual que en el caso anterior, en las bandas de cabeza/auriculares variará el número de sensores disponibles, o en este caso el número de electrodos; y en general, penalizando el confort de su uso cuantos más electrodos ofrezca el dispositivo. Algunos ejemplos son Muse S2 (4 electrodos), EMOTIV EPOC X (con 14 electrodos) y los auriculares Neurable ya mencionados (con 16 electrodos, superando en ergonomía a las bandas). Gracias a los avances en micro y nanotecnología, los dispositivos *wearables* cada vez son más pequeños, poco o nada intrusivos e invasivos, menos costosos y cada vez es más la cantidad y variedad disponible de éstos en el mercado [5], [6]. De hecho, según las previsiones de aumento de la demanda y el incremento de concienciación de la población en la salud (para 2028, 118,16 billones de dólares estadounidenses en *wearables*), en el futuro cabe esperar que formen parte de nuestra vida cotidiana —como ya lo hacen las pulseras (*wristbands*) y relojes inteligentes (*smartwatches*)—[5]. En ese futuro no muy lejano, será muy probable que las personas llevemos puestos distintos *wearables* a la vez. En nuestra rutina diaria, podríamos usar el teléfono móvil junto al reloj (o pulsera) inteligente (como hacemos en la actualidad), tendríamos un anillo que pasaría desapercibido —pero, controlando nuestra fisiología—, del mismo modo que podríamos utilizar unos auriculares para escuchar música —mientras nos monitorizan nuestro estado mental y/o emocional—. Además, normalmente, se desarrollan soluciones que permiten interconectar estos dispositivos para poder interactuar entre sí, lo cual permite que se compartan e integren los datos que generan acerca de nuestro estado de salud.

En la actualidad, el paradigma por excelencia que contempla la interconexión de una vasta heterogeneidad de dispositivos de todo tipo (ordenadores, móviles, sensores de diversa índole, etc.) se conoce como «Internet de las cosas» (IoT, por sus siglas en inglés). En 1999, Kevin Ashton propuso por primera vez el concepto IoT como: los objetos interconectados (nodos) identificables de forma única con tecnología de identificación por radiofrecuencia (RFID) [7]. En general, IoT se define como [8]:

> «La presencia omnipresente a nuestro alrededor de una variedad de cosas u objetos —como etiquetas de identificación por radiofrecuencia (RFID), sensores, actuadores, teléfonos móviles, etc.— que, mediante esquemas de direccionamiento únicos, son capaces de interactuar entre sí y cooperar con sus vecinos para alcanzar objetivos comunes.»

A partir de aquí, aunque inicialmente IoT estaba orientado sólo a esas «cosas» con RFIDs, Atzori *et al.* [8] introduce la concepción del IoT como el resultado de la convergencia entre distintas visiones: las «cosas» (RFID, sensores, etc.), Internet (*Web of things*, Internet, servicios, etc.) y la semántica (ontologías,

[1] Headphones – Neurable: https://neurable.com/headphones

etc.). Dentro del campo de la salud, IoT ha demostrado su potencial y capacidad en varias aplicaciones como, por ejemplo, la monitorización de funciones vitales y enfermedades crónicas [9], entre otras [10]. Para diseñar un sistema de salud *m-Health* compuesto por una amplia variedad de sensores en la parte BAN y que ofrezca un servicio integral de salud ubicua en la parte BESys, podemos beneficiarnos de las ventajas técnicas que aporta el paradigma IoT. En particular, nos centraremos en la visión orientada a servicios [8], [11] gracias a las arquitecturas de servicios, las cuales permiten abstraerse de las tecnologías subyacentes y acceder al servicio en cuestión a través de uno de los nodos del *m-Health*.

Desde el punto de vista de la arquitectura software, uno de los enfoques que se están proponiendo para crear sistemas *m-Health* se basa en arquitecturas de microservicios [12], [13]. Se entienden por microservicios aquellas pequeñas partes de software (por ello el prefijo *micro*) responsables de una única tarea diferenciada. Éstos se pueden programar para que sean capaces de adaptarse y modificarse en tiempo de ejecución [14] a situaciones cambiantes que puedan ocurrir como, por ejemplo, elegir entre distintos modelos de evaluación o diagnóstico bajo demanda. Se trata de una evolución de las arquitecturas tradicionales orientadas a servicios (SOA) que, entre otras cuestiones, mejora a SOA en dotar a los microservicios de autonomía e independencia en su desarrollo, implementación y tiempo de ejecución. Además, un servicio en SOA se implementa de forma más generalista que un microservicio, por lo que presentará una granularidad mucho mayor en comparación con un microservicio —concebido para el desempeño de tareas pequeñas y específicas—. En la literatura, siempre se habla de microservicios *pequeños* [14], [15], pero no se especifica el tamaño de su granularidad, así que debemos tener presente este dilema y pensar en términos del desacoplamiento de funcionalidades del propio sistema, para que los microservicios sean suficientemente autónomos y puedan cooperar unos con otros. Por ello, distintos microservicios se pueden combinar sin necesidad de reescribir código cada vez que se quieran utilizar sus funcionalidades en distintos escenarios. Por ejemplo, un servicio en SOA encargado de ofrecer un diagnóstico sobre el estado de salud físico de un paciente puede que recoja los datos, los procese y genere un modelo de evaluación a partir de éstos. Sin embargo, siguiendo el enfoque basado en microservicios, podemos establecer una granularidad lo suficientemente fina como para desacoplar el servicio SOA, más generalista, en pequeñas tareas desempeñadas por microservicios: recogida de datos, distintas fases de preprocesamiento, diferentes tipos de evaluación y un microservicio que reporte el diagnóstico final. De esta forma, si en el futuro se desea ampliar el sistema *m-Health* para evaluar otra situación, se podrán reutilizar microservicios puntuales como los de recogida y reutilizar o añadir microservicios con distintas técnicas de procesamiento de datos, entre otros. Así, desde el punto de vista técnico, cuanto más pequeño es el microservicio, más fácil será su desarrollo (agilidad), iteración y despliegue. En un sistema IoT, donde su infraestructura es tan cambiante, esta granularidad en el diseño convierte a la arquitectura de microservicios en un claro candidato a tener en cuenta. Por contra, el sistema será mucho más complejo que en SOA, y se requerirá prestar atención a la coordinación y orquestación [12] entre los distintos microservicios. Otras ventajas técnicas [15] que los caracterizan son el desarrollo de su código autónomo (por ejemplo, un repositorio de código por cada microservicio [14]), despliegue independiente y descentralizado, soporte de múltiples lenguajes de programación y *frameworks*, flexibilidad, agilidad, extensibilidad, composición y orquestación [12] de sus elementos y fácil reemplazo. Así, la construcción de un sistema *m-Health* basado en microservicios, podrá beneficiarse de todas estas características. Y, en este sentido, podemos concebir la implementación de la BAN y del BESys en distintos microservicios encargados de tareas específicas y suficientemente granulares. Dado que en un sistema *m-Health* basado en microservicios necesitaremos trabajar en todo momento con los datos de los sensores de una BAN —provenientes de los dispositivos de las personas— para que el BESys los procese y analice, empiezan a surgir nuevos conceptos al respecto como el *framework Human Microservices* [16], [17], concibiendo a las personas como proveedores de datos.

En el ámbito del análisis de datos, el aprendizaje automático (en inglés, *machine learning* —ML) se ha convertido en una técnica importante para aprovechar el potencial de los datos en diversos campos [18]–[20]. Asimismo, en los últimos años se ha generalizado la adopción del ML en diversas aplicaciones sanitarias que van desde la predicción de paros cardíacos —a partir de señales cardíacas unidimensionales— hasta el diagnóstico asistido por ordenador (CADx) —mediante imágenes médicas multidimensionales— [21]. Normalmente, cuando trabajamos con aprendizaje automático seguimos un flujo de trabajo secuencial (*pipeline* o *workflow*), consistente en distintas etapas como: procesamiento de datos, extracción y selección de características, ajuste de *hiperparámetros*, generación y validación del modelo. Siguiendo la filosofía basada en microservicios, podremos diseñar este flujo de trabajo donde cada etapa es responsabilidad de un microservicio único, pudiendo reutilizarse en distintos modelos según la necesidad. Adicionalmente, aunque son muchos los proyectos de ML que se están desarrollando —no sólo en el ámbito sanitario, sino en todos los campos—, en la mayoría de ocasiones estos desarrollos quedan relegados a construir modelos de aprendizaje, compararlos con *benchmarks* y obtener el que mejor precisión reporta [22]. Por tanto, se deja a un lado el diseño de sistemas complejos de ML que estén listos para la producción y despliegue, y que se proporcione la coordinación necesaria de sus componentes e infraestructura [23]. De hecho, un nuevo paradigma empieza a emerger para arrojar luz a esta cuestión: Machine Learning Operations (MLOps) [24]. Sin embargo, no se conocen esfuerzos suficientes, que persigan una estrategia de diseño arquitectónico de aplicación en salud, la cual contribuya a que el proceso de construcción y puesta en marcha de estos modelos se conciba para desplegarlos y llevarlos a producción. O lo que es lo mismo, se precisa de una solución tecnológica integral de *m-Health* flexible a los cambios y que considere el despliegue de un producto final o servicio de ML para evaluar el estado de salud de una persona, desde las dimensiones física y mental.

En esta tesis se propone un sistema integral *m-IoTHealth* —*m-Health* bajo el paradigma IoT—, basado en una arquitectura de microservicios. Dicha arquitectura concibe la recolección de datos de una BAN (provenientes de diferentes dispositivos *wearables*) y la transmisión de éstos, via *gateway* al sistema *back-end* (BESys) para procesarlos, analizarlos y evaluar las dimensiones de salud física y mental. Como resultados, se han construido modelos de ML con una alta precisión, tanto para evaluar la dimensión física como la mental. También se ha tenido en cuenta la realización de un diseño MLOps, separando en *microservicios* tanto cada una de las tareas involucradas en un proyecto ML (*pipeline* o *workflow*), como la generación y la adquisición de los datos de cada sensor y los distintos modelos generados. La solución *m-IoTHealth* se ha diseñado bajo una arquitectura de microservicios, para tener en cuenta los posibles cambios o exigencias que puedan darse. Por ejemplo, para evaluar un aspecto específico de salud, se pueden seleccionar distintos sensores con sus microservicios de recogida de datos asociados y distintos modelos de ML, que se han generado según el dominio de aplicación. Además, se pueden reutilizar microservicios del *pipeline* de ML independientemente de los modelos que se usen.

Las implicaciones de nuestra propuesta se traducen en un beneficio para las personas y gobiernos, ahorrándoles distintos recursos, de tiempo y dinero, y visitas al hospital innecesarias. Esto se debe gracias a la ubicuidad del sistema *m-IoTHealth*, el diseño arquitectónico propuesto basado en microservicios —tanto de la BAN como del *pipeline* de ML en el BESys— y los modelos de ML generados —que evalúan distintas dimensiones de salud—. Además, en el futuro inmediato —donde se prevé un aumento demográfico de las personas mayores— puede contribuir a promover un envejecimiento activo y saludable, monitorizando a los mayores en su día a día desde cualquier parte, de forma preventiva, y ayudando a los profesionales de salud a detectar problemas cuanto antes para tomar medidas de intervención inmediatas. Motivados por esta problemática demográfica, nuestro propósito es investigar en la aplicación de sistemas de *m-Health* para personas mayores, con objeto de promover el envejecimiento activo y saludable, y en concreto identificar problemas de fragilidad y dependencia. A continuación, se van a describir estos términos y la problemática asociada a nuestro caso de uso.

En los últimos años, se han creado varias líneas de actuación centradas en promover el envejecimiento activo y saludable [25], lo cual implica optimizar el bienestar físico, social, mental e, incluso, nutricional [26] durante toda la vida. Envejecimiento activo y saludable significa ayudar a las personas, a medida que envejecen, a mantenerse con la autonomía de sus propias vidas durante el mayor tiempo posible y, cuando sea asumible, contribuir a la economía y la sociedad [27]. Trabajar en esta línea es importante porque ayudará a mejorar la calidad de vida de las personas —desde el bienestar de los recién jubilados [28] hasta los más mayores— y, en definitiva, su salud, lo que repercutirá en un impacto positivo en la estructura socioeconómica de los estados. Este creciente interés en promover un envejecimiento activo y saludable se debe a que, según la Organización Mundial de la Salud (OMS), se prevé que para 2030 una de cada seis personas en el mundo tendrá más de 60 años, lo que supondrá casi duplicar el porcentaje de mayores en 2050 —pasando de un 12% al 22%— y el 80% vivirá en países de ingresos bajos y medios [29]. En los próximos años, este envejecimiento creciente de la población supondrá un impacto socioeconómico para todas las sociedades y sistemas nacionales sanitarios [30], [31]. De hecho, hoy en día, tenemos evidencias de que el envejecimiento masivo se está produciendo también entre algunos de los países más pobres, lo cual contribuirá en una doble carga para su estructura nacional de salud y social [32], [33]. Desde un punto de vista tecnológico, la tecnología *m-Health* ha demostrado ser eficaz en la prevención de enfermedades y en los cambios de estilo de vida, resultando ser una herramienta adecuada para las personas mayores [33]. Gracias a la ubicuidad de un sistema *m-Health*, los mayores y los cuidadores se mantienen informados en todo momento del estado de salud y problemas que puedan surgir, por lo que se pueden tomar decisiones inmediatas. Asimismo, la persona mayor conoce y es responsable de su propia salud, viendo los resultados de sus hábitos saludables [34] y empoderándose [9], [35] para mejorar su autonomía, apoyo social, etc. Entre las acciones que se están tomando en la promoción de un envejecimiento activo y saludable, cabe destacar poner el foco en la prevención de la fragilidad y el mantenimiento de la independencia de las personas mayores, para lo cual deben tenerse en cuenta el estilo de vida, incluyendo la actividad física y la dieta [25]. El objetivo de la prevención[2] es evitar o retrasar la aparición de fragilidad o dependencia, cuando exista posibilidad de reversión, mejorar su pronóstico impidiendo su empeoramiento.

El síndrome de la **fragilidad** se define como el deterioro en una o más dimensiones del funcionamiento humano (físico, mental, social, nutricional, etc.) provocando causas adversas [36]. Afecta aproximadamente al 11% de las personas mayores de 65 años que viven en la comunidad, y entre el 30% y el 70% de los pacientes quirúrgicos de edad avanzada, y se asocia con otros problemas como la autonomía o dependencia, la pérdida del equilibrio y las caídas, comorbilidades, la hospitalización, e

[2] Ley 9/2016, de 27 de diciembre, de Servicios Sociales de Andalucía.

incluso puede llegar a causar la muerte [37]. Su diagnóstico viene determinado por la evaluación de algunas pruebas físicas (velocidad de la marcha, fuerza de agarre de la mano) y la administración de un cuestionario validado en el que se le hacen una serie de preguntas al encuestado y, según el cuestionario utilizado, se evalúan una o varias dimensiones humanas [36], [37]. Lo mismo sucede con la evaluación de la **dependencia** en las personas mayores —entendida como la ausencia de autonomía para poder desempeñar actividades de la vida diaria (AVD) [38] (estrechamente relacionada con la fragilidad)—, cuya evaluación se realiza con sus cuestionarios correspondientes y varias visitas de profesionales cualificados a los solicitantes [39], [40]. La contrapartida de estos procesos de evaluación, de fragilidad o dependencia, es el arduo retraso que arrastran las comunidades autónomas en el caso de España. Por ejemplo, en el caso de la evaluación de la dependencia, la espera de la resolución del proceso evaluativo por completo en promedio es de 430 días, superándola Andalucía en 700 días (casi 2 años) [41]. Esto se debe a que los instrumentos de evaluación empleados son cuestionarios y pruebas físicas que tienen asociados un consumo de tiempo en su realización y una supervisión de profesionales sanitarios cualificados (como enfermeros, terapeutas ocupacionales, etc.), suponiendo para los sistemas sanitarios altos costes y tiempos prolongados de atención a los ciudadanos que esperan a ser evaluados. Además, los instrumentos de evaluación son cualitativos y muchos de ellos de carácter subjetivo, ya que la mayoría de los ítems que evalúan son preguntas que responden los sujetos evaluados. Exceptuando algunas pruebas físicas medibles de forma objetiva en tiempo de ejecución o con instrumentos de fuerza. Por tanto, su fiabilidad dependerá de la disposición del encuestado a responder la verdad y, quizás, resulten conclusiones erróneas debido a que no se hayan considerado otras preguntas que podrían ser relevantes. Además de todo esto, cabe destacar que no suelen estar programadas las consultas o visitas a los domicilios de los mayores —lo que favorecería la prevención—; sino que la iniciativa de la evaluación parte de la propia persona mayor, o de sus familiares (generalmente, con el rol de cuidadores informales [42]) cuando detectan algún problema, lo que podría derivar en un diagnóstico tardío, imposibilitando incluso una intervención adecuada. Para tratar de subsanar esto se tendrían que programar visitas periódicas (anuales o bienales) en personas con edad cercana a los 60 años. Sin embargo, atendiendo a los costes que supondría y al retraso administrativo que ya supone sólo evaluar la dependencia, se considera inasumible esta iniciativa hoy en día. Por tanto, teniendo en cuenta las dificultades que presentan las actuales evaluaciones de fragilidad y dependencia, surge la necesidad de considerar nuevas estrategias y tecnologías para que los servicios sanitarios puedan ofrecerse a precios asequibles, mejoren la calidad de su servicio y, en definitiva, la calidad de vida de las personas y nuestros mayores. Con tecnologías móviles como los sistemas *m-Health* se considera a la persona como proveedor de datos de salud [43], para monitorizar —de forma automática— su actividad diaria y, así, reducir la necesidad de aumentar los recursos humanos que se necesitarían para una iniciativa realista de prevención bajo la premisa tradicional. En cualquier caso, sería deseable involucrar a los usuarios y demás interesados (familiares, cuidadores, clínicos, etc.) en el diseño del sistema [44]. En este sentido, explorar una evaluación más objetiva y holística de la fragilidad o dependencia sería una que se llevara a cabo durante el desempeño de las actividades diarias que realizan las personas en su día a día, lo que se conoce como una evaluación ecológica o ubicua.

Tanto la fragilidad como la dependencia afectan directamente al desempeño de las actividades instrumentales de la vida diaria (AIVD) de las personas [45]. Se entienden como AIVD aquellas que suponen una complejidad multidimensional, es decir, física, mental, social, etc., que normalmente suelen implicar la interacción con otros objetos o herramientas [46]. Ejemplos de estas pueden ser la limpieza de la casa, cocinar, ir a comprar, etc. Así, la monitorización y evaluación del desempeño de las AIVD pueden contribuir a identificar la fragilidad y la dependencia [46], puesto que las personas mayores son vulnerables a la complejidad intrínseca de las AIVD. En este sentido, explorar una evaluación más objetiva y holística sería una que se llevara a cabo durante el desempeño de las AIVD, es decir, una evaluación ecológica (transparente para el usuario o ubicua). Para ello, el uso de los dispositivos *wearables* más confortables disponibles en el mercado (relojes, pulseras, etc.) pueden ayudar en esta monitorización no intrusiva del desempeño de las AIVD.

Con el fin de proporcionar una solución tecnológica que recopile datos para evaluar la fragilidad y la dependencia, se han investigado una variedad de dispositivos *wearables* que se pueden llevar puestos en el día a día sin resultar excesivamente intrusivos. Esto significa que ofrezcan la cualidad de ser ecológicos o ubicuos (que los puedan llevar y no molesten a las personas en su rutina diaria), nada invasivos y recopilen señales —como son las pulseras o relojes inteligentes—. Además de esto, teniendo en cuenta cómo afecta la fragilidad y la dependencia a diferentes dimensiones de salud en los mayores, los sensores a considerar deben poder medir señales tanto de dimensión física como mental. Así que, como eje central de esta tesis para medir la dimensión física, y después de considerar las distintas alternativas del mercado, se eligieron tanto la pulsera Empatica E4 como el reloj inteligente, Samsung Gear. La pulsera cuenta con algunas certificaciones [47], [48] que avalan su uso en la clínica y dispone de los sensores: acelerómetro, pulsómetro por fotopletismografía (PPG) —mide el pulso de volumen sanguíneo (BVP), del cual se puede derivar la variabilidad de la frecuencia cardíaca—, temperatura y actividad electrodérmica (EDA). Sin embargo, pensando que sensores como el giroscopio, contador de pasos o GPS también podían ayudar a monitorizar las AIVD, se consideraron otros dispositivos como el reloj inteligente Samsung Gear que cuenta con: acelerómetro, giroscopio, ritmo cardíaco, GPS y contador de pasos. Por otra parte, y en un segundo plano de la tesis, como un primer acercamiento hacia el tratamiento de señales EEG y mejorar las mediciones de la dimensión mental, se complementó a los a los anteriores *wearables* con una diadema de cabeza EEG, Muse 2, que cuenta con 4 electrodos y sensores de acelerómetro y giroscopio. Así que se obtuvieron distintos tipos de ondas cerebrales, según sus frecuencias, para explorar la detección de la actividad mental en intención motora (en inglés, *motor imagery*). Para extraer información valiosa de los datos en bruto que podemos obtener de los dispositivos con sensores, la aplicación de técnicas de análisis de datos —como el aprendizaje automático—, nos permitirán evaluar la fragilidad y la dependencia y adentrarnos a explorar la actividad mental.

En la literatura, existen algunos trabajos que intentan sobrepasar las limitaciones de la evaluación actual de fragilidad o dependencia aplicando técnicas de aprendizaje automático (ML) y, así, poder automatizar la evaluación. Tanto en la detección de la fragilidad [20], [49]–[51] como en la evaluación de la dependencia [52], [53] diferentes algoritmos de ML supervisado (k-Nearest Neighbour (kNN), Naïve Bayes (NB), Neural Networks (NN), Decision Tree (DT), Random Forests (RF)) se han estado aplicando usando los datos de dispositivos *wearables*. En el caso de la fragilidad, podemos destacar algunos estudios [54], [55] donde se le da importancia a que dichos dispositivos sean ubicuos y no intrusivos, imprescindible para no molestar el ecosistema social de los mayores [56]. Además, instan a que no solo se creen modelos automáticos de evaluación, sino que los futuros trabajos se encaminen hacia soluciones integrales —en la línea de nuestra propuesta—. Sin embargo, no hemos encontrado soluciones tecnológicas integrales de *m-Health* —flexibles a los cambios y que consideren el despliegue de un producto final de ML— para evaluar la fragilidad o la dependencia desde un punto de vista de la prevención atendiendo a su carácter holístico, ecológico, dinámico y multidimensional.

Como primer paso hacia el estudio de las señales EEG (dimensión mental), las cuales pueden revelar patrones de comportamiento de un usuario o anomalías e, incluso, medir actividades de pensamiento. En este sentido, se estudió la detección de la imaginación motora o el pensamiento orientado a objetivos, más conocido como *motor imagery*. Aunque este campo de estudio no es nuevo, sí lo es la utilización de dispositivos *wearables* que midan la EEG. Tradicionalmente, se han estado utilizando los cascos EEG para estudiar la intención de movimiento, mediante la aplicación de técnicas de ML [57]–[66]. Sin embargo, estos dispositivos son muy intrusivos y poco ergonómicos para su uso en la vida cotidiana, ya que están equipados con muchos electrodos y cables que se conectan a ordenadores o equipos específicamente diseñados para recoger estas señales [67]. Además, es habitual la aplicación de líquidos o geles a los electrodos para que faciliten la conductividad eléctrica de las señales a través del cuero cabelludo. Aunque ya existen estudios de EEG con *headsets* (*wearables* menos intrusivos que los cascos) el tiempo de respuesta de la detección suele ser elevado, suponiendo un retardo de entre unos 10 y 30 segundos [68]–[70]; y en nuestro caso, se consiguió reducir a una ventana de 2 segundos de señales EEG. Con todo ello, es interesante explorar el uso de dichos dispositivos nuevos en la detección de *motor imagery* y su combinación con los datos fisiológicos de otros *wearables* e intentar acercarnos a una evaluación de la dimensión de salud mental más ubicua que la tradicional.

Para terminar, este trabajo se estructura en dos partes diferenciadas. En la primera parte, en el capítulo 2, se presentan los fundamentos que soportan esta tesis, en el capítulo 3 los objetivos, en el capítulo 4 la metodología de este trabajo. Un resumen de los principales resultados de las publicaciones se describe en el capítulo 5. Las conclusiones extraídas de la tesis y las posibles líneas futuras de trabajo se presentan en los capítulo 6 y 7, respectivamente. En la segunda parte de este documento, se presentan las 3 publicaciones que forman parte de esta tesis:

- A Microservices e-Health System for Ecological Frailty Assessment Using Wearables.
- A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors.
- Reducing Response Time in Motor Imagery Using A Headband and Deep Learning.

1.1. Introduction

A mobile-health (m-Health) system consists of using mobile or wireless devices, equipped with different sensors, to collect patient data (physical, mental, etc.), store and send them to servers connected to the Internet for further processing, analysis and knowledge extraction [1], [2]. An m-Health is able to monitor patients remotely in a ubiquitous way and even, in many occasions, replace costly traditional procedures [2]. Generally, the architecture of an m-Health [3] (Figura 1) is composed of two distinct parts: the body area network (BAN) and the back-end system (BESys). To perform patient monitoring, the sensors are organized in the BAN communication network, which, in turn, sends the collected data to other nearby devices—mobile base units (MBUs)—that act as gateways. In this way, the gateway is responsible for connecting the BAN with an external network, where the back-end system, BESys, is located, waiting to receive the sensory data for further processing and analysis. The BESys consists of one or more back-end servers and complementary applications to process the biosignals and other data received by the servers. All this data could be used in healthcare systems to monitor different dimensions of human health, such as the physical dimension and the cognitive or mental dimension, among others. The adoption of m-Health systems and the use of wireless devices in healthcare systems will contribute to a reduction in social and healthcare costs [4].

The BAN can include wireless devices such as cell phones or wearables. These devices can offer different types of sensors capable of measuring signals of various kinds. Although the most widespread ones capture physical signals—from sensors such as the accelerometer, gyroscope, skin perspiration, heart rate, blood oxygen, etc.—, there are also wearables capable of measuring mental signals - electroencephalography (EEG). On the one hand (excluding the widely known smartphone), some examples of wearable devices that measure physical dimension signals are: smartwatches (Samsung Gear, Apple Watch, Fitbit Versa, Mobvoi TicWatch, etc.), wristbands (Empatica E4, Xiaomi Smart Band, etc.), chest bands (Aidlab, BioHarness, etc.), arm bands (BodyMedia SenseWear), leg or waist bands (DataLog), rings (e.g., Oura Ring, Moodmetric, etc.). Of all of them, the most complete are usually watches and some specific wristbands (such as the Empatica E4), since the same device can have different sensors (more than 3 health sensors, geolocation and environmental sensors such as pressure or humidity). However, chest, arm, waist and leg bands are more limited, having a reduced number of real sensors (between 1 and 3), and if they include more sensors, they are not real, they are derived from data calculated from other sensors, being called virtual sensors. On the other hand, we find different headbands that measure EEG mental signals (in addition to some physical ones) and, even recently, a headphone, Neurable[3], has been marketed that is capable of such a feat. As in the previous case, headbands/headphones will vary in the number of sensors available, or in this case the number of electrodes; and in general, the more electrodes the device offers, the more the device penalizes the comfort of use. Some examples are Muse S2 (4 electrodes), EMOTIV EPOC X (with 14 electrodes) and the aforementioned Neurable headphones (with 16 electrodes, surpassing bands in ergonomics). Thanks to advances in micro- and nanotechnology, wearable devices are becoming smaller, less or non-intrusive and invasive, less expensive, and the number

[3] Headphones – Neurable: https://neurable.com/headphones.

and variety of these devices available on the market is increasing [5], [6]. In fact, according to forecasts of increasing demand and increasing health awareness of the population (in 2028, USD 118.16 billion in wearables), in the future they can be expected to be part of our daily lives—as wristbands and smartwatches already are—[5]. In the not-too-distant future, it will be very likely that people will be wearing different wearables at the same time. In our activities of daily living, we could use the smartphone together with the smartwatch or wristband (as we do today), we would have a smart ring that would go unnoticed—but, monitoring our physiology—, just as we could use headphones to listen to music—while monitoring our mental and/or emotional state—. In addition, solutions are usually developed to allow these devices to be interconnected, they can interact with each other, allowing the data they generate about our state of health to be shared and integrated.

Today, Internet of Things (IoT) is the paradigm which contemplates the interconnection of a vast heterogeneity of devices of all types (computers, cell phones, sensors of various kinds, etc.). In 1999, Kevin Ashton first proposed the IoT concept as: interconnected objects (nodes) uniquely identifiable with radio frequency identification (RFID) technology [7]. In general, IoT can be defined as [8]:

> "The pervasive presence around us of a variety of things or objects—such as Radio-Frequency IDentification (RFID)tags, sensors, actuators, mobile phones, etc.—which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals."

From here, although initially IoT was oriented only to those "things" with RFIDs, Atzori *et al.* [8] introduces the conception of IoT as the result of the convergence between different visions: "things" (RFID, sensors, etc.), Internet (Web of things, Internet, services, etc.) and semantics (ontologies, etc.). Within the healthcare field, IoT has demonstrated its potential and capability in several applications such as, monitoring of vital functions and chronic diseases [9], among others [10]. To design an m-Health system composed of a wide variety of sensors in the BAN and offering a comprehensive ubiquitous health service in the BESys, we can benefit from the technical advantages brought by the IoT paradigm. In particular, we will focus on the service-oriented view [8], [11] thanks to service architectures, which allow to abstract from the underlying technologies and to access the service in question through one of the m-Health nodes.

From the point of view of software architecture, one of the approaches being proposed to create m-Health systems is based on microservice architectures [12], [13]. Microservices are understood as those small pieces of software (hence the prefix micro) responsible for a single task. They can be programmed to be able to adapt and modify themselves at runtime [14] to changing situations that may occur, such as choosing between different evaluation or diagnostic models on demand. This is an evolution of traditional service-oriented architectures (SOA) that, among other issues, improves SOA in providing microservices with autonomy and independence in their development, implementation and runtime. In addition, a service in SOA is implemented more generalist than a microservice, so it will present a much greater granularity compared to a microservice—conceived for the performance of small and specific tasks—. In the literature, we always talk about small microservices [14], [15], but the size of their granularity is not specified, so we should keep this dilemma in mind and think in terms of decoupling functionalities from the system itself, so that microservices are sufficiently autonomous and can cooperate with each other. Therefore, different microservices can be combined without the need to rewrite code every time you want to use their functionalities in different scenarios. For example, an SOA service in charge of providing a diagnosis of a patient's physical health status may collect data, process them and generate an assessment model from them. However, following the microservices-based approach, we can establish a granularity fine enough to decouple the more generalist SOA service into small tasks performed by microservices: data collection, different preprocessing phases, different types of assessment, and a microservice that reports the final diagnosis. In this way, if in the future the m-Health system is to be extended to evaluate another disorders, it will be possible to reuse specific microservices such as those for data collection and reuse or to add microservices with different data processing techniques, among others. Thus, from a technical point of view, the smaller the microservice, the easier its development (agility), iteration and deployment will be. In an IoT system, where its infrastructure is so changeable, this granularity in design makes microservices architecture a clear candidate to consider. On the downside, the system will be much more complex than in SOA, and attention to coordination and orchestration [12] between the different microservices will be required. Other technical advantages [15] of microservices are the development of their autonomous code (e.g., one code repository per microservice [14]), independent and decentralized deployment, support of multiple programming languages and frameworks, flexibility, agility, extensibility, composition and orchestration [12] of their elements, and easy replacement. Thus, the construction of an m-Health system based on microservices will be able to benefit from all these features. And, in this sense, we can conceive the implementation of BAN and BESys in different microservices in charge of specific and fine-grained tasks. Since m-Health systems based on microservices we will always need to work with data from the sensors of a BAN—coming from people's devices—to be processed and analyzed by BESys, new concepts are beginning to emerge in this regard, such as the Human Microservices framework [16], [17], conceiving people as data providers.

In the field of data analytics, machine learning (ML) has become an important technique for harnessing the potential of data in various fields [18]–[20]. Also, in recent years, ML has been widely adopted in various healthcare applications ranging from cardiac arrest prediction—from one-dimensional cardiac signals—to computer-aided diagnosis (CADx)—using multidimensional medical images—[21]. Normally, when working with machine learning we follow a sequential workflow (pipeline), consisting of different stages such as: data processing, feature extraction and selection, hyper-parameter tuning, model generation and validation. Following the philosophy based on microservices, we can design this workflow where each stage is the responsibility of a single microservice and can be reused in different models as needed. Additionally, although many ML projects are being developed -not only in the healthcare field, but in all fields-, in most cases these developments are relegated to building learning models, comparing them with benchmarks and obtaining the one that reports the best accuracy [22]. Thus, the design of complex ML systems that are ready for production and deployment, and that provide the necessary coordination of their components and infrastructure, is left aside [23]. In fact, a new paradigm is beginning to emerge to shed light on this issue: Machine Learning Operations (MLOps) [24]. However, not enough efforts are known, pursuing an architectural design strategy for healthcare application, which contributes to the process of building and commissioning these models is conceived to deploy them and bring them to production. In other words, what is needed is an integral technological solution for m-Health that is flexible to changes and considers the deployment of a final product or ML service to evaluate the state of health of a person, from the physical and mental dimensions.

This thesis proposes an integral m-IoTHealth system—e.g., an m-Health system under the IoT paradigm—, based on a microservices architecture. This architecture considers the collection of data from a BAN (coming from different wearable devices) and the transmission of these data via gateway to the back-end system (BESys) to process, analyze and evaluate the dimensions of physical and mental health. As a result, ML models have been built with high accuracy, both for the physical and mental dimensions. We have also considered the MLOps design, splitting each of the MLOps tasks involved in a ML pipeline into microservices, as well as the generation and acquisition of data from each sensor and the different models generated. The m-IoTHealth solution has been designed under a microservices architecture, to consider possible changes or requirements that may arise. For example, to evaluate a specific health aspect, different sensors can be selected with their associated data collection microservices and different ML models, which have been generated according to the application domain. Furthermore, microservices from the ML pipeline can be reused regardless of the models used.

The implications of our proposal translate into a benefit for individuals and governments, saving them various resources, time and money, and unnecessary hospital visits. This is due to the ubiquity of the m-IoTHealth system, the proposed microservices-based architectural design of both the BAN and the ML pipeline in BESys, and the generated ML models that evaluate different dimensions of health. Moreover, in the immediate future—where a demographic increase of elderly people is expected—it can contribute to promote an active and healthy aging, monitoring the elderly in their day-to-day life from anywhere, in a preventive way, and helping health professionals to detect problems as soon as possible to take immediate intervention measures. Motivated by this demographic problem, our purpose is to investigate the application of m-Health systems for the elderly, to promote active and healthy aging, and specifically to identify problems of frailty and dependence. In the following, we will describe these terms and the problems associated with our use case.

In recent years, several lines of action have focused on promoting active and healthy aging [25], which involves optimizing physical, social, mental, and even nutritional well-being [26] throughout life. Active and healthy aging means helping people, as they age, to remain autonomous in their own lives for as long as possible and, when feasible, to contribute to the economy and society [27]. Working along these lines is important because it will help to improve people's quality of life-from the well-being of the newly retired [28] to the oldest old-and ultimately their health, which will have a positive impact on the socioeconomic structure of states. According to the World Health Organization (WHO), this growing interest in promoting active and healthy aging is due to it is predicted that by 2030 one in six people in the world will be over 60 years of age, which will almost double the percentage of older people in 2050-from 12% to 22%-and 80% will live in low- and middle-income countries [29]. In the coming years, this growing aging of the population will have a socio-economic impact on all societies and national healthcare systems [30], [31]. In fact, today, we have evidence that massive aging is also occurring among some of the poorest countries, which will contribute to a double burden on their national health and social structure [32], [33]. From a technological point of view, m-Health technology has proven to be effective in disease prevention and lifestyle changes, proving to be a suitable tool for the elderly [33]. Thanks to the ubiquity of an m-Health system, the elderly and caregivers are kept informed of health status and problems that may arise anytime, so that immediate decisions can be made. Likewise, the older adult knows and is responsible for their own health, seeing the results of their healthy habits [34] and becoming empowered [9], [35] to improve their autonomy, social support, etc. In the promotion of active and healthy aging, we emphasize the prevention of frailty and the maintenance of independence of the elderly, considering lifestyle, including physical activity and diet [25]. The aim of prevention[4] is to avoid or delay the onset of frailty or dependence, when there is a possibility of reversal, to improve its prognosis by preventing its worsening.

Frailty syndrome is defined as a deterioration in one or more dimensions of human functioning (physical, mental, social, nutritional, etc.) leading to adverse causes [36]. It affects approximately 11% of people over 65 years of age living in the community, and between 30% and 70% of elderly surgical patients, and is associated with other problems such as loss of autonomy or dependence, loss of balance and falls, comorbidities, hospitalization, and may even lead to death [37]. Its diagnosis is determined by the evaluation of some physical tests (gait speed, hand grip strength) and the administration of a validated questionnaire in which the respondent is asked a series of questions and, depending on the questionnaire used, one or several human dimensions are evaluated [36], [37]. The same applies to the assessment of **dependency** in the elderly—understood as the absence of autonomy to be able to perform activities of daily living (ADLs) [38] (closely related to frailty)—whose assessment is carried out with their corresponding questionnaires and several visits by qualified professionals to the applicants [39], [40]. The downside of these assessment processes, of frailty or dependency, is the arduous delay that the autonomous communities in the case of Spain are dragging behind. For example, in the case of the evaluation of dependency, the average wait for the resolution of the entire evaluation process is 430 days, with Andalusia exceeding it by 700 days (almost 2 years) [41]. This is due to the assessment instruments used are questionnaires and physical tests that are associated with time-consuming completion and supervision by qualified health professionals (such as nurses, occupational therapists, etc.), which means high costs for the health systems and long waiting times for citizens waiting to be assessed. In addition, the evaluation instruments are qualitative and many of them are subjective in nature, since most of the items they evaluate are questions answered by the subjects being evaluated. Some exceptions are the physical tests, which can be measured objectively in execution time or with strength instruments. Therefore, their reliability will depend on the willingness of the respondent to answer truthfully and, perhaps, erroneous conclusions may result because other questions that could be relevant have not been considered. Furthermore, it should be noted that consultations or home visits are not usually scheduled—which would favor prevention—. Rather, the initiative for the evaluation comes from the elderly person themselves, or from their relatives (generally in the role of informal caregivers [42]) when they detect a problem, which could lead to a late diagnosis and making it impossible to intervene appropriately. To try to remedy this, periodic visits (annual or biennial) would have to be scheduled for people close to 60 years of age. However, in view of the costs involved and the administrative delay involved in just assessing dependency, this initiative is considered unfeasible today. Therefore, considering the difficulties presented by current frailty and dependency assessments, the need arises to focus on new strategies and technologies so that health services can be offered at affordable prices, to improve the quality of their service and, ultimately, the quality of life of people and our elders. With mobile technologies such as m-Health systems, the person is considered as the provider of health data [43], to monitor—automatically—their daily activity and, thus, reduce the need to increase the human resources that would be required for a realistic prevention initiative under the traditional premise. In any case, it would be desirable to involve

[4] Ley 9/2016, de 27 de diciembre, de Servicios Sociales de Andalucía.

users and other stakeholders (family members, caregivers, clinicians, etc.) in the design of the system [44]. In this sense, exploring a more objective and holistic assessment of frailty or dependency would be one that is conducted during the performance of the daily activities that people perform in their day-to-day lives, known as an ecological or ubiquitous assessment.

Both frailty and dependence directly affect people's performance of instrumental activities of daily living (IADLs) [45]. IADLs are understood as those that involve multidimensional complexity, i.e., physical, mental, social, etc., usually involving interaction with other objects or tools [46]. Examples of these may include house cleaning, cooking, shopping, etc. Thus, monitoring and assessment of IADL performance can contribute to identify frailty and dependency [46], as older people are vulnerable to the intrinsic complexity of IADLs. In this sense, a more objective and holistic assessment would be one that is conducted during the performance of IADLs, i.e., an ecological (user-transparent or ubiquitous) assessment. To this end, the use of the most comfortable wearable devices available on the market (watches, bracelets, etc.) can assist in this non-intrusive monitoring of IADL performance.

A variety of wearable devices have been analyzed, which can be worn daily without being overly intrusive, to provide a technological solution that collects data to assess frailty and dependency. This means that wearables offer the quality of being eco-friendly or ubiquitous (wearable and do not disturb people in their daily routine), non-invasive and collect signals—such as wristbands or smartwatches—. Furthermore, considering how frailty and dependence affect different dimensions of health in the elderly, the sensors to be considered should be able to measure signals of both physical and mental dimensions. So, on the one hand, to measure the physical dimension and after considering the different alternatives on the market, both the Empatica E4 bracelet and the smartwatch Samsung Gear, were chosen. The bracelet has some certifications [47], [48] that validate its use in the clinic setting and has the sensors: accelerometer, photoplethysmography pulsometer (PPG)—measures blood volume pulse (BVP), from which heart rate variability can be derived—, temperature and electrodermal activity (EDA). However, thinking that sensors such as gyroscope, step counter or GPS could also help to monitor IADV, other devices were considered, such as the Samsung Gear smart watch that has accelerometer, gyroscope, heart rate, GPS and step counter. On the other hand, as a first approach towards the treatment of EEG signals and improve the measurements of the mental dimension, the previous wearables were complemented with an EEG headband, Muse 2, which has 4 electrodes and accelerometer and gyroscope sensors. Thus, different types of brain waves were obtained, according to their frequencies, to explore the detection of mental activity in motor imagery. To extract valuable information from the raw data we can obtain from sensor devices, the application of data analysis techniques—such as machine learning—will allow us to assess frailty and dependence and to explore mental activity.

In the literature, there are some works that try to overcome the limitations of the current frailty or dependency assessment by applying machine learning (ML) techniques and, thus, to automate the assessment. Both in frailty detection [20], [49]–[51] and in dependency assessment [52], [53] different supervised ML algorithms (k-Nearest Neighbour (kNN), Naïve Bayes (NB), Neural Networks (NN), Decision Tree (DT), Random Forests (RF)) have been applied using data from wearable devices. In the case of frailty, we can highlight some studies [54], [55] where importance is given to these devices being ubiquitous and non-intrusive, essential not to disturb the daily environment of the elderly people [56]. Furthermore, authors of these studies urge that not only automatic assessment models should be created, but future work should move towards integral solutions along the lines of our proposal. However, we have not found comprehensive m-Health technological solutions—flexible to changes and considering the deployment of a final ML product—to assess frailty or dependence from a prevention point of view attending to its holistic, ecological, dynamic and multidimensional nature.

As a first step towards the study of EEG signals (mental dimension), which can reveal a user's behavioral patterns or anomalies and even measuring thinking activities. In this regard, the detection of motor imagery or goal-oriented thinking, better known as motor imagery, was studied. Although this field of study is not new, the use of wearable devices that measure EEG is. Traditionally, EEG helmets have been used to study movement intention by applying ML techniques [57]–[66]. However, these devices are very intrusive and unergonomic for use in everyday life, as they are equipped with many electrodes and wires that are connected to computers or equipment specifically designed to collect these signals [67]. In addition, it is common to apply liquids or gels to the electrodes to facilitate the electrical conductivity of the signals through the scalp. Although there are already EEG studies with headsets (wearables less intrusive than helmets) the detection response time is usually high, assuming a delay of between about 10 and 30 seconds [68]–[70]; and in our case, we managed to reduce to a 2-second window of EEG signals. All in all, it is interesting to explore the use of such new devices in motor imagery detection and their combination with physiological data from other wearables and try to approach a more ubiquitous than traditional assessment of the mental health dimension.

Finally, this work is structured in two distinct parts. In the first part, in Chapter 2, the fundamentals that support this thesis are presented, in chapter 3 the objectives, in Chapter 4 the methodology of this work. A summary of the main results of the publications is described in Chapter 5. The conclusions drawn from the thesis and possible future lines of work are presented in Chapters 6 and 7, respectively. In the second part of this document, the 3 publications that are part of this thesis are presented:

- A Microservices e-Health System for Ecological Frailty Assessment Using Wearables.
- A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors.
- Reducing Response Time in Motor Imagery Using A Headband and Deep Learning.

2. Fundamentos

Esta sección presenta los conceptos principales en los que se sustenta la presente tesis doctoral. En primer lugar, la sección 2.1 introduce los sistemas *m-Health*, la tecnología que subyace a su implementación, funcionamiento, recopilación de datos de sensores y aprovisionamiento del servicio de salud. En la sección 2.3, se describen los fundamentos de las técnicas de aprendizaje automático que servirán para generar modelos de detección y evaluación del estado de salud. En la sección 2.2, abordaremos los conceptos fundamentales de la arquitectura de microservicios. Seguidamente, para entender el impacto de estos sistemas en los casos de uso aplicados en este trabajo, se presentarán los instrumentos de evaluación tradicional tanto de fragilidad como de dependencia en la sección 2.4. Y, por último, en la sección 2.5, se introduce el concepto *motor imagery* como un primer paso al estudio de las señales EEG.

2.1. Sistema m-Health

El concepto *m-Health* tiene su origen en un término más amplio como es «electronic-Health» (e-Health). La *e-Health* indica la utilización de las TIC para proveer un servicio de salud [2]. El desarrollo de la *e-Health* ha ido moviéndose hacia sistemas inalámbricos, móviles y *wearables*, lo cual nos brinda la oportunidad para apoyar propuestas de prestación de servicios de salud basados en estas nuevas tecnologías. A este fenómeno se le conoce como *m-Health* [3].

Un sistema *m-Health* hace referencia a la utilización de dispositivos móviles (dotados de un sistema de comunicación inalámbrico, *wireless*) para el aprovisionamiento de un servicio de salud ubicuo [3], para recopilar datos de salud de un paciente en tiempo real, almacenarlos y enviarlos a servidores accesibles desde Internet, así como el análisis y extracción de conocimiento de dichos datos [2]. Los componentes básicos [3] de un sistema *m-Health* son dos: la *body area network* (BAN) y el sistema *back-end* (BESys). En Figura 1 (en el capítulo anterior) se presentaban dichos elementos, los cuales se describen de forma más detallada en las secciones siguientes.

2.1.1. Body Area Network

La Body Area Network (BAN) es la red de comunicación de los dispositivos que se encuentran dentro del alcance del paciente (en el cuerpo o alrededor), capaces de realizar medidas fisiológicas del cuerpo, procesarlas y almacenarlas [4], [71]–[73]. También podemos encontrar el término WBAN, que hace hincapié en que es una red inalámbrica (en inglés, la *W* es acrónimo de *wireless*). No obstante, también se pueden considerar otro tipo de sensores, mientras estén cercanos o del entorno del paciente, dentro de una BAN como el GPS, o sensores ambientales (presión, humedad), que enriquezcan el sistema *m-Health* [4]. El radio de cobertura de comunicación efectiva de una BAN para *m-Health* [74] dependerá del tipo de protocolo que se utilice. Para ayudar a los sensores corporales de una BAN a transmitir información dentro de un entorno como, por ejemplo, el hogar e incluso los centros médicos, existen arquitecturas *ad hoc* que despliegan múltiples puntos de acceso, al crearse la red las comunicaciones dan saltos y aumentan la distancia en cientos de metros [4]. En la Figura 2, se presentan las entidades que conforman esta: los sensores y las unidades base móviles o MBU (del inglés, Mobile Base Unit) [3]. Dentro de la BAN (ver Figura 2), nos referiremos a las MBU como MBUP, es decir, las MBU del paciente.

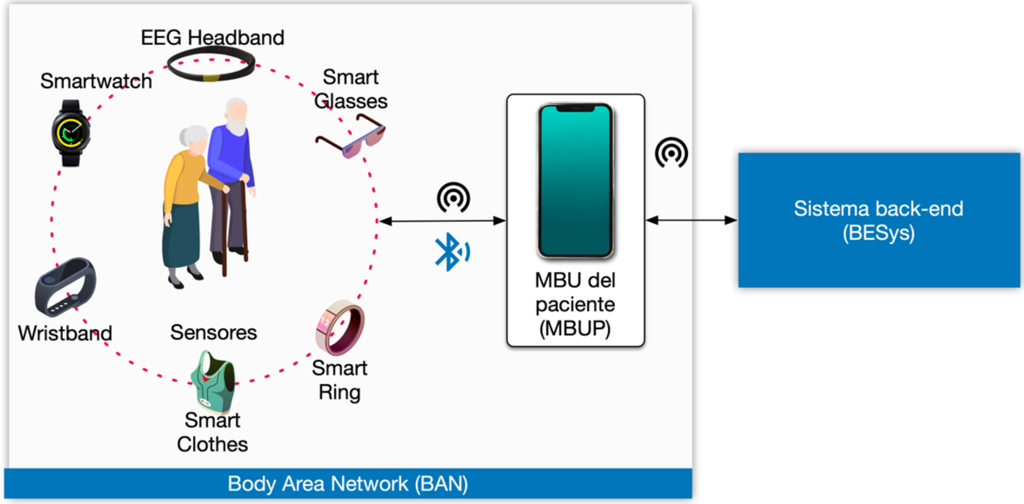

Figura 2. Ejemplo de entidades de una BAN y su comunicación

Los sensores son dispositivos que detectan señales de diversa naturaleza (mecánica, electrofísica, etc.). A los sensores de las BANs se les denomina *Sensor front-end* (SFE) y son los encargados de enviar esas señales que registran, de forma continua, hacia las MBU [3], [4], [71]–[73]. Algunos ejemplos de sensores que recogen señales fisiológicas (o bioseñales) pueden ser: sensor ECG (electrocardiograma) para medir la actividad del corazón; sensor EDA (Electro Dermal Activity) para determinar la actividad eléctrica de la piel; y sensor de temperatura para medir la temperatura de la piel en un paciente. Con el avance de la nanotecnología se ha conseguido una miniaturización de los componentes electrónicos de los dispositivos *wearables* o vestibles (en español), permitiendo que éstos incluyan varios sensores incorporados. Así, tenemos *wearables* que suelen ser aparatos cotidianos y ergonómicos como un reloj inteligente, una pulsera o un anillo. Estas características hacen que sean dispositivos no intrusivos, es decir, que no molestan al desempeño de las actividades de la vida diaria de sus usuarios. Sin embargo, existen otros dispositivos *wearables* que aún deben mejorar su funcionalidad, su fiabilidad, especialmente, y su ergonomía, como las bandas de cabeza que miden señales EEG (electroencefalograma) o las gafas inteligentes, para que se conviertan en dispositivos de uso diario. Por ello, empiezan a emerger nuevas iniciativas que buscan poner en el mercado dispositivos con sensores complejos como el EEG, por ejemplo, pero manteniendo el confort en su uso. Algún ejemplo reciente son los auriculares Neurable que miden EEG y algunas señales físicas. Se espera que en los próximos años haya un avance a nivel tecnológico en el campo de los *wearables*.

Las unidades base móviles, o MBU, son dispositivos móviles que ofrecen la funcionalidad de una puerta de enlace (o *gateway*, en inglés) —normalmente hacia Internet—. Una *gateway* tiene la capacidad de establecer la comunicación de múltiples dispositivos ubicados en diferentes redes. Así, las MBUP sirven de puente de comunicaciones entre el paciente y el BESys, es decir, conectando la BAN del usuario con el sistema *back-end* o BESys, permitiendo el intercambio de datos y mensajes.

2.1.2. Sistema back-end (BESys)

El BESys de un *m-Health* es la plataforma de control donde se pueden conectar una o varias BANs, a través de sus MBUs, para recibir los datos sensoriales de cada paciente [3]. Algunos ejemplos de entidades que pueden considerarse dentro del BESys del *m-Health* se presentan en la Figura 3: un servidor *backend* y las distintas aplicaciones (en las MBUC) para usuarios como personal médico o clínicos y cuidadores informales de las personas mayores.

El servidor back-end se encarga de recibir y procesar los datos y mensajes recibidos de una o varias BANs, analizarlos y extraer conocimiento a partir de éstos. Sus diferentes funcionalidades se pueden separar para que las realicen componentes especializados en cada una de ellas. Los esenciales pueden ser: recopilación, procesamiento, y almacenamiento de los datos sensoriales, análisis e informe de resultados (Figura 3). Además de estos componentes, podemos tener en el BESys aplicaciones disponibles en MBUCs para cuidadores y clínicos, donde se muestren los datos recopilados de los pacientes y se reporten informes acerca de los análisis llevados a cabo. Todo ello dependerá de la aplicación particular que se quiera proveer a estos usuarios.

Por último, para que el sistema back-end, BESys, sea de utilidad, tiene que estar accesible desde el exterior. Para ello, es habitual exponer su servicio mediante una API (Application Programming Interface) gateway y, así, se podrá comunicar el back-end con el exterior y los usuarios externos, como el paciente, con ésta.

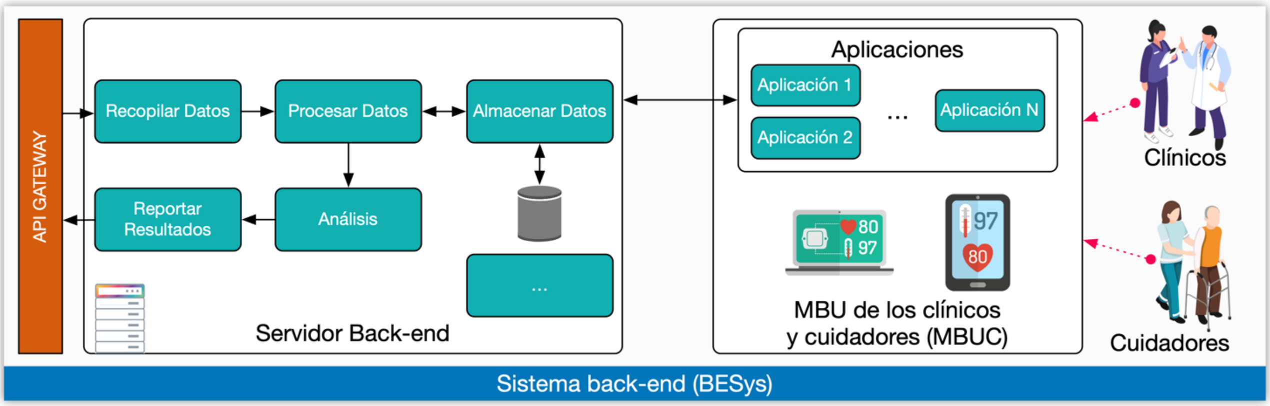

Figura 3. Ejemplo de arquitectura para el sistema back-end BESys

2.2. Arquitectura de microservicios

Para diseñar e implementar un sistema *m-Health* en IoT (*m-IoTHealth*), basado en servicios accesibles en Internet, existen diferentes alternativas, tantas como estilos arquitectónicos de software existan y dependiendo de la estrategia de diseño que se persiga. En esta línea, se presentará la arquitectura Service Oriented Architecture (SOA) y una especialización de ésta, la *Arquitectura de Microservicios*, por ser ésta última la más adecuada para el diseño de un *m-IoTHealth*.

2.2.1. Service Oriented Architecture (SOA)

En IoT, es habitual el uso de *arquitecturas basadas en servicios*, en las que el servicio (trozo de código reducido [74]) es un componente arquitectónico principal [8]. Éstas promueven una abstracción de las tecnologías subyacentes en la vasta infraestructura que puede tener un sistema IoT. Para ello, uno de los nodos de la red se expone al exterior para que el servicio desplegado sea consumido.

SOA se concibe como solución a los desafíos de los sistemas distribuidos anteriores que se desarrollaban como aplicaciones monolíticas, donde el sistema incumplía principios de diseño como el *acoplamiento débil*, puesto que contenía completamente toda la funcionalidad para realizar la tarea para la cual fue diseñado [75]–[78]. En el estilo arquitectónico SOA, cada componente de la arquitectura se implementa como un servicio independiente, el cual se concibe como un «pequeño» trozo de código que realiza una o varias tareas [79]. Sin embargo, es habitual que, en función de las necesidades, estos servicios no sean tan «pequeños» debido a que tengan que asumir varias tareas u operaciones. Los principios que SOA promueve son la escalabilidad, reusabilidad, autonomía y *acoplamiento débil*, entre otros [80]. Generalmente, SOA se utiliza en la integración de aplicaciones software independientes, proporcionando la funcionalidad de *middleware* para el intercambio de datos entre los servicios de estas aplicaciones. Cada servicio comparte el mismo mecanismo de comunicación común llamado *bus de servicio empresarial* (ESB). En este sentido, SOA puede implementarse tanto como *servicios web*, como servicios completamente aislados e independientes (*servicios stand-alone*).

Estos *servicios web* representan un enfoque importante en la implementación del estilo arquitectónico SOA en la web. Cabe mencionar que, a veces, en la literatura se confunde el *servicio web* con la arquitectura SOA, llamándolos indistintamente para referirse a SOA. A colación, podemos rescatar la definición que el World Wide Web Consortium (W3C) hace de los *servicios web*[5]:

> «Un sistema de software diseñado para soportar la interacción interoperable de máquina a máquina a través de una red. Dispone de una interfaz descrita en un formato procesable por la máquina (concretamente WSDL). Otros sistemas interactúan con el servicio web de la manera prescrita por su descripción mediante mensajes SOAP, normalmente transmitidos mediante HTTP con serialización XML junto con otros estándares relacionados con la web.»

Es por ello que, aunque los *servicios web* se han usado ampliamente como implementación de la arquitectura SOA, éstos no cumplen totalmente los principios SOA [80]. Por ejemplo, el principio de *acoplamiento débil* (minimizar dependencias) en los *servicios web* que implementan SOA se aplica parcialmente, ya que los servicios pueden ser responsables de varias funcionalidades al mismo tiempo. Eso provoca más incumplimientos parciales de otros principios SOA, como el principio de reusabilidad de los servicios. Sólo los servicios que sean genéricos e independientes de la funcionalidad, pueden ser reutilizados [81]. Cuando los servicios están *débilmente acoplados*, habrá margen de reutilización. Por contra, cuanto mayor sea el *acoplamiento* entre los servicios, menos escalable será la aplicación (principio de escalabilidad), lo que afectará al rendimiento de la misma. Por ello, para lograr una mayor escalabilidad

[5] W3C. ¿Qué es un servicio web? https://www.w3.org/TR/ws-arch/#whatis

en la aplicación de software, los servicios deben estar *débilmente acoplados*. Y en este sentido, un nuevo concepto de *microservicios* ha evolucionado para superar los problemas que tienen los *servicios web*.

2.2.2. Microservicios

La *Arquitectura de Microservicios* también es una *arquitectura basada en servicios* como SOA. Y del mismo modo que los *servicios web* son el origen de la implementación SOA, los *microservicios* son una implementación moderna de los conceptos de SOA en la implementación de aplicaciones de nivel empresarial [75]. Una definición muy completa de *Arquitectura de Microservicios* es la que dieron Martin Fowler y James Lewis[6]:

> «El estilo arquitectónico de microservicios es un enfoque para desarrollar una aplicación única como un conjunto de pequeños servicios, cada uno de los cuales se ejecuta en su propio proceso y se comunica con mecanismos ligeros, a menudo una API de recursos HTTP. Estos servicios se construyen en torno a las capacidades empresariales y se pueden desplegar de forma independiente mediante una maquinaria de despliegue totalmente automatizada. Hay un mínimo de gestión centralizada de estos servicios, que pueden estar escritos en diferentes lenguajes de programación y utilizar diferentes tecnologías de almacenamiento de datos.»

Según Fowler y Lewis, el término «*microservicio*» tiene su origen en una discusión de arquitectos de software en un *workshop* cerca de Venecia (en mayo de 2011). Muchos de los participantes del *workshop* pusieron en común, y trataron de describir, el nuevo estilo arquitectónico que habían estado explorando recientemente. En marzo de 2012, James Lewis presentó algunas de estas ideas como un caso de estudio en la 33ª Degree Conference, en su charla «Microservices - Java, the Unix Way[7]», al igual que la presentación de Fred George[8] más o menos al mismo tiempo. Y así, el mismo grupo del *workshop*, finalmente en mayo de 2012 se decidió por acuñar el término «microservicios» como el nombre más apropiado para esa arquitectura. Merece especial mención Adrian Cockcrof [9], describiendo este enfoque como un «SOA detallado» (*fine-grained SOA*), quien fue un pionero en la implementación del estilo arquitectónico de microservicios en Netflix a escala web.

Teniendo esto presente, los *microservicios* se conciben con una granularidad más fina que los *servicios web* SOA. Recordemos que los *servicios web* SOA pueden ser responsables de una o varias funcionalidades, lo que reduce su reutilización. Sin embargo, la granularidad de los *microservicios* se concreta en el desempeño de una única responsabilidad (o principio responsabilidad única, SRP), es decir, implementa muy bien una única tarea u operación [82], o que contribuirá a que estén muy especializados, aumentando su reutilización. Además, el desarrollo de cada *microservicio* es completamente independiente del resto de *microservicios*, incluso en repositorios de código distintos [14], [74], y normalmente con sus propios

[6] Microservices - Martin Fowler. https://martinfowler.com/articles/microservices.html

[7] Micro services - Java, the Unix Way - 33rd Degree. http://2012.33degree.org/talk/show/67

[8] MicroService Architecture - SlideShare. https://www.slideshare.net/fredgeorge/micro-service-architecure

[9] Talking microservices with the man who made Netflix's cloud famous. http://shorturl.at/aNZ24

protocolos de comunicación [83]. Por ejemplo, se puede decidir que un *microservicio* se exponga con REST mediante HTTP, otro *microservicio* podría usar *buffers* de protocolo y un tercer *microservicio* podría usar Java RMI. Teniendo esto presente, la integración de los distintos *microservicios* en un sistema podría complicarse, ya que los *microservicios* tendrían que soportar múltiples estilos de intercambio. Por ello, en la medida de lo posible, una práctica recomendada es restringir el intercambio a los protocolos establecidos [14]. Un ejemplo común es usar REST vía HTTP con mensajes JSON en el *payload* [84].

En la literatura relacionada, podemos encontrar distintas comparaciones entre las arquitecturas SOA y de *microservicios,* donde se han presentado las ventajas y desventajas de ambas arquitecturas.

2.2.3. SOA versus Microservicios

Existen distintos trabajos de investigación que comparan las arquitecturas SOA y *microservicios*. Por ejemplo, Cerny *et al.* señalan las ventajas y desventajas de ambas arquitecturas, como que los *microservicios* se basan en el patrón *share-as-little-as-possible* (compartir lo mínimo posible), mientras que SOA en *share-as-much-as-possible* (compartir lo máximo posible) [85]. En su lugar, Bogner *et al.* también analizaron las diferencias entre ambas arquitecturas basándose en los resultados de encuestas, que distribuyeron entre sus contactos profesionales que usaban microservicios, mediante listas de correo y redes sociales [86]. Además, posteriormente los mismos autores examinaron las cualidades de la evolución de los sistemas SOA y *microservicios* y afirmaron que existen diferencias en términos de *acoplamiento débil* [87]. A la misma conclusión llegaron los autores Raj y Ravichandra [76].

Tanto la elección de las *arquitecturas de microservicios* en lugar de SOA al inicio de un desarrollo, como la migración de sistemas SOA existentes a *microservicios*, no se han estudiado en profundidad. Existen estudios como el de Bhallamudi *et al.*, que justificaron la migración de una aplicación web existente y monolítica a una basada en SOA, con el aumento del retorno de la inversión (ROI) [77]. Otros estudios, como Levco-vitz *et al.* y Balalaie *et al.* propusieron técnicas para migrar sistemas heredados existentes (generalmente, monolíticos) a *microservicios* [78], [88]. Recientemente, se publicó el estudio de Raj y Sandam, donde demuestran que la elección de una granularidad muy fina en *microservicios* asegura el principio de *acoplamiento débil* frente a la alternativa de *servicios web* SOA. Este principio está estrechamente relacionado con otros como la escalabilidad, rendimiento, autonomía, etc. [75]. Así, el *acoplamiento débil* minimiza las dependencias entre servicios, lo que contribuye a conseguir la autonomía de los servicios, mejorar la reusabilidad de éstos, establecer servicios *statelessness* (sin información de estado) y lograr la escalabilidad [89]. Que las aplicaciones sean más escalables favorece a la evolución de los negocios contemporáneos, que necesitan sistemas más escalables [89]. Muchos sistemas SOA están basados en el protocolo de comunicación SOAP y su escalabilidad se puede ver afectada, ya que los *servicios web* deben registrar información de su estado (o transacciones) en el servidor (service *stateful*) [75]. Así, como cada servicio depende de otros servicios para obtener información del estado, existe un estrecho *acoplamiento* entre los servicios, lo que a su vez reduce la escalabilidad [90].

Raj y Sandam demostraron que elegir *microservicios* frente a *servicios web* SOA puede ser más beneficioso, en términos del principio de *acoplamiento débil* [75]. Mediante un grafo de servicios capturaban las interacciones entre los diferentes servicios de una aplicación. Elaboraron un grafo similar para el caso de los *microservicios*. Extrajeron métricas de *acoplamiento débil* en función de las interacciones entre los servicios (y *microservicios*) en cada arquitectura. Observaron que tanto los valores de *acoplamiento* de los servicios individuales, como el factor global de *acoplamiento*, son menores con *microservicios* que con la misma aplicación implementando SOA. No obstante, todo dependerá de cada situación y caso de estudio, aún se necesitan más estudios empíricos que comparen ambas arquitecturas siguiendo otras métricas como, por ejemplo, la reusabilidad, la tolerancia a fallos, etc., incluso estudios que valoren combinar lo mejor de cada arquitectura, en lugar de ponerlas a competir.

En cualquier caso, esta tesis se ha centrado en aprovechar las bondades de las arquitecturas basadas en *microservicios* para la construcción de un sistema *m-IoTHealth* de monitorización ubicua, flexible y con la presencia de componentes especializados y reutilizables.

2.3. Aprendizaje automático para m-Health

Entendemos por aprendizaje automático (en adelante, ML —por sus siglas en inglés *machine learning*) el campo de la inteligencia artificial (IA) que utiliza técnicas estadísticas para dotar a los sistemas de computación la capacidad de aprender a partir de un conjunto de datos [91]. Su objetivo es el estudio y la construcción de modelos computacionales de los procesos de aprendizaje en sus múltiples manifestaciones. El ML ha resuelto muchos problemas, para los cuales no existían algoritmos que lo solucionaran, desarrollando algoritmos que aprenden por sí mismos. Formalmente, Tom Michael Mitchell definió que los algoritmos de ML aprenden de la *experiencia* E con respecto a alguna *tarea* T y alguna medida de *rendimiento* P, si su rendimiento en T medido por P mejora con la experiencia E [92].

Los algoritmos de ML se clasifican principalmente en dos categorías:

- **Aprendizaje supervisado:** se trata de un aprendizaje basado en ejemplos conocidos, donde a partir de dicho conjunto de ejemplos (datos) y unas etiquetas (clases) asociadas a cada uno de ellos, se aprende a inferir esas etiquetas [93]. Concretamente, el algoritmo de aprendizaje supervisado recibe como entrada unos ejemplos y unas etiquetas, que identifican a cada ejemplo (pertenencia a una clase). A modo ilustrativo, podríamos pensar en ejemplos como los datos de los tres ejes de un acelerómetro (x, y, z) medidos en distintos instantes de tiempo. Las etiquetas asociadas a cada ejemplo (en cada instante de tiempo) podrían ser, por ejemplo, «sano» o «enfermo». Formalmente, dada la tupla (ejemplo, etiqueta), existe una función f para ese ejemplo que genera la etiqueta: (ejemplo, etiqueta) = (ejemplo, f(ejemplo)). En el caso del acelerómetro, podríamos escribirlo como ((x, y, z), etiqueta) = ((x, y, z), f((x, y, z))). Por tanto, el objetivo es que el algoritmo encuentre patrones en los datos (ejemplos) y descubra una función que aproxime a la función f, que es la que generó la etiqueta de los ejemplos conocidos. Ejemplos de estos algoritmos son los algoritmos de regresión (la etiqueta es un número continuo) y la clasificación (la etiqueta es discreta).

- **Aprendizaje no supervisado:** en este caso, se trata de un aprendizaje por observación donde a partir de un conjunto de ejemplos no etiquetados, se construyen descripciones, hipótesis y teorías acerca de dicho conjunto [93]. Los algoritmos de aprendizaje no supervisado más habituales son los de *clustering*, que agrupan los ejemplos observando los valores de sus datos, correlaciones entre ellos, estructura intrínseca, etc. A modo ilustrativo, si tenemos ejemplos del acelerómetro de personas *sanas* y *enfermas*, pero se desconocen dichas etiquetas, un algoritmo *clustering* es capaz de observar los datos de los ejemplos y proponer su hipótesis de agrupación de los ejemplos en sanos y enfermos, es decir, es capaz de distribuir los ejemplos dados en dos grupos distintos, aunque realmente desconoce si son sanos o enfermos; por lo que hará falta un experto que interprete a posteriori las agrupaciones.

2.3.1. Etapas del ML supervisado para m-Health

Hoy en día, los algoritmos de ML se han aplicado con éxito en diversos campos como la recomendación de productos, detección de *spam*, diagnóstico de enfermedades, etc. Asimismo, también se han aplicado en estudios relacionados con nuestros ámbitos de interés en el dominio de salud de personas mayores, la detección de fragilidad [20], [49]–[51], [55] y la evaluación de la dependencia [52], [53], a partir de los datos de dispositivos *wearables*. Dada la naturaleza dinámica y temporal de los datos que recogen los *wearables*, en la mayoría de estos estudios, se establece un sistema ML que sigue el flujo de trabajo representado en la Figura 4 con sus diferentes etapas, pero donde la etapa fundamental suele ser la de *procesamiento de los datos*, con tareas comunes como *alineamiento de señales*, *segmentación* de los datos y *extracción de características* de los mismos. A continuación, en los siguientes apartados se dan los detalles de cada una de estas tareas.

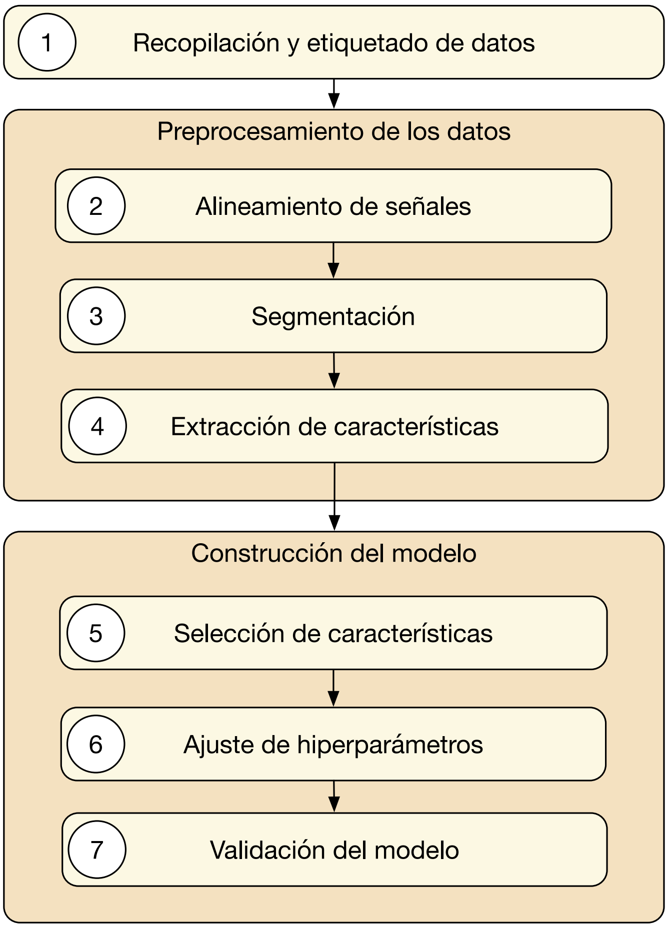

Figura 4. Flujo de trabajo en ML con datos de wearables

2.3.1.1. Recopilación y etiquetado de los datos

En esta primera etapa, se realiza el almacenamiento de los datos provenientes de los distintos sensores disponibles, en un formato adecuado para poder trabajar con ellos. Normalmente, es habitual utilizar ficheros de tipo texto con formatos como JSON, JSON-LD (o RDF) o CSV (*comma sepated values*). El formato CSV, por ejemplo, estructura los valores separados por coma donde, habitualmente, cada columna representa una medida del sensor, y cada fila es la medición en un tiempo determinado (marca de tiempo según la tasa de muestreo correspondiente a ese sensor).

Adicionalmente, si el problema en cuestión con el que va a lidiar el sistema *m-Health* utiliza los algoritmos de ML supervisado, en esta etapa se necesitará que los datos estén etiquetados (clase) para poder descubrir qué relación se da entre la entrada y salida del algoritmo. La clase puede ser binaria (dos posibles valores como, por ejemplo, positivo y negativo) o múltiple (más de dos valores). En particular, en *m-Health* este proceso se realiza bajo la supervisión y experiencia de los profesionales sanitarios. Por ejemplo, si el modelo a construir debe clasificar a las personas en *frágiles* frente a *no frágiles*, los ejemplos que se le proporcionarán a los algoritmos de aprendizaje supervisado tendrán que estar etiquetados con estas clases. Dichas etiquetas son el resultado de una evaluación tradicional previa de la fragilidad mediante cuestionarios *ad-hoc* suministrados por el personal cualificado para ello.

2.3.1.2. **Preprocesamiento de datos wearables**

En un sistema *m-Health,* principalmente, contaremos con datos de los sensores de los dispositivos personales que lleven los pacientes en su BAN. Sin embargo, también podremos contar con datos de registro médicos del paciente, como su historial clínico, resultados de pruebas o cuestionarios, etc. Los datos provenientes de sensores suelen ser muy ruidosos. Una sesión de grabación de datos de un sensor puede presentar datos perdidos en ciertos intervalos de tiempo, por ejemplo, debido a desconexiones o interferencias del dispositivo. También puede haber otros problemas como datos que caducan, algunos que provienen de fuentes poco fiables, o sensores no calibrados que pueden dar valores fuera de los rangos establecidos, etc. Del mismo modo, los datos del registro médico, cuestionarios o pruebas del paciente también pueden presentar valores en blanco como, por ejemplo, resultados vacíos en una prueba concertada, pero a la que el paciente finalmente decidió no asistir.

Algunas de las técnicas de preprocesamiento de datos ruidosos pueden ser la interpolación para datos perdidos (o imputación de valores perdidos), la aplicación de filtros de frecuencias para limpiar el ruido en ciertas bandas no deseadas (o seleccionar las bandas de interés para el estudio), etcétera.

Asimismo, para poder analizar y comparar todas las señales entre sí, será necesario realizar un preprocesamiento particular consistente en el *alineamiento de las señales* hacia una tasa de muestreo homogénea. Asimismo, el preprocesamiento incluye la *segmentación en ventanas* de datos y la *extracción de características* (o ingeniería de características) a partir de esos datos. A continuación, describiremos estos procesos en detalle.

Alineamiento de señales

Las señales de un *m-Health* son heterogéneas y se capturan a distintas frecuencias, algunas se reciben antes y otras con más retardo. Por ejemplo, en las diademas que miden la EEG es muy común que las tasas de muestreo sean elevadas, por encima de 128Hz. En cambio, en sensores de movimiento como el acelerómetro o el giroscopio, se suelen usar tasas en torno a 25 o 30 Hz. Para un sensor que mida el ritmo cardíaco, también es habitual ver tasas bajas, aproximadamente de 5 Hz.

El alineamiento de señales consiste en transformar los datos de señales temporales heterogéneas, a partir de sus frecuencias originales y distintas unas de otras, hacia una frecuencia común. Para ello, y dado que este procedimiento generará valores perdidos (Figura 5-A), suele usarse la interpolación (Figura 5-B) para así poder aproximar la señal original a una señal en la nueva frecuencia objetivo. En concreto, la Figura 5-A muestra la señal del acelerómetro (eje x) después de haber sido alineada, por lo cual presenta valores perdidos como puede observarse. A continuación, en la Figura 5-B se han imputado los valores perdidos insertando nuevos valores mediante interpolación (basada en el cómputo de la media aritmética entre el valor anterior de la señal y el siguiente).

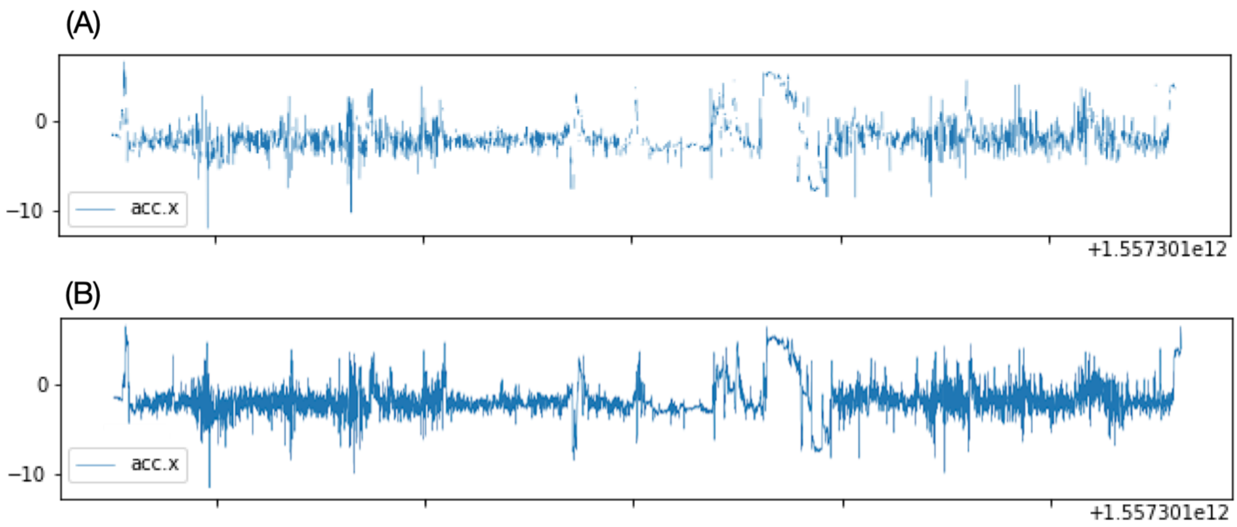

Figura 5. Interpolación en el alineamiento de distintas señales

Segmentación en ventanas

Disponemos de series de datos temporales para una actividad o evaluación con una duración concreta. Una técnica de preprocesamiento muy común es segmentar la serie en trozos llamados ventanas (*sliding windows*), y cuya longitud se mide en segundos. Normalmente, estas ventanas suelen ser de igual longitud temporal dado que serán la entrada que recibirán los algoritmos de ML que trabajan con una longitud fija. Aunque, ya hay algunas excepciones con el uso de algoritmos más sofisticados, que permiten longitudes variables.

Asimismo, con el objetivo de aumentar el número de datos disponibles, también se suele segmentar la serie solapando datos entre dos ventanas consecutivas, también conocido como *overlapping*, siendo un valor común el solapamiento del 50% de los datos de una ventana predecesora con la sucesora.

Extracción de características

Otra tarea fundamental es la extracción (o ingeniería) de características, que consiste en aumentar el número de variables independientes, creando nuevas variables mediante técnicas estadísticas, a partir de los datos en crudo disponibles [94], [95]. Dichas características se pueden derivar en función de su carácter temporal (como la media, desviación típica, máximo, mínimo, etc. [94]) o, en cambio, desde una perspectiva frecuencial (como: frecuencia de pico, potencia de pico, potencia espectral en diferentes bandas de frecuencia, entropía espectral, etc. [95]). Normalmente, si se ha seguido un enfoque previo de segmentación de los datos, por cada una de las ventanas se extraen las características deseadas; y si no se segmentan los datos, se extraerán las características de los datos para la duración completa.

2.3.1.3. **Construcción del modelo**

En la última etapa del flujo de trabajo se construye el modelo de ML supervisado, incluyendo tareas como seleccionar de todas las características cuáles son las más explicativas (o de mayor importancia para la clasificación), ajustar los *hiperparámetros* de los propios algoritmos de ML y evaluar el modelo construido.

Selección de características

En la tarea de selección de características, además de reducir la dimensionalidad del conjunto de datos, se persigue detectar cuáles son las variables más influyentes en la clasificación, es decir, seleccionar las que mejor explicabilidad del modelo proporcionen [96]. Por ejemplo, una estrategia común [50] para buscar esas variables es la Random Feature Elimination (RFE) [97]. Consiste en realizar una búsqueda de esas variables partiendo del conjunto completo, ordenarlas por una cierta relevancia inicial (usando algún método estadístico como, por ejemplo, chi-cuadrado; o métodos de *feature importance*, como Random Forest, etc.), evaluar nuestro modelo y eliminar la menos relevante. Este proceso de ordenación-evaluación-eliminación es iterativo hasta encontrar el mejor modelo posible.

Ajuste de hiperparámetros

En el *pipeline*, otra etapa a abordar es la búsqueda de los mejores *hiperparámetros* asociados a cada algoritmo de ML [96]. Estos *hiperparámetros* son valores que no se aprenden directamente en los algoritmos de ML, sino que constituyen los valores de los parámetros propios (variables de configuración) de dichos algoritmos. Ejemplos de *hiperparámetros* pueden ser la *capacidad de aprender* del algoritmo, el *número de hojas* a considerar en el algoritmo Random Forest, el número de *kernels* en el algoritmo Support Vector Machines (SVM) o el número de *k* vecinos más cercanos en k-Nearest Neighbors. La elección de los algoritmos de ML a utilizar dependerá del tipo de aprendizaje que estemos realizando (supervisado, no supervisado…). Además, es habitual no decantarse por ninguno a priori y, así, en esta etapa se prueban diferentes algoritmos y sus diferentes *hiperparámetros*, con el objetivo de encontrar el algoritmo de ML y sus *hiperparámetros* asociados más óptimos.

Para realizar este ajuste u optimización de *hiperparámetros*, es habitual usar la técnica de búsqueda en rejilla (*Grid Search*) [96]. Consiste en definir un espacio de búsqueda de los posibles valores que tomarán los *hiperparámetros*, construir un modelo de ML con el algoritmo indicado y los *hiperparámetros* obtenidos y evaluar el rendimiento del algoritmo, mediante alguna técnica como las que se verán en la siguiente subsección. Se repetirá la búsqueda para cada algoritmo de ML que se desee probar.

Ejemplos de posibles espacios de búsqueda para *hiperparámetros* asociados a distintos algoritmos ML son [95], [98], [99]:

- k-NN: 1) k: {1, 3, 5, ..., $\sqrt{\text{total registros}}$ } (números impares, para evitar empates en la selección de los vecinos).
- SVM: 1) C (coste): {0.1, 1, 10, 100}; 2) gamma value: {0.5, 1, 2}; 3) y, tipo de *kernel*: {"radial", "polynomial", "linear", "sigmoid"}.

En esta optimización, se encontrará cuál de los modelos obtiene la mejor puntuación de rendimiento y cuáles son los valores sus *hiperparámetros* asociados que contribuyeron a ello.

Evaluación del modelo

Por último, para saber la eficacia y la confianza que nos aportará nuestro modelo necesitamos métricas que evalúen su rendimiento. Por ejemplo, en un problema de clasificación de etiquetas en aprendizaje supervisado, un buen rendimiento supone comprobar que la clasificación realizada coincide con las etiquetas iniciales de los ejemplos, resultando en un 100% de precisión.

Las métricas más utilizadas en los trabajos de *m-Health* son las que se enumeran a continuación para un problema de clasificación binaria, donde tenemos dos clases o etiquetas a las cuales nos referiremos como la clase positiva y la clase negativa. En *m-Health*, la clase positiva se corresponde con los ejemplos de datos asociados a un paciente que está enfermo. Por el contrario, la clase negativa será de ejemplos de un paciente sano. En este contexto, para medir el grado de acierto de un algoritmo de ML en la clasificación de la clase positiva y negativa, se definen los conceptos TP, TN, FP y FN. Los TP (*true positive*) significa la condición *verdaderos positivos*, es decir el número de ejemplos clasificados correctamente como positivos (enfermos). Por el contrario, TN (*true negative*) son los *verdaderos negativos* o lo que es lo mismo, el número de ejemplos clasificados correctamente como la clase negativa (sanos). También se contabilizan los FP (*false positive*) que simbolizan el número de *falsos positivos* o, dicho de otra forma, número de ejemplos de la clase negativa clasificados erróneamente como positivos; cuando debían ser clasificados como negativos (los ejemplos, realmente no eran enfermos). Y, por último, los FN representan los *falsos negativos*, es decir, los ejemplos de la clase positiva que fueron clasificados erróneamente como negativos; cuando debían ser clasificados como positivos (los ejemplos, realmente no eran sanos, sino que eran enfermos). Asimismo, las métricas que evalúan la bondad o rendimiento de un algoritmo de ML de clasificación, utilizan estos recuentos tal y como se enumera, a continuación, para proporcionar índices de *accuracy* (exactitud: proporción de ejemplos clasificados correctamente, tanto los positivos como los negativos); *sensitivity* (sensibilidad: proporción de la clasificación correcta de los ejemplos positivos, de todo el conjunto de ejemplos positivos); *specificity* (especificidad: proporción de la clasificación correcta de los ejemplos negativos, de todo el conjunto de ejemplos negativos); y F1-score (valor-F1: una métrica más completa que promedia el acierto de la clase positiva teniendo en cuenta tanto los FP como los FN):

- Exactitud (*accuracy*) = (TP + TN) / (TP + TN + FP + FN)
- Sensibilidad (*sensitivity*) = TP / (TP + FN)
- Especificidad (*specificity*) = TN / (TN + FP)
- Valor-F1 (F1-Score) = 2TP / (2TP + FP + FN)

Aunque todas estas métricas se utilizan en *m-Health* para construir un modelo robusto, merece especial mención la relación entre la sensibilidad y la especificidad de un modelo. Se prefiere una sensibilidad mayor que la especificidad de un modelo, dado que es asumible un clasificador más errático en falsos positivos (FP) que en falsos negativos (FN). En la práctica, esto se traduce en que más adecuado que el clasificador proporcione un diagnóstico positivo de una enfermedad —aunque realmente el paciente esté sano (FP)—, a que se equivoque en el diagnóstico de un paciente que realmente está enfermo (FN). Por lo tanto, en *m-Health* se prefiere un modelo con la sensibilidad mayor que la especificidad.

Teniendo en cuenta estas métricas, para finalmente evaluar el modelo existen distintas estrategias como *leave-one-out*, *hold-out* o *k-fold cross-validation*. Todas comparten el mismo principio, utilizar una cantidad de datos (ejemplos) para entrenar el modelo de ML y dejar otra cantidad de ejemplos para validar el rendimiento del modelo con los nuevos ejemplos reservados. Por ejemplo, *leave-one-out* consiste ir entrenando un modelo con todos los ejemplos disponibles excepto uno, que se reserva para validar el rendimiento del modelo y repitiendo esta validación con todos los ejemplos disponibles, lo cual supone un cómputo elevado. Por su parte, la validación *hold-out* es menos costosa, puesto que se trata de dividir, aleatoriamente, todo el conjunto de ejemplos disponibles en dos subconjuntos: uno para entrenar el modelo y otro para validarlo y, aunque el proceso se suele repetir varias veces (por ejemplo, 10-*hold-out indica que se repite 10 veces*), no suele igualar las repeticiones de *leave-one-out* (que itera sobre el número de ejemplos). Otra técnica habitual es la validación *k-fold cross-validation*, que consiste en dividir el conjunto de ejemplos en *k* trozos (*folds*), de los cuales uno de los trozos se reserva para validar el modelo y el resto (*k*-1 trozos) se usa para entrenar el modelo, repitiendo el proceso *k* veces y usando cada vez un trozo distinto para la validación. Finalmente, en todas las técnicas presentadas se reporta el rendimiento promediando las distintas iteraciones realizadas de la métrica utilizada (por ejemplo, el promedio de la *accuracy*).

En resumen, en la etapa de optimización o búsqueda de mejores *hiperparámetros*, se elegirá una de las métricas —o una combinación de ambas— para centrar la optimización en dichas métricas, se construirán los modelos, y se evaluará su rendimiento (por ejemplo, con *k-fold cross-validation*), reportando el mejor modelo y configuración de *hiperparámentros* en función de las métricas a usar en la optimización. Por ejemplo, un criterio de optimización podría ser maximizar el *accuracy* y la sensibilidad.

2.3.2. El paradigma MLOps

Como hemos visto, el flujo de trabajo en ML se divide en distintas etapas y siguiendo un orden determinado (Figura 4). Recientemente, el paradigma MLOps promueve que la implementación del flujo de trabajo de ML se realice de forma independiente (piezas de software distintas), diferenciando cada tarea de forma unívoca para facilitar su mantenimiento, y centrándose en proveer la generación del modelo de ML como un servicio o producto final [23], [24].

Como veremos en la siguiente sección, un diseño arquitectónico basado en *microservicios* puede ayudar a seguir el MLOps gracias a sus características particulares, como la especificidad de las tareas de cada *microservicio* (esa granularidad fina). Asimismo, la flexibilidad y adaptabilidad que nos ofrece la arquitectura de *microservicios*, supone contar con diferentes *microservicios* especializados en cada paso del flujo de trabajo de ML, de tal forma que se puede escoger usarlos o no, dependiendo del conjunto de datos o del dominio del problema (o del que se haya usado en el paso previo o de la salida del paso previo).

Como se mencionó en subsecciones anteriores, para el ML supervisado se necesita etiquetar los datos y, para nuestros casos de estudio de evaluación de la fragilidad y la dependencia, este etiquetado se hace mediante una evaluación tradicional que consiste en completar cuestionarios validados y hacer pruebas específicas que nos dan datos y nos ayudan a clasificar.

2.4. Evaluación tradicional de fragilidad y dependencia

En primer lugar, la evaluación del **síndrome de la fragilidad** se viene realizando de la forma tradicional: suministrando un cuestionario validado como el de los autores Fried *et al.* (del año 2001) [37], dando nombre al cuestionario del «Fenotipo de Fried». Este cuestionario consta de cinco criterios: tres preguntas y dos pruebas físicas. Por un lado, las preguntas que deben responder los encuestados son acerca de: cuántas kilocalorías queman a la semana realizando actividades físicas; si han perdido más de 4,5 kg en el último año, de forma no intencionada; y una doble pregunta (puntúa como una sola respuesta) acerca de su estado de salud mental (1: ¿cuántas veces sintió que todo le suponía un esfuerzo?; 2: ¿cuántas veces pensó que no se podría ni mover?). Por otro lado, las dos pruebas físicas que se evalúan son: el tiempo en realizar una caminata de 4,5 metros y la fuerza de agarre de la mano. Según unos puntos de corte validados por los autores del *Fenotipo de Fried*, se establece la fragilidad si se cumplen tres o más criterios (del total de cinco); y uno o dos criterios, se considera como prefrágil. Asimismo, han ido surgiendo otras iniciativas que tratan de evaluar otras dimensiones humanas como la social y la psicológica en mayor medida que el cuestionario de *Fried*, como es el cuestionario Tilburg Frailty Indicator (TFI) [36], el cual cuenta con 25 preguntas divididas en dos partes: 1) los determinantes de fragilidad (6 preguntas, sociodemográficas y 4 acerca de la percepción propia y satisfacción); y 2) los componentes de la fragilidad: 15 preguntas donde 8 son relacionadas con la dimensión física, 4 sobre aspectos psicológicos y 3 sobre apoyo social.

En segundo lugar, la evaluación del **estado de dependencia** se realiza tradicionalmente usando el cuestionario original de Lawton and Brody [100]. Se trata de un cuestionario que evalúa la capacidad de realización o desempeño de 8 AIVD: usar el teléfono, ir de compras, preparación de la comida, cuidado de la casa, lavado de la ropa, uso de medios de transporte, responsabilidad respecto a su medicación, manejo de asuntos económicos. Normalmente, se suele utilizar una versión del cuestionario original validada en la población objetivo como, por ejemplo, en la española [101].

En conclusión, a excepción de algunos de los datos demográficos y algunos datos físicos, medibles objetivamente, este tipo de instrumentos tienen un carácter subjetivo tanto del evaluador como del encuestado, dependiendo de si éste último dirá la verdad o cuál es el grado o puntuación que el evaluador considera adecuado. Una evaluación algo menos subjetiva, holística y con carácter preventivo, sería aquella que requiere de la supervisión de un profesional sanitario durante el desempeño de las AIVD para poder evaluar la medida en que considera que el sujeto evaluado es capaz de realizarla de forma autónoma. Esto sabemos que sería altamente costoso para los sistemas sociosanitarios.

Un sistema *m-Health* sería una solución como se ha venido sosteniendo en esta tesis, ya que aporta sensores para poder registrar datos objetivos —en cualquier lugar y en cualquier momento— y mecanismos de análisis de esos datos, para clasificar a las personas según su estado de fragilidad o dependencia. Asimismo, adoptar una *arquitectura de microservicios* permitirá que el sistema *m-Health* sea escalable, flexible y adaptativo a un entorno cambiante como es el sanitario.

2.5. Estudio de las señales EEG

Analizar señales EEG (dimensión mental) puede revelar patrones de comportamiento de un usuario y medir su actividad mental. Así que, como un primer paso en este estudio, se realizó una detección de actividad mental que tiene que ver con la *imaginación motora*, o más conocida como *motor imagery* [57]–[66]. El proceso consiste en utilizar dispositivos que capten señales EEG, durante una sesión donde los participantes del experimento tienen que imaginarse cómo mueven alguna de sus extremidades, pero sin realizar el movimiento físico. A partir de las señales EEG, el objetivo es lograr clasificar correctamente cada tipo de movimiento mediante algoritmos de ML (por ejemplo, detectar la imaginación motora de la mano derecha frente a la izquierda).

Dado que en un *m-Health* se desea que los dispositivos sean inalámbricos para aportar ubicuidad, hemos estudiado cómo detectar la intención motora utilizando dispositivos *wearables* que proporcionan señales EEG (diademas os *headsets*). Y, en esta línea, existen pocos estudios que usen este tipo de dispositivos. Además, los que existen, para realizar una detección precisa necesitan una duración de la grabación de señales EEG entre 10 y 30 segundos [68]–[70]. Hay estudios anteriores con cascos EEG que reducen la detección a 2 segundos [57]–[66]. Sin embargo, sabemos que estos cascos EEG son muy intrusivos y nada ubicuos, además que necesitan la aplicación de geles en el cuero cabelludo para favorecer la conductividad de sus electrodos, lo cual no es adecuado para un *m-Health*.

Incluir este tipo de dispositivos *wearable* de EEG en un sistema *m-Health*, en combinación con los datos fisiológicos de otros *wearables*, nos permitirá acercarnos a una evaluación de la dimensión de salud mental más ubicua que la tradicional. Además, también enriquecerán los análisis de datos de las enfermedades que sufren los mayores que afectan a varias de las dimensiones humanas (no solo a la física).

3. Objetivos

Una vez introducidos los fundamentos que sustentan este trabajo, presentaremos los principales objetivos que han impulsado esta tesis. Incluyen el diseño de un sistema *m-Health* y el uso de técnicas de análisis de datos para abordar la evaluación de la fragilidad, la dependencia y el estudio de las señales EEG. A continuación, se presentan los objetivos planteados:

1) **Diseñar un sistema *m-Health* para que sea flexible, reutilizable y extensible**, que contemple los procesos de recopilación de datos de dispositivos *wearables*, procesamiento y análisis, promoviendo la especialización de los componentes involucrados en dichos procesos, para contribuir a su reutilización y extensión de las funcionalidades del sistema *m-Health*, aplicable en diferentes casos de estudio que mejoren la calidad de vida de las personas mayores.

2) **Diseñar un sistema de recopilación y federación de datos de *wearables***, acelerómetro, giroscopio, ritmo cardíaco, contador de pasos, sudoración, temperatura corporal, etc., y otras fuentes —como cuestionarios validados, informes clínicos, datos de contexto, etc.—, que involucre la monitorización de una actividad instrumental de la vida diaria completa, como realizar la compra, y el etiquetado de las diferentes etapas de la actividad.

3) **Generar modelos de evaluación del estado de fragilidad y dependencia** mediante técnicas de aprendizaje automático, que permitan una evaluación precisa, objetiva, holística y ubicua, al tiempo que se explora el número mínimo de datos, de sensores y de dispositivos *wearables* necesarios.

4) **Estudiar otros factores que influyan en el envejecimiento activo**, como los relacionados con la dimensión humana cognitiva y mental, mediante el análisis de las señales EEG de dispositivos *wearables*., . Este objetivo se restringe a tener una primera experiencia de tratamiento de señales EEG, ya que la pandemia nos ha limitado los experimentos con mayores. Esta primera aproximación se plasma a través de una detección de la imaginación motora.

4. Metodología

Esta tesis requiere la aplicación de una metodología tanto teórica como práctica. Por lo tanto, necesitamos una estrategia que, al tiempo que mantiene las directrices del método científico tradicional, sea capaz de satisfacer las necesidades especiales del problema específico estudiado. En particular, se aplican las siguientes adaptaciones del método científico para el trabajo de investigación y los experimentos:

1. **Observación**: Se van a estudiar el síndrome de fragilidad, el estado de dependencia y la imaginación motora, mediante la obtención de datos de cuestionarios, pruebas y sensores de dispositivos *wearables*.

2. **Formulación de hipótesis**: Nuestro objetivo será comprobar la utilidad de nuevos métodos de evaluación de la fragilidad y dependencia en personas mayores. Los nuevos métodos deben implementar las características descritas en los objetivos, mencionados anteriormente, para hacer frente al problema de la subjetividad, los elevados costes sociosanitarios y diagnósticos tardíos que suponen las evaluaciones tradicionales.

3. **Recogida de observaciones**: Se obtendrán los resultados mediante la aplicación de los nuevos métodos en la evaluación de fragilidad y dependencia utilizando diferentes tipos de medidas de rendimiento como la exactitud (*accuracy*), sensibilidad, especificidad y valor-F1.

4. **Contraste de hipótesis**: Se compararán los resultados obtenidos con los publicados por otros métodos relacionados con el estado del arte en evaluación de fragilidad y dependencia.

5. **Validación de la hipótesis o rechazo**: Se aceptarán o rechazarán y modificarán, en su debido caso, las técnicas desarrolladas como consecuencia de los experimentos realizados y los resultados obtenidos. Si es necesario, se repetirán los pasos anteriores para formular nuevas hipótesis que puedan probarse.

6. **Tesis científica**: Se aceptarán o rechazarán y modificarán, en su debido caso, las técnicas desarrolladas como consecuencia de los experimentos realizados y los resultados obtenidos. Si es necesario, se repetirán los pasos anteriores para formular nuevas hipótesis que puedan probarse.

Asimismo, merece especial mención el proceso de diseño y recogida de los datos del experimento realizado en este trabajo, para la consecución del objetivo nº 2. En este experimento se prestará especial atención y cuidado a su diseño minucioso, consistiendo en diferentes etapas involucradas en el proceso de la realización de la AIVD *comprar* en un supermercado: desde que la persona se dirige al supermercado; cuando llega a éste; después se dirige a buscar el producto que tiene que comprar; localizarlo; ir a pagarlo; y, finalmente, regresar al punto de inicio. Todas estas etapas de la *compra* se planifican, valoran y acuerdan con profesionales sanitarios. Para ello, realizamos un experimento piloto con 5 participantes, con objeto de validar técnicamente este diseño inicial, el cual nos sirvió para perfeccionar el protocolo de recogida y el propio diseño de la propuesta tecnológica [102]. Además, personalmente, me involucré en la recolección de datos observables *in situ*, acompañando a las 79 personas mayores participantes del estudio mientras realizaban la compra y ayudé a los compañeros sanitarios a suministrar los diferentes cuestionarios y pruebas clínicas, como tomar medidas, peso, cronometrar la caminata, etc. Asimismo, se estudiaron y eligieron varios dispositivos *wearables* y se optó por no limitar, a priori, la captura de datos, sino obtener el máximo posible de todos los datos disponibles que ofrecían sus sensores y analizar cada uno de ellos junto con los de los cuestionarios y los observables, posteriormente con las técnicas oportunas.

5. Resultados

A continuación, se presentan los resultados más relevantes de la presente tesis, de cada uno de los casos de estudio como son, la evaluación de la fragilidad, la dependencia y la exploración de la dimensión cognitiva mediante las señales EEG.

5.1. A Microservices e-Health System for Ecological Frailty Assessment Using Wearables

Como se ha mencionado anteriormente, en el capítulo 1 y en la sección 2.4, la evaluación del síndrome de la fragilidad en las personas mayores se viene realizando de forma manual, suministrando cuestionarios que suponen un carácter subjetivo a su validación, diagnósticos tardíos y unos costes sociosanitarios derivados elevados. En aras de superar estas limitaciones, se propone el diseño de un sistema *m-Health* basado en *microservicios* para evaluar la fragilidad de forma holística y ubicua, utilizando dispositivos *wearables* (dispuestos en una BAN) y un modelo de aprendizaje automático implementado, específicamente, para evaluar el estado de fragilidad (como parte del BESys).

Primero, diseñamos un protocolo experimental para monitorizar la realización de la actividad instrumental de la *compra* por parte de personas mayores. El protocolo incluía la recogida y el posterior tratamiento de los datos provenientes de fuentes heterogéneas: distintos *wearables*, cuestionarios validados, pruebas físicas, datos sociodemográficos, etc. En particular los datos de los participantes del experimento se clasificaron según el resultado del cuestionario de Fried [37], para etiquetarlos en *frágiles*, *pre-frágiles* y *no frágiles*. Además, como se ha mencionado en el capítulo anterior de Metodología, también se etiquetaron las distintas etapas de la *compra*. En paralelo, se diseñó la *arquitectura de microservicios* para el sistema integral *m-Health*, que contemplaba *microservicios* para la recopilación de esos datos heterogéneos, su procesamiento, análisis y construcción de distintos modelos de aprendizaje automático para fragilidad. Como parte de esta arquitectura, se implementaron dos aplicaciones software y una biblioteca de utilidad, publicadas conjuntamente bajo un mismo nombre y repositorio de código abierto: *Tizensor*[10]. Constaba de una aplicación y biblioteca para la recopilación de datos de un dispositivo *wearable* con sistema operativo Tizen (como el que utilizamos en este trabajo: el reloj Samsung Gear S3); y otra aplicación para móviles con sistemas iOS capaz de controlar el flujo de datos del reloj y el etiquetado de las distintas etapas de la actividad de la *compra* (ir al supermercado, llegar allí, buscar el producto, etc.). Asimismo, se desarrollaron *scripts* para federar los datos del reloj, ya que cada sensor podía trabajar a distintas frecuencias, lo que hacía necesario alinearlos para los consiguientes análisis de los datos. Por último, de la arquitectura *m-Health*, se implementó el software del *back-end* encargado del análisis de los datos, y las

[10] Tizensor: https://github.com/frangam/tizensor

diferentes etapas del *pipeline* de ML (MLOps), hasta obtener el modelo de ML capaz de evaluar la fragilidad con un 99,2% de confianza.

Para obtener el modelo de ML, se realizaron varios experimentos de análisis de los datos, exclusivamente, de los *wearables* (excluyendo variables observacionales y subjetivas). Por un lado, se crearon modelos para las diferentes fases de la compra y, los resultados reportaron la mejor precisión cuando se construía un modelo que tenía en cuenta todas las etapas de la actividad *comprar* y no por separado. No obstante, se observó que la etapa de la *compra* más sensible en la clasificación de fragilidad fue «buscar el producto», lo cual indica su importancia central en dicha evaluación. Esta etapa tiene una clara implicación en el desempeño de *compra*, desde una visión multidimensional de las funciones humanas, puesto que implica una actividad motora, cognitiva importante (incluso cierta planificación), e incluso social —puesto que se observó y anotó, cómo algunos participantes socializaban con los dependientes u otros clientes del supermercado para preguntar dónde podían localizar lo que buscaban—. Por otro lado, y en aras de cumplir con el tercer objetivo de esta tesis, se extrajeron diversidad de variables independientes o características a partir de los sensores (como la media, desviación típica, etc.) y se llevó a cabo una exploración exhaustiva para obtener las más influyentes en la detección de fragilidad. Esto nos llevó a comprobar que solo utilizando el reloj Tizen con sus sensores de acelerómetro, giroscopio y ritmo cardíaco era suficiente era suficiente para hacer una buena clasificación. Además, se compararon distintos algoritmos de aprendizaje automático, hasta encontrar que 1-NN reportaba la mejor clasificación, implicando una clara separación entre los sujetos frágiles frente a los no frágiles.

Estos resultados suponen la realización de una evaluación automática de la fragilidad, de forma ubicua, monitorizando una actividad compleja como es ir a *comprar* y mediante la extracción y procesamiento de los datos de un *wearable*, mediante sus sensores de acelerómetro, giroscopio y ritmo cardíaco. En comparación a las evaluaciones tradicionales, el despliegue de nuestro modelo de ML para fragilidad puede suponer la detección temprana, la cual ayudaría a los profesionales médicos a realizar un diagnóstico precoz de la misma, y la reducción de tiempos y costes a los sistemas sociosanitarios, para que mejore la calidad de vida de las personas mayores y promover el envejecimiento activo y saludable.

5.2. A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors

Esta publicación continúa con el trabajo anterior sobre fragilidad con el objetivo de extender el sistema *m-Health* hacia la evaluación de otros estados de salud que afectan a los mayores, como es el caso de la dependencia.

En este punto, ya contábamos con la *arquitectura de microservicios*, las aplicaciones de recopilación de datos y un *pipeline* o flujo de ML listo para reutilizarse en otros casos de uso. Así, el objetivo principal de esta publicación fue la implementación de un modelo de evaluación de la dependencia ubicuo, utilizando datos de dispositivos *wearables*. En este caso, además del reloj (con acelerómetro, giroscopio y ritmo cardíaco) se incluyó la pulsera Empatica E4 que ampliaba el número de sensores a explorar (con actividad electrodérmica de la piel —EDA— y temperatura). También se incluyeron variables observacionales recogidas en el mismo experimento de la *compra*, tanto de carácter motor, como cognitivas, sociales o del propio rendimiento en la actividad *comprar*. Algunos ejemplos son, si el sujeto utilizaba alguna ayuda para caminar (como un bastón), si necesitaba ayuda para localizar el producto, si preguntaba dónde localizar el producto a algún dependiente o cliente del supermercado, etc. Con todo ello, se programaron *scripts* para federar tanto los datos de los distintos sensores de *wearables* como estos datos observacionales, constituyendo una base de datos heterogénea. Además, se etiquetaron los datos de cada participante según los resultados del cuestionario de Lawton y Brody en *dependientes* y *no dependientes* [100]. Asimismo, se realizaron varios experimentos con distintos algoritmos de ML y se buscaron diferentes combinaciones de variables, tanto incluyendo como excluyendo las observacionales, utilizando algoritmos de selección de características como RFE para reducir la gran dimensionalidad de los datos federados y hallar las más significativas para nuestro caso de uso.

Los resultados de este trabajo reportaron un modelo muy competitivo, con un 97% de precisión en la evaluación de la dependencia, considerando sólo 10 variables provenientes de un único dispositivo *wearable* y sin incluir ninguna observacional. Esas variables incluían medidas estadísticas de los sensores acelerómetro, ritmo cardíaco, temperatura y EDA de la pulsera Empatica. Al igual que los resultados de la publicación sobre fragilidad, presentada en la sección anterior, estos resultados suponen una evaluación de la dependencia en personas mayores automática y ubicua, mediante la monitorización de datos *wearables* mientras realizan una actividad compleja como es ir a *comprar*. Desplegar este modelo basado en ML, puede contribuir a una detección precoz del estado de dependencia, tan importante para nuestras sociedades, promoción del envejecimiento activo y reducción de tiempos y costes.

5.3. Reducing Response Time in Motor Imagery Using A Headband and Deep Learning

La última publicación, que forma parte de esta tesis, tiene que ver con el estudio de la dimensión mental humana mediante las señales EEG. Como hemos explicado, no solo es importante el estudio de la dimensión física de las personas para promover un envejecimiento activo y saludable, sino que también es necesario analizar la dimensión mental en el desempeño de las actividades de la vida diaria de las personas mayores. Y en este sentido, este trabajo es un primer esfuerzo por iniciar esta labor y se basa en el estudio de las señales EEG para detectar la intención motora de movimiento o *motor imagery*. Sin embargo, debido a la pandemia de la COVD-19 no fue posible el reclutamiento de personas mayores para poder validar el modelo en dicha población, por lo que contamos con población más joven.

En este trabajo logramos la detección de la intención de movimiento (mover la mano izquierda o derecha), imaginándolo mentalmente, y analizando las señales EEG que provee el dispositivo *wearable* Muse, en su versión 2. Se usaron los datos que provee el dispositivo de las diferentes ondas α, β, δ, γ, θ en cada uno de sus cuatro canales TP9, TP10, AF7 y AF8; conformando un total de 20 características. El conjunto de datos constó de 60 grabaciones de EEG a 256Hz. La arquitectura de la red neuronal que se diseñó consistió en: una capa 1D-CNN para extraer las características más relevantes; y una LSTM para realizar la clasificación de intención motora basada, en las grabaciones recibidas en forma de serie temporal. Esta detección la logramos reducir a solo 2 segundos, superando los trabajos previos que utilizaban *wearables* [68]–[70].

Además, se ha aprendido a trabajar con las señales EEG, preprocesarlas, visualizarlas, extraer características, etc., con objeto de poder aplicarlas a estudios cognitivos (en lugar de intención motora). Estos resultados tienen el potencial de incrementar la autonomía de las personas, utilizando sólo su pensamiento y una interfaz de interacción-persona-ordenador que los traduzca en acciones. Por lo tanto, seguir explorando esta línea resultaría muy útil para ampliar nuestro sistema *m-Health* holístico, hacia la dimensión cognitiva. Adicionalmente, también tiene interés mencionar la utilidad que puede tener el EGG para la detección de emociones y problemas de salud emocional, como el estrés, la depresión y la ansiedad.

6. Conclusiones

En la presente tesis se ha presentado una propuesta tecnológica *m-Health*, con el objetivo de proveer un sistema holístico y ubicuo que permita evaluar el deterioro del estado de salud de las personas a medida que envejecen, el cual se entiende con carácter multidimensional, y no sólo a nivel físico, sino también mental, social, etc. El sistema presenta un diseño arquitectónico de *microservicios* que incluye recopilación de datos sensoriales mediante dispositivos *wearables* (ubicuidad), la monitorización de la salud durante la realización de actividades de la vida diaria y la creación de modelos de aprendizaje automático —a partir de esos datos—. Con esta propuesta, se abordan los diferentes problemas que presentan las evaluaciones tradicionales de fragilidad y dependencia, las cuales dificultan una detección precoz de estos deterioros del estado de salud.

El objetivo inicial fue diseñar un sistema *m-Health* para que fuera flexible, reutilizable y extensible, por lo que se decidió usar una arquitectura basada en *microservicios*, especializados en los procesos de recopilación de datos de dispositivos *wearables*, procesamiento y análisis. Se ha propuesto esta arquitectura teniendo muy presente la granularidad de dichos *microservicios* siguiendo el principio de responsabilidad única, lo que contribuye al principio de *acoplamiento débil*. En este sentido, se puso el foco en tomar las ideas del MLOps para diseñar los *microservicios* del *back-end* del *m-Health* como especializaciones de cada una de las tareas del *pipeline* o flujo de ML, permitiendo su reutilización en diferentes casos de estudio, concretamente en fragilidad, dependencia y futuras extensiones.

El segundo objetivo consistía en diseñar un sistema de recopilación y federación de datos heterogéneos, principalmente provenientes de distintos sensores presentes en los dispositivos *wearables*, pero también de otras fuentes observacionales como resultados de cuestionarios, informes, etc. Se realizó un estudio para recoger datos de la actividad (instrumental) de la vida diaria *hacer la compra* (en un entorno real) y se consideraron varias etapas en esta actividad, las cuales se etiquetaron. En este sentido, se diseñó e implementó un flujo de trabajo de ML donde la primera etapa consistió en la recopilación, federación, preprocesamiento y etiquetado de los datos. En esta recopilación, se utilizó el software *Tizensor* desarrollado en este trabajo para recopilar datos de los sensores del reloj Samsung Gear S3, con Tizen, y se etiquetaron las distintas etapas de la *compra*. Así, los datos de sensores de estas señales heterogéneas se recopilaron y almacenaron y, posteriormente, se alinearon a una misma frecuencia de muestreo. Además, estos datos de sensores se federaron también tanto con otro *wearable* —Empatica E4— como con datos observacionales, resultados de cuestionarios y otras pruebas.

El tercer objetivo fue generar un modelo de evaluación del estado de fragilidad y dependencia mediante técnicas de aprendizaje automático y, a ser posible, con el menor número de variables independientes posible. Se logró el desarrollo de un modelo de evaluación de la fragilidad con un único dispositivo *wearable* —un reloj con sistema operativo Tizen—. Para llegar a éste, se realizó un análisis en profundidad de los datos de diferentes sensores y dispositivos hasta comprobar que sólo con el reloj era suficiente para nuestro caso de estudio, caracterizándose el modelo generado por ser preciso, objetivo, holístico y ubicuo. Siguiendo los pasos exitosos en el caso de estudio previo sobre la fragilidad, se volvieron a analizar en profundidad los diferentes datos disponibles, aunque esta vez se incluyeron en los análisis variables observacionales. Mediante la aplicación de técnicas de aprendizaje automático y selección de características, se obtuvo un modelo de evaluación de dependencia preciso, también con un solo dispositivo y finalmente sin presencia de variables observacionales.

Por último, el cuarto objetivo fue estudiar otros factores que influyen en el envejecimiento activo, relacionados con la función humana cognitiva y mental. Se realizó un análisis de las señales EEG que genera el dispositivo *wearable* Muse (versión 2), el cual consiste en una diadema inalámbrica que podemos llevar puesta sobre la frente. Las señales EEG pueden utilizarse para evaluar la actividad cognitiva, pero debido a la situación acontecida de la pandemia COVID-19, que impidió la participación de mayores, el estudio se centró en una primera aproximación a recopilar, filtrar, analizar (trabajar con) las señales EEG y se empezó por una detección de intención de movimiento con voluntarios más jóvenes. En esta línea, se generó un modelo de detección preciso sobre la intención motora, mediante la aplicación de técnicas de aprendizaje automático nuevamente, el cual supuso una detección de intención de movimiento de la mano derecha o la izquierda en una ventana de solo 2 segundos.

6.1. Conclusions

In this thesis, an m-Health technological proposal has been presented, with the aim of providing a holistic and ubiquitous system to assess the deterioration of the health status of people as they age, which is understood as multidimensional human functioning and not only physical, but also mental, social, etc. The system presents a design based on microservices architecture, which includes sensory data collection through wearable devices (ubiquity), health monitoring during the performance of activities of daily living and the creation of machine learning models from that data. This proposal addresses the different problems presented by traditional frailty and dependency assessments, which hinder early detection of these deteriorations in health status.

The initial objective was to design an m-Health system to be flexible, reusable and extensible, so it was decided to use a software architecture based on microservices, where each component is specialized in a single m-Health task, such as data collection from wearable devices, processing and analysis. This architecture has been proposed keeping in mind the granularity of these microservices following the single responsibility principle, which contributes to the principle of loose coupling. In this sense, the focus was put on taking the ideas of MLOps to design the m-Health back-end microservices as specializations of each of the tasks of the ML pipeline, allowing their reuse in different case studies, specifically in fragility, dependency and future extensions.

The second objective was to design a system for collecting and federating heterogeneous data, mainly from different sensors built-in wearable devices, but also from other observational sources such as results from questionnaires, reports, etc. A study was conducted to collect data on the (instrumental) activity of daily life shopping (in a real environment) and several stages in this activity were considered, which were labeled. In this sense, a ML pipeline was designed and implemented where the first stage is the data collection, federation, preprocessing and labeling. The Tizensor software—developed in this work—was used to collect sensor data from the Samsung Gear S3 smartwatch sensors, running Tizen, and the different stages of the purchase were labeled. Thus, sensor data from these heterogeneous signals were collected, stored and subsequently aligned to the same sampling rate. In addition, this sensor data was also federated both with another wearable—Empatica E4—and with observational data, questionnaire results and other tests.

The third objective was to generate a frailty and dependency status assessment model using machine learning techniques and, if possible, with as few independent variables as possible. The development of a frailty assessment model was achieved with a single wearable device—a smartwatch with Tizen operating system—. To achieve this model, an in-depth analysis of data from different sensors and devices was carried out, which reported that the use of only one smartwatch was sufficient for our case study, characterizing the model generated as accurate, objective, holistic and ubiquitous. Following the successful steps in the previous case study on frailty, the different available data were again analyzed in depth, although this time observational variables were included in the analyses. By applying machine learning and feature selection techniques, an accurate dependency assessment model was obtained, also with a single device and finally without the presence of any observational variables.

Finally, the fourth objective was to study other factors that influence active aging, related to human cognitive and mental function. An analysis of the EEG signals generated by the Muse wearable device (version 2), which consists of a wireless headband that can be worn on the forehead, was performed. EEG signals can be used to assess cognitive activity, but due to the situation that occurred during the COVID-19 pandemic, which prevented the participation of older people, the study focused on a first approach to collect, filter, analyze (work with) EEG signals and started with a motor imagery detection with younger volunteers. Along these research lines, an accurate detection model on motor intention was generated by applying machine learning techniques again, which involved a detection of movement intention of the right or left hand in a window of only 2 seconds.

7. Trabajo futuro / Future work

De las conclusiones extraídas de esta tesis, se pueden proponer interesantes y prometedoras nuevas líneas de investigación. Entre ellas, destacamos las siguientes:

Modelos de evaluación multidimensionales. Sabemos que los estados de fragilidad y dependencia deben entenderse de forma holística y multidimensional. Aunque, a priori se pueda pensar que los dispositivos *wearables* usados en los modelos generados en esta tesis —tanto de fragilidad como de dependencia—, se corresponden con señales alineadas con la dimensión física humana, estos modelos se han generado en un contexto multidimensional basado en la monitorización de la actividad de la *compra*, la cual implica que el sujeto de estudio se involucre en una serie de procesos no solo físicos, sino también cognitivo y sociales. Además, también hemos usado sensores como el EDA, el cual está correlacionado con la detección de cuestiones emocionales, como el estrés. Es por ello que, de forma latente, los modelos generados están teniendo en cuenta esta multidimensionalidad, de hecho, observábamos cómo algunas etapas de la *compra* —como buscar el producto— eran más decisivas que otras en la evaluación final de fragilidad. No obstante, debe estudiarse más a fondo esta multidimensionalidad —física, cognitiva, social e incluso emocional— introduciendo nuevos dispositivos *wearables* como el que utilizamos aquí también, como son las diademas EEG. Estos dispositivos, que pueden medir la actividad cerebral —como hemos comprobado en este trabajo— podrían arrojar más luz al carácter multidimensional de la fragilidad y la dependencia. Asimismo, explorar el potencial de otros dispositivos EEG como los auriculares de *Neurable*, también puede ser interesante desde el punto de vista de la ergonomía, ya que hemos observado que las diademas EEG sufren ciertos problemas de desconexión en movimiento.

Generación de datos sintéticos. En el campo de la salud, es muy habitual encontrarse con la dificultad en la recopilación de datos de las personas para investigación. Normalmente, nos encontramos con conjuntos de datos de pocas muestras (pocos sujetos) y, sin embargo, una alta dimensionalidad (grandes cantidades de medidas de cada sujeto). Esto se puede deber a varios factores, pero, especialmente, uno de éstos tiene que ver con cuestiones de privacidad y seguridad. Además de esto, hay que sumarle la presencia en los datos de salud de valores perdidos o ruidosos. Esto último también es habitual en los datos de las señales de los sensores *wearables*, debido a posibles desconexiones en las comunicaciones, agotamiento de la batería, etc. Abordar el problema de los valores perdidos y aumentar el conjunto de datos es esencial para mejorar las predicciones con ML, puesto que estos algoritmos suelen necesitar conjuntos de datos de gran tamaño. Asimismo, también es importante generar datos sintéticos porque ayudarían a preservar la privacidad, sustituyendo a datos de pacientes reales. Recientemente, y concretamente este año 2022 podríamos considerar que ha sido la explosión de los modelos generativos de datos. Ejemplos como las redes generativas adversarias (GAN) o los modelos de difusión, están teniendo gran repercusión en la generación de datos sintéticos, especialmente, generación de imágenes, texto, audio y recientemente incluso vídeos. En este sentido, se abre una nueva línea de investigación que centre sus esfuerzos en explorar este tipo de algoritmos en la generación sintética de los datos de carácter temporal de sensores, que hemos venido utilizando en la presente tesis para el ámbito sanitario. Esto podrá ayudar a lidiar con los problemas mencionados de acceso a los datos sanitarios por su privacidad o imputación de valores perdidos.

Tecnología *Blockchain*. La tecnología *Blockchain* se caracteriza por promover la descentralización, persistencia, anonimato y auditabilidad en las transacciones de datos, lo cual podría suponer mejorar en gran medida el proceso de gestión de los datos sanitarios [103]. Aunque en sus inicios se empleó exclusivamente en el mundo de las criptomonedas, ya se empieza a usar en otros dominios como el IoT. En un futuro, la combinación de los sistemas *m-Health* con la disruptiva tecnología *Blockchain*, podría promover el intercambio de datos sanitarios —centrado en el paciente— de forma más ubicua y privada que la actual Es por ello que se abre un nuevo camino para explorar el diseño de los sistemas *m-Health* basados en la *Blockchain*, que aseguren la anonimización en los intercambios de los datos de los pacientes, los cuales proveerán cada vez más datos mediante sus dispositivos *wearables* y móviles.

Identificación de emociones. Mencionar, además, mi interés en identificar emociones o estado de salud mental, utilizando dispositivos *wearables* como las diademas EEG u otros dispositivos. La reutilización de la *arquitectura de microservicios*, gracias a su diseño adaptativo y granular, será esencial para su aplicabilidad a otros problemas de salud y a este, en particular, de la dimensión cognitiva. Asimismo, permitiría profundizar en la creación de modelos de ML específicos de análisis y detección de esos problemas de salud y emocionales, trazando un puente hacia la computación afectiva en este último caso.

7.1. Future work

From the conclusions drawn from this thesis, interesting and promising new lines of research can be proposed. Among them, we highlight the following:

Multidimensional assessment models. We know that states of frailty and dependence must be understood in a holistic and multidimensional way. A priori, it may be thought that the wearable devices used in the models generated in this thesis—both of frailty and dependence—, correspond to signals aligned with the human physical dimension. However, these models have been generated in a multidimensional context based on the monitoring of the shopping activity, which implies that the subject of study is involved in a series of processes not only physical, but also cognitive and social. In addition, we have also used sensors such as the EDA, which is correlated with the detection of some aspects related to emotional health, such as stress. That is why, in a latent way, the models generated are considering this multidimensionality, in fact, we observed how some stages of the purchase—such as looking for the product—were more decisive than others in the final assessment of frailty. However, this multidimensionality—physical, cognitive, social and even emotional—should be further studied by introducing new wearable devices such as the one we use here as well, e.g., EEG headsets. These devices, which can measure brain activity—as we have tested in this work—could shed lighter on the multidimensional nature of frailty and dependence. Also, exploring the potential of other EEG devices, such as Neurable headphones, may also be interesting from an ergonomics point of view, as we have observed that EEG headsets suffer from certain disconnection problems in motion.

Synthetic data generation. In the health field, it is very common to find difficulty in collecting data from people for research. Usually, although we have high dimensionality datasets (large numbers of measurements of each subject), we have small datasets with few samples (few subjects). This may be due to several factors, but, especially, one of these has related to privacy and security issues. In addition to this, the presence of missing or noisy values is common in health dataset. The latter is also common in wearable sensor signal data, due to possible communication disconnections, battery depletion, etc. Addressing the problem of missing values and increasing the dataset is essential to improve ML predictions, as these algorithms often require large datasets. Likewise, it is also important to generate synthetic data because it would help to preserve privacy by replacing real patient data. Recently, and specifically this year 2022 could be considered the explosion of generative data models. Examples such as generative adversarial networks (GAN) or diffusion models, are having a great impact on the generation of synthetic data, especially, generation of images, text, audio and recently even videos. In this sense, a new line of research is opened that focuses its efforts on exploring this type of algorithms in the synthetic generation of temporal sensor data, which we have been using in this thesis for the health field. This may help to deal with the aforementioned problems of access to health data due to their privacy or imputation of missing values.

Blockchain technology. Blockchain technology is characterized by promoting decentralization, persistence, anonymity, and auditability in data transactions, which could mean greatly improving the healthcare data management process [103]. Although it was initially used exclusively in the world of cryptocurrencies, it is already starting to be used in other domains such as IoT. In the future, the combination of m-Health systems with the disruptive Blockchain technology, could promote the exchange of healthcare data - patient-centric - in a more ubiquitous and private way than the current one This is why a new path is opening up to explore the design of m-Health systems based on the Blockchain, which ensure anonymization in the exchanges of patient data, which will increasingly provide data through their wearable and mobile devices.

Emotion Recognition. Mentioning my interest in identifying emotions or state of mental health, using wearable devices such as EEG headsets or other devices. The reusability of the microservices architecture, thanks to its adaptive and fine-grained design, will be essential for its applicability to other health problems, such as related problems to the cognitive dimension. It would also allow to deepen the creation of specific ML models for analysis and detection of these health and emotional problems, building a bridge to affective computing in the latter case.

Bibliografía

[1] S. H. Almotiri, M. A. Khan, and M. A. Alghamdi, "Mobile health (m-Health) system in the context of IoT," in *Proceedings - 2016 4th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2016*, Oct. 2016, pp. 39–42. doi: 10.1109/W-FiCloud.2016.24.

[2] G. Sannino, G. de Pietro, and L. Verde, "Healthcare systems: An overview of the most important aspects of current and future M-health applications," in *Connected Health in Smart Cities*, Springer International Publishing, 2019, pp. 213–231. doi: 10.1007/978-3-030-27844-1_11.

[3] E. M. Almahdi, A. A. Zaidan, B. B. Zaidan, M. A. Alsalem, O. S. Albahri, and A. S. Albahri, "Mobile-Based Patient Monitoring Systems: A Prioritisation Framework Using Multi-Criteria Decision-Making Techniques," *J Med Syst*, vol. 43, no. 7, p. 219, Jul. 2019, doi: 10.1007/s10916-019-1339-9.

[4] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body Area Networks: A Survey," *Mobile Networks and Applications*, vol. 16, no. 2, pp. 171–193, Apr. 2011, doi: 10.1007/s11036-010-0260-8.

[5] Research and Markets, "Wearable Technology Market Size, Share & Trends Analysis Report By Product (Wrist-Wear, Eye-Wear & Head-Wear, Foot-Wear, Neck-Wear, Body-wear), By Application, By Region, and Segment Forecasts, 2021-2028," 2021. https://www.researchandmarkets.com/reports/5124989/wearable-technology-market-size-share-and-trends?utm_source=BW&utm_medium=PressRelease&utm_code=wgq329&utm_campaign=1640832+-+Global+Wearable+Technology+Market+Trends+%26+Analysis+Report+2021-2028%3a+Adoption+of+Fitness+Trackers+and+Health-based+Wearables+is+Anticipated+to+Propel+Growth&utm_exec=chdo54prd (accessed Jun. 19, 2022).

[6] Yahoo, "Wearable Electronics Global Market 2022 Report: Wristwear Segment to Grow at a CAGR of 17.5% from 2021 to 2030," 2022. https://finance.yahoo.com/news/wearable-electronics-global-market-2022-102800051.html?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAADagnt299rUwjs_Me1wYEpN0bhNCzrFNmzW5lTD50A3b8GAMFy9nSnvOebUeaji1-qCPL-DRYPUsJlW0rbNSJh3GlUVF244wfEPoU2aVrMifNvzZF3mTaGaY_H7PDUBSQR9LMJGp3eUE3THv6Wx1ROGUrKYM2XiitcKQy1OxC4LL&guccounter=2 (accessed Jun. 19, 2022).

[7] S. Li, L. da Xu, and S. Zhao, "The internet of things: a survey," *Information Systems Frontiers*, vol. 17, no. 2, pp. 243–259, Apr. 2015, doi: 10.1007/s10796-014-9492-7.

[8] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010, doi: 10.1016/j.comnet.2010.05.010.

[9] P. J. Mccullagh and J. C. Augusto, "The internet of Things: The potential to facilitate health and wellness," *UPGRADE: The European Journal for the Informatics Professional*, vol. XII, no. 1, pp. 59–68, 2011.

[10] P. Zadtootaghaj, A. Mohammadian, B. Mahbanooei, and R. Ghasemi, "Internet of Things: A Survey for the Individuals' E-Health Applications," *Journal of Information Technology Management*, vol. 11, no. 1, 2019, doi: 10.22059/jitm.2019.288695.2398.

[11] C. C. Aggarwal, N. Ashish, and A. Sheth, "The Internet of Things: A Survey from the Data-Centric Perspective," in *Managing and Mining Sensor Data*, Boston, MA: Springer US, 2013, pp. 383–428. doi: 10.1007/978-1-4614-6309-2_12.

[12] M. Alam, J. Rufino, J. Ferreira, S. H. Ahmed, N. Shah, and Y. Chen, "Orchestration of Microservices for IoT Using Docker and Edge Computing," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 118–123, Sep. 2018, doi: 10.1109/MCOM.2018.1701233.

[13] L. Sun, Y. Li, and R. A. Memon, "An open IoT framework based on microservices architecture," *China Communications*, vol. 14, no. 2, pp. 154–162, Feb. 2017, doi: 10.1109/CC.2017.7868163.

[14] S. Newman, *Building Microservices*, First Edit., vol. 52, no. 15. Sebastopol: O'Reilly Media, 2015.

[15] I. Nadareishvili, R. Mitra, M. McLarty, and M. Amundsen, *Microservice Architecture: Aligning Principles, Practices, and Culture*, First Edition. Sebastopol: O'Reilly Media, 2016.

[16] S. Laso, J. Berrocal, J. Garcia-Alonso, C. Canal, and J. M. Murillo, "Service Oriented Computing for Humans as Service Providers," 2021, pp. 111–122. doi: 10.1007/978-3-030-73203-5_9.

[17] S. Laso, J. Berrocal, J. García-Alonso, C. Canal, and J. Manuel Murillo, "Human microservices: A framework for turning humans into service providers," *Softw Pract Exp*, vol. 51, no. 9, pp. 1910–1935, Sep. 2021, doi: 10.1002/spe.2976.

[18] M. K. Gourisaria, R. Agrawal, G. Harshvardhan, M. Pandey, and S. S. Rautaray, "Application of Machine Learning in Industry 4.0," 2021, pp. 57–87. doi: 10.1007/978-981-33-6518-6_4.

[19] A. D. las Heras, A. Luque-Sendra, and F. Zamora-Polo, "Machine learning technologies for sustainability in smart cities in the post-covid era," *Sustainability (Switzerland)*, vol. 12, no. 22, pp. 1–25, Nov. 2020, doi: 10.3390/su12229320.

[20] T. Tegou, I. Kalamaras, M. Tsipouras, N. Giannakeas, K. Votis, and D. Tzovaras, "A low-cost indoor activity monitoring system for detecting frailty in older adults," *Sensors (Switzerland)*, vol. 19, no. 3, Feb. 2019, doi: 10.3390/s19030452.

[21] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and Robust Machine Learning for Healthcare: A Survey," *IEEE Rev Biomed Eng*, vol. 14, pp. 156–180, 2021, doi: 10.1109/RBME.2020.3013489.

[22] Rob van der Meulen and Thomas McCall, "Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence," 2018. ttps://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence (accessed Jun. 20, 2022).

[23] A. Posoldova, "Machine Learning Pipelines: From Research to Production," *IEEE Potentials*, vol. 39, no. 6, pp. 38–42, Nov. 2020, doi: 10.1109/MPOT.2020.3016280.

[24] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine Learning Operations (MLOps): Overview, Definition, and Architecture," May 2022, [Online]. Available: http://arxiv.org/abs/2205.02302

[25] L. Carretero, E. Navarro-Pardo, and A. Cano, "Progression in healthy ageing: frailty, cognitive decline and gender in the European Innovation Partnership for Active and Healthy Ageing," *Eur J Psychiatry*, vol. 29, no. 4, pp. 231–237, 2015, doi: 10.4321/S0213-61632015000400001.

[26] V. Espín, M. V. Hurtado, and M. Noguera, "Towards Holistic Support of Active Aging through Cognitive Stimulation, Exercise and Assisted Nutrition," in *Lecture Notes in Computer Science*, vol. 8868, 2014, pp. 312–319. doi: 10.1007/978-3-319-13105-4_45.

[27] European Commission, "Employment, Social Affairs & Inclusion: Active ageing." https://ec.europa.eu/social/main.jsp?langId=en&catId=1062 (accessed Jun. 23, 2022).

[28] C. O. Acosta Quiroz, J. J. Vales García, and R. R. Palacio Cinco, "Ajuste psicosocial, bienestar subjetivo y ocio en adultos mayores jubilados mexicanos," *Enseñanza e Investigación en Psicología*, vol. 20, no. 3, pp. 316–325, 2015.

[29] World Health Organization, "Ageing and health," 2021. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed May 25, 2022).

[30] G. Onder and D. Fialová, "Ageism In The Health Care System," *Innov Aging*, vol. 1, no. suppl_1, pp. 1072–1072, Jul. 2017, doi: 10.1093/geroni/igx004.3927.

[31] M. Jakovljevic, K. M. Janicijevic, and M. Stepovic, "Book Review: The New Public Health 3rd Edition," *Front Public Health*, vol. 6, Sep. 2018, doi: 10.3389/fpubh.2018.00265.

[32] J. Dieleman *et al.*, "Evolution and patterns of global health financing 1995-2014: Development assistance for health, and government, prepaid private, and out-of-pocket health spending in 184 countries," *The Lancet*, vol. 389, no. 10083, pp. 1981–2004, May 2017, doi: 10.1016/S0140-6736(17)30874-7.

[33] M. Changizi and M. H. Kaveh, "Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population—a systematic review," *Mhealth*, vol. 3, pp. 51–51, Nov. 2017, doi: 10.21037/mhealth.2017.08.06.

[34] F. P. Tajudeen, N. Bahar, T. Maw Pin, and N. I. Saedon, "Mobile Technologies and Healthy Ageing: A Bibliometric Analysis on Publication Trends and Knowledge Structure of mHealth

Research for Older Adults," *Int J Hum Comput Interact*, vol. 38, no. 2, pp. 118–130, Jan. 2022, doi: 10.1080/10447318.2021.1926115.

[35] N. P. Rocha, M. Rodrigues dos Santos, M. Cerqueira, and A. Queirós, "Mobile Health to Support Ageing in Place," *International Journal of E-Health and Medical Communications*, vol. 10, no. 3, pp. 1–21, Jul. 2019, doi: 10.4018/IJEHMC.2019070101.

[36] R. J. Gobbens, K. G. Luijkx, M. T. Wijnen-Sponselee, and J. M. Schols, "Toward a conceptual definition of frail community dwelling older people," *Nurs Outlook*, vol. 58, no. 2, pp. 76–86, Mar. 2010, doi: 10.1016/j.outlook.2009.09.005.

[37] L. P. Fried *et al.*, "Frailty in Older Adults: Evidence for a Phenotype," *Journal of Gerontology: MEDICAL SCIENCES America*, vol. 56, no. 3, pp. 146–156, 2001.

[38] J. Devi, "The scales of functional assessment of Activities of Daily Living in geriatrics," *Age Ageing*, vol. 47, no. 4, pp. 500–502, Jul. 2018, doi: 10.1093/ageing/afy050.

[39] Barthel, "Escalas de valoración funcional, psicoafectiva y sociofamiliar".

[40] M. J. Cabañero-Martínez, J. Cabrero-García, M. Richart-Martínez, and C. L. Muñoz-Mendoza, "The Spanish versions of the Barthel index (BI) and the Katz index (KI) of activities of daily living (ADL): A structured review," *Arch Gerontol Geriatr*, vol. 49, no. 1, pp. e77–e84, Jul. 2009, doi: 10.1016/j.archger.2008.09.006.

[41] F. Alonso-Trujillo *et al.*, "III Informe de Progreso y Gestión. I Bienal 2017-2018 y II BIENAL 2019. I Plan Andaluz de Promoción de la Autonomía Personal y Prevención de la Dependencia (2016-2020)," Sevilla, 2020.

[42] F. J. Gutierrez and S. F. Ochoa, "Making visible the invisible: understanding the nuances of computer-supported cooperative work on informal elderly caregiving in Southern Cone families," *Pers Ubiquitous Comput*, vol. 25, no. 2, pp. 437–456, Apr. 2021, doi: 10.1007/s00779-020-01404-4.

[43] J. Garcia-Alonso, J. Berrocal, J. M. Murillo, D. Mendes, C. Fonseca, and M. Lopes, "Situational-Context for Virtually Modeling the Elderly," 2019, pp. 298–305. doi: 10.1007/978-3-030-01746-0_35.

[44] S. M. M. Ali, J. C. Augusto, and D. Windridge, "A Survey of User-Centred Approaches for Smart Home Transfer Learning and New User Home Automation Adaptation," *Applied Artificial Intelligence*, vol. 33, no. 8, pp. 747–774, Jul. 2019, doi: 10.1080/08839514.2019.1603784.

[45] J. M. Pérez Mármol, M. L. Flores Antigüedad, A. M. Castro Sánchez, R. M. Tapia Haro, M. del C. García Ríos, and M. E. Aguilar Ferrándiz, "Inpatient dependency in activities of daily living predicts informal caregiver strain: A cross-sectional study," *J Clin Nurs*, vol. 27, no. 1–2, pp. e177–e185, Jan. 2018, doi: 10.1111/jocn.13900.

[46] C. O. Quiroz Acosta and A. L. M. González-Celis Rangel, "Actividades de la vida diaria en adultos mayores: la experiencia de dos grupos focales," *Psicología y Salud*, vol. 19, no. 2, pp. 289–293, 2009.

[47] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the Empatica E4 wristband," in *2016 IEEE EMBS International Student Conference (ISC)*, May 2016, pp. 1–4. doi: 10.1109/EMBSISC.2016.7508621.

[48] M. Ragot, N. Martin, S. Em, N. Pallamin, and J.-M. Diverrez, "Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors," 2018, pp. 15–22. doi: 10.1007/978-3-319-60639-2_2.

[49] H. Rahemi, H. Nguyen, H. Lee, and B. Najafi, "Toward Smart Footwear to Track Frailty Phenotypes—Using Propulsion Performance to Determine Frailty," *Sensors*, vol. 18, no. 6, p. 1763, Jun. 2018, doi: 10.3390/s18061763.

[50] H. Lee, B. Joseph, A. Enriquez, and B. Najafi, "Toward Using a Smartwatch to Monitor Frailty in a Hospital Setting: Using a Single Wrist-Wearable Sensor to Assess Frailty in Bedbound Inpatients," *Gerontology*, vol. 64, no. 4, pp. 389–400, 2018, doi: 10.1159/000484241.

[51] M. Bernardo-Filho *et al.*, "Effects of Remotely Supervised Physical Activity on Health Profile in Frail Older Adults: A Randomized Controlled Trial Protocol. Front," *Aging Neurosci*, vol. 14, p. 807082, 2022, doi: 10.3389/fnagi.2022.807082.

[52] G. Sacco *et al.*, "Detection of activities of daily living impairment in Alzheimer's disease and mild cognitive impairment using information and communication technology," *Clin Interv Aging*, p. 539, Dec. 2012, doi: 10.2147/CIA.S36297.

[53] A. König *et al.*, "Ecological Assessment of Autonomy in Instrumental Activities of Daily Living in Dementia Patients by the Means of an Automatic Video Monitoring System," *Front Aging Neurosci*, vol. 7, Jun. 2015, doi: 10.3389/fnagi.2015.00098.

[54] F. Anabitarte-García *et al.*, "Early diagnosis of frailty: Technological and non-intrusive devices for clinical detection," *Ageing Res Rev*, vol. 70, p. 101399, Sep. 2021, doi: 10.1016/j.arr.2021.101399.

[55] G. Vavasour, O. M. Giggins, J. Doyle, and D. Kelly, "How wearable sensors have been utilised to evaluate frailty in older adults: a systematic review," *J Neuroeng Rehabil*, vol. 18, no. 1, p. 112, Dec. 2021, doi: 10.1186/s12984-021-00909-0.

[56] A. Gaete, F. J. Gutierrez, S. F. Ochoa, P. Guerrero, and A. Wyzykowski, "Visitrack: A Pervasive Service for Monitoring the Social Activity of Older Adults Living at Home," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10586 LNCS, Springer Verlag, 2017, pp. 520–530. doi: 10.1007/978-3-319-67585-5_52.

[57] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, "EEG-based Intention Recognition from Spatio-Temporal Representations via Cascade and Parallel Convolutional Recurrent Neural Networks," Aug. 2017, Accessed: Jul. 16, 2020. [Online]. Available: www.aaai.org

[58] S. Bhattacharyya, A. Khasnobish, A. Konar, D. N. Tibarewala, and A. K. Nagar, "Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms," in

2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), Apr. 2011, pp. 1–8. doi: 10.1109/CCMB.2011.5952111.

[59] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik (Stuttg)*, vol. 130, pp. 11–18, Feb. 2017, doi: 10.1016/j.ijleo.2016.10.117.

[60] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J Neural Eng*, vol. 14, no. 1, p. 016003, Feb. 2017, doi: 10.1088/1741-2560/14/1/016003.

[61] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-Based EEG Classification in Motor Imagery Tasks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018, doi: 10.1109/TNSRE.2018.2876129.

[62] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum Brain Mapp*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017, doi: 10.1002/hbm.23730.

[63] K.-W. Ha and J.-W. Jeong, "Motor Imagery EEG Classification Using Capsule Networks," *Sensors*, vol. 19, no. 13, p. 2854, Jun. 2019, doi: 10.3390/s19132854.

[64] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *J Neural Eng*, vol. 15, no. 5, p. 056013, Oct. 2018, doi: 10.1088/1741-2552/aace8c.

[65] B. Blankertz *et al.*, "The Berlin Brain–Computer Interface: EEG-Based Communication Without Subject Training," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 147–152, Jun. 2006, doi: 10.1109/TNSRE.2006.875557.

[66] R. Tomioka, K. Aihara, and K.-R. Müller, "Logistic Regression for Single Trial EEG Classification," in *Twentieth Annual Conference on Neural Information Processing Systems*, Dec. 2006, vol. 19, pp. 1377–1384.

[67] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, Jan. 2012, doi: 10.3390/s120201211.

[68] F. M. Garcia-Moreno, M. Bermudez-Edo, M. J. Rodriguez-Fortiz, and J. L. Garrido, "A CNN-LSTM Deep Learning Classifier for Motor Imagery EEG Detection Using a Low-invasive and Low-Cost BCI Headband," in *2020 16th International Conference on Intelligent Environments (IE)*, Jul. 2020, pp. 84–91. doi: 10.1109/IE49459.2020.9155016.

[69] Z. Li, J. Xu, and T. Zhu, "Recognition of Brain Waves of Left and Right Hand Movement Imagery with Portable Electroencephalographs," Sep. 2015, Accessed: Jan. 27, 2020. [Online]. Available: http://arxiv.org/abs/1509.08257

[70] P. Rodriguez, A. O. Zezzatti, and J. Mejía, "Machine Learning Algorithms Based on the Classification of Motor Imagination Signals Acquired with an Electroencephalogram," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13067 LNAI, Springer Science and Business Media Deutschland GmbH, 2021, pp. 239–249. doi: 10.1007/978-3-030-89817-5_18.

[71] N. Dokovsky, A. van Halteren, and I. Widya, "BANip: Enabling Remote Healthcare Monitoring with Body Area Networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2952, 2004, pp. 62–72. doi: 10.1007/978-3-540-24639-8_6.

[72] M. S. Hajar, M. O. Al-Kadri, and H. K. Kalutarage, "A survey on wireless body area networks: architecture, security challenges and research opportunities," *Comput Secur*, vol. 104, p. 102211, May 2021, doi: 10.1016/j.cose.2021.102211.

[73] S. Ullah *et al.*, "A Comprehensive Survey of Wireless Body Area Networks," *J Med Syst*, vol. 36, no. 3, pp. 1065–1094, Jun. 2012, doi: 10.1007/s10916-010-9571-3.

[74] M. Richards, *Microservices vs. Service-Oriented Architecture*. Sebastopol, CA: O'Reilly Media, 2015.

[75] V. Raj and R. Sadam, "Evaluation of SOA-Based Web Services and Microservices Architecture Using Complexity Metrics," *SN Comput Sci*, vol. 2, no. 5, p. 374, Sep. 2021, doi: 10.1007/s42979-021-00767-6.

[76] V. Raj and S. Ravichandra, "Microservices: A perfect SOA based solution for Enterprise Applications compared to Web Services," in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, May 2018, pp. 1531–1536. doi: 10.1109/RTEICT42901.2018.9012140.

[77] P. Bhallamudi, S. Tilley, and A. Sinha, "Migrating a Web-based application to a service-based system - an experience report," in *2009 11th IEEE International Symposium on Web Systems Evolution*, Sep. 2009, pp. 71–74. doi: 10.1109/WSE.2009.5630392.

[78] A. Levcovitz, R. Terra, and M. T. Valente, "Towards a Technique for Extracting Microservices from Monolithic Enterprise Systems," May 2016, [Online]. Available: http://arxiv.org/abs/1605.03175

[79] T. Erl, *SOA Principles of Service Design*, 1st edition. Pentrice Hall, 2007. [Online]. Available: www.soamag.com

[80] S. Bhiri, W. Gaaloul, M. Rouached, and M. Hauswirth, "Semantic Web Services for Satisfying SOA Requirements," in *LNCS*, vol. 4891, 2008, pp. 374–395. doi: 10.1007/978-3-540-89784-2_15.

[81] B. Hutchison, M.-T. Schmidt, D. Wolfson, and M. Stockton, "SOA programming model for implementing Web services, Part 4: An introduction to the IBM Enterprise Service Bus," *IBM Developerworks: IBM*, 2005.

[82] M. Rahman and J. Gao, "A Reusable Automated Acceptance Testing Architecture for Microservices in Behavior-Driven Development," in *2015 IEEE Symposium on Service-Oriented System Engineering*, Mar. 2015, vol. 30, pp. 321–325. doi: 10.1109/SOSE.2015.55.

[83] "SOA vs. Microservicios: ¿Cuál es la diferencia? | IBM." https://www.ibm.com/cloud/blog/soa-vs-microservices (accessed Oct. 09, 2022).

[84] Xinyang Feng, Jianjing Shen, and Ying Fan, "REST: An alternative to RPC for Web services architecture," in *2009 First International Conference on Future Information Networks*, Oct. 2009, pp. 7–10. doi: 10.1109/ICFIN.2009.5339611.

[85] T. Cerny, M. J. Donahoo, and J. Pechanec, "Disambiguation and Comparison of SOA, Microservices and Self-Contained Systems," in *Proceedings of the International Conference on Research in Adaptive and Convergent Systems*, Sep. 2017, vol. 2017-January, pp. 228–235. doi: 10.1145/3129676.3129682.

[86] J. Bogner, J. Fritzsch, S. Wagner, and A. Zimmermann, "Limiting technical debt with maintainability assurance," in *Proceedings of the 2018 International Conference on Technical Debt*, May 2018, pp. 125–133. doi: 10.1145/3194164.3194166.

[87] J. Bogner, S. Wagner, and A. Zimmermann, "Using architectural modifiability tactics to examine evolution qualities of Service- and Microservice-Based Systems," *SICS Software-Intensive Cyber-Physical Systems*, vol. 34, no. 2–3, pp. 141–149, Jun. 2019, doi: 10.1007/s00450-019-00402-z.

[88] A. Balalaie, A. Heydarnoori, and P. Jamshidi, "Migrating to Cloud-Native Architectures Using Microservices: An Experience Report," vol. 567, A. Celesti and P. Leitner, Eds. Cham: Springer International Publishing, 2016, pp. 201–215. doi: 10.1007/978-3-319-33313-7_15.

[89] P. Leitão, A. W. Colombo, and S. Karnouskos, "Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges," *Comput Ind*, vol. 81, pp. 11–25, Sep. 2016, doi: 10.1016/j.compind.2015.08.004.

[90] L. Srinivasan and J. Treadwell, "An Overview of Service-oriented Architecture, Web Services and Grid Computing," *HP Software Global Business Unit*, 2005.

[91] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, "An Overview Of Machine Learning," in *Machine Learning: An Artificial Intelligence Approach*, Elsevier, 1983, pp. 3–23. doi: 10.1016/B978-0-08-051054-5.50005-4.

[92] T. M. Mitchell, *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.

[93] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine Learning: An Artificial Intelligence Approach*. Berlin, Heidelberg: Elsevier, 1983. doi: 10.1016/C2009-0-27563-5.

[94] C. Guo, M. Lu, and J. Chen, "An evaluation of time series summary statistics as features for clinical prediction tasks," *BMC Med Inform Decis Mak*, vol. 20, no. 1, p. 48, Dec. 2020, doi: 10.1186/s12911-020-1063-x.

[95] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical Human Activity Recognition Using Wearable Sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, Dec. 2015, doi: 10.3390/s151229858.

[96] I. Guyon and A. Elisseeff, "An Introduction of Variable and Feature Selection," *CrossRef Listing of Deleted DOIs*, vol. 1, pp. 1157–1182, 2000, doi: 10.1162/153244303322753616.

[97] I. Guyon, J. Weston, and S. Barnhill, "Gene Selection for Cancer Classification using Support Vector Machines," 2002. doi: 10.1023/A:1012487302797.

[98] N. Twomey *et al.*, "A Comprehensive Study of Activity Recognition Using Accelerometers," *Informatics*, vol. 5, no. 2, p. 27, May 2018, doi: 10.3390/informatics5020027.

[99] S. Patel *et al.*, "Using wearable sensors to predict the severity of symptoms and motor complications in late stage Parkinson's Disease," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, pp. 3686–3689. doi: 10.1109/IEMBS.2008.4650009.

[100] M. P. Lawton and E. M. Brody, "Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living," *Gerontologist*, vol. 9, no. 3 Part 1, pp. 179–186, Sep. 1969, doi: 10.1093/geront/9.3_Part_1.179.

[101] I. Vergara, A. Bilbao, M. Orive, S. Garcia-Gutierrez, G. Navarro, and J. Quintana, "Validation of the Spanish version of the Lawton IADL Scale for its application in elderly people," *Health Qual Life Outcomes*, vol. 10, no. 1, p. 130, 2012, doi: 10.1186/1477-7525-10-130.

[102] F. M. García-Moreno *et al.*, "Designing a Smart Mobile Health System for Ecological Frailty Assessment in Elderly," in *13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019*, Nov. 2019, vol. 31, no. 1, p. 41. doi: 10.3390/proceedings2019031041.

[103] M. V. Hurtado-Torres, F. L. Benítez-Martínez, and C. Rodríguez Domínguez, "Privacy and Securization of Patient Data in a Neural Blockchain: The Case of the Ehealth Phisicos Platform," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4133271.

II Publicaciones

1. Publicaciones

A continuación, se enumeran todas las disertaciones más relevantes realizadas durante el transcurso de la presente tesis doctoral, las cuales se presentarán a lo largo de este capítulo.

En particular, en [A] se propone el sistema *m-Health* basado en *microservicios* y se presenta el modelo de detección de fragilidad holístico y automático; así como en [B], se presenta el modelo de detección de dependencia. Y, por último, en [C] se presenta el modelo *motor imagery* en el estudio de señales EEG con *wearables*.

[A] A Microservices e-Health System for Ecological Frailty Assessment Using Wearables. F.M. García-Moreno, M. Bermúdez-Edo, J.L. Garrido, E. Rodríguez-García, J.M. Pérez-Mármol, M.J. Rodríguez-Fórtiz. *Sensors* 2020 vol: pp: 114-123. . [JCR 2020 Q1; IF 4,046]. DOI: 10.3390/s20123427. Nº de citas: 19 (Fuente, Google Scholar).

[B] A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors. F.M. García-Moreno, M. Bermúdez-Edo, E. Rodríguez-García, J.M. Pérez-Mármol, J.L. Garrido, M.J. Rodríguez-Fórtiz. *International Journal of Medical Informatics* 2022 vol: 157 pp: 104625. [JCR 2020 Q1; IF 4,046]. DOI: 10.1016/j.ijmedinf.2021.104625. Nº de citas: 5 (Fuente, Google Scholar).

[C] Reducing Response Time in Motor Imagery Using A Headband and Deep Learning. F.M. García-Moreno, M. Bermúdez-Edo, J.L. Garrido, M.J. Rodríguez-Fórtiz. *Sensors* 2020 vol: 20 (23) pp: 6730. [JCR 2020 Q1; IF 3,576]. DOI: 10.3390/s20236730. Nº de citas: 8 (Fuente, Google Scholar).

2. A Microservices e-Health System for Ecological Frailty Assessment Using Wearables

[A] Garcia-Moreno, F. M., Bermudez-Edo, M., Garrido, J. L., Rodríguez-García, E., Pérez-Mármol, J. M., Rodríguez-Fórtiz, M. J. (2020). A Microservices e-Health System for Ecological Frailty Assessment Using Wearables. *Sensors*, *20*(12), 3427. DOI: 10.3390/s20123427

a. Estado: publicado.

b. Factor de impacto (JCR 2020): 3.576.

c. Categoría: Instruments & Instrumentation. Posición: 14 / 64 (**Q1**).

d. Categoría: Engineering, Electrical & Electronic. 82 / 273 (**Q2**).

e. Nº de citas: 19 (Fuente, Google Scholar).

Este artículo se puede encontrar en acceso abierto en el siguiente enlace:

https://doi.org/10.3390/s20123427

A continuación, se proporciona un borrador del mismo para cumplir con los derechos de autor.

A Microservices e-Health System for Ecological Frailty Assessment using Wearables

Francisco M. García-Moreno [1], María Bermúdez-Edo [1], José Luis Garrido [1], Estefanía Rodríguez-García [2], María José Rodríguez-Fórtiz [1], and José Manuel Pérez-Mármol [2]

[1]Department of Software Engineering, Computer Sciences School, University of Granada.
[2]Department of Physiology, Faculty of Health Sciences, University of Granada.

Abstract. The population in developed countries is aging and this fact involves high costs in elderly health, as well as a decrease in the number of active working members to support these costs. This could lead to a collapse of the current systems. One of the first insights of the decline in elderly people is frailty, which could be decelerated if it is detected at an early stage. Nowadays, health professionals measure frailty manually through questionnaires and tests of strength or gait focused on the physical dimension. Sensors are increasingly used to measure and monitor different e-health indicators while the user is performing Basic Activities of Daily Life (BADL). In this paper, we present a system based on microservices architecture, which collects sensory data while the older adults perform Instrumental ADLs (IADLs) in combination with BADLs. IADLs involve physical dimension, but also cognitive and social dimensions. With the sensory data we built a machine learning model to assess frailty status which outperforms the previous works that only used BADLs. Our model is accurate, ecological, non-intrusive, flexible and can help health professionals to automatically detect frailty.

***Keywords**: wearable devices; sensors; mobile health systems; microservices architecture, IoT, machine learning; elderly frailty assessment, e-health.*

1. Introduction

In a global ageing society, it is important to detect the frailty in the early stages to be able to decelerate the decline of elderly people and to keep them active and healthy for as long as possible. Frailty is a syndrome of older adults that increases the risk of falls, hospitalizations and even death [1,2]; affecting 11% of the home-dwelling elderly population without hospitalization and increasing drastically to 30-70% of the surgical patients over 65 years old [3]. The early detection of frailty has been proved to increase the independence of the older adults and the decrease on health care costs [4,5].

The current detection of frailty is performed manually via independent tests of strength, gait or self-report questionnaires, being Fried [4] the most used test, which only assess the physical dimension of frailty. Some authors have recently accepted that frailty involves not only the physical dimension, but also social and cognitive dimensions [6]. Accordingly, some wider tests have appeared which include the three dimensions of frailty, such as the Tilburg Frailty Indicator (TFI) test [7]. However, the tests are time consuming, and require investment in health resources and physical interaction between patients and doctors. Indeed, the detection and manual assessment of the frailty for all older people is not currently affordable, especially due to the organizational and financial issues mentioned above [8].

Frailty also affects the performance of the Instrumental Activities of Daily Living (IADLs) [9]. The Activities of Daily Living (ADL) are classified as: Basic (BADL), such as toilet hygiene or functional mobility; Instrumental (IADL), such as house cleaning or shopping; and Advance (AADL), such as travelling and social events. BADLs are crucial for the human being´s survival and they involve less demanding activities in terms of physical or cognitive resources in comparison with IADLs. These activities are commonly named as self-activities performed by the person without the need of help or interaction with other people. On the other hand, IADLs and AADLs are more complex activities that involve more resources than BALDs in terms of physical, cognitive and social functions. These activities usually also involve the interaction with other objects or tools. Monitoring of the performance of the IADLs could potentially be a possible marker of the frailty status, since the complexity of the IADLs make elders them more vulnerable to show dependence in their performance. Therefore, the monitoring of the IADLs could replace the manual tests or questionnaires, alleviating the health financial and organizational requirements to globally asses frailty. This alleviation could be possible if the monitorization is performed with low-cost or well-spread devices, such as sensors, that are currently built-in in smartphones or wearables.

Wearable devices, together with data analysis techniques, can seamlessly monitor and assess the frailty status during the performance of the IADLs. This could be even less intrusive in the near future, because it is foreseen that the majority of the population, including elderly people, will wear wearables with built-in sensors [10], especially after the pandemic coronavirus of 2019 [11]. Sensors have been successfully used in several e-health scenarios [12]. They can provide efficient and precise psychophysiological and location values [13]. Some research works have started to measure frailty indicators [3,14–18], being the most used devices electronic stickers, bands, bracelets, smartphones and smartwatches [15,19–23]. However, these approaches only measure the physical dimension of the frailty during a BADL with only one sensor, either the accelerometer or the gyroscope. None of them monitored IADLs, such as a real shopping in the supermarket, which involves not only physical activity, but also cognitive and social interactions. Furthermore, existing approaches are also limited in terms of flexibility, extension and evolution, as they do not take advantage (at design and deployment levels) of software architectures that allow developers to easily integrate current and future sensors and functionalities. Hence, to the best of our knowledge, there is no flexible approach which assesses the frailty status through the monitoring of IADLs in an ecological way, i.e. in the daily living environment of older adults

To overcome these limitations, we expanded our microservices architecture proposed in [24], with a choreography solution and a use case that collects sensory data from wearables and builds a Machine Learning (ML) model to assess frailty in an ecological manner —while the older adult is performing an IADL, in combination with BADLs. Service-Oriented Architectures (SOA) are well-known for its generalization, and therefore for its reuse in different scenarios. They can also deal with mobility, integration interoperability, fault tolerance and self-adaptation [25–29]. A microservices architecture improves the SOA approach, following the concept share-as-little-as-possible [30], with additional system characteristics such as decoupling, flexibility, extension, scalability and evolution [31,32]. Microservices architectures allow to collect a wide variety of sensory data under the Internet of Things (IoT) paradigm. We also explored several ML algorithms in order to build the most accurate model to assess frailty. In our proposal we use microservices for data collection, data analysis and to expose the frailty model as a microservice, with the aim of reusing the microservices in future use cases.

Our proposal is ecological and effective in terms of costs and time-consuming reduction, using an accurate frailty assessment model that outperforms the related works. We collect sensory data in a non-intrusive and transparent manner while the older adult is performing the IADL of shopping; without disturbing neither the person daily life nor his environmental conditions (same supermarket, time, pathway, etc.), i.e. ecologically. Measuring the performance of IADL we can gain a comprehensive vision of elderly people disorders at physical, cognitive and/or social levels. Hence, health professionals can design specific interventional programs to improve the frailty, increasing the independence of the older adults and decreasing health care costs. Furthermore, with the microservices architecture, the system could be easily extended [33] to address other use cases, by adding new microservices with the corresponding ML algorithms, models and sensors. For example, we could add microservices for the ML models built for a different population, with specific pathologies, ethnicities; or we could add microservices for collecting data coming from new sensors, or data sources. Our results outperformed the ecological related works, which use only the physical data from the performing of BADLs (instead of IADLs) and are non-low-cost. These findings have important implications for healthcare systems, because our automatic frailty detection system can reduce health costs and health professionals time assessing frailty ecologically with low-cost devices.

The rest of the paper is organized as follows. Section 2 introduces the related works. Section 3 describes the design of the microservices system architecture to collect and analyze data, as well as a description of the data analysis pipeline. Section 4 presents a system validation and the results of the ML frailty model performance. Section 5 discusses the findings. And the last section summarizes conclusions and future work.

2. Related work

In order to monitor a variety of sensory data, we need to integrate and combine several sensors through systems, which usually use IoT and/or microservices architectures [34–36]. IoT health systems allow doctors to monitor patients constantly and to suggest corrective actions according to the analyzed data [29,37,38]. For example, one of the goals of City-4Age and ACTIVAGE H2020 projects is the assessment of frailty using IoT technology based on rules [39]. Furthermore, there are a few efforts to apply microservices in the health domain. For example, some studies focused on security and privacy of microservices architectures for the sensible data in the health systems [40,41]; other work explored monitorization of depression disorder through IoT and microservices [42]. However, to the best of our knowledge, there are no ecological approaches that use microservices to perform detailed ML pipelines.

Sensors in combination with ML algorithms have been used to recognize different BADLs such as sitting, standing, walking, running, stair climbing, eating and sleeping [43–50]; and IADLS such as preparing a meal, bathing, using the phone, driving, traveling, and even shopping [46,51]. Some works asses the physical component of frailty in BADLs using wearable devices in combination with ML algorithms (Table 1). Tegou et al. [14] applied several ML algorithms, such as k-Nearest

Neighbour (kNN), Random Forests (RF) and Naïve Bayes (NB), to classify the three frailty status based on Fried criteria [4] (non-frail, pre-frail, frail), by measuring the number of transitions between rooms inside the older adult home. Schwenk et al. [15] and Kumar et al. [16] applied Multinomial Logistic Regression (MLR) model to discriminate between the three frailty status, by measuring the gait, balance, and physical activity (PA). Toosizadeh et al. [3] used Ordinal Logistic Regression (OLR) for a frailty detection in 20 seconds elbow flexion. Greene et al. [17,52] designed a digital assessment based on a Logistic Regression (LR) model for detecting falls, frailty and mobility impairment. Razjouyan et al. [18] applied ML embedded feature selection method for remotely monitoring frailty status, during walking and sleeping BADLs. These studies focused on assessing only physical components of frailty (based on Fried [4]) during physical BADLs (walking, sitting, standing or sleeping), or simple activities such as the elbow flexion or number of transitions from one room to another at home. They neither consider other human dimensions such as cognitive or social functions, nor assess frailty during an IADL (which include these functions). In addition, most of these logistics models did not focus directly on detecting frailty status, instead they looked for the correlation between risk factors (such as falls, balance, physical performance and gait) and the frailty status. Also, the ecological approaches reviewed [16,18] use expensive wearables and questionnaire data such as demographic and clinical data, which do not allow the automation of the data collection.

Therefore, to the best of our knowledge there are no works that asses the frailty status during the performance of IADLs, combining low-cost sensors with a flexible and extensible system.

Table 1. Review of previous works related to assessing frailty with wearables and ML.

Work	Aim	Eco	Data Sources	System	Frailty Status	Best ML
[14]	To assess frailty by a system based on Bluetooth RSSI fingerprints using beacons, collecting data derived from transitions among rooms	Yes. Transitions between rooms.	Smartphone Beacons (low-cost)	RSS	Three[2] and two[3]	RF[2]: Accuracy: 82,33% Sensibilty: 83.83% RF[3]: Accuracy: 97.92% Sensibilty: 94.2%
[15][2]	To discriminate between frailty status with gait, balance or during a physical activity.	No	LEGSys[1] ($10.000) BalanSens[1] ($4.450)	None	Three[2]	MLR: AUC: 85.7%
[16]	To implement a wearable to characterize the quantity and quality of everyday walking, and to establish associations between gait impairment and frailty	Yes. Walking ADL during 2 days	PAMSys[1] Demographic Clinical	None	Two[4]	MLR: Accuracy: 77.7% Sensibilty: 76.8% Specificity: 80%
[3]	To assess frailty by a wearable during the flexibility of upper-extremity movements	No	Gyroscope[1]	None	Three[2]	OLR: Accuracy: 69%
[17,52]	To design a digital assessment protocol and algorithm for prediction of falls, frailty and mobility impairment	No	Shimmer ($495) Demographic Clinical	None	Two[4]	LR: Accuracy: 72.8% Sensibilty: 72,99%
[18]	To remotely monitor the frailty status using an accelerometer.	Yes. Walking & Sleeping ADLs during 2 days	PAMSys[1] Demographic Clinical	None	Two[5]	EFS: Accuracy: 84.7% Sensibilty: 91.8% Specificity: 81.4%

Eco: Is it ecological? Is it measuring frailty in the users' daily living environment?

RSS: Received Signal Strength; RSSI: Received Signal Strength Indicator; AUC: Area under the curve.

RF: Random Forest; MLR: Multinomial Logistic Regression; OLR: Ordinal Logistic Regression; EFS: Embedded Feature Selection

1 BioSensics LLC manufacturs non-low-cost wearables devices.

2 Considered Fried 3 classes: non-frail, pre-frail, frail status.

3 Considered Fried 2 classes: identification of frail participants against non-frail and prefrail participants together as a single class.

4 Considered Fried 2 classes: identification of non-frail participants against pre-frail and frail participants together as a single class.

5 Considered Fried 2 classes: identification of pre-frail participants against non-frail and frail participants together as a single class.

3. Materials and Methods

This section presents the participants' sample, the considered variables, the system architecture proposed and the data analysis performed to assess participant's frailty status, during the shopping IADL performance. First of all, it describes the inclusion and exclusion criteria for recruiting the participants. Secondly, it presents the dependent variable, which is the result of the traditional assessment of frailty by the Fried test. Thirdly, it presents the variables identified through the collected raw data coming from sensory data. Fourthly, it introduces the system architecture designed following the microservices paradigm.

And finally, it explains the pipeline workflow designed to apply data analysis, considering: data collection and labelling; data preprocessing; and machine learning model generation and evaluation.

3.1 Sample description

We conducted a cross-sectional study in three community day centers of Granada, Spain. We recruited a total of 79 participants (69 women and 10 men) older than 65 years (average age equals 75). The inclusion criteria for this study were: 1) ages ranged from 65 to 90 years old; 2) non severe cognitive decline (using a cut-off score of ≥ 24 points in the Mini-Mental State Examination test [53]); 3) non perceptual alterations, determined by medical diagnosis report. Thus, the participants were included with perceptual alterations corrected with a support device, such as pair of glasses or hearing aid; 4) walking with/without help (cane or walker); 5) community-dwelling older adults. Exclusion criteria: 1) severe mental disorder; 2) severe language alterations; 3) medical instability; 4) pathology in acute stage; 5) hospitalized; 6) serious behavior alterations or motor risks.

Our study was approved by the Investigation Ethical Committee of Granada province with reference 0111-N-19 on May 31, 2019 (Andalusian Health Service, Granada, Spain).

3.2 Fried test and frailty status variable

ML models need a training phase and a test phase. The model is trained with labelled data and once the model is created it is tested with different data. Finally, with the model created and tested new data could be automatically classified with one label. We used the score of Fried test [4] to label the data because it is the most accepted test to asses Frailty. This test consists on the assessment of 5 components: slowness in mobility, low energy, low physical activity, muscle weakness, and involuntary weight loss. Depending on the previous factors, elderly people are classified into: "frail" if they score positive in three or more components of the test; "pre-frail" if the score is positive in 2 components and "non-frail" if they do not meet any component. Hence, the Fried test classification ("frail", "pre-frail" and "non-frail") will be our dependent variable in the training and test phases of the model.

3.3 Wearable sensors variables

The independent variables considered come from the raw data of wearable sensors. Previous works used accelerometer, gyroscope and heart rate sensors for monitoring physical activity, health conditions, cognitive, social and other factors [49,54–57]. We have used similar sensors built-in a low-cost smartwatch, the Samsung Gear S3 [58]. Samsung Gear S3 supports development of software applications on its open access Tizen operating system. This characteristic allows to collect the sensory raw data and to develop software applications in a flexible manner. The built-in sensors we used are: 1) the triaxial accelerometer, which measures changes in the device velocity; 2) the triaxial gyroscope, which detects angular velocity and orientation of the device; 3) and the heart rate sensor, which measures the heart beats per minute. In particular, we collected the raw data from these wearable sensors obtaining 7 variables, detailed in Table 2. The value of these variables was provided in integer number (without a decimal point), or in float (floating-point number, which is a number that has a decimal place).

Table 2. Wearable sensors variables from raw data.

Variable Description	Type
Accelerometer X-axis value	Float
Accelerometer Y-axis value	Float
Accelerometer Z-axis value	Float
Gyroscope X-axis value	Float
Gyroscope Y-axis value	Float
Gyroscope Z-axis value	Float
Heart Rate value	Integer

3.4 Microservices system architecture

Microservice architecture is the next generation of SOA architecture, improving this approach on decoupling, flexibility and extensibility, by sharing services and resources as little as possible. A microservice is a miniature software responsible of a unique and single task (i.e., the minimum level of responsibility and loosely coupled) [30,59,60]. Furthermore, since microservices architecture is based on the concept of share-as-little-as-possible [30], they try to minimize on sharing, whereas SOA tries to maximize on sharing. Consequently, microservices only expose well-defined interfaces. Hence, microservices ensure the maximum decoupling and extensibility, and therefore it is possible to update a microservice independently of the rest of the microservices in an architecture. Indeed, while SOA uses centralized dependencies (such as a centralized data storage), which makes it hard to avoid decoupling, in microservices architectures each microservice manage all of its own dependencies (such as database, key-value store, search index and queue), in order to develop and deploy each microservice independently [30,59]. However, in real scenarios could be necessary to share data between microservices. In our architecture, the data collected by one microservice, need to be stored in a shared database that other microservice could access in order to process the data [60].

Microservices are classified in two groups: functional microservices (FM) and infrastructure microservices (IM) [30]. FMs implement business functions (related with the business domain, e.g., frailty assessment) and can be accessed from external clients (such as wearables, mobile phones and web applications), via an Application Programming Interface (API) gateway. In contrast, IMs implement nonfunctional business tasks such as logging, auditing and monitoring. The main difference between FMs and IMs is that IMs provide a local context and they are not publicly accessible from the outside by client requests, while FMs expose their services publicly. Figure 1 illustrates the microservices taxonomy. In this example, FMs and IMs have been deployed in the cloud, but they can be deployed elsewhere (such as in mobile phones or in wearable devices).

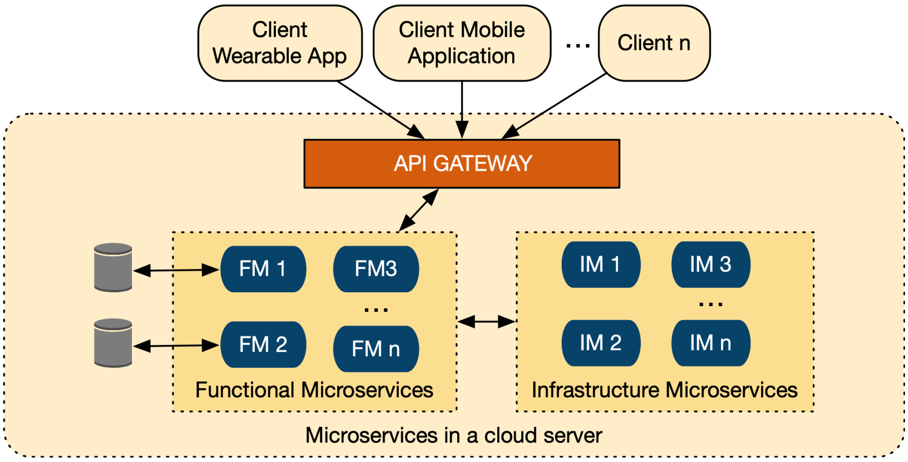

Figure 1. Microservices architecture taxonomy (adapted from [30]).

The functional microservices can communicate with external clients using an API gateway, and with infrastructure microservices via internal requests. For example, in our use case (Figure 2), the application in the mobile phone (Frailty Status App in Figure 2), used by end users, can access the frailty assessment (Fried Frailty model in Figure 2) through the API gateway in the cloud server. The API gateways are the interface components of our architecture, which aggregate microservices. They can distribute information to the external clients (i.e., functional microservices, web and applications) by abstracting the real microservice endpoints. With this abstraction, modifications in the internal microservices are transparent to the clients. For instance, if a microservice is splited into two microservices, clients do not need to change anything [30,59].

The API Gateway could be implemented with different protocols, such as REST API [61], CoAP [62], AMQP [63], or MQTT [64]. In particular, we use two API Gateways: a REST API for a one-time communication, and a MQTT for continuous communications. The API gateway in the smartwatch (hereafter, API-s) is a REST API (see Figure 2). API-s setup and start the microservices deployed in the wearable device by invocations from the smartphone. The second API gateway (hereafter, API-c) is deployed in the cloud server and it is based on the MQTT Publish/Subscribe protocol [65] (see Figure 2). API-c supports continuous communication of data between the smartwatch and the cloud. For example, sensor microservices send sensory data and microservices deployed in the cloud stored and processed said data. API-c also supports the communication of events informing the app in the smartphone about the frailty status assessment carried out by a single specific frailty model or even more than one model simultaneously (extensibility), e.g., Fried and TFI. In particular, API-c follows the publish/subscribe communication pattern [65]. The publishers (sensor microservices) push messages (sensory data) through a specific topic (e.g., accelerometer), and the microservices subscribed to that topic (functional microservices in the cloud) receive it. The Publish/subscribe pattern used by the API-c provides decoupling between sender and receiver, allowing asynchronous communications and mobility support. Subscriptions can be established and abandoned dynamically (i.e., Publish/subscribe pattern supports adaptability). Furthermore, MQTT has low overhead, which is essential in sensor environments. In our proposal microservices can publish messages to different topics (communication channels) asynchronously, and others microservices can subscribe to these topics and react accordingly [59].

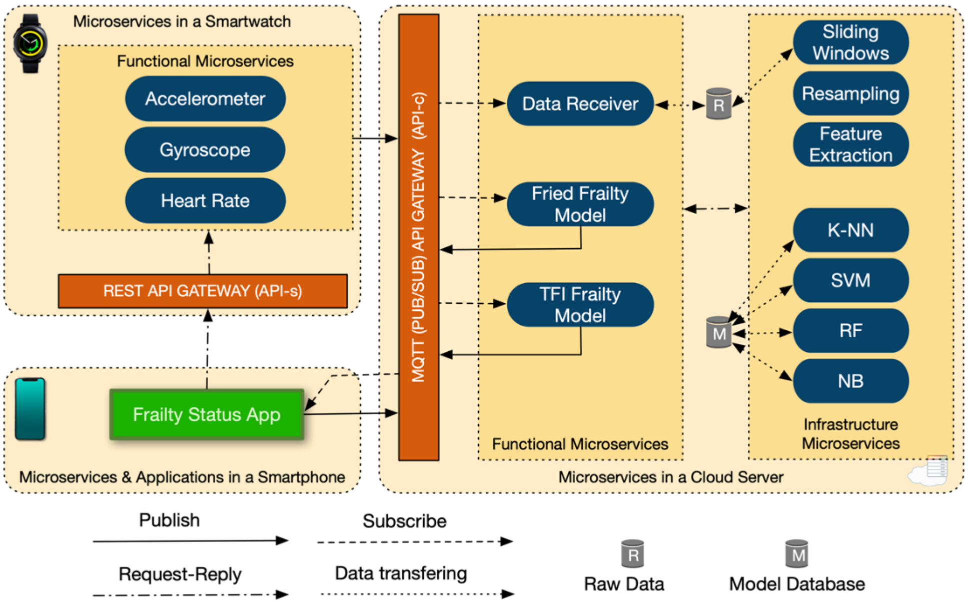

Figure 2. Microservices architecture for frailty assessment.

3.4.1 Microservices deployed in wearable devices

In the wearable device, we deploy only functional microservices, dealing each one with the collection of data from one sensor, namely Accelerometer, Gyroscope and Heart Rate microservices (see Figure 2). The smartphone clients interact with all of the sensor microservices, through the API-s, to setup and start the production of data. Likewise, sensor microservices interact with other microservices deployed in the cloud server through the API-c. In particular, when the Frailty Status App in the smartphone sends the initial setup with the sampling rate to work with and starts the sensors microservices for collecting data (through API-s), the sensor microservices stream the collected data to the API-c in the cloud server.

Due to the adoption of a microservice architecture, our system is ready to add new wearable sensors in the future (if needed or convenient) for the assessment of frailty or other pathologies. For example, in order to add a new sensor, we need to implement the corresponding sensor microservice and include it into the architecture to collect the sensory data. This sensor microservice should publish data through a new topic using the MQTT protocol (API-c). Likewise, we need to expose the new sensor microservice in the API-s of the smartwatch.

3.4.2 App deployed in the Smartphone

The purpose of the Frailty Status App deployed in the smartphone is fourfold. First, it requests the sensor microservices to setup and start the data forwarding. Second, it publishes (through API-c) the event that starts one or more frailty model microservices. Third, it subscribes to the topic that allows the reception of events informing about the frailty status of a person. Four, it shows the frailty status to the user (on the mobile application).

Our system can be extended easily with several frailty models to assess frailty and they can even operate simultaneously. To that end, we only need to include the microservice for the new model and create its own topics (one for setup and start process; and another to expose the frailty status). The Frailty Status App will need to subscribe to these new topics. For example, we could include a frailty model which considers different datasets, which could lead to a different model; i.e. using a different sample, for example selecting elderly people with a different age ranges or ethnicities. In fact, when having several frailty models, the Frailty Status App could subscribe to any of the frailty model and status subtopics (e.g. frailty/model1 for starting to work; and frailty/model1/status for getting the frailty assessment result). For example, one Frailty Status App can subscribe to the model-1 subtopic; another Frailty Status App, running in another smartphone, could subscribe to the model-2 subtopic; and even another Frailty Status App could receive results from more than one frailty model at the same time when it is subscribed to the main topic (frailty/), that includes the subtopics.

3.4.3 Microservices deployed in the cloud server

The first microservice we implemented in the cloud server is the functional Data Receiver, which is in charge of receiving and storing the sensory data coming from all the sensor microservices. The Data Receiver receives these raw data without applying any preprocessing. We have also implemented two functional microservices for assessing frailty: Fried Frailty Model microservice (hereafter, Fried microservice) and TFI Frailty Model microservice (hereafter, TFI microservice). Fried and TFI microservices are in charge of the choreography of the corresponding ML pipeline (explained in detail in the next subsection). Both, Fried and TFI microservices assess the frailty status and send the resulting value ("pre-frail", "frail" or "non-frail") through the API-c. The pre-built ML models are stored in the Model Database (this offline creation is common in ML domain), in order to use them later by Fried and TFI microservices automatically online. The building process of these models will be explained in section 3.5.

The infrastructure microservices implement several data analytics techniques. We have implemented three microservices related with preprocessing algorithms (top right of Figure 2): Sliding Windows microservice; Resampling microservice and Feature Extraction microservice, which we will described in detail in section 3.5.3. We have also implemented as microservices (bottom right of Figure 2), four ML algorithms: k-NN, SVM, RF and NB. These algorithms require the pre-built frailty models stored in Model Database (see Figure 2) in order to calculate the frailty assessment ("frail", "pre-frail" or "non-frail"). They require as well, the preprocessed data from the previous microservices (sliding windows, resampling and Feature Extraction microservices).

Adding a new frailty model to our system, requires the following: 1) to pre-build the new ML model, and to store it in the Model Database; 2) to create the new frailty microservice for the new model, together with its own subtopics (frailty/newmodel; and frailty/newmodel/status) and ML pipeline setup; and 3) if we need a new ML algorithm not yet implemented, such as Artificial Neural Network algorithm, we need to create it as an infrastructure microservice.

3.4.4 Workflow

Figure 3 shows the workflow communication during the assessment of frailty. In this section, we show the potential of our architecture in terms of extension of future business functionalities, illustrating a scenario with two frailty models —Fried and TFI. In particular, we will explain in detail the workflow communication between microservices and app through the APIs and internal requests and events.

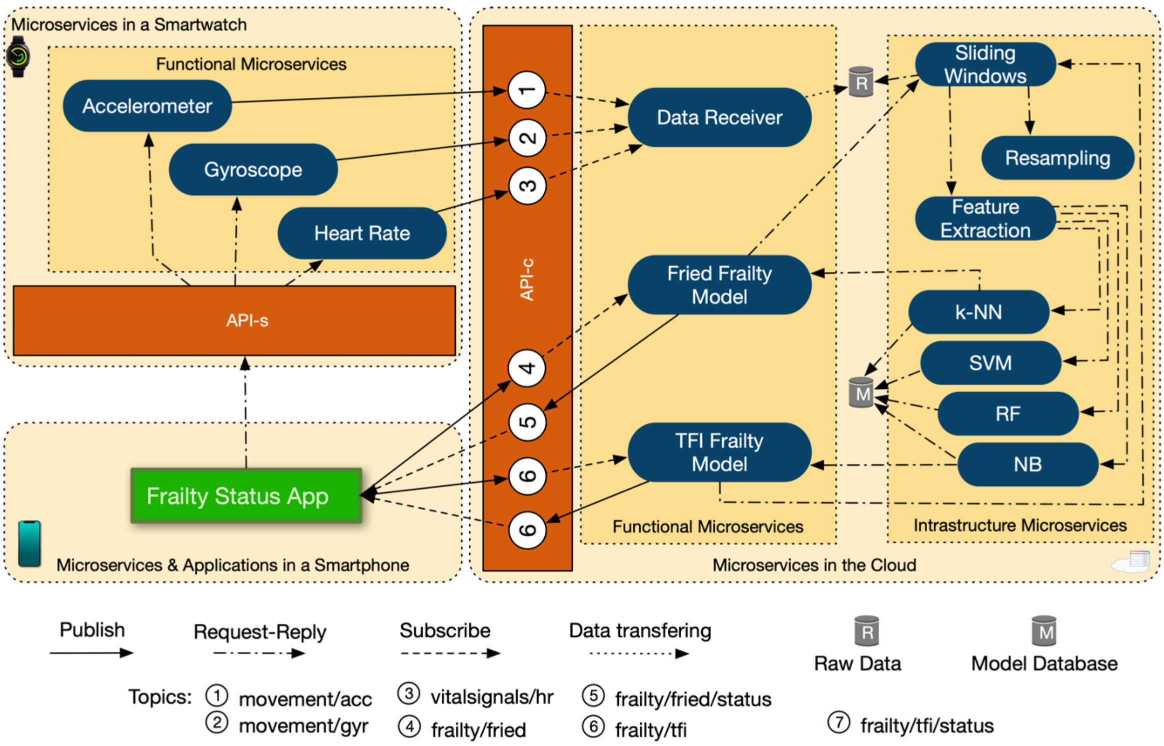

Figure 3. Workflow communication details of the microservice architecture for frailty assessment.

First, the Frailty Status App sends a REST POST request to start data forwarding from each sensor microservice available in the smartwatch (in particular, the Accelerometer, Gyroscope and Heart Rate microservices). As a response, all the sensor microservices publish their raw data into the API-c gateway to the corresponding topics: “movement/acc”, “movement/gyr”, and “vitalsignals/hr”, respectively. Then, the Frailty Status App publishes an event associated to the topic “frailty/fried” with a value of “1”, and another event “frailty/tfi” with a value of “1”, in order to start the Fried and TFI microservices, respectively. When the app wants to stop the model microservices it will send the same event with a value of “0”.

Second, the Data Receiver microservice, which is subscribed to all of sensory data topics (“movement/acc”, “movement/gyr”, and “vitalsignals/hr”), receives the raw data from the sensor microservices, and stores the data. Simultaneously, the Fried and TFI microservices, which are subscribed to the topics “frailty/fried” and “frailty/tfi”, respectively, and start their choreographies. These choreographies consist of a list of sequential ML pipeline microservices with their respective parameters, which are sent to the first IM in the pipeline (in our case Sliding Windows microservice). For example, for Fried microservice the choreography consist of: 1) sensor microservices to use (accelerometer, gyroscope and heart rate) ; 2) data sampling rate (25 Hz); 3) sliding windows size (0.5 seconds); 4) subset of feature extractions (mean, standard deviation, minimum and maximum values, kurtosis, skewness and energy —explained in section 3.5 -); 5) pre-built frailty model (e.g. pre-built Fried model database path); 6) the ML algorithm to make the frailty assessment (e.g. k-NN); 7) and the returning Frailty model microservice where to send the frailty status assessed in step 6 (e.g. Fried microservice).

Third, the ML pipeline, which will be described in detail in next subsection, starts. The Sliding Windows microservice receives the ML pipeline setting, and each microservice of the pipeline is activated sequentially by the previous one. Finally, the ML microservices, which are the last of the pipeline, send the resulting frailty status to the corresponding Fried or TFI microservice.

Fourth, the Fried microservices receive the frailty status predicted by the ML microservice and publish the result of the frailty assessment—“non-frail”, “pre-frail’ or “frail”—to the topic “frailty/fried/status”. Likewise, TFI microservice publishes the result to the topic “frailty/tfi/status”. Then, Frailty Status App on the smartphone receive this result and notifies the user.

3.5 The Data Analysis pipeline to build a predictive model for frailty assessment

In order to build a frailty model, we have designed a data analysis pipeline, which includes three main phases: 1) data collection and data labelling; 2) data preprocessing; 3) and frailty model building with ML techniques. In the first phase, we collect the data and label it in order to apply machine learning algorithms in supervised manner, which need labelled data. Our health experts classify each older adult with a label (“frail”, “pre-frail” and “non-frail”) according to the traditional assessment of frailty based on Fried test [4]. This label becomes the dependent variable for assessing frailty status. In the second phase we preprocess the data focusing on the enhancement of the model performance. Usually, preprocessing techniques include segmentation, feature extraction and dimensionality reduction (i.e., feature selection) [43]. In our case, the preprocessing techniques applied consist in the feature extraction combined with segmentation, and feature selection to reduce these features extracted. In the third phase, we build a predictive model using the labelled and preprocessed data. We build the model outside the system architecture as we mentioned above, but we sored the model in a database accessible by our Fried microservice. Figure 4 shows all the phases and steps inside the pipeline, which will be described in detail bellow.

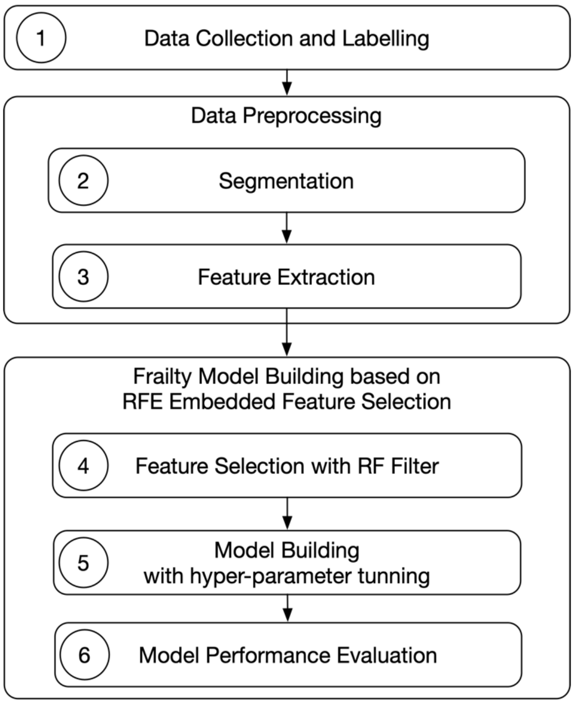

Figure 4. Data analysis pipeline for frailty status assessment.

3.5.1 Data collection and labelling process

In order to make the frailty assessment, we collected the results of the participants' frailty status based on Fried classification. Thus, we had labelled all of these participants with their corresponding frailty status ("frail", "pre-frail" and "non-frail").

Participants were asked to buy a specific product in the nearest supermarket following the next protocol. The participant starts in a sitting position on a chair without armrests, with €1 coin and wearing the Gear S3 smartwatch in the non-dominant hand. The supervisor, who is responsible of the experiment supervision, starts the Frailty Status App to start capturing wearable sensory data. The participant stands up and walks to the supermarket; looks for and pick up a 1kg salt pack; goes to the check out and pays; comes back to the starting point; and sits down on the chair. The supervisor stops the collecting of data with the Frailty Status App.

The performance of the shopping activity was divided into several tasks or sub-activities doing an activity analysis using the Occupational Therapy Practice Framework [66]. During this whole process, the evaluator labeled the different tasks of the shopping activity during the performance of it: 1) sitting; 2) standing; 3) walking to the supermarket; 4) in the supermarket; 5) looking for the product to purchase; 6) picking the product; 7) going to the checkout; 8) in the checkout; 9) paying; 10) go to the exit; 11) in the outside; 12) coming back; 13) standing at the start point; 14) and sitting back. This labelling was performed by observation, and with the help of an app connected to the smartwatch.

The smartwatch transmits (offline) the data collected to the smartphone (to the Frailty Status App) via Bluetooth. The sampling rates used in the data collection is normally greater than 10 Hz, and in most cases ~20 Hz, as similar studies in the field of activity recognition suggest [50,51,67]. In particular, we recorded data at 25 Hz as in Genovese et al. [55] and close to the 20 Hz used by Garcia-Ceja et al. [51]. Nonetheless, we have performed several experiments with different sampling rates between 10Hz and 100 Hz, obtaining similar results. Moreover, we excluded anomalies in heart rate values based on the formula: HR maximum = 220 − participant age [68,69]. Furthermore, as we recruited participants in three community-centers, the distance from the center to the supermarket were different (two centers: 50m; and the third 100m). This fact does not carry any inconvenience as the activities are labelled (i.e., when the participant starts and end walking to the supermarket) and the distance for each participant is also recorded.

3.5.2 Data preprocessing

In order to get accurate results assessing frailty [70], we performed two well-known data preprocessing techniques sequentially: 1) segmentation of raw data (without preprocessing) with 50% overlapped windows strategy; 2) feature engineering, for every window, in order to extract relevant variables.

3.5.2.1 Segmentation

We implemented 50% overlapped sliding windows approach to segment the data, as in previous works [50,67,71]. This is a common technique used with wearable sensory data, which implies an increase of the sample size due to the overlap, and therefore the re-used of the data. Hence, this technique improves the accuracy of the results [43]. We test different window sizes to get which of them reports the best model performance: 0.5s; 1s; 1.5s; 2s; and 2.5s.

3.5.2.2 Feature Extraction

We also apply feature selection, as in other similar works [12,49,50,72,73]. In particular, we extract eight statistical features for every wearable variable (we identified 7 wearable variables in Table 2) and for each window. Related with time-domain, the extracted features are: mean, standard deviation (SD), skewness (the probability distribution asymmetry), kurtosis (the probability distribution, a.k.a. tailedness), maximum, minimum and amplitude (the absolute difference between the maximum and minimum). The only frequency-domain feature extracted is the energy, which is the sum of the squared Fast Fourier Transform (FFT) components [74,75]. In order to normalize the energy feature, it was divided by each window size ($|w|$). Let $x_1, x_2, \dots, x_n$ be the i-th FFT components for the windows $w_1, w_2, \dots, w_m$, then Energy [74] expression is:

$$\text{Energy} = \frac{\sum_{i=1}^{|w|} |x_i|^2}{|w|} \tag{1}$$

3.5.3 Frailty model

3.5.3.1 Feature Selection

Feature selection is a technique that selects only the most relevant features to train the model and reduces the dimensionality of the dataset. As we could train the model with only a selection of features: the model will train faster; the complexity of a model is lower and therefore it makes it easier to interpret; the accuracy of a model could improve if the right subset is chosen; and the overfitting is also lower. Feature selection methods apply statistics to identify what are the most important features and remove the redundant features. A feature is redundant when another relevant feature exists with a similar power of prediction [76]. The feature selection techniques are the filter, wrapper and embedded methods [16,77]. Filter methods identify the relevant features based on statistics, without considering a machine learning classifier to build a model. Wrappers use a subset of features to measure the performance of a machine learning classifier; and then they iteratively add or remove features to the subset according to the results, until they reach an optimum. Embedded methods include feature selection as part of the machine learning model building. Therefore, both wrapper and embedded methods are specific to the machine learning algorithm used.

In our case, we have used embedded methods (such as in [18]), hence using feature selection in the building process of the frailty assessment model. Our embedded method consists in the combination of a filter method (based on Random Forest to rank features by their importance) with the Recurrent Feature Elimination (RFE) strategy [78] to build (also called train) one ML model per feature selected. First, RF sort the features by their relevance in descending order, then RFE starts considering all of the sorted feature set. Then, we build the model with that sorted feature subset and evaluate different performance metrics (we will detail these metrics in the next subsection). In each iteration, we eliminated the least relevant variable (or redundant); recalculate the importance rank of the new subset with RF filter; create a model with this subset; and evaluate the frailty model performance. At the end, we select the model with the best performance and identify what are the resulting features subset. Considering that the total wearable variables coming from the raw data are 7 (see Table 2) and that from each variable we extract 8 features, our initial set of features is 56

3.5.3.2 Frailty model building

To build an accurate machine learning model it is necessary to tune its hyper-parameters, considered high level concepts of the model. For instance, some hyper-parameters are complexity, capacity to learn, number of leaves in trees of RF, or kernels in SVM [79]. As described in the related work, several machine learning algorithms have been applied successfully to assess physical frailty, such as k-NN, RF and NB. We used these algorithms and Support Vector Machines (SVM), which is widely used for ADL recognition. Therefore, we tried them all, and identify which one of them reported the best performance for our use case. We tested these algorithms, with R programming language, tuning its specific hyper-parameters based on related studies [45,49,80]:

- k-NN: 1) k: {1, 3, 5, ..., square of number of rows} (only odd numbers).
- SVM: 1) cost function: {0.1, 1, 10, 100}; 2) gamma value: {0.5, 1, 2}; 3) and, kernel type: {"radial", "polynomial", "linear", "sigmoid"}.
- RF: 1) number of trees: {10, 100, 200, 500, 1000}; 2) number of variables randomly sampled: {10, 25, 50}.
- NB: 1) use of kernel: {True, False}; 2) use of poisson: {True, False}.

Regarding the computational load, it is known that in the creation of the model (training phase), some algorithms are slower (SVM and RF) than others (k-NN and NB). However, this process is offline and it does not require tight responsiveness. The complexity of the offline building process for every ML algorithm is as follows. Let n be the training size, m the number of features, n_{trees} the number of trees (for RF algorithm), then the computational complexity is:

- k-NN: O(nm)
- SVM: $O(n^2m+n^3)$
- RF: $O(n^2mn_{trees})$
- NB: O(nm)

However, once we have our model built, to know the prediction of a particular record (i.e., to assess frailty of a person) the computational complexity is reduced. Specifically, the complexity of each ML algorithm is:

- k-NN: O(nm)
- SVM: $O(n_{sv}m)$, where n_{sv} is the number of support vectors, which is the resulting points of the SVM model, close to the decision boundary.
- RF: O(ntreesm)
- NB: O(m)

3.5.3.3 Model performance evaluation

In order to validate our ML model, we used 5-fold stratified cross-validation [3,14], in which the dataset was randomly divided into k=5 equal parts keeping the proportion of samples of the three frailty labels ("frail", "pre-frail" and "non-frail"). At each k-th iteration, the k-1 partitions were used to train (build) the model, and the left-out partition (hidden data for the trained model) was used for validating the model. The performance metrics for this validation were: accuracy, f1-score, sensitivity (true positive rate) and specificity (true negative rate). These metrics are related with these concepts: 1) true positive (TP) condition, which is the number of true positive cases in the data and rightly predicted as positive by model; 2) true negative (TN) condition, which is the total number of true negative cases and rightly predicted as negative; 3) false positive (FP) condition, which is the total number of positive cases classified wrongly as negatives; 4) and, false negative (FN) condition, which is the total number of negative cases but classified wrongly as positives. Taking into account these considerations, the expressions of the four metrics used are the following:

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN) \tag{2}$$

$$\text{Sensitivity} = TP / (TP + FN) \tag{3}$$

$$\text{Specificity} = TN / (TN + FP) \tag{4}$$

$$\text{F1-Score} = 2TP / (2TP + FP + FN) \tag{5}$$

4. Results

Before using the system, we performed a system validation, with only 5 participants, in which we validate the technological solution, as well as the usability and acceptance, following a user-centered design approach. After this preliminary validation, we were ready to validate the complete frailty assessment system with 78 participants.

4.1 System validation results

We carried out a preliminary study [24] as a proof of concept for the technical validation of the system architecture and the usability and acceptance by end-users, i.e., experts and the older adults. Debes et al. [46] suggested that the acceptance of wearables for Health assessment can be increased with a user-centered design, that guarantees privacy and transparence. Therefore, we not only design at the beginning for elderly people, but also adapt the system to them. Taking this into account, we performed this proof of concept in order not only to validate the technical proposal, but also to adapt the system to the users.

In order to validate the technological solution, older adults were assessed in a real environment (performing a shopping activity at the supermarket). The participants recruited for the proof of concept were five adults coming from a community center in Granada (Spain); three of them were women (1 frail; and 2 non-frails); and two were men (1 frail and 1 non-frail). The average age was 84 years old. All of them performed the shopping IADL, as explain in section 3.5.1, while we collect sensory data using our system architecture. After that, we performed a preliminary Exploratory Data Analysis (EDA) with the collected data revealing that heart rate data could be a relevant feature to distinguish between non-frail and frail participants (Figure 5).

In particular, frail individuals reported a wider range of heart rate over the non-frail individuals. Specifically, heart rate values in frail individuals move between 50 and 90/100; whereas, in non-frail individuals the values move only between 80 and 100. In addition, the mean heart rate values were minor than 88 in frail participants, and greater than 88 in non-frail participants. However, these observations are a proof of concept of the system, and the sample is not representative (only 5 participants) for inferring that wider range of heart rate is a feature exclusive of non-frail individuals. Nonetheless, this proof of concept represents that our system is technically validated, and ready to perform a frailty assessment with a larger dataset.

We also validated our system in terms of usability, acceptance. The supervisor of this preliminary experiment observed the participants and interviewed them about this activity. Participants answered whether the wearable affected their comfort and mobility. The main findings are in line to those of Ehmen et al. [81]: 1) older adults with low sensorimotor abilities can use our system; 2) our system requires little time of use and it is transparent for older adults, because they only have to wear the smartwatch; 3) our system is easy to use for older adults, they do not require any technological competence; and 4) our system does not require extra personnel to train the users or to use the system. In conclusion, all of the participants reported a satisfactory experience. They said that the system was non-intrusive. Participants also reported that: the wearable is easy to wear; the app is easy to learn; and the collected data are secured because they are anonymized.

Therefore, the technological solution was viable, and the system was adapted to the end users (user design centered) without the need to any further tuning. Hence, we started the assessment of the frailty with a larger sample.

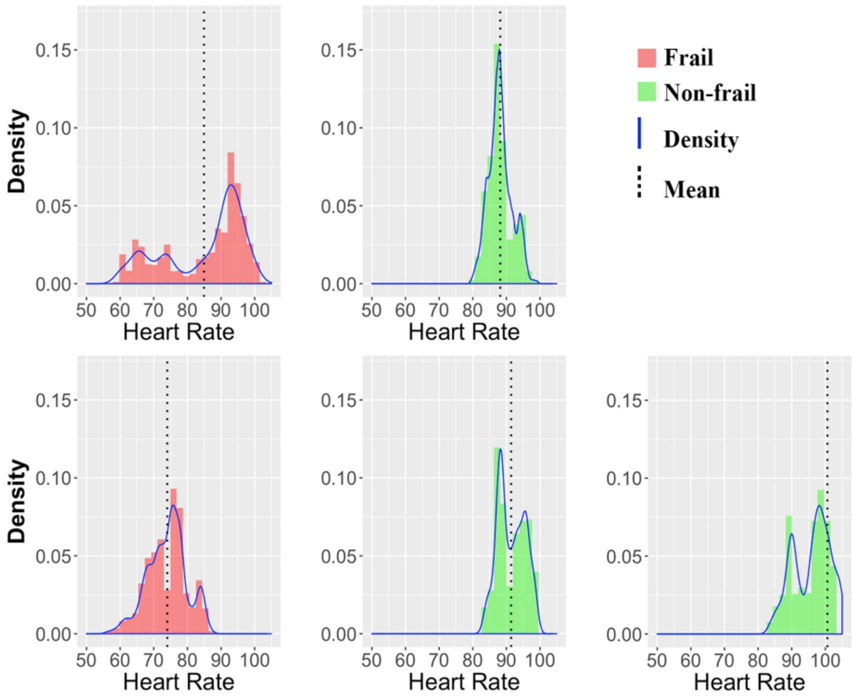

Figure 5. Comparison between frail and non-frail individuals by heart rate [24].

4.2 Frailty Model results

In this section, we present the results of the frailty model building: first, the recount of frailty classes of the recruited sample, with/without considering shopping phases; second, a summary of the ML algorithms used and total features explored; third, the best ML model built and its metrics; and finally, a comparative about how shopping IADL influences in the model performance for frailty assessment.

Wearable sensory data of 79 participants (69 women; 10 men) were collected. However, the exploratory data analysis reported one male with some missing values, thus we discarded this record. Then, the final sample consisted in 78 participants, 12 of them classified as "frail", 47 "pre-frail" and 19 as "non-frail". In addition, we labelled the global shopping activity into 14 shopping tasks or sub-activities (as described in the Materials and Methods section), with the help of an occupational therapists' analysis to define the phases. The final distribution sample is: 1) 168 phases for "non-frail" individuals; 2) 658 phases for "pre-frail" individuals; 3) and 266 phases for "frail" individuals. In total 1,092 samples.

We built four frailty classification models using the machine learning algorithms k-NN, SVM, RF and NB, following the embedded feature selection method (with RF filter to rank features) and using RFE strategy. Then, we used 5-fold stratified cross-validation to evaluate all of models built. Figure 6 shows the metric F1-score over features from 1 to 56 of total amount of features extracted.

Table 3 shows the results. The k-NN model reported the best performance with k=1 (i.e., 1-NN) at 25 Hz, 0.5 seconds windows size and considering only 29 features followed closely by SVM. The best accuracy was 0.99, i.e., it classifies correctly the frailty status in most than 99% of the participants recruited, using only 29 features extracted from the wearable sensors. In 1-NN, F1-score was 0.98. Sensitivity was 0.97 and specificity was 0.99. In addition, we found that for 1-NN (see Table 4), the best frailty status detected was "pre-frail", followed by "non-frail" and, then, "frail", giving priority to sensitivity metrics rather than specificity, which is preferred in health problems.

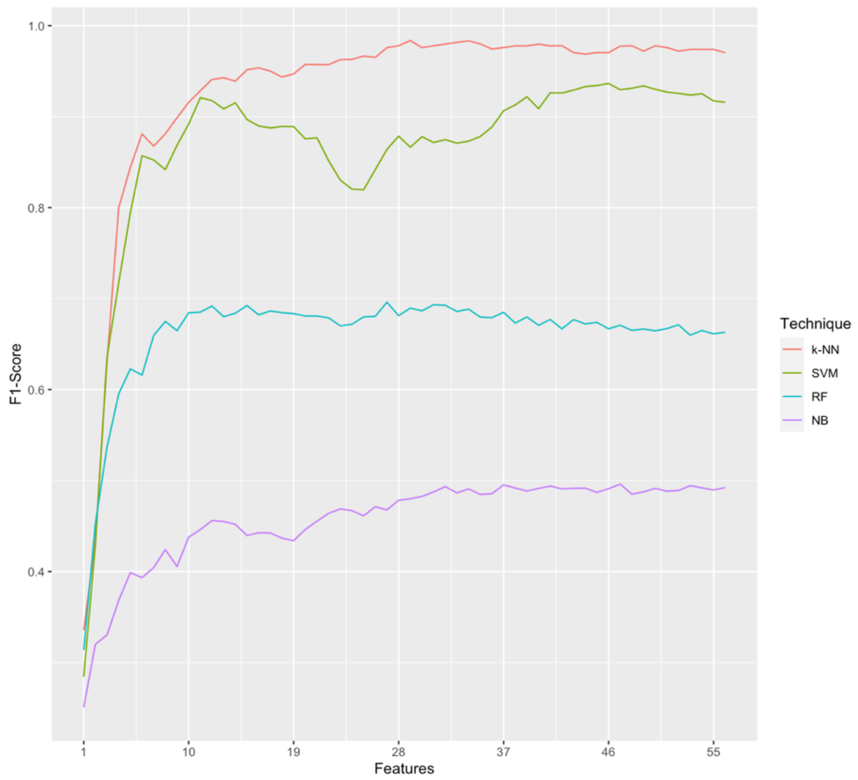

Figure 6. Performance of different machine learning algorithms by RFE embedded feature selection.

Table 3. Performance of different Machine Learning algorithms to assess frailty status.

Algorithm	Features	Accuracy	F1-Score	Sensitivity	Specificity
k-NN [1]	29	0.9917641	0.9837171	0.9764216	0.9947197
SVM	46	0.9670102	0.9364576	0.9108271	0.9779242
RF	27	0.8461648	0.6960141	0.6244533	0.8733734
NB	47	0.6621256	0.4960688	0.4353061	0.7659894

[1] k-NN reached the best performance with k=1 (1-NN).

Table 4. Performance of 1-NN for the three frailty status.

Frailty Status	Sensitivity	Specificity
Frail	0.9375	0.9946237
Pre-frail	0.9851852	0.9879518
Non-frail	0.962963	0.9939024

Furthermore, we used these findings in order to test 1-NN to know which are the most relevant shopping phases for the final assessment of frailty status. To that end, we perform several experiments with different phases or grouping of phases in them. For example, automatic detection of the shopping phases inside the supermarket could pose a challenge. Therefore, we performed an experiment consisting in grouping all these stages in one single phase in order to compare the impact of packaging phases or not. This package process (packed shopping) consisted in reducing the number of all shopping phases samples by computing the arithmetic mean of all the shopping phases values (row 4 in Table 5). As expected, the results indicate that the performance is worst, by 2.6%, if we consider the shopping phases packaged (row 6 in Table 6) than if we considered one single phase separately (rows 5 in Table 6). Figure 7 shows these results and additionally, we can see how the F1-score becomes stable after 20 features approximately for most of the phases. In particular, for the phase "Walking & Sitting/Standing & Shopping" we can see that it increases linearly with the number of features until 15 features, and then improves slowly until 29 features where it reaches its maximum. Furthermore, the most influent shopping phase is "looking for the product" with a 93.2% accuracy, followed by "walking to the checkout".

Table 5. Experiment names and phases (tasks or sub-activities) of shopping considered.

Experiment (Phases)	Tasks or sub-activities
Walking	1) Walking to the supermarket 2) Coming back
Sitting/Standing	1) Sitting 2) Standing 3) Standing at start point 4) Sitting back.
Shopping	1) Participant is in the supermarket 2) Looking for the product to purchase 3) Picking the product 4) Going to the checkout 5) In the checkout 6) Paying 7) Go to the exit 8) In the outside
Packed Shopping	1) Same phases as Shopping experiment but considered as a unique phase by computing the arithmetic mean of the values.

Table 6. Performance of 1-NN in different experiment phases.

Algorithm	Features	Accuracy	F1-Score	Sensitivity	Specificity
Walking [1] & Sitting/Standing [2] & Shopping [3]	29	0.9917641	0.9837171	0.9764216	0.9947197
Walking [1] & Sitting/Standing [2] Packed Shopping [4]	31	0.9503722	0.9036798	0.8792540	0.9705433
Walking [1]	19	0.9425269	0.8927655	0.8656145	0.9650742
Sitting/Standing [2]	21	0.9325653	0.8722884	0.8430592	0.9594234
Shopping [3]	42	0.9359852	0.8785180	0.8397436	0.9588485
Packed Shopping [4]	18	0.9091168	0.8450966	0.8034157	0.9488836
Looking for the product	28	0.9322537	0.8944003	0.8754377	0.9671364
Picking up the product	15	0.8940171	0.8218771	0.7855828	0.9441125
Walking to checkout	8	0.9245014	0.8669296	0.8295743	0.9552347
Waiting for their turn	17	0.851007	0.7638345	0.7256926	0.9190001
Paying	48	0.8863248	0.8151564	0.7741049	0.9367538

[1] Walking: 1) walking to the supermarket; 2) coming back;

[2] Sitting/Standing: 1) sitting; 2) standing; 3) standing at start point; 4) and sitting back.

[3] Shopping: 1) in the supermarket; 2) looking for the product to purchase; 3) picking the product; 4) going to the checkout; 5) in the checkout; 6) paying; 7) go to the exit; 8) in the outside.

[4] Packed Shopping: all phases of Shopping3, but packed in only one phase.

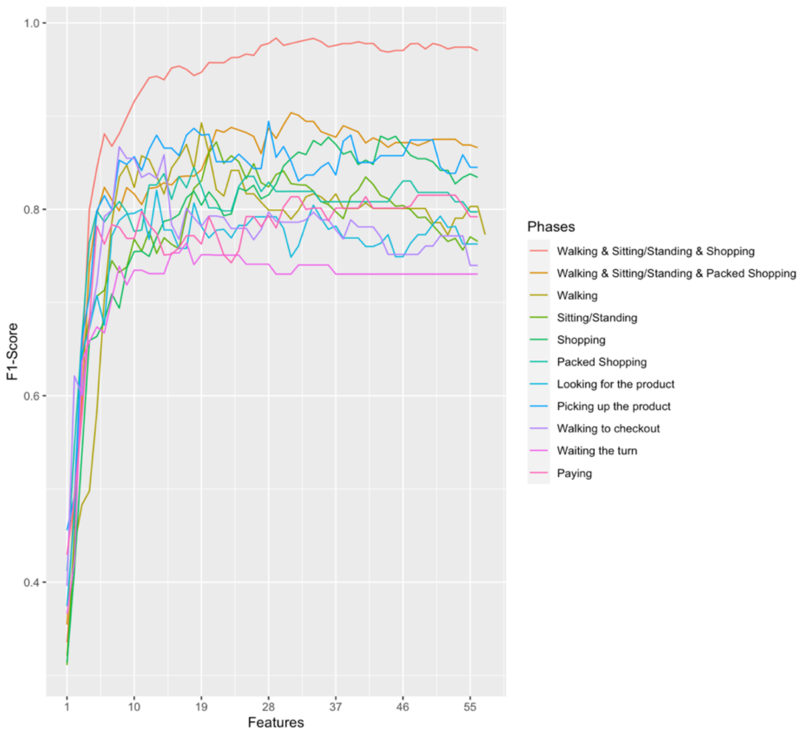

Figure 7. Performance of 1-NN by RFE embedded feature selection over shopping phases. Walking: 1) walking to the supermarket; 2) coming back. Sitting/Standing: 1) sitting; 2) standing; 3) standing at start point; 4) and sitting back. Shopping: 1) in the supermarket; 2) looking for the product to purchase; 3) picking the product; 4) going to the checkout; 5) in the checkout; 6) paying; 7) go to the exit; 8) in the outside. Packed Shopping: all phases of Shopping but packed in only one phase.

5. Discussion

We have created a platform based on microservices, which allows the generalization, extension and reuse of our system functionalities. In particular, we can easily include new sensors coming from different wearables, such as smartwatches, wristbands, bracelets and headbands. Likewise, we can reuse or include new microservices to predict other pathologies (e.g. the dependency status).

We found that the shopping activity (IADL) is equally important than individual BADLs in the assessment of frailty, but the combination of both improves the accuracy. The prediction of frailty status has a similar accuracy considering only the shopping IADL (row 5 in Table 6) or considering only the BADLs (rows 3 and 4 of Table 6) such as sitting, standing and walking outside the supermarket. However, the addition of shopping IADL data to sitting, standing and walking BADLs data (see row 1 in Table 6) increases the accuracy in 5% (from rows 3 and 4 to row 1 in Table 6). This combination reaches the best accuracy of 99.2% using k-NN algorithm with k=1 (1-NN). Then, since k-NN algorithm works with euclidean distance, and provides better results than other ML algorithms tested, this suggests our classes (three frailty status) are quite separable. Furthermore, the most influent shopping phase is “looking for the product”, with 93.2% accuracy followed by “walking to the checkout”. These results could be due to the fact that both phases involve the physical activity of walking, which is the BADL with the best results. The worst accuracy was reached by the “waiting in turn” phase, which probably is due to the fact that it does not involve any physical or cognitive activity and the sensors could not discerned frailty form non-frailty older adults. The second worst results (although still a good results) was provided by the “paying” phase. This result is probably due to the fact that paying involves cognitive and social factors, that we do not considered when we labelled the data with Fried test. Thus, we will explore in the future the use of sensors, such as Electrodermal Activity (EDA), which could sense physiological signals which are known to be directly correlated to cognitive and social functions; as well as labelling the data with other frailty tests which considered social and cognitive functions. Nonetheless, our findings confirm the relationship between the IADLs performance with the frailty status of older adults [9] and state that the combination of BADLs and IADLs improve the accuracy.

Our findings with “shopping packed” are relevant, because the detection of a person inside of a supermarket (and therefore performing the “shopping packed” phase), could be easily detected, for example, using a GPS location sensor or some IADL recognition methods available in the literature. Whereas the detection of each phase inside the supermarket could be a challenge or require more resources, such as fixing sensors inside the supermarket.

Our frailty model outperforms the previous works. In particular, it outperforms the most relevant study we found compared to ours [14], as mention in section 2, which has an accuracy of 82.33% and sensitivity of 83.83% classifying the three classes of frailty status (frail, pre-frail and non-frail), but using only BADLs. Also, our model has an accuracy 1.26% higher than the other experiment performed by Tegou et al. [14], which divided the sample in two classes (group 1: dentification of frail participants; group 2: non-frail and prefrail participants together as a single class). In addition, we outperform [18] from 85% to 99% accuracy, which used embedded feature selection approach as our proposal.

Our proposal is novel because it is ecological, low-cost, wearable-based to automatize the data collection non-intrusively —applying ML techniques to classify the three classes of frailty syndrome—, microservices based architecture for healthcare systems. Furthermore, our proposal outperformed similar studies, which measure Frailty with ML, but do not have all the above characteristics.

This paper may have some limitations. Firstly, the recruited sample had an unequal gender distribution, which is common in this kind of experiments with elder people. Likewise, the sample is imbalance, due to the bigger number of "pre-frail" people in this age range and able to perform IADLs, than the number of "frail" or "non-frail".

The benefits of our results consist on the automatic assessment of frailty status during IADLs, which involves an ecological approach. This could reduce health professionals' time, health costs, and could be the basis for a preventive intervention to reduce frailty. Our proposal is ecological because, we can measure frailty of older adults in their own environment (their daily living), without disturbing them, in a non-intrusive and transparent manner. A consequence of this ecological point of view, is the reduction of costs and time in the assessment of frailty, in comparison with traditional frailty tests. Furthermore, due to the adoption of microservices approach in our system architecture we can take advantage of all of its properties, such as extensibility and reusability. For that reason, we are able to extend the system with different frailty model microservices or including different sensors. Moreover, our frailty model contributes to an early detection of frailty status, which is a key point for prevention [4,5], allowing health professionals to make early decisions for performing specific interventions.

6. Conclusions and future work

In this paper, we presented a novel proposal for e-health, which is a platform based on microservices for assessing frailty status of older adults, ecologically during the performance of an IADL, such as shopping, which involves physical, cognitive and social human functions. To that end, we designed a microservice architecture to support sensory data collection from wearables, and ML analytics. Hence, our proposal take advantage of the microservices characteristics, such as extensibility, reusability and scalability. The main potential of the microservices architecture is that it could be extended with little effort. For example, we could easily include new data sources, new sensors, new business functionalities, or new frailty models based on different assessment methods. Our ML model has an accuracy of 99.2%, a 98.4% F1-score, a 97.7% sensitivity and a 99.5% specificity, outperforming those of the previous works. Our system is ecological, because it does not disturb the daily living of elderly people. Our frailty assessment is performed in a transparent and non-intrusive way, thanks to the use of mobile computing technologies.

Our proposed model can help healthcare professionals in the early detection of frailty, reducing costs and time. It can help clinicians in the early detection of frailty, because the monitorization could be done automatically and continuously, even before any symptoms appear. Hence, clinicians could design personalized intervention plans to prevent or revert frailty. Traditional frailty assessments require to perform several tests of gait, balance, physical activities and self-report questionnaires. Thus, healthcare services could adopt our solution, because its implementation implies a more objective assessment than the traditional tests, which involves questions about the feelings of your personal health and independence. Our solution also implies a time reduction and a decrease of cost with respect to the conventional assessments.

Author Contributions

"Conceptualization, All authors; methodology, F.M.G.M., J.L.G. and M.B.E.; software, F.M.G.M.; validation, F.M.G.M.; formal analysis, F.M.G.M.; investigation, F.M.G.M., M.J.R.F. and M.B.E.; resources, All authors; data curation, E.R.G., J.M.P.M., F.M.G.M. and M.B.E.; writing—original draft preparation, F.M.G.M.; writing—review and editing, M.B.E., J.L.G. and M.J.R.F.; supervision, M.J.R.F.; project administration, J.L.G. and M.J.R.F.; funding acquisition, J.L.G. and M.J.R.F. All authors have read and agreed to the published version of the manuscript."

Funding

This research was funded by the Spanish Ministry of Economy and Competitiveness - Agencia Estatal de Investigación - with European Regional Development Funds (AEI/FEDER, UE) through the project ref. TIN2016-79484-R, and the Scholarship Program FPU Ref. FPU18/00287 granted by the Spanish Ministry of Science, Innovation and Universities.

References

1. Carretero, L.; Navarro-Pardo, E.; Cano, A. Progression in healthy ageing: frailty, cognitive decline and gender in the European Innovation Partnership for Active and Healthy Ageing. Eur. J. Psychiatry 2015, 29, 231–237.

2. Xue, Q.-L. The Frailty Syndrome: Definition and Natural History. Clin. Geriatr. Med. 2011, 27, 1–15.

3. Toosizadeh, N.; Wendel, C.; Hsu, C.-H.; Zamrini, E.; Mohler, J. Frailty assessment in older adults using upper-extremity function: index development. BMC Geriatr. 2017, 17, 117.

4. Fried, L.P.; Tangen, C.M.; Walston, J.; Newman, A.B.; Hirsch, C.; Gottdiener, J.; Seeman, T.; Tracy, R.; Kop, W.J.; Burke, G.; et al. Frailty in Older Adults: Evidence for a Phenotype. J. Gerontol. Med. Sci. Am. 2001, 56, 146–156.

5. van Velsen, L.; Illario, M.; Jansen-Kosterink, S.; Crola, C.; Di Somma, C.; Colao, A.; Vollenbroek-Hutten, M. A Community-Based, Technology-Supported Health Service for Detecting and Preventing Frailty among Older Adults: A Participatory Design Development Process. J. Aging Res. 2015, 2015, 1–9.

6. Gobbens, R.J.; Luijkx, K.G.; Wijnen-Sponselee, M.T.; Schols, J.M. Toward a conceptual definition of frail community dwelling older people. Nurs. Outlook 2010, 58, 76–86.

7. Gobbens, R.J.J.; van Assen, M.A.L.M.; Luijkx, K.G.; Wijnen-Sponselee, M.T.; Schols, J.M.G.A. The Tilburg Frailty Indicator: Psychometric Properties. J. Am. Med. Dir. Assoc. 2010, 11, 344–355.

8. De Witte, N.; Gobbens, R.; De Donder, L.; Dury, S.; Buffel, T.; Schols, J.; Verté, D. The comprehensive frailty assessment instrument: Development, validity and reliability. Geriatr. Nurs. (Minneap). 2013, 34, 274–281.

9. Pérez Mármol, J.M.; Flores Antigüedad, M.L.; Castro Sánchez, A.M.; Tapia Haro, R.M.; García Ríos, M. del C.; Aguilar Ferrándiz, M.E. Inpatient dependency in activities of daily living predicts informal caregiver strain: A cross-sectional study. J. Clin. Nurs. 2018, 27, e177–e185.

10. Kekade, S.; Hseieh, C.-H.; Islam, M.M.; Atique, S.; Mohammed Khalfan, A.; Li, Y.-C.; Abdul, S.S. The usefulness and actual use of wearable devices among the elderly population. Comput. Methods Programs Biomed. 2018, 153, 137–159.

11. Harari, Y.N. Yuval Noah Harari: the world after coronavirus Available online: https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75 (accessed on Apr 1, 2020).

12. Majumder, S.; Mondal, T.; Deen, M. Wearable Sensors for Remote Health Monitoring. Sensors 2017, 17, 130.

13. Wang Member, F.; Stone Student Member, E.; Skubic Member, M.; Keller Fellow, J.M.; Abbott, C.; Rantz Member, M. Towards a Passive Low-Cost In-Home Gait Assessment System for Older Adults. IEEE J Biomed Heal. Inf. 2013, 17.

14. Tegou, T.; Kalamaras, I.; Tsipouras, M.; Giannakeas, N.; Votis, K.; Tzovaras, D. A Low-Cost Indoor Activity Monitoring System for Detecting Frailty in Older Adults. Sensors 2019, 19, 452.

15. Schwenk, M.; Mohler, J.; Wendel, C.; D"Huyvetter, K.; Fain, M.; Taylor-Piliae, R.; Najafi, B. Wearable Sensor-Based In-Home Assessment of Gait, Balance, and Physical Activity for Discrimination of Frailty Status: Baseline Results of the Arizona Frailty Cohort Study. Gerontology 2015, 61, 258–267.

16. Kumar, D.P.; Toosizadeh, N.; Mohler, J.; Laksari, K. FRAILTY ASSESSMENT BASED ON THE QUALITY OF DAILY WALKING. Innov. Aging 2019, 3, S85–S85.

17. Greene, B.R.; McManus, K.; Redmond, S.J.; Caulfield, B.; Quinn, C.C. Digital assessment of falls risk, frailty, and mobility

impairment using wearable sensors. npj Digit. Med. 2019, 2, 125.

18. Razjouyan, J.; Naik, A.; Horstman, M.; Kunik, M.; Amirmazaheri, M.; Zhou, H.; Sharafkhaneh, A.; Najafi, B. Wearable Sensors and the Assessment of Frailty among Vulnerable Older Adults: An Observational Cohort Study. Sensors 2018, 18, 1336.

19. Fontecha, J.; Navarro, F.J.; Hervás, R.; Bravo, J. Elderly frailty detection by using accelerometer-enabled smartphones and clinical information records. Pers. Ubiquitous Comput. 2013, 17, 1073–1083.

20. González, I.; Fontecha, J.; Hervás, R.; Bravo, J. Estimation of Temporal Gait Events from a Single Accelerometer Through the Scale-Space Filtering Idea. J. Med. Syst. 2016, 40, 251.

21. Zhu, L.; Mark Speechley, S. Measuring Community Mobility in Older Adults with Parkinson' s Disease Using A Wearable GPS Sensor And Self-report Assessment Tools Graduate Program in Epidemiology and Biostatistics, 2017.

22. DeMasi, O.; Feygin, S.; Dembo, A.; Aguilera, A.; Recht, B. Well-Being Tracking via Smartphone-Measured Activity and Sleep: Cohort Study. JMIR mHealth uHealth 2017, 5, e137.

23. Do, T.M.T.; Gatica-Perez, D. Human interaction discovery in smartphone proximity networks. Pers. Ubiquitous Comput. 2013, 17, 413–431.

24. García-Moreno; Rodríguez-García; Rodríguez-Fórtiz; Garrido; Bermúdez-Edo; Villaverde-Gutiérrez; Pérez-Mármol Designing a Smart Mobile Health System for Ecological Frailty Assessment in Elderly. Proceedings 2019, 31, 41.

25. Esposito, M.; Minutolo, A.; Megna, R.; Forastiere, M.; Magliulo, M.; De Pietro, G. A smart mobile, self-configuring, context-aware architecture for personal health monitoring. Eng. Appl. Artif. Intell. 2018, 67, 136–156.

26. Schmidt, M.; Obermaisser, R. Adaptive and technology-independent architecture for fault-tolerant distributed AAL solutions. Comput. Biol. Med. 2018, 95, 236–247.

27. Harous, S.; El Menshawy, M.; Serhani, M.A.; Benharref, A. Mobile health architecture for obesity management using sensory and social data. Informatics Med. Unlocked 2018, 10, 27–44.

28. Qi, J.; Yang, P.; Min, G.; Amft, O.; Dong, F.; Xu, L. Advanced internet of things for personalised healthcare systems: A survey. Pervasive Mob. Comput. 2017, 41, 132–149.

29. Guerrero-Contreras, G.; Navarro-Galindo, J.L.; Samos, J.; Garrido, J.L. A Collaborative Semantic Annotation System in Health: Towards a SOA Design for Knowledge Sharing in Ambient Intelligence. 2017.

30. Richards, M. Microservices vs. Service-Oriented Architecture; O'Reilly Media, 2016; ISBN 978-1-491-95242-9.

31. Lu, D.; Huang, D.; Walenstein, A.; Medhi, D. A Secure Microservice Framework for IoT. In Proceedings of the 2017 IEEE Symposium on Service-Oriented System Engineering (SOSE); IEEE, 2017; pp. 9–18.

32. Dragoni, N.; Giallorenzo, S.; Lafuente, A.L.; Mazzara, M.; Montesi, F.; Mustafin, R.; Safina, L. Microservices: Yesterday, Today, and Tomorrow BT - Present and Ulterior Software Engineering. In; Mazzara, M., Meyer, B., Eds.; Springer International Publishing: Cham, 2017; pp. 195–216 ISBN 978-3-319-67425-4.

33. Thönes, J. Microservices. IEEE Softw. 2015, 32, 116–116.

34. Kolozali, S.; Kuemper, D.; Tonjes, R.; Bermudez-Edo, M.; Farajidavar, N.; Barnaghi, P.; Gao, F.; Intizar Ali, M.; Mileo, A.; Fischer, M.; et al. Observing the Pulse of a City: A Smart City Framework for Real-Time Discovery, Federation, and Aggregation of Data Streams. IEEE Internet Things J. 2019, 6, 2651–2668.

35. Guth, J.; Breitenbucher, U.; Falkenthal, M.; Leymann, F.; Reinfurt, L. Comparison of IoT platform architectures: A field study based on a reference architecture. In Proceedings of the 2016 Cloudification of the Internet of Things (CIoT); IEEE, 2016; pp. 1–6.

36. Carranza-García, F.; García-Moreno, F.M.; Rodriguez-Dominguez, C.; Garrido, J.L.; Bermúdez-Edo, M.; Rodriguez-Fortiz, M.J.; Pérez-Mármol, J.M. Supporting Active Ageing Interventions with Web and Mobile/Wearable Technologies and Using

Microservice Oriented Architectures. In First International Workshop, IWoG 2018; García-Alonso, J., Fonseca, C., Eds.; Cáceres, Spain, and Évora, Portugal, 2019; pp. 1–10 ISBN 978-3-030-16028-9.

37. Srinivasa K G; Sowmya BJ; Shikhar, A.; Utkarsha, R.; Singh, A. Data Analytics Assisted Internet of Things Towards Building Intelligent Healthcare Monitoring Systems. J. Organ. End User Comput. 2018, 30, 83–103.

38. Elsaleh, T.; Enshaeifar, S.; Rezvani, R.; Acton, S.T.; Janeiko, V.; Bermudez-Edo, M. IoT-Stream: A Lightweight Ontology for Internet of Things Data Streams and Its Use with Data Analytics and Event Detection Services. Sensors 2020, 20, 953.

39. Abril-Jiménez, P.; Javier, ·; Lacal, R.; Silvia De Los Ríos Pérez, ·; Páramo, M.; Bautista, J.; Colomer, M.; María, ·; Arredondo Waldmeyer, T. Ageing-friendly cities for assessing older adults' decline: IoT-based system for continuous monitoring of frailty risks using smart city infrastructure. Aging Clin. Exp. Res.

40. Esposito, C.; Castiglione, A.; Tudorica, C.-A.; Pop, F. Security and privacy for cloud-based data management in the health network service chain: a microservice approach. IEEE Commun. Mag. 2017, 55, 102–108.

41. Jita, H.; Pieterse, V. A Framework to Apply the Internet of Things for Medical Care in a Home Environment. In Proceedings of the Proceedings of the 2018 International Conference on Cloud Computing and Internet of Things - CCIOT 2018; ACM Press: New York, New York, USA, 2018; pp. 45–54.

42. Jarwar, M.A.; Ali, S.; Chong, I. Exploring Web Objects enabled Data-Driven Microservices for E-Health Service Provision in IoT Environment. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC); 2018; pp. 112–117.

43. Avci, A.; Bosch, S.; Marin-Perianu, M.; Marin-Perianu, R.; Havinga, P. Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. In Proceedings of the 23th International Conference on Architecture of Computing Systems 2010; Hannover, Germany, 2010; pp. 1–10.

44. Mannini, A.; Intille, S.S. Classifier Personalization for Activity Recognition Using Wrist Accelerometers. IEEE J. Biomed. Heal. Informatics 2019, 23, 1585–1594.

45. Twomey, N.; Diethe, T.; Fafoutis, X.; Elsts, A.; McConville, R.; Flach, P.; Craddock, I. A Comprehensive Study of Activity Recognition Using Accelerometers. Informatics 2018, 5, 27.

46. Debes, C.; Merentitis, A.; Sukhanov, S.; Niessen, M.; Frangiadakis, N.; Bauer, A. Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior. IEEE Signal Process. Mag. 2016, 33, 81–94.

47. Watanabe, Y.; Sara, S. Toward an Immunity-based Gait Recognition on Smart Phone: A Study of Feature Selection and Walking State Classification. Procedia Comput. Sci. 2016, 96, 1790–1800.

48. Gjoreski, M.; Gjoreski, H.; Luštrek, M.; Gams, M. How Accurately Can Your Wrist Device Recognize Daily Activities and Detect Falls? Sensors 2016, 16, 800.

49. Attal, F.; Mohammed, S.; Dedabrishvili, M.; Chamroukhi, F.; Oukhellou, L.; Amirat, Y. Physical Human Activity Recognition Using Wearable Sensors. Sensors 2015, 15, 31314–31338.

50. Fida, B.; Bernabucci, I.; Bibbo, D.; Conforto, S.; Schmid, M. Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer. Med. Eng. Phys. 2015, 37, 705–711.

51. Garcia-Ceja, E.; Brena, R.; Carrasco-Jimenez, J.; Garrido, L. Long-Term Activity Recognition from Wristwatch Accelerometer Data. Sensors 2014, 14, 22500–22524.

52. Greene, B.R.; Doheny, E.P.; O'Halloran, A.; Anne Kenny, R. Frailty status can be accurately assessed using inertial sensors and the TUG test. Age Ageing 2014, 43, 406–411.

53. Lobo, A.; Ezquerra, J.; Gómez Burgada, F.; Sala, J.M.; Seva Díaz, A. [Cognocitive mini-test (a simple practical test to detect intellectual changes in medical patients)]. Actas Luso. Esp. Neurol. Psiquiatr. Cienc. Afines 1979, 7, 189–202.

54. Patel, S.; Park, H.; Bonato, P.; Chan, L.; Rodgers, M. A review of wearable sensors and systems with application in rehabilitation.

J. Neuroeng. Rehabil. 2012, 9, 21.

55. Genovese, V.; Mannini, A.; Sabatini, A.M. A Smartwatch Step Counter for Slow and Intermittent Ambulation. IEEE Access 2017, 5, 13028–13037.

56. Schwenk, M.; Hauer, K.; Zieschang, T.; Englert, S.; Mohler, J.; Najafi, B. Sensor-Derived Physical Activity Parameters Can Predict Future Falls in People with Dementia. Gerontology 2014, 60, 483–492.

57. Hernandez, J.; Riobo, I.; Rozga, A.; Abowd, G.D.; Picard, R.W. Using electrodermal activity to recognize ease of engagement in children during social interactions. In Proceedings of the Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct; ACM Press: New York, New York, USA, 2014; pp. 307–317.

58. Tizen Device Sensors Available online: https://developer.tizen.org/development/guides/.net-application/location-and-sensors/device-sensors (accessed on Jul 3, 2019).

59. Newman, S. Building Microservices; Loukides, M., MacDonald, B., Eds.; First Edit.; O'Reilly Media: Sebastopol, 2015; Vol. 52; ISBN 978-1-491-95035-7.

60. Nadareishvili, I.; Mitra, R.; McLarty, M.; Amundsen, M. Microservice Architecture: Aligning Principles, Practices, and Culture; O'Reilly Media, Inc, 2016; ISBN 9781491956328.

61. Fielding, R.T. REST: Architectural Styles and the Design of Network-Based Software Architectures, University of California, Oakland, CA, USA, 2000.

62. Shelby, Z.; Sensinode; Hartke, K.; Bormann, C. Constrained application protocol (CoAP); 2013;

63. AMQP Available online: https://www.amqp.org/ (accessed on Apr 4, 2019).

64. MQTT.org MQ Telemetry Transport Available online: http://mqtt.org/ (accessed on Apr 4, 2019).

65. Eugster, P.T.; Felber, P.A.; Guerraoui, R.; Kermarrec, A.-M. The many faces of publish/subscribe. ACM Comput. Surv. 2003, 35, 114–131.

66. Amini, D.A.; Kannenberg, K.; Bodison, S.; Chang, P.F.; Colaianni, D.; Goodrich, B.; Mahaffey, L.; Painter, M.; Urban, M.; Handley-More, D.; et al. Occupational Therapy Practice Framework: Domain and Process (3rd Edition). Am. J. Occup. Ther. 2017, 68, S1.

67. Cleland, I.; Kikhia, B.; Nugent, C.; Boytsov, A.; Hallberg, J.; Synnes, K.; McClean, S.; Finlay, D. Optimal Placement of Accelerometers for the Detection of Everyday Activities. Sensors 2013, 13, 9183–9200.

68. Ogliari, G.; Mahinrad, S.; Stott, D.J.; Jukema, J.W.; Mooijaart, S.P.; Macfarlane, P.W.; Clark, E.N.; Kearney, P.M.; Westendorp, R.G.J.; de Craen, A.J.M.; et al. Resting heart rate, heart rate variability and functional decline in old age. Can. Med. Assoc. J. 2015, 187, E442–E449.

69. Hamedinia, M.R.; Sardorodian, M.; Haghighi, A.H.; Vahdat, S. The Effects of Moderate Swimming Training on Blood Pressure Risk Factors in Hypertensive Postmenopausal Women. Iran. J. Heal. Phys. Act. 2010, 1.

70. Hassler, A.P.; Menasalvas, E.; García-García, F.J.; Rodríguez-Mañas, L.; Holzinger, A. Importance of medical data preprocessing in predictive modeling and risk factor discovery for the frailty syndrome. BMC Med. Inform. Decis. Mak. 2019, 19, 33.

71. Bermudez-Edo, M.; Barnaghi, P.; Moessner, K. Analysing real world data streams with spatio-temporal correlations: Entropy vs. Pearson correlation. Autom. Constr. 2018, 88, 87–100.

72. Godfrey, A. Wearables for independent living in older adults: Gait and falls. Maturitas 2017, 100, 16–26.

73. Zak, I.; Klein, I.; Katz, R. A Feasibility Study of Machine Learning Based Coarse Alignment. Proceedings 2018, 4, 50.

74. Ravi, N.; Dandekar, N.; Mysore, P.; Littman, M.L. Activity Recognition from Accelerometer Data; 2005;

75. Preece, S.J.; Goulermas, J.Y.; Kenney, L.P.J.; Howard, D. A Comparison of Feature Extraction Methods for the Classification of

Dynamic Activities From Accelerometer Data. IEEE Trans. Biomed. Eng. 2009, 56, 871–879.

76. Zhao, Z.; Morstatter, F.; Sharma, S.; Alelyani, S.; Liu, H. Advancing Feature Selection Research − ASU Feature Selection Repository;

77. Zhang, M.; Sawchuk, A.A. USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In Proceedings of the Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12; ACM Press: New York, New York, USA, 2012; pp. 1036–1043.

78. Lee, H.; Joseph, B.; Enriquez, A.; Najafi, B. Toward Using a Smartwatch to Monitor Frailty in a Hospital Setting: Using a Single Wrist-Wearable Sensor to Assess Frailty in Bedbound Inpatients. Gerontology 2018, 64, 389–400.

79. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. J. Mach. Learn. Res. 2003, 3, 1157–1182.

80. Patel, S.; Hughes, R.; Huggins, N.; Standaert, D.; Growdon, J.; Dy, J.; Bonato, P. Using wearable sensors to predict the severity of symptoms and motor complications in late stage Parkinson's Disease. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; IEEE, 2008; pp. 3686–3689.

81. Ehmen, H.; Haesner, M.; Steinke, I.; Dorn, M.; Gövercin, M.; Steinhagen-Thiessen, E. Comparison of four different mobile devices for measuring heart rate and ECG with respect to aspects of usability and acceptance by older people. Appl. Ergon. 2012, 43, 582–587.

3. A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors

[B] Garcia-Moreno, F. M., Bermudez-Edo, M., Rodríguez-García, E., Pérez-Mármol, J. M., Garrido, J. L., Rodríguez-Fórtiz, M. J. (2022). A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors. *International Journal of Medical Informatics*, *157*, 104625. DOI: 10.1016/j.ijmedinf.2021.10462.

a. Estado: publicado.

b. Factor de impacto (JCR 2020): 4.046.

c. Categoría: Health Care Sciences & Services. Posición: 26 / 107 (**Q1**).

d. Categoría: Computer Science, Information Systems. 50 / 161 (**Q2**).

e. Categoría: Medical Informatics. 13 / 30 (**Q2**).

f. Nº de citas: 5 (Fuente, Google Scholar).

Este artículo se puede encontrar en acceso abierto en el siguiente enlace:

https://doi.org/10.1016/j.ijmedinf.2021.10462

A continuación, se proporciona un borrador del mismo para cumplir con los derechos de autor.

Assessment of IADL Dependence in Older Adults with Machine Learning and Wearables

Francisco M. García-Moreno [1], María Bermúdez-Edo [1], Estefanía Rodríguez-García [2], José Manuel Pérez-Mármol [2], María José Rodríguez-Fórtiz [1] and José Luis Garrido [1]

[1]Department of Software Engineering, Computer Sciences School, University of Granada.
[2]Department of Physiology, Faculty of Health Sciences, University of Granada.

Abstract. Background and Objective: The assessment of dependence in older adults currently requires a manual collection of data taken from questionnaires. This process is time consuming for the clinicians and intrudes the daily life of the elderly. This paper aims to semi-automate the acquisition and analysis of health data to assess and predict the dependence in older adults while executing one instrumental activity of daily living (IADL).

Methods: In a mobile-health (m-Health) scenario, we analyze whether the acquisition of data through wearables during the performance of IADLs, and with the help of machine learning techniques could replace the traditional questionnaires to evaluate dependence. To that end, we collected data from wearables, while older adults do the shopping activity. A trial supervisor (TS) labelled the different shopping stages (SS) in the collected data. We performed data pre-processing techniques over those SS and analyzed them with three machine learning algorithms: k-Nearest Neighbors (k-NN), Random Forest (RF) and Support Vector Machines (SVM).

Results: Our results confirm that it is possible to replace the traditional questionnaires with wearable data. In particular, the best learning algorithm we tried reported an accuracy of 97% in the assessment of dependence. We tuned the hyperparameters of this algorithm and used embedded feature selection technique to get the best performance with a subset of only 10 features out of the initial 85. This model considers only features extracted from four sensors of a single wearable: accelerometer, heart rate, electrodermal activity and temperature. Although these features are not observational, our current proposal is semi-automatic, because it needs a TS labelling the SS (with a smartphone application). In the future, this labelling process could be automatic as well.

Conclusions: Our method can semi-automatically assess the dependence, without disturbing daily activities of elderly people. This method can save clinicians' time in the evaluation of dependence in older adults and reduce healthcare costs.

***Keywords**: dependence assessment, IADL, older adults, machine learning, wearable sensors, e-health, prediction.*

1. Introduction

Activities of Daily Living (ADLs) play an important role in the health status, well-being and the prevention of dependence [1]. Basic ADLs (BADLs) are survival and self-care activities [2], while instrumental ADLs (IADLs) require cognitive and motor complexity and imply an interaction with the social environment that surrounds the persons [3]. IADLs performance is considered a direct index of the health status because IADLs involve motor, cognitive or social functions. Also, IADL refers to activities to support daily living within the home and community that, depending on the situation, require more complex interactions than those used in ADLs. The performance of IADLs is an important health indicator that can predict mild or several cognitive impairments, such as dementia, and mortality in older adults [4]. The early detection of the state of IADL dependence in older adults and the activation of a rehabilitation plan avoids the establishment of functional dependence and several additional disorders in older adults, such as musculoskeletal problems, hearing impairments, cataract, falls rate, depression [5] and even dementia (with a high conversion rate from IALD dependence to dementia) [6]. Hence, the early detection of dependence in elderly could reduce socioeconomic costs in healthcare services, hospitalizations, deterioration in some chronic diseases, comorbidities, and even mortality rates. In particular, shopping IADL involves interaction with different tools, devices and other people [7]. For these reasons, shopping usually has higher complexity than other IADLs, and therefore, shopping may represent the gold standard to evaluate the performance in IADLs.

Traditionally, there are different scales to evaluate the performance in IADLs, being the Lawton and Brody scale (LBS) [3] the most used. LBS is holistic, because it evaluates cognitive, motor, and social components. It is not ecological because it is not automated; it needs observation of IADLs by clinicians during a long period of time.

From an ecological perspective (without disturbing the elderly life), the monitorization of older people using wearables contributes to an early detection and prevention of disorders [8,9]. In addition, to reduce the time for in-situ observations, wearables avoid inter- and intra-observer biases. In recent years, wearables have already been used to effectively monitor ADL (such as walking, jogging, sitting or standing) [10,11], due to its low cost, size, weight and energy consumption [12,13].

To address the automation of the data collection and analysis in the health domain, previous works have used mobile-health (m-Health) systems [14–18]. In m-Health mobile technology helps in the monitorization of health status, while the patient is walking or performing other movements. M-Health systems follow a patient-centric approach, collecting and processing data from wearables [14,16,18]. The implementation involves a medical Body Area Network (BAN) and a Mobile Base Unit (MBU) [19,20]. A BAN is a computer network consisting of different wearables located on the body of a person [16,21]. The (MBU) is the central element and acts as a gateway aggregating different sensors (intra-BAN communication) and transmitting the collected data to a back-end system (extra-BAN communication) (see Fig. 1). MBUs can be smartphones, laptops, or other devices with processing and transmission capabilities.

The data collected by the m-Health needs to be analyzed in order to monitor the health status of patients. Previous literature has used Machine Learning (ML) algorithms to select relevant variables and to analyze data. ML also provide better results (higher performance) than other data analytics techniques, when working with sensory data [7,11,21–24]. ML has already been used in ADL recognition (detecting which ADL is executed) [26,27], but not to evaluate the level of dependence during the performance of the ADLs.

The aim of our proposal is to create a machine learning model of IADL dependence in older adults, using data from wearables during shopping. Hence, we could substitute the traditional manual assessment of dependency by an automatic assessment while the elder is performing an IADL. This automatic assessment could be repeated frequently, allowing the early detection of dependence. In particular, we propose a m-Health system to assess IADL dependence automatically and ubiquitously, thus ecologically. Our BANs consist of different wearables transmitting data to the back-end located in the m-Health Service System (mSS). The mSS receives these data and federates them with observational variables supplied by healthcare professionals through their MBUs. Then, our mSS analyses and classifies IADL dependence in older adults using ML. Our holistic and ecological approach would save clinicians' time.

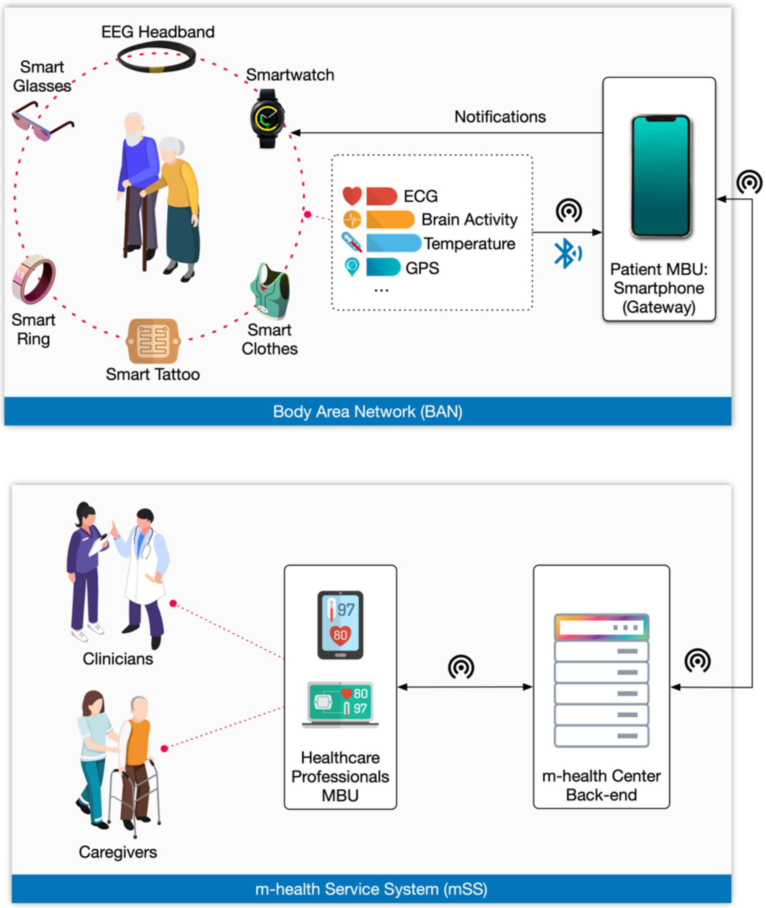

Fig. 1. A generic m-Health scenario overview.

2. Related work

Different tools have been used to evaluate IADL assessment manually, such as Lawton and Brody Scale (LBS), Health and Retirement Study Care Questionnaire, and Pfeffer Functional Activities Questionnaire. Although, the last two have been applied for detection of mild cognitive impairment in IADL [28]. Moreover, LBS is the most used tool used for dependence assessment during the IADL performance and provides an early warning of functional decline [4,29,30].

Introducing wearables with low-cost sensors in common spaces would help older adults in their daily activities and in turn facilitates healthy aging within the community. In addition, new wearable devices through detection, data transmission, algorithms, and IoT applications (m-Health) generate new opportunities for the exploration of movements and activities of daily life in individuals. All this can help health workers with the diagnosis, prevention, intervention and evaluation of the results obtained during the progress of each patient [31].

Most studies in the domain of m-Health focus on ADLs recognition (detecting which ADL a person is performing) such as sitting, walking [10,22,24,26,27]; or even IADLs such as making a phone call, managing money [6] and shopping [32]; or they focus on physiological changes detection [33–35]. Monitoring the performance of these activities daily can be used to know if the elderly have a healthy lifestyle or not [36]. However, none of the previous works assess dependence during an IADL such as shopping. There is another important research line using wearables and ML, which is focused on the prediction or assessment of risk factors in diseases or disorders, such as fall detection [33] and assessment or prediction of motor skills [33, 34]. However, as they detect motor patterns, they use only the data of one or two sensors.

Previous studies have assessed autonomy of IADL in an ecological manner by using ML and video event monitoring systems [37,38]. However, they do not use a validated test to assess dependence (e.g., LBS), require an infrastructure of video-cameras, and were applied for patients with a specific pathology (Alzheimer). Other study [39] assessed the performance of two IADLs using a smartwatch wearable (preparation of a cup of tea and replanting a plant), but with a small sample size of 17 subjects and correlating the performance of ADL with frailty score—which is not the most suitable score for ADL assessment (such as LBS). Other recent study [6] assessed IADLs such as handling money and making a phone call, using a smart home equipped with sensors and a camera. However, they focused on the relationship with the cognitive impairment, not in IADL dependence and without a validated scale.

3. Materials and methods

3.1 Protocol and m-Health scenario for assessing IADL dependence

The protocol, defined by our health experts, consists of 14 Shopping Stages (SS) (see Table 1). Each subject with two wearables in the dominant hand [10,40] sits in a chair without armrests. Next, the trial supervisor (TS) pairs via Bluetooth these devices to a smartphone application (MBU gateway in Fig. 2) and starts capturing data through the sensors [18–20]. The subject performs all the SS (while the TS labels them) and may ask for help for finding the shopping product. The m-Health Service System (or back-end) receives the data in the "Analysis of Data" back-end. After the model is created, in a real scenario, we could assess the dependency without the need to label the data (i.e., without the need to fill in traditional questionnaires). But for training the model, this first time, our health experts used the LBS to evaluate the performance of the subjects in the instrumental ADLs, and the results were received also by the "Analysis of Data" back-end. Finally, with all of these data we built the ML model.

Table 1. Summary of Shopping Stages.

Shopping Stages
1) Sitting
2) Standing
3) Going the supermarket
4) In the supermarket
5) Looking for the product to purchase
6) Picking up the product
7) Going to the checkout
8) In the checkout
9) Paying
10) Going to the exit
11) Going out of the supermarket
12) Coming back to the star point
13) Standing at start point
14) Sitting back

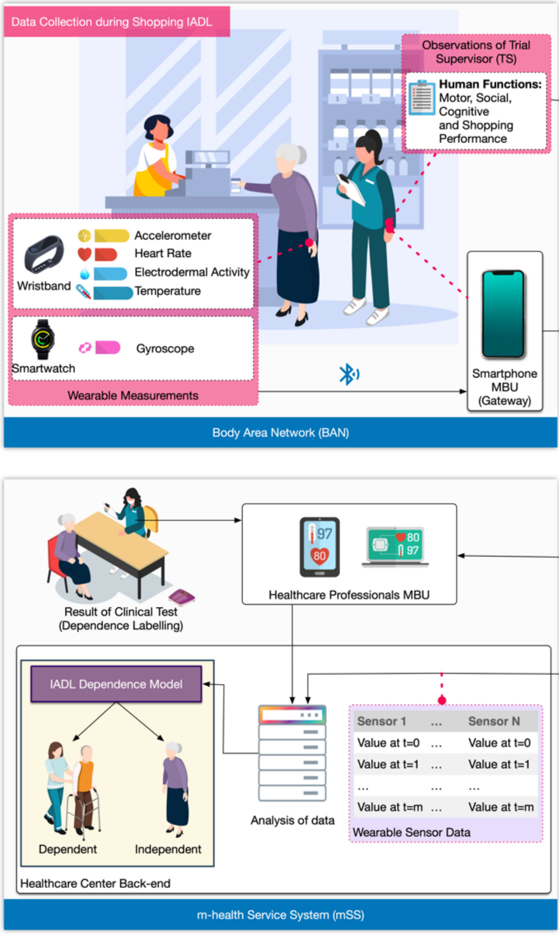

Fig. 2. Overview of the proposed m-Health scenario.

3.2 Recruitment

We conducted a cross-sectional study in two community day centers for active participation of older people, because they assist people who may be at risk to have dependence (or with a dependence already established), and this was our target population. The sampling was executed in a consecutive manner. Sample comprises seventy-nine subjects (69 women and 10 men) aged 65 years or over, with an average age of 75 years. Inclusion criteria of the subjects were: 1) age from 65 to 90; 2) without severe cognitive impairment according to the Spanish validation of the Pfeiffer test [41]; 3) without perceptual alterations; 4) walking with or without help; 5) community-dwelling older adults. Exclusion criteria were: 1) severe mental disorder; 2) disability or severe language alterations; 3) medical instability; 4) pathology in acute period; 5) hospitalized; 6) serious behavior alterations or motor risk.

3.3 Clinical scale

We have used LBS [3] as the dependent variable. LSB has eight questions to assess the ability to perform the tasks of using the telephone, shopping, food preparation, housekeeping, laundry, travelling via car or public transportation, ability to take own medication and ability to handle finances. The index ranges from 0 (dependent) to 8 (independent). Vergara et al. [30] showed a satisfactory validity and reliability of the scale with Cronbach's alpha coefficient 0.94. We stored the answer to each item as 0 (unable) or 1 (able).

3.4 Sensors and observational variables

Our independent variables are collected from different sources: wearables (Table 2), questionnaires and observational inputs (Table 3). IADLs assessed with LBS involve measuring physical, cognitive and social functions. These functions can also be registered through different wearables and some observational variables. We need the observational variables because our current wearables are not able to measure some cognitive and social functions. In the future, we hope to have low-cost sensors that can measure said functions.

The wearables with built-in sensors used in this study are an Empatica E4 wristband [42] and a Samsung Gear S3 smartwatch [43]. Empatica E4 is certified for obtaining accurate and precise physiological data [44], but it lacks a gyroscope sensor, which we think it is important for our study. Both of the devices have open Software Development Kit (SDK) to develop custom applications.

Table 2. Sensors used from two wearable devices.

Device	Signal description
Empatica E4 wristband	Accelerometer x-axis Accelerometer y-axis Accelerometer z-axis Heart rate Electrodermal activity Infrared Thermopile
Samsung Gear S3 smartwatch	Gyroscope x-axis Gyroscope y-axis Gyroscope z-axis

Table 3. Description of observational features considered per experiment.

Human function	ID	Observational feature	Value	Ex1	Ex2
Motor	Ob1	Technical help for walking	Cane / Walker / None	X	X
	Ob2	Average speed during the walk to supermarket	Number	X	
	Ob3	Average speed during the walk back	Number	X	
	Ob4	Average speed per Shopping Stage	Number		X
Cognitive	Ob5	The subject needs help to find the product	Yes / No	X	X
	Ob6	The subject has great difficulty to complete the experiment	Yes / No	X	X
Social	Ob7	How many times per week the subject needs someone to go shopping	Number	X	X
	Ob8	The subject asks for help	Yes / No	X	X
Shopping Performance	Ob9	How many times per week the subject goes shopping	Number	X	X
	Ob10	The subject is tired at the end of the experiment	Yes / No	X	X
	Ob11	The subject has previously shopped at the chosen supermarket	Yes / No	X	X
	Ob12	The subject knows the location of the product	Yes / No	X	X
	Ob13	Distance to the supermarket where the subject usually shops	Number	X	X
	Ob14	Time to find the product	Number	X	
	Ob15	Shopping Stage Identifier	Number		X
	Ob16	Duration of Shopping Stage	Number		X

The experiments consider a total of 13 observational variables.

Note Ob14 is included in Ob16 (it is the duration of a particular shopping stage). The same for Ob2 and Ob3, which are included in Ob4.

3.5 Data Analysis

Fig. 3 shows the proposed ML pipeline. ML needs a training phase with labelled data (box 1 in Fig. 3). With these labelled data the ML algorithm creates a binary classification model (box 7 in Fig. 3), where the output specifies the dependent or independent status. In order to increase the performance of the classification, we pre-process the data (box 2 to 6 in Fig. 3). In addition, we applied IJMEDI checklist [45] (listed in Appendix A) for assessment of the quality of the work on medical AI.

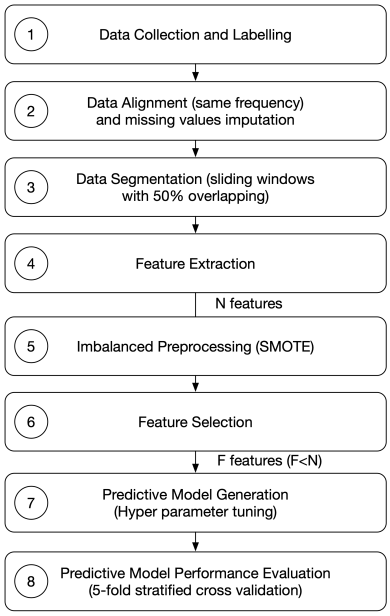

Fig. 3. Machine learning pipeline of the IADL dependence classification process. N features are the total features extracted from the Data Segmentation step. The feature selection step selects the F most relevant features for this study out of the N total features collected (F < N).

3.5.1 Data collection and labelling

We used sampling rates greater than 10 Hz [22]. For the triaxial gyroscope sensor we used 25 Hz [46]. We used the default values from the Empatica E4 wristband: EDA and IT at 4 Hz, heart rate at 1 Hz and accelerometer at 32 Hz. Anomalous beat per minute values from the heart rate sensor were excluded [47,48] based on Equation 1.

$$\text{maximum} = 220 - \text{subject age} \qquad (1)$$

We divide the sample into two groups (dichotomization) [49]. A subject is independent if LBS equals 8; and dependent if it is lower than 7 [50].

3.5.2 Signals alignment and segmentation

We stored all the signals to the highest rate (25 Hz) and interpolated the missing values in signals with lower rates. Fig. 4 shows an example of the alignment results.

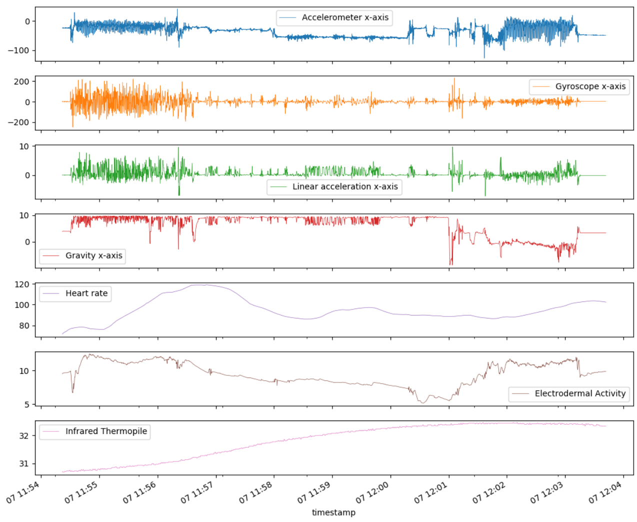

Fig. 4. Some aligned wearable signals from an anonymous subject at 25 Hz during the performance of the shopping IADL.

We used sliding windows to segment the physiological time series of sensory data [18,24,25,51–53] (see Fig. 5). The reason behind using these portions of the data instead of the entire sensor signal is based on distribution and trends of time series [54]. If we only use the entire signal, we will lose the fine granularity of the time intervals when that signal significantly fluctuates or becomes constant. We segmented the raw data with different window sizes measured in seconds (0.5s, 1s, 1.5s, 2s, 2.5s and 3s). The exploration of different ranges of seconds ensures a high precision to capture the physical movements, heart rate variation, and other physiological signals in the subjects. In addition, a 50% overlap between consecutive windows increases the number of samples in a virtual manner.

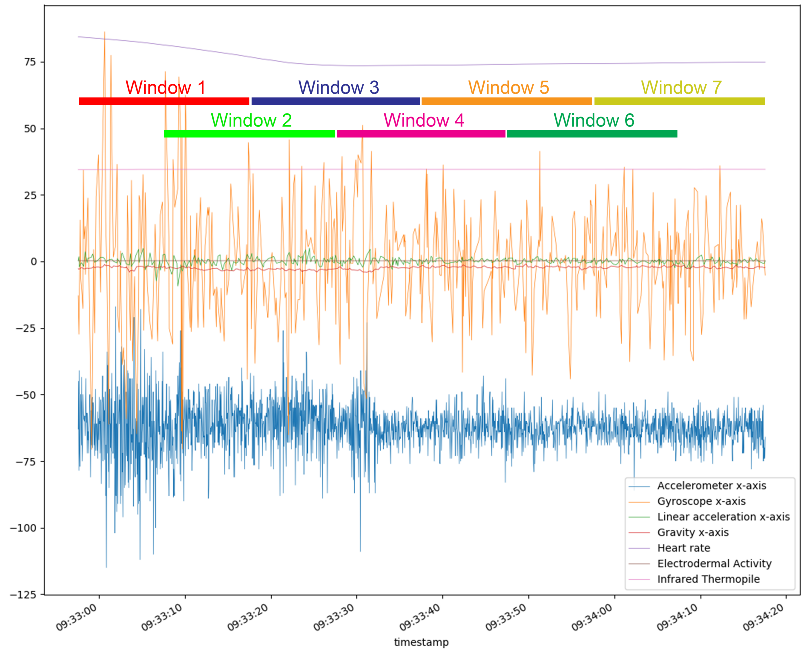

Fig. 5. Sliding windows approach using 50% overlapping.

3.5.3 Sensor Feature extraction

Most of the previous models [10,22–25,54] have shown that using statistics as a form of summary values, concatenating them as subject vectors, can reflect the nature of time series. For example, the standard deviation of a measure—such as heart rate—indicates the degree of dispersion in the distribution, thus a larger value reflects the subject's heart rate fluctuates widely. In fact, a study reported that lower heart rate variability indicates worse IADL dependence [55]. The minimum and maximum observations reflect the range of the measure change, indicating the trend of the data and the centre value. Moreover, the shape of distribution could represent the evolution of the measure [54]: 1) Skewness indicates the symmetry of data distribution, which is the stability of the measure change (e.g., the heart rate change); 2) Kurtosis, reflects the peak sharpness and peak degree, which reveals the trend of fluctuation and the subjects' physiologic state. Hence, to keep the trends and temporal characteristics of our data, the extraction of these features in portions of data is better than using the entire signal data without segmentation.

Therefore, as in previous works [10,22–25], for each sensor we extracted time- and frequency-domain features (summary statistics — F1..Fn, in Fig. 6) from every 50% overlapped window (see Table 4), such as mean, standard deviation, minimum, maximum, skewness, kurtosis and entropy. Since each subject takes different times to finish the experiment, the number of windows (k, in Fig. 6) can be different in every shopping stage and subject. Per each subject and per each shopping stage, we averaged the values of each feature of all the windows inside each shopping stage. The purpose of this processing is double: 1) to reduce the high dimensionality of using thousands of windows separately as the model input; and 2) to capture the distribution and fine granularity trends of our physiological time series in order to improve detection stability and avoiding loss of temporal resolution (central and dispersion tendencies and distribution shape) [54].

The result is a matrix (yellow box in Fig. 6) of all the subjects and all the averaged features per SS. This result will be the input of the ML algorithm. Hence, the number of input samples of the model is determined by the number of shopping stages (14) times the features extracted times the number of subjects.

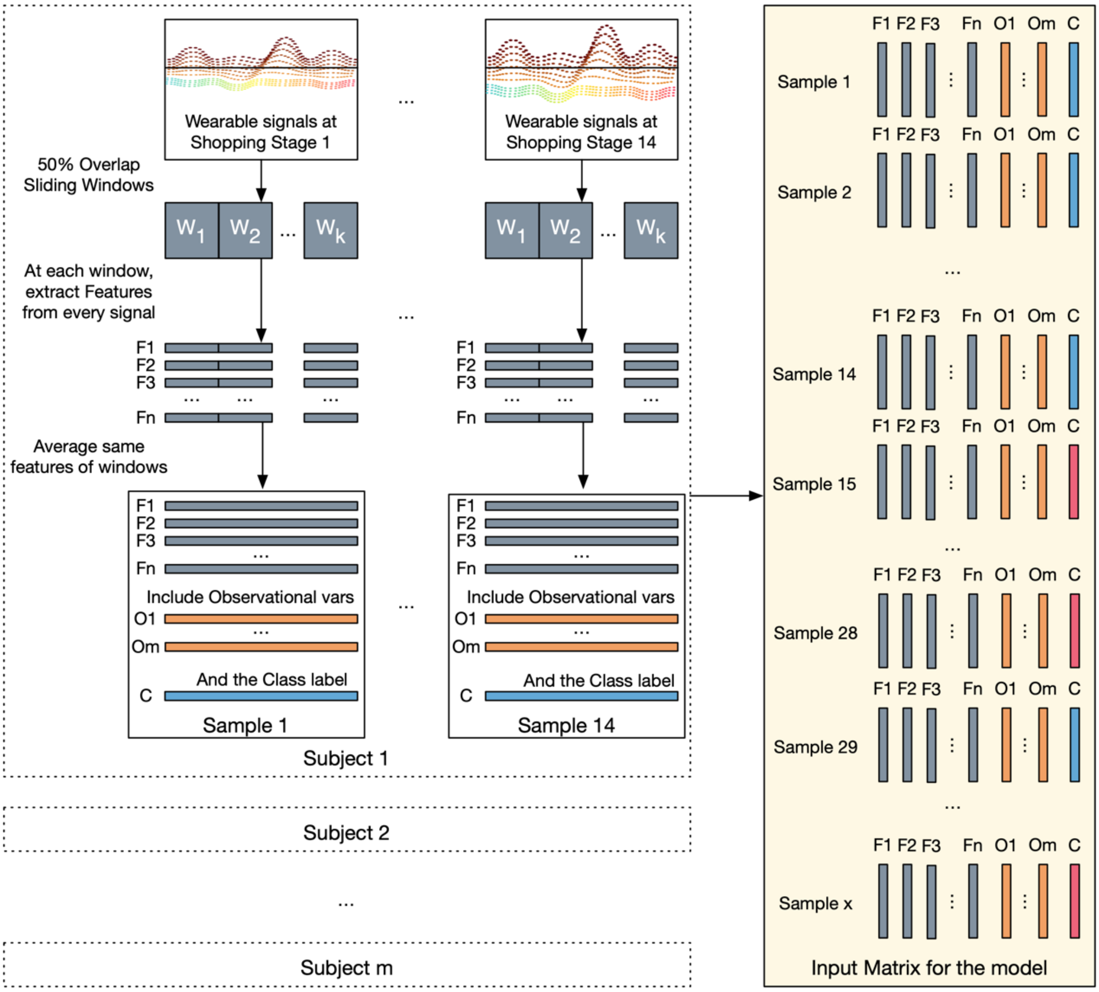

Fig. 6. Creation of the input matrix for each model (k-NN, RF or SVM). The blue color of the field C (label of the sample) represents a dependent subject, while the red color represents an independent subject.

Table 4. Description of sensor features extracted by each sensor axis.

Axis	**Feature**
By each axis (x^a, y^a, z^a and uniaxial [b])	Mean Standard deviation Minimum Maximum Amplitude Skewness Kurtosis Energy

Each feature was extracted by each axis specified and per each sensor (triaxial and uniaxial).

aTriaxial sensor: accelerometer and gyroscope.

bUniaxial sensor: heart rate, EDA and IT.

Total features: (8·3)·2 + (8·3) = 72 wearable features

3.5.4 Imbalance preprocessing

Due to the imbalance of the labelled data (see Fig. 10), we applied the Synthetic Minority Oversampling Technique (SMOTE) for obtaining balanced data and therefore, a higher model performance [56].

3.5.5 Feature selection

Feature selection algorithms are used to deal with the curse of dimensionality in ML, by reducing the number of features with the selection of the most relevant and non-redundant features. We follow an embedded approach [57], based on Random

Forest (RF) algorithm [58] for getting features ranked by their importance. Then, we use a Recursive Feature Elimination (RFE) method for building different models with different subsets of features [40] and select the best one.

3.5.6 Predictive model generation with hyperparameter tuning

Our study applies three different machine learning algorithms: k-NN, RF and SVM, which had good results in previous works on ADL. We created the classifiers with these algorithms and the input matrices (samples), which are the result of the data pre-processing explained in subsection 3.5.3 and Fig. 6.

3.5.7 Predictive Model performance evaluation

To validate our proposal, we use the Cross Validation (CV) technique with the most used indexes: accuracy, sensitivity, specificity and the F1-score. Although we reported all of these indexes, as we have imbalance data, we will focus on F1-score. The second most important index in the health field is the sensitivity, because it better detects the positives although with a higher rate of wrongly classified negatives. For example, it is better to detect a disease when it is there than to not detect it, although we can have a higher rate of healthy subjects detected as ill subjects.

In particular, we use 5-Fold Stratified CV (5-FSCV). The idea behind the k-fold CV is to virtually increase the number of samples by creating different models with different folds (samples) for training and testing the classifiers; allowing testing the model with different test samples in k iterations contributing to obtain a robust model. This technique consists of training the model with k-1 folds (samples) and testing it with the remnant fold. Then, we repeat the train and validation k times selecting each time a different fold for validation. When we have imbalance data it is important to keep the proportion of each class in each fold (stratified CV).

For instance, Fig. 7 shows an example dataset with 20 samples (A-T) and 50% of dependent (A-J) and 50% independent (K-T) subjects—coming from 5 shopping stages (SS 1-5)—, thus we create folds of 4 SS: 2 from dependent subjects and 2 for independent subjects. At each iteration (total iterations: 5, the number of folds), one of the k folds is used to test (see bottom Fig. 7) the model with the selected metric, while the rest of k-1 folds are used to train the model. Since we use 5-fold, at the end of the process we will have 5 performances of every index (such as 5 F1-scores and 5 accuracies), which will be average to report a single test performance for our model.

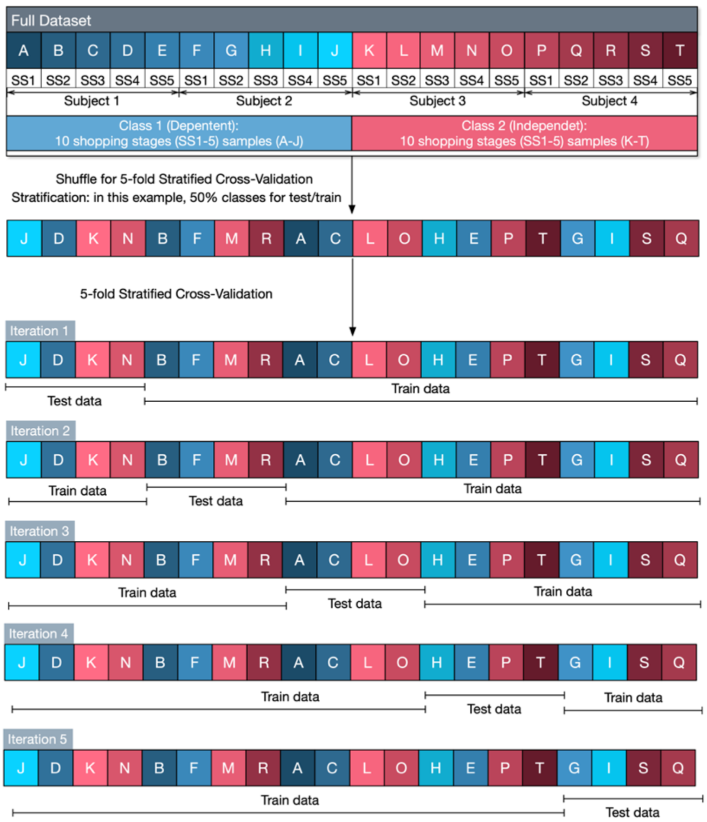

Fig. 7. Example of 5-Fold Stratified Cross Validation with 20 samples.

We are aware of the recent open debate on CV approaches in the health field, record-wise CV (RWCV) vs subject-wise (SWCV) [59,60]. RWCV consists of including samples of the same subject for both train and test sets, while SWCV uses all the samples of the same subject for either train or test. In general, RWCV should be avoided when the aim is to build a model for classifying new subjects, because it could introduce a bias of identity confounding [60,61]. However, Little et al. [60] claim SWCV is not always a valid substitute for RWCV, especially when splitting the dataset in a way that the feature distributions are different per subject and per split (as our shopping stages splits). These differences in the feature distributions can be seen in Fig. 8. The top three graphs in Fig. 8 show an example in which three subjects have different data distribution (histogram shape) of the same feature (accelerometer x-axis) at the same split (shopping stage: paying). The bottom three graphs in Fig. 8 present a new split (shopping stage: going the supermarket) where the feature distributions are different of the feature distributions of the shopping stage paying of the same subject (top graphs vs bottom graphs). Hence, in our case, we used RWCV because we take advantage of the insights of inter- and intra-subject distributions [59]. When our model will try, in the future, to classify a new subject (unseen in the original sample), the model will need a recalibration to reach better results with the new subject. Namely, the first time a new subject uses our model, the model will be recalibrated in order to consider not only the common aspects of dependence (inter-subject detection), but also the personal aspects of dependence (intra-subject).

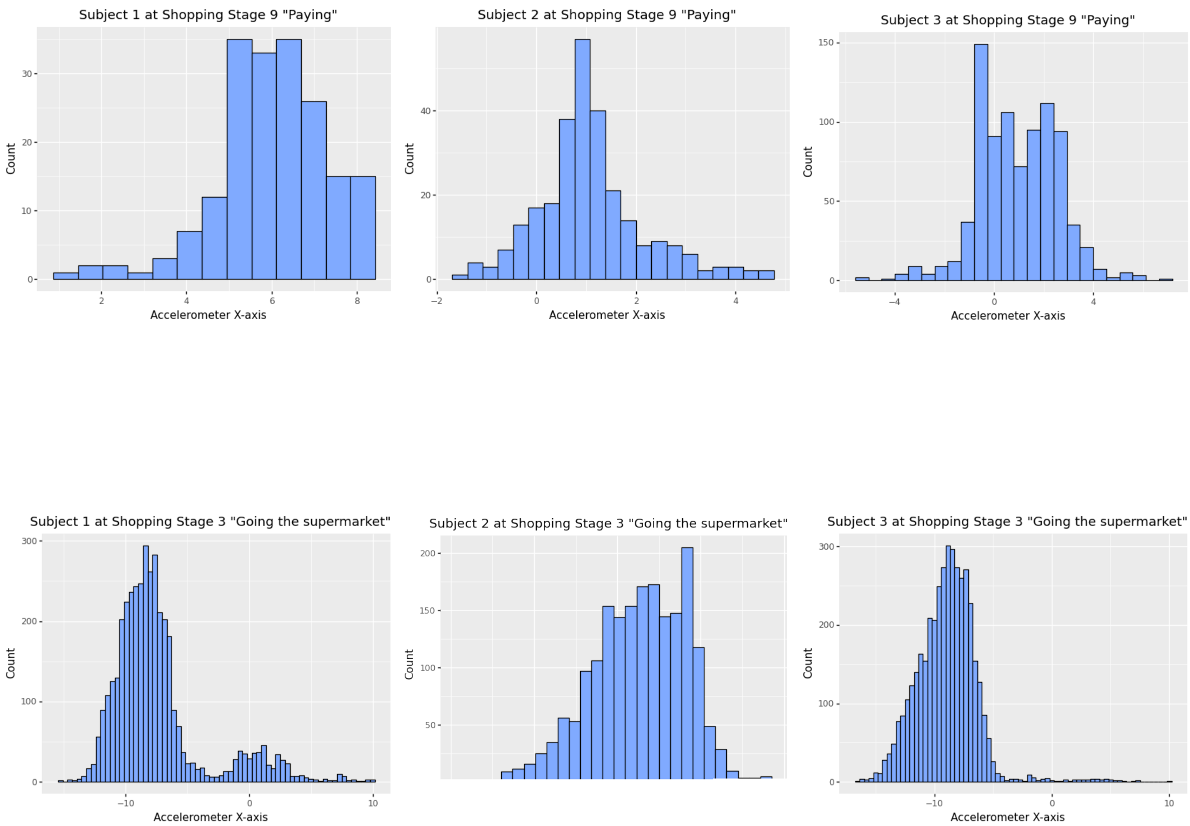

Fig. 8. Example of three subjects' data distributions where Accelerometer x-axis feature is different per subject and per split.

4. Results

We discarded one subject's data due to missing values in several SS. Therefore, the total number of subject's was 78 (69 females and 9 males). We performed a previous experiment with all the SS considered as a unique SS, but in order to improve the results we split the data distinguishing between the SS, which increases the original sample size, because each subject has 14 SS (Table 1). LBS [3] reported: 420 samples classified as dependent in IADLs and 672 classified as independent (1092 samples in total) (see Fig. 9, Fig. 10 and Fig. 11). This distribution is imbalanced, thus applying SMOTE we obtained a proportion of 630 to 630 dependent and independent samples (see Fig. 11).

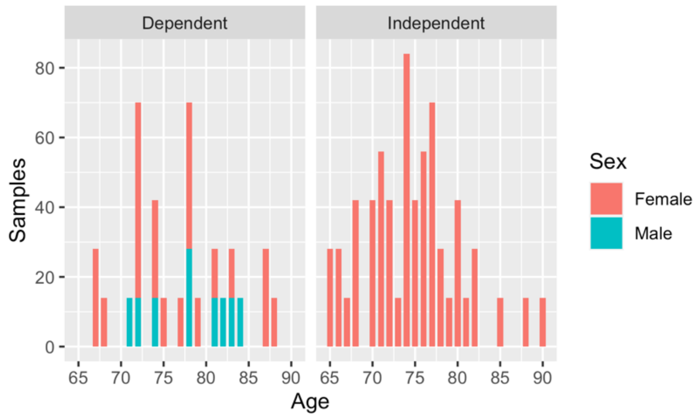

Fig. 9. Ex2: Distribution of subject's age and sex grouped by the result of Lawton & Brody scale.

Fig. 10. Ex2: Distribution of subject's age and sex after SMOTE and grouped by the result of Lawton & Brody scale.

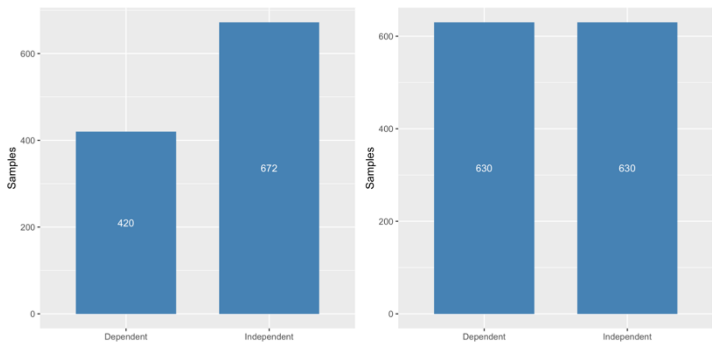

Fig. 11. Ex2: Class proportion comparison between original imbalance data (left) and preprocessed data with Smote to balance the original data (right).

The results were stabilized when we took over 10 features and we achieved the best performance for windows of 0.5 seconds (see the sliding window and feature extraction steps explained in sections 3.5.2 and 3.5.3). As can be appreciated in Fig. 12, with 10 features or more, all the algorithms reported an F1-score above 90%. Additionally, we obtained the perfect performance without SMOTE—100% in every metric with 11 features for 1-NN (see Table 5). The best model was built with 1-NN with only 11 features, followed by SVM (with 65 features) and RF (with 69 features). Among the 11 features selected by 1-NN, there are 3 observational features: one social component (Ob7), and two shopping performance functions (Ob9 and Ob11). Likewise, RF and SVM also include observational features related to motor (Ob1), cognitive (Ob5 and Ob6), social (Ob7 and Ob8) and shopping performance functions (Ob9, Ob10, Ob11 and Ob12).

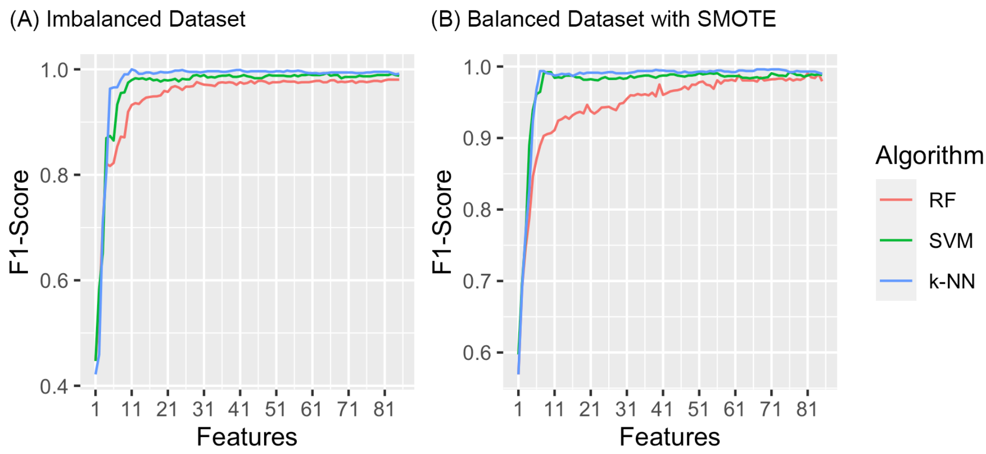

Fig. 12. Ex2: F1-score over all features with the original Dataset (A) and after SMOTE (B)

We also explored models with less features to know if there exists one without observational features, but we did not find anyone. Additionally, there is a model with only 5 features that reaches over 96% in every metric (Table 6). In particular, this model includes the features "acc.x.min" (minimum value of accelerometer x-axis), "temp.min" (minimum value of temperature sensor), "Ob7" (related to social functions), "Ob11" (shopping performance) and "eda.max" (maximum value of EDA sensor). Thus, gyroscope is not present in this model. Therefore, this model with 5 features is the best candidate if we want to use only one wearable, the E4

Table 5. Experiment 2 (split shopping stages): performance of different algorithms without SMOTE

Condition	**Random Forest**	**Support Vector Machines**	**k-Nearest Neighbors**
F1-score	0.9806108	0.992785	**1.0**
Accuracy	0.9853588	0.9945164	1.0
Sensitivity	0.9644629	0.9859331	1.0
Specificity	0.9984848	1.0	1.0
Total Features	82	65	11
Features	Ob7 + Ob9 + acc.x.min + hr.skew + gyr.z.mean + acc.x.energy + eda.max + eda.min + Ob6 + gyr.x.kurtosis + acc.z.energy + eda.mean + Ob11 + gyr.y.kurtosis + temp.max + acc.y.skew + temp.energy + temp.mean + hr.kurtosis + temp.sd + acc.x.mean + acc.z.skew + temp.min + eda.energy + temp.range + acc.y.min + Ob1 + acc.y.max + acc.x.skew + acc.y.mean + hr.range + acc.x.sd+ acc.z.min + hr.sd + acc.x.max + acc.y.kurtosis + gyr.y.skew + gyr.y.mean + eda.range + eda.kurtosis + gyr.z.energy + hr.max + eda.sd + gyr.y.min + acc.z.kurtosis + gyr.x.min + temp.kurtosis + temp.skew + gyr.x.mean + gyr.z.max+ acc.x.range + hr.mean + gyr.x.max + gyr.z.skew + hr.min + hr.energy + acc.y.energy + eda.skew + acc.x.kurtosis + gyr.z.min + gyr.x.skew + gyr.x.sd + gyr.x.energy + acc.z.max + acc.z.mean + acc.y.sd + gyr.y.energy+ gyr.y.range + gyr.z.kurtosis + acc.y.range + gyr.x.range + event_time + gyr.z.range + acc.z.sd + gyr.y.sd + gyr.z.sd + gyr.y.max + Ob8 + acc.z.range + Ob10 + Ob5 + Ob12	Ob7 + Ob9 + acc.x.min + Ob6 + acc.x.energy + hr.skew + gyr.z.mean + eda.mean + gyr.x.kurtosis + eda.min + Ob11 + eda.max + temp.mean + acc.z.skew + acc.z.energy + acc.y.skew+ temp.range + temp.min + acc.x.skew + temp.sd + Ob1 + gyr.y.kurtosis + temp.max + temp.energy + acc.z.min + hr.kurtosis + eda.energy + acc.y.kurtosis + acc.y.mean + acc.x.mean + gyr.z.energy + acc.x.sd + gyr.y.min + eda.sd + acc.y.max + acc.y.min + gyr.z.max + acc.x.max + temp.kurtosis + hr.min + eda.kurtosis + hr.range + gyr.y.mean + acc.z.kurtosis + hr.sd + gyr.z.min + gyr.y.skew + hr.max + eda.range + acc.x.range+ gyr.x.energy + gyr.y.energy + gyr.x.range + acc.z.max + gyr.x.min + acc.z.mean + hr.energy + acc.y.energy + gyr.x.max + hr.mean + gyr.z.kurtosis + acc.x.kurtosis + temp.skew + acc.y.range + acc.y.sd	Ob7 + Ob11 + acc.x.min + Ob9 + eda.min + acc.z.energy + eda.mean + eda.max + temp.min + temp.mean + acc.x.max
Observational	Yes: 9 features (underlined)	Yes: 5 features (underlined)	Yes: 3 features (underlined)
Hyperparameters	number of trees: 500 variables randomly sampled: 10	cost: 0.1 gamma: 0.5 kernel: "polynomial"	k: 1

Best performance with 1-NN of 100% F1-score (bold letters). Empatica sensors: acc (accelerometer); eda (EDA sensor); hr (heart rate); temp (temperature). Gear S3 sensors: gyr (gyroscope)

Table 6. Metrics for 1-NN between 1 and 15 features

Features	**F1-Score**	**Accuracy**	**Sensitivity**	**Specificity**
1	0.4216014	0.6895564	0.2972841	0.9346149
2	0.4590805	0.7307779	0.2997788	1.0000000
3	0.7094287	0.7893343	0.6733311	0.8618040
4	0.8060638	0.8570734	0.7736693	0.9097372
5	**0.9633385**	**0.9716099**	**0.9739039**	**0.9703216**
6	0.9657163	0.9734448	0.9739039	0.9732413
7	0.9659479	0.9734531	0.9762849	0.9716936
8	0.9802166	0.9844414	0.9906690	0.9805949
9	0.9905123	0.9926731	0.9928571	0.9925362
10	0.9905123	0.9926731	0.9928571	0.9925362
11	1.0000000	1.0000000	1.0000000	1.0000000
12	0.9952076	0.9963345	1.0000000	0.9941167
13	0.9940381	0.9954212	0.9976744	0.9941167
14	0.9928391	0.9945038	0.9976744	0.9926347
15	0.9939954	0.995417	1.0000000	0.9926461

Best performance with 11 features (underlined).

Selection of features subset over 96% in every metric and sensitivity greater than specificity (bold letters)

As our final aim is to completely automate the process, we performed additional experiments in which we excluded the observational features (Table 7). Obviously, we need observation with a TS that labels each SS. We believe that in the future, this labelling could be done automatically, either with location sensors or with ML, learning in which SS the subject is at any time. 1-NN performed the best: 99.51% F1-score with 39 wearable features. However, we obtained competitive performance (over 94%) with less features as we can see in Fig. 12 and Table 8. With only 5 features, 1-NN is over 94% in every metric only with the presence of accelerometer, EDA, heart rate and temperature sensors. Since the gyroscope is not present in this model, we can use only the Empatica wristband device with this model. Models with more than 10 features do include gyroscope (see Table 9).

Table 7. Experiment 2 (split shopping stages): performance of different algorithms without SMOTE and non-observational variables

Condition	Random Forest	Support Vector Machines	k-Nearest Neighbors
F1-score	0.9542469	0.981914	0.9951802
Accuracy	0.9661179	0.9862637	0.9963303
Sensitivity	0.9286523	0.9785293	1.0
Specificity	0.9895499	0.9910763	0.9941279
Total Features	69	65	39
Features	hr.skew + acc.x.min + acc.y.kurtosis + gyr.z.mean + gyr.x.kurtosis + acc.z.skew + eda.min + eda.mean + acc.y.skew + acc.z.min + eda.max + temp.mean + temp.max + temp.energy + gyr.y.kurtosis + acc.x.energy + temp.min + gyr.y.skew + acc.x.skew + eda.energy + hr.kurtosis + acc.y.min + temp.range + acc.y.max + acc.z.energy + acc.y.mean + hr.max + temp.sd + acc.z.kurtosis + acc.x.sd + event_time + eda.range + gyr.z.energy+ gyr.z.skew + acc.x.mean + acc.z.max + gyr.y.min+ temp.kurtosis + acc.x.max + eda.sd + acc.z.mean + gyr.z.max + eda.kurtosis + hr.energy + gyr.y.mean + hr.sd + gyr.x.energy + acc.x.kurtosis + hr.min + acc.z.range + gyr.x.max + hr.range + hr.mean + gyr.x.skew + acc.y.energy + acc.y.sd + gyr.z.min + gyr.x.min + gyr.x.sd + gyr.y.energy + acc.x.range + gyr.x.mean + temp.skew + eda.skew + acc.y.range + acc.z.sd + gyr.z.range + gyr.x.range + gyr.z.kurtosis	hr.skew + acc.x.min + acc.y.kurtosis + gyr.z.mean + gyr.x.kurtosis + acc.z.skew + acc.x.energy + acc.z.min + eda.max+ eda.mean + eda.min + temp.min + temp.mean + temp.max + acc.y.skew + gyr.y.kurtosis + acc.x.skew + hr.kurtosis + gyr.y.skew + temp.energy + temp.sd + acc.y.mean + temp.range + acc.y.max + hr.max + acc.x.sd + acc.y.min + eda.energy + acc.z.energy + gyr.y.min + acc.z.max + gyr.z.max + gyr.z.skew + acc.z.kurtosis + event_time + acc.x.kurtosis + temp.kurtosis + acc.x.mean + gyr.z.energy + eda.range + gyr.x.energy + acc.z.mean + gyr.y.mean + hr.sd + acc.x.max + gyr.z.min + eda.kurtosis + hr.mean + gyr.x.min + gyr.x.max + hr.range + eda.sd + hr.energy + acc.y.sd + hr.min + gyr.x.sd + acc.y.energy + eda.skew + gyr.x.skew + acc.x.range + gyr.y.energy + gyr.x.mean + temp.skew + acc.z.sd + acc.z.range	hr.skew + acc.x.min + acc.y.kurtosis + gyr.z.mean + acc.z.min + gyr.x.kurtosis + acc.y.skew + eda.mean + eda.min + acc.z.skew + acc.x.sd + acc.x.energy + hr.kurtosis + acc.x.skew + hr.max + eda.max + temp.min + acc.y.max + gyr.y.kurtosis + eda.range + temp.max + hr.mean + gyr.y.skew + temp.energy + temp.mean + gyr.x.max + gyr.x.range + acc.y.min + gyr.z.max + hr.range + temp.range + eda.energy + temp.sd + acc.z.max + acc.z.energy + acc.x.mean + gyr.y.min + acc.y.mean + acc.z.kurtosis
Hyperparameters	number of trees: 500 variables randomly sampled: 10	cost: 0.1 gamma: 0.5 kernel: "polynomial"	k: 1

Best performance with 1-NN of 99.52% F1-score (bold letters)

Empatica sensors: acc (accelerometer); eda (EDA sensor); hr (heart rate); temp (temperature). Gear S3 sensors: gyr (gyroscope)

Table 8. Metrics for 1-NN between 1 and 15 features

Features	F1-Score	Accuracy	Sensitivity	Specificity
1	0.5299375	0.6913745	0.4525743	0.8411121
2	0.6195547	0.7353525	0.5624456	0.8438088
3	0.8600445	0.8928239	0.8637154	0.9109086
4	0.8711083	0.9019815	0.8709224	0.9212700
5*	0.9428919	0.9560303	0.9451741	0.9629114
6	0.9555843	0.9660927	0.9545607	0.9734273
7	0.9630037	0.9715973	0.9642888	0.9763790
8	0.9633316	0.9716015	0.9691675	0.9734372
9	0.9633316	0.9716015	0.9691675	0.9734372
10	**0.9704221**	**0.9770977**	**0.9786673**	**0.9763903**
11	**0.9796175**	**0.9844162**	**0.9878547**	**0.9823069**
12	**0.9856647**	**0.9890034**	**0.9927930**	**0.9867080**
13	**0.9915874**	**0.9935780**	**1.0000000**	**0.9897161**
14	**0.9940089**	**0.9954170**	**0.9975309**	**0.9940845**
15	0.9891905	0.9917473	0.9975309	0.9882451
<u>39</u>	<u>0.9951802</u>	<u>0.9963303</u>	<u>1.0000000</u>	<u>0.9941279</u>

Reference best performance with 39 features (underlined)

Candidate features subsets over close to 97% in every metric and sensitivity greater than specificity (bold letters)

*This subset also has a decent performance over 94% in every metric

Table 9. Description of candidate 1-NN models without observational features

Features	F1-Score	Accuracy	Sensitivity	Specificity	Features
5	0.9428919	0.9560303	0.9451741	0.9629114	acc.x.min + eda.min + eda.max + temp.max + acc.z.min
10	0.9704221	0.9770977	0.9786673	0.9763903	acc.x.min + temp.mean + temp.max + hr.max + acc.z.min + hr.skew + eda.max + eda.mean + acc.y.min + eda.min
11	0.9796175	0.9844162	0.9878547	0.9823069	acc.x.min + acc.z.min + temp.max + temp.mean + hr.max + hr.skew + acc.y.min + eda.mean + eda.min + eda.max + gyr.y.kurtosis
12	0.9856647	0.9890034	0.9927930	0.9867080	acc.x.min + acc.z.min + temp.mean + temp.max + hr.max + hr.skew + eda.min + acc.y.min + gyr.y.kurtosis + eda.mean + acc.z.skew + eda.max
13	0.9915874	0.9935780	1.0000000	0.9897161	acc.x.min + acc.z.min + temp.max + temp.mean + hr.max + hr.skew + eda.min + eda.mean + eda.max + acc.z.skew + acc.y.min+ gyr.y.kurtosis + acc.x.sd

14	0.9940089	0.9954170	0.9975309	0.9940845	acc.x.min + hr.skew + acc.z.min + temp.mean + temp.max + hr.max + eda.min + eda.max + gyr.y.kurtosis + acc.y.min + eda.mean + acc.x.sd + acc.z.skew + hr.kurtosis
39	0.9951802	0.9963303	1.0000000	0.9941279	hr.skew + acc.x.min + acc.y.kurtosis + gyr.z.mean + acc.z.min + gyr.x.kurtosis + acc.y.skew + eda.mean + eda.min + acc.z.skew + acc.x.sd + acc.x.energy + hr.kurtosis+ acc.x.skew + hr.max + eda.max+ temp.min + acc.y.max + gyr.y.kurtosis + eda.range + temp.max + hr.mean + gyr.y.skew+ temp.energy + temp.mean + gyr.x.max + gyr.x.range + acc.y.min + gyr.z.max + hr.range + temp.range + eda.energy + temp.sd+ acc.z.max + acc.z.energy + acc.x.mean + gyr.y.min + acc.y.mean + acc.z.kurtosis

Reference best performance with 39 features (underlined)

Empatica sensors: acc (accelerometer); eda (EDA sensor); hr (heart rate); temp (temperature). Gear S3 sensors: gyr (gyroscope)

5. Discussion

The results showed that IADLs dependence assessment in elderly population could be performed using wearables during the activity of shopping and analyzing the data with ML, supporting previous literature that encourage the use of IoT to detect health issues [31], and the use of wearables during ADLs [6,36,39]. The experiment performs over 96% in every metric in all the models that include more than 5 features. In addition, our best ML model (i.e., 1-NN with 39 features) used only wearable data (hence, it is completely ecological) and learned to classify dependent and independent subjects with an F1-score and accuracy over 99%, using few features (Table 7): 8 from the smartwatch and 31 from the wristband. Hence, we could confirm that our proposal could be completely automatic (with 2 wearables), without the need of an observer (observational variables), with accuracies over 99%, which is an excellent accuracy in m-Health [6,10,22,24,26,27,32]. The only observation variable we need with this model is the annotation of the shopping stages, which we believe could also be annotated automatically in the future. We also built another model with the competitive performance of 94% in every metric, using only 5 features and all of them coming from one device (the wristband). However, in this model, sensitivity was lower than specificity, which is against the common rule in the health domain. This fact implies that our model will detect better true negatives (independent subjects) than true positives (dependent subjects), missing some people with dependency. Hence, we also obtained another model keeping sensitivity greater than specificity, with 10 features from the wristband and a score of 97% in every metric. Therefore, we could confirm that our solution is completely automatic, and uses only one wearable, with accuracies ranging from 94% to 97% and 5 to 10 features (see Table 9). These results are competitive in m-Health [6,10,22,24,26,27,32]. Summarizing, for the sample analysed in this paper—keeping sensitivity greater than specificity—, we created an automatic model with data from only 2 wearables and an accuracy over 99%, and another model with data from only one wearable and an accuracy of 97%. Therefore, it is possible to substitute/replace the manual questionnaires by the automatic assessment of dependency proposed in this paper with high accuracy.

The benefits of our results focus on the automatization of the process of a holistic and ecological assessment of the IADL dependence. Our proposal is holistic because the shopping task used as an evaluation process involves the main human functions to be independent (physical, cognitive and social) [3,7]. We have proved that the dependence of IADLs can be generalized evaluating only an IADL (shopping). Our proposal is ecological because it is unobtrusive and transparent for the subjects and performed in their daily life. Hence, it can reduce the time the health professionals need to assess the dependence.

As the aim of the research is empiric, i.e., to study the feasibility of evaluating dependence in older adults in an ecological way by means of a method based on m-Health systems, there are no direct theoretical implications in the work. The main strength of this study is related to the practical implications of the findings in a real clinical setting assisting older adults. Recognizing in early stages the potential of being dependent in the near future before it happens is crucial to implement rehabilitation strategies to reverse or delay the dependence or increase the quality of life [62]. The implications of our work are inline with the literature supporting the idea that wearables contribute to an early detection and prevention of disorders [8,9,63]. General benefits (e.g., cost reduction) and risks (e.g., lack of regulation) of the mobile technology are reported in [64]. The technological solution that we have developed and tested by mixing wearables sensors, the performance of an instrumental activity of daily living and the use of machine learning provides a novel approach to evaluate the possible dependence saving time for the health professionals in a daily practice and in an easier manner than traditional assessments. Therefore, our protocol, using wearables, has the power of saving time and improving the evaluation process when the clinician aims to assess dependency in potential dependent people.

However, the present study has some limitations that have to be considered in the interpretation of the results. The first limitation is that our proposal is not completely ecological. Although our best ML model with 10 wearable features does not include any observational feature, it still needs the TS to label the SS. However, the labelling process is automated in the mobile application paired to the wearable device, thus the supervisor only has to press the labelling button.

The second limitation is that the pragmatic integration and adoption of wearable technologies in the healthcare services is a challenge [65], especially in primary care [66], for several reasons. First, healthcare systems have to change their models of care for using these devices and sharing information. Second, wearable developers have to consider constraints of standardization, data privacy and security. Third, the solutions need to be low-cost and enable their use at large scale. And, fourth, the acceptance of the technology is still a challenge [67], although it is becoming common among the elderly [68].

The third limitation is that there is an imbalance in terms of sex distribution of the sample. Nevertheless, on the one hand, the sex imbalance is representative of the real scenario in terms of participation in the IADLs, with a higher participation of women, due to cultural factors [69,70]. Furthermore, since all male subjects belong to one single class (dependent individuals), this could introduce a bias in the algorithm performance because the sex of an individual may involve differences in physiological functions. Therefore, significant differences could exist in the physiological parameters of the sample between female and males. On the other hand, the aim of the present study was focused on validating the technical solution, based on the information recruited from wearables and on the use of Machine Learning techniques, needed to validate this technical solution. Additionally, this study aimed to predict the dependence in IADLs of older adults while executing a shopping activity without taking into consideration sex parameters or differences in the physiological changes across sexes. Hence, future studies with the aim of predicting the dependence in male and female separately should recruit a balanced sample in terms of sex, assuring an equal distribution of dependent and independent participants for IADLs in both groups.

The fourth limitation is that we conducted a cross-sectional study in only two community day centers, offering social activities to older people to increase and promote an active life in this stage of their lives. These centers were chosen because these organizations assist people that may be at risk or have dependence and this was our target population. The results of the present work represent a first step to test the functioning of the wearables that automatizes the assessment process in the clinical setting, saving time and complexity of the evaluation of the possible dependence. Future studies should conduct new studies to increase the generalization of the results to the whole older people population.

6. Conclusions

In this work, we proposed a novel approach for assessing the elderly IADL dependence with an m-Health system. We achieved this goal using wearables for collecting data, transparently to the subjects during the performance of one single IADL (shopping). With a sample size of 78 subjects, the resulting model (k-NN) was validated with 5-fold stratified cross-validation technique, reporting an F1-score of 97% and a similar accuracy—keeping sensitivity higher than specificity. This model uses only 10 features extracted from four sensors of a single wearable placed on the subject's non-dominant-hand wrist.

The use of wearables and ML contribute to create a holistic and semi-ecological model that improves the traditional assessment of IADL performance. The proposed m-Health system could help clinicians to evaluate and monitor the independence and autonomy of older adults, assessing all the human functions with one IADL (shopping).

Regarding future directions, we will focus on the generation of a fully ecological model for evaluating, monitoring and assessing IADL dependence. In particular, we will try to avoid the intervention of an external observer (TS). These would further reduce economic and human costs in public and private health systems. The proposed model will be integrated in a generalized m-Health system, which takes advantage of the decoupling, flexibility, extension, scalability and evolution of microservices and cloud technologies [18–20]. These technologies also allow interoperability with different devices, sensors and applications.

References

[1] J. Devi, The scales of functional assessment of Activities of Daily Living in geriatrics, Age and Ageing. 47 (2018) 500–502. https://doi.org/10.1093/ageing/afy050.

[2] K. Prakoso, Vitriana, A. Ong, Correlation between Cognitive Functions and Activity of Daily Living among Post-Stroke Patients, Althea Medical Journal. 3 (2016) 329–333. https://doi.org/10.15850/amj.v3n3.874.

[3] M.P. Lawton, E.M. Brody, Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living, The Gerontologist. 9 (1969) 179–186. https://doi.org/10.1093/geront/9.3_Part_1.179.

[4] M. Pashmdarfard, A. Azad, Assessment tools to evaluate activities of daily living (ADL) and instrumental activities of daily living (IADL) in older adults: A systematic review, Medical Journal of the Islamic Republic of Iran. 34 (2020). https://doi.org/10.34171/mjiri.34.33.

[5] N. Bhardwaj, M. Kashyap, Study of the Association of Geriatric syndromes with Functional Dependence in the Elderly, International Journal of Medical and Health Sciences. 7 (2018) 16–20.

[6] I. Rawtaer, K. Abdul Jabbar, X. Liu, T.T.H. Ying, A.T. Giang, P.L.K. Yap, R.C.Y. Cheong, H.P. Tan, P. Lee, S.L. Wee, T.P. Ng, Performance-based IADL evaluation of older adults with cognitive impairment within a smart home: A feasibility study, Alzheimer's & Dementia: Translational Research & Clinical Interventions. 7 (2021) 1–8. https://doi.org/10.1002/trc2.12152.

[7] C. Bottari, P.L. Wai Shun, G. Le Dorze, N. Gosselin, D. Dawson, Self-Generated Strategic Behavior in an Ecological Shopping Task, American Journal of Occupational Therapy. 68 (2014) 67–76. https://doi.org/10.5014/ajot.2014.008987.

[8] S. Kekade, C.-H. Hseieh, M.M. Islam, S. Atique, A. Mohammed Khalfan, Y.-C. Li, S.S. Abdul, The usefulness and actual use of wearable devices among the elderly population, Computer Methods and Programs in Biomedicine. 153 (2018) 137–159. https://doi.org/10.1016/j.cmpb.2017.10.008.

[9] S. Malwade, S.S. Abdul, M. Uddin, A.A. Nursetyo, L. Fernandez-Luque, X. (Katie) Zhu, L. Cilliers, C.-P. Wong, P. Bamidis, Y.-C. (Jack) Li, Mobile and wearable technologies in healthcare for the ageing population, Computer Methods and Programs in Biomedicine. 161 (2018) 233–237. https://doi.org/10.1016/j.cmpb.2018.04.026.

[10] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, Y. Amirat, Physical Human Activity Recognition Using Wearable Sensors, Sensors. 15 (2015) 31314–31338. https://doi.org/10.3390/s151229858.

[11] J.-Y. Yang, J.-S. Wang, Y.-P. Chen, Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers, Pattern Recognition Letters. 29 (2008) 2213–2220. https://doi.org/10.1016/J.PATREC.2008.08.002.

[12] C.V.C. Bouten, K.T.M. Koekkoek, M. Verduin, R. Kodde, J.D. Janssen, A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity, IEEE Transactions on Biomedical Engineering. 44 (1997) 136–147. https://doi.org/10.1109/10.554760.

[13] L. Atallah, B. Lo, R. King, G.-Z. Yang, Sensor Positioning for Activity Recognition Using Wearable Accelerometers, IEEE Transactions on Biomedical Circuits and Systems. 5 (2011) 320–329. https://doi.org/10.1109/TBCAS.2011.2160540.

[14] T. Broens, A. Van Halteren, M. Van Sinderen, K. Wac, Towards an application framework for context-aware m-Health applications, International Journal of Internet Protocol Technology. 2 (2007) 109. https://doi.org/10.1504/IJIPT.2007.012374.

[15] S. Majumder, T. Mondal, M. Deen, Wearable Sensors for Remote Health Monitoring, Sensors. 17 (2017) 130. https://doi.org/10.3390/s17010130.

[16] V. Jones, A. van Halteren, N. Dokovsky, G. Koprinkov, J. Peuscher, R. Bults, D. Konstantas, I. Widya, R. Herzog, Mobihealth: Mobile Services for Health Professionals, in: M-Health, Springer US, Boston, MA, 2006: pp. 237–246. https://doi.org/10.1007/0-387-26559-7_17.

[17] D. Lu, D. Huang, A. Walenstein, D. Medhi, A Secure Microservice Framework for IoT, in: 2017 IEEE Symposium on Service-Oriented System Engineering (SOSE), IEEE, San Francisco, CA, USA, : pp. 9–18. https://doi.org/10.1109/SOSE.2017.27.

[18] F.M. Garcia-Moreno, M. Bermudez-Edo, J.L. Garrido, E. Rodríguez-García, J.M. Pérez-Mármol, M.J. Rodríguez-Fórtiz, A Microservices e-Health System for Ecological Frailty Assessment Using Wearables, Sensors. 20 (2020) 3427. https://doi.org/10.3390/s20123427.

[19] A. Van Halteren, R. Bults, K. Wac, N. Dokovsky, G. Koprinkov, I. Widya, D. Konstantas, V. Jones, R. Herzog, Wireless body area networks for healthcare: The mobihealth project, Studies in Health Technology and Informatics. 108 (2005) 181–193.

[20] N. Dokovsky, A. van Halteren, I. Widya, BANip: Enabling Remote Healthcare Monitoring with Body Area Networks, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2004: pp. 62–72. https://doi.org/10.1007/978-3-540-24639-8_6.

[21] D. Peri, Body Area Networks and Healthcare, in: S. Gaglio, G. Lo Re (Eds.), Advances in Intelligent Systems and Computing, Springer International Publishing, Cham, 2014: pp. 301–310. https://doi.org/10.1007/978-3-319-03992-3_21.

[22] A. Godfrey, Wearables for independent living in older adults: Gait and falls, Maturitas. 100 (2017) 16–26. https://doi.org/10.1016/j.maturitas.2017.03.317.

[23] I. Zak, I. Klein, R. Katz, A Feasibility Study of Machine Learning Based Coarse Alignment, Proceedings. 4 (2018) 50. https://doi.org/10.3390/ecsa-5-05735.

[24] B. Fida, I. Bernabucci, D. Bibbo, S. Conforto, M. Schmid, Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer, Medical Engineering & Physics. 37 (2015) 705–711. https://doi.org/10.1016/j.medengphy.2015.04.005.

[25] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, D. Finlay, Optimal Placement of Accelerometers for the Detection of Everyday Activities, Sensors. 13 (2013) 9183–9200. https://doi.org/10.3390/s130709183.

[26] A. Mannini, S.S. Intille, Classifier Personalization for Activity Recognition Using Wrist Accelerometers, IEEE Journal of Biomedical and Health Informatics. 23 (2019) 1585–1594. https://doi.org/10.1109/JBHI.2018.2869779.

[27] N. Twomey, T. Diethe, X. Fafoutis, A. Elsts, R. McConville, P. Flach, I. Craddock, A Comprehensive Study of Activity Recognition Using Accelerometers, Informatics. 5 (2018) 27. https://doi.org/10.3390/informatics5020027.

[28] H.J. Guo, A. Sapra, Instrumental Activity of Daily Living, 2021.

[29] I. Martín-Lesende, I. Ortiz-Lebaniegos, E. Montalvillo-Delgado, M. Pérez-Abad, P. Sánchez-Junquera, C. Rodríguez-Andrés, Identification of Items for Creating a Questionnaire for the Assessment of Instrumental Activities of DailyLiving (IADL) in Elderly Patients, Atención Primaria. 37 (2006) 313–318. https://doi.org/10.1016/S0212-6567(06)70366-3.

[30] I. Vergara, A. Bilbao, M. Orive, S. Garcia-Gutierrez, G. Navarro, J. Quintana, Validation of the Spanish version of the Lawton IADL Scale for its application in elderly people, Health and Quality of Life Outcomes. 10 (2012) 130. https://doi.org/10.1186/1477-7525-10-130.

[31] Z. Meng, M. Zhang, C. Guo, Q. Fan, H. Zhang, N. Gao, Z. Zhang, Recent Progress in Sensing and Computing Techniques for Human Activity Recognition and Motion Analysis, Electronics. 9 (2020) 1357. https://doi.org/10.3390/electronics9091357.

[32] E. Garcia-Ceja, R. Brena, J. Carrasco-Jimenez, L. Garrido, Long-Term Activity Recognition from Wristwatch Accelerometer Data, Sensors. 14 (2014) 22500–22524. https://doi.org/10.3390/s141222500.

[33] R. Delgado-Escaño, F.M. Castro, J.R. Cózar, M.J. Marín-Jiménez, N. Guil, E. Casilari, A cross-dataset deep learning-based classifier for people fall detection and identification, Computer Methods and Programs in Biomedicine. 184 (2020) 105265. https://doi.org/10.1016/j.cmpb.2019.105265.

[34] J.C. Vasquez-Correa, T. Arias-Vergara, J.R. Orozco-Arroyave, B. Eskofier, J. Klucken, E. Noth, Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach, IEEE Journal of Biomedical alth Informatics. 23 (2019) 1618–1630. https://doi.org/10.1109/JBHI.2018.2866873.

[35] X. Li, X. Zhang, J. Zhu, W. Mao, S. Sun, Z. Wang, C. Xia, B. Hu, Depression recognition using machine learning methods with different feature generation strategies, Artificial Intelligence in Medicine. 99 (2019) 101696. https://doi.org/10.1016/j.artmed.2019.07.004.

[36] S. Paraschiakos, R. Cachucho, M. Moed, D. van Heemst, S. Mooijaart, E.P. Slagboom, A. Knobbe, M. Beekman, Activity recognition using wearable sensors for tracking the elderly, User Modeling and User-Adapted Interaction. 30 (2020) 567–605. https://doi.org/10.1007/s11257-020-09268-2.

[37] A. König, C.F. Crispim-Junior, A.G.U. Covella, F. Bremond, A. Derreumaux, G. Bensadoun, R. David, F. Verhey, P. Aalten, P. Robert, Ecological Assessment of Autonomy in Instrumental Activities of Daily Living in Dementia Patients by the Means of an Automatic Video Monitoring System, Frontiers in Aging Neuroscience. 7 (2015). https://doi.org/10.3389/fnagi.2015.00098.

[38] G. Sacco, Joumier, Darmon, Dechamps, Derreumaux, Lee, Piano, Bordone, Konig, Teboul, David, Guerin, Bremond, Robert, Detection of activities of daily living impairment in Alzheimer's disease and mild cognitive impairment using information and communication technology, Clinical Interventions in Aging. (2012) 539. https://doi.org/10.2147/CIA.S36297.

[39] S. Schmidle, P. Gulde, B. Jansen, S. Herdegen, J. Hermsdörfer, Frailty Assessment in Daily Living (FRAIL) - Assessment of ADL Performance of Frail Elderly with IMUs, in: Communications in Computer and Information Science, 2020: pp. 92–101. https://doi.org/10.1007/978-3-030-60703-6_12.

[40] H. Lee, B. Joseph, A. Enriquez, B. Najafi, Toward Using a Smartwatch to Monitor Frailty in a Hospital Setting: Using a Single Wrist-Wearable Sensor to Assess Frailty in Bedbound Inpatients, Gerontology. 64 (2018) 389–400. https://doi.org/10.1159/000484241.

[41] J. Martínez de la Iglesiaa, R. DueñasHerrerob, M. Carmen Onís Vilchesa, C. Aguado Tabernéa, C. Albert Colomerc, R. Luque Luquec, Adaptación y validación al castellano del cuestionario de Pfeiffer (SPMSQ) para detectar la existencia de deterioro cognitivo en personas mayores e 65 años, Medicina Clínica. 117 (2001) 129–134. https://doi.org/10.1016/S0025-7753(01)72040-4.

[42] Empatica, E4 wristband, (n.d.). https://www.empatica.com/en-gb/research/e4/ (accessed June 27, 2019).

[43] Tizen, Device Sensors, (n.d.). https://developer.tizen.org/development/guides/.net-application/location-and-sensors/device-sensors (accessed July 3, 2019).

[44] C. McCarthy, N. Pradhan, C. Redpath, A. Adler, Validation of the Empatica E4 wristband, in: 2016 IEEE EMBS International Student Conference (ISC), IEEE, 2016: pp. 1–4. https://doi.org/10.1109/EMBSISC.2016.7508621.

[45] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies, International Journal of Medical Informatics. 153 (2021). https://doi.org/10.1016/j.ijmedinf.2021.104510.

[46] V. Genovese, A. Mannini, A.M. Sabatini, A Smartwatch Step Counter for Slow and Intermittent Ambulation, IEEE Access. 5 (2017) 13028–13037. https://doi.org/10.1109/ACCESS.2017.2702066.

[47] G. Ogliari, S. Mahinrad, D.J. Stott, J.W. Jukema, S.P. Mooijaart, P.W. Macfarlane, E.N. Clark, P.M. Kearney, R.G.J. Westendorp, A.J.M. de Craen, B. Sabayan, Resting heart rate, heart rate variability and functional decline in old age, Canadian Medical Association Journal. 187 (2015) E442–E449. https://doi.org/10.1503/cmaj.150462.

[48] M.R. Hamedinia, M. Sardorodian, A.H. Haghighi, S. Vahdat, The Effects of Moderate Swimming Training on Blood Pressure Risk Factors in Hypertensive Postmenopausal Women, Iranian Journal of Health and Physical Activity. 1 (2010) 24–28. https://doi.org/10.22067/ijhpa.v1i1.3537.

[49] R.C. MacCallum, S. Zhang, K.J. Preacher, D.D. Rucker, On the practice of dichotomization of quantitative variables., Psychological Methods. 7 (2002) 19–40. https://doi.org/10.1037/1082-989X.7.1.19.

[50] T.A. Dombrowsky, Relationship between engagement and level of functional status in older adults, SAGE Open Medicine. 5 (2017) 1–9. https://doi.org/10.1177/2050312117727998.

[51] M. Bermudez-Edo, P. Barnaghi, K. Moessner, Analysing real world data streams with spatio-temporal correlations: Entropy vs. Pearson correlation, Automation in Construction. 88 (2018) 87–100. https://doi.org/10.1016/j.autcon.2017.12.036.

[52] F.M. Garcia-Moreno, M. Bermudez-Edo, M.J. Rodríguez-Fórtiz, J.L. Garrido, A CNN-LSTM Deep Learning Classifier for Motor Imagery EEG Detection Using a Low-invasive and Low-Cost BCI Headband, in: 16th International Conference on Intelligent Environments, 2020.

[53] F.M. Garcia-Moreno, M. Bermudez-Edo, J.L. Garrido, M.J. Rodríguez-Fórtiz, Reducing Response Time in Motor Imagery Using A Headband and Deep Learning, Sensors. 20 (2020) 6730. https://doi.org/10.3390/s20236730.

[54] C. Guo, M. Lu, J. Chen, An evaluation of time series summary statistics as features for clinical prediction tasks, BMC Medical Informatics and Decision Making. 20 (2020) 48. https://doi.org/10.1186/s12911-020-1063-x.

[55] G. Ogliari, S. Mahinrad, D.J. Stott, J.W. Jukema, S.P. Mooijaart, P.W. Macfarlane, E.N. Clark, P.M. Kearney, R.G.J. Westendorp, A.J.M. de Craen, B. Sabayan, Resting heart rate, heart rate variability and functional decline in old age, Canadian Medical Association Journal. 187 (2015) E442–E449. https://doi.org/10.1503/cmaj.150462.

[56] N. V Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research. 16 (2002) 321–357. https://doi.org/10.1613/jair.953.

[57] T.N. Lal, O. Chapelle, J. Weston, A. Elisseeff, Embedded Methods, in: Feature Extraction, Springer Berlin Heidelberg, Berlin,

Heidelberg, 2006: pp. 137–165. https://doi.org/10.1007/978-3-540-35488-8_6.

[58] Y. Watanabe, S. Sara, Toward an Immunity-based Gait Recognition on Smart Phone: A Study of Feature Selection and Walking State Classification, Procedia Computer Science. 96 (2016) 1790–1800. https://doi.org/10.1016/j.procs.2016.08.228.

[59] S. Saeb, L. Lonini, A. Jayaraman, D.C. Mohr, K.P. Kording, The need to approximate the use-case in clinical machine learning, GigaScience. 6 (2017) 1–9. https://doi.org/10.1093/gigascience/gix019.

[60] M.A. Little, G. Varoquaux, S. Saeb, L. Lonini, A. Jayaraman, D.C. Mohr, K.P. Kording, Using and understanding cross-validation strategies. Perspectives on Saeb et al., GigaScience. 6 (2017) 1–6. https://doi.org/10.1093/gigascience/gix020.

[61] E. Chaibub Neto, A. Pratap, T.M. Perumal, M. Tummalacherla, P. Snyder, B.M. Bot, A.D. Trister, S.H. Friend, L. Mangravite, L. Omberg, Detecting the impact of subject characteristics on machine learning-based diagnostic applications, Npj Digital Medicine. 2 (2019) 99. https://doi.org/10.1038/s41746-019-0178-x.

[62] T. Crocker, J. Young, A. Forster, L. Brown, S. Ozer, D.C. Greenwood, The effect of physical rehabilitation on activities of daily living in older residents of long-term care facilities: systematic review with meta-analysis, Age and Ageing. 42 (2013) 682–688. https://doi.org/10.1093/ageing/aft133.

[63] J. a Fraile, J. Bajo, J.M. Corchado, A. Abraham, Applying wearable solutions in dependent environments, IEEE Transactions on Information Technology in Biomedicine. 14 (2010) 1459–1467. https://doi.org/10.1109/TITB.2010.2053849.

[64] W. Currie, Health Organizations' Adoption and Use of Mobile Technology in France, the USA and UK, Procedia Computer Science. 98 (2016) 413–418. https://doi.org/10.1016/j.procs.2016.09.063.

[65] H. Lewy, Wearable technologies – future challenges for implementation in healthcare services, Healthcare Technology Letters. 2 (2015) 2–5. https://doi.org/10.1049/htl.2014.0104.

[66] S.-T. Liaw, A. Georgiou, H. Marin, Evaluation of Digital Health & Information Technology in Primary Care, International Journal of Medical Informatics. 144 (2020) 104285. https://doi.org/10.1016/j.ijmedinf.2020.104285.

[67] B. Rahimi, H. Nadri, H. Lotfnezhad Afshar, T. Timpka, A Systematic Review of the Technology Acceptance Model in Health Informatics, Applied Clinical Informatics. 09 (2018) 604–634. https://doi.org/10.1055/s-0038-1668091.

[68] S.J. Czaja, Current Findings and Issues in Technology and Aging, Journal of Applied Gerontology. 40 (2021) 463–465. https://doi.org/10.1177/0733464821998579.

[69] T. da S. Alexandre, L.P. Corona, D.P. Nunes, J.L.F. Santos, Y.A. de O. Duarte, M.L. Lebrão, Disability in instrumental activities of daily living among older adults: gender differences, Revista de Saúde Pública. 48 (2014) 379–389. https://doi.org/10.1590/S0034-8910.2014048004754.

[70] F. Nourhashemi, S. Andrieu, S. Gillette-Guyonnet, B. Vellas, J.L. Albarede, H. Grandjean, Instrumental Activities of Daily Living as a Potential Marker of Frailty: A Study of 7364 Community-Dwelling Elderly Women (the EPIDOS Study), The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 56 (2001) M448–M453. https://doi.org/10.1093/gerona/56.7.M448.

4. Reducing Response Time in Motor Imagery Using A Headband and Deep Learning

[C] Garcia-Moreno, F. M., Bermudez-Edo, M., Garrido, J. L., Rodríguez-Fórtiz, M. J. (2020). Reducing Response Time in Motor Imagery Using A Headband and Deep Learning. *Sensors*, *20*(23), 6730. DOI: 10.3390/s20236730.

g. Estado: publicado.

h. Factor de impacto (JCR 2020): 3.576.

i. Categoría: Instruments & Instrumentation. Posición: 14 / 64 (**Q1**).

j. Categoría: Engineering, Electrical & Electronic. 82 / 273 (**Q2**).

k. Nº de citas: 8 (Fuente, Google Scholar).

Este artículo se puede encontrar en acceso abierto en el siguiente enlace:

https://doi.org/10.3390/s20236730

A continuación, se proporciona un borrador del mismo para cumplir con los derechos de autor.

Reducing Response Time in Motor Imagery Using A Headband and Deep Learning

Francisco M. García-Moreno [1], María Bermúdez-Edo [1], and José Luis Garrido [1], María José Rodríguez-Fórtiz [1]

[1]Department of Software Engineering, Computer Sciences School, University of Granada.

Abstract. Electroencephalography (EEG) signals to detect motor imagery have been used to help patients with low mobility. However, the regular brain computer interfaces (BCI) capturing the EEG signals usually require intrusive devices and cables linked to machines. Recently, some commercial low-intrusive BCI headbands have appeared, but with less electrodes than the regular BCIs. Some works have proved the ability of the headbands to detect basic motor imagery. However, all of these works have focused on the accuracy of the detection, using session sizes larger than 10 s, in order to improve the accuracy. These session sizes prevent actuators using the headbands to interact with the user within an adequate response time. In this work, we explore the reduction of time-response in a low-intrusive device with only 4 electrodes using deep learning to detect right/left hand motion imagery. The obtained model is able to lower the detection time while maintaining an acceptable accuracy in the detection. Our findings report an accuracy above 83.8% for response time of 2 s overcoming the related works with both low- and high-intrusive devices. Hence, our low-intrusive and low-cost solution could be used in an interactive system with a reduced response time of 2 s.

Keywords: *Neural networks; deep learning; motor imagery; wearable; EEG; BCI; users' interaction; response time.*

1. Introduction

Health care professionals are beginning to use wearable devices for patient monitoring and clinical practice. The wearable data can be analyzed with machine learning (ML) algorithms and consequently predict, prevent, or design an intervention plan [1]. Wearables with different types of built-in sensors (mechanical, physiological, biochemical, and bioimpedance) can measure the physical and mental state of the individuals. They can also act as actuators (output devices), allowing users to interact with the environment. For example, some wearables, such as smart glasses and vibration bracelets (sometimes in combination with other static sensors in the environment, e.g. camera, presence, etc.), provide feedback about users' interactions or guidance during different activities [2–4]. Some commercial wearables could replace expensive medical devices under specific circumstances. Wearables lead to new possibilities in research. Their main benefits are that they are not tied to clinical environments. They can rather be used anytime and anywhere, even outside, i.e., in an ecological way. Additionally, they gather big quantities of data in an objective and precise way. Wearables have been used in evaluation and intervention systems in several healthcare areas, such as healthy ageing, to help the elderly living independently, increasing their autonomy and improving their quality of life [5,6]; and impaired persons [2,7,8].

In particular, some sensors record the brain activity through electroencephalography (EEG) signals to make decisions or to control the environment by evaluating the mind state of the user. Different analyses of data from EEG signals can discover patterns or anomalies of the user's behavior. Even in the absence of movement, a headband can measure thinking activity, such as motor imagery or goal-oriented thinking, which can be used to interact with a computer system [9]. For example, if the user is thinking about pressing a button on a PC screen, the analysis of the EEG signals could detect such intention, and with the help of an application, actually press the button.

Most of the work performed with EEG signals use intrusive brain computer interfaces (BCI) devices, normally inside hospitals [9]. These traditional BCIs have up to 256 channels (electrodes) and are uncomfortable to wear in outdoor activities, because they need to be in direct contact to the head scalp (it needs even a conductive gel) at different locations, involving expensive machinery and cables paired to the head of the user [10]. The low-cost and low-intrusive EEG headbands, with only a few electrodes, are more comfortable (without cap, cables, and conductive gels) and portable than intrusive EEG devices. Figure 1 summarizes the differences between traditional EEG devices and EEG headbands. EEG headbands could replace traditional EEG machinery for some tasks, especially in combination with ML techniques, which could learn different states of the mind. These headbands prioritize the usability (user-friendly), and the adoption of the headbands by the general public, due to its low-price, versus the accuracy and precision of the data.

Our aim is to explore the potential of low-cost and low-intrusive commercial BCIs on the classification of motor imagery within a short period of time, i.e., that our classifier could detect the motor imagery within a few seconds (less than 10 s) after the subject starts thinking on that movement. In particular, the Muse headband, made by InteraXon [11], of which sensors record electroencephalography (EEG) signals through four channels located around the head of a person. These positions follow the 10-20 international standards [12]. In our previous work [13], we explored the possibilities that the Muse headband offers on the detection of motor imagery. Specifically, we implemented a deep learning (DL) model that could detect left- and right-hand motor imagery with an accuracy of 98.9%, using data from several users and with 20 s session size. These results, although quite accurate, means that it is not possible for an actuator or application to respond before the session ends. For example, if we need to detect the brain activity of a user in order to press a button (left or right bottom) on a screen, the system needs to wait the 20 s of the session and then classify it (left or right) before responding. Most of the previous works on motor imagery use session sizes bigger than 10 s to ensure they capture the activity of the brain. Only few works use session sizes smaller than 10 s but with intrusive BCI devices. This paper aims at extending our previous work by exploring different session sizes, and therefore to check the response time of an actuator. Specifically, our aim is to check what is the smallest response time that still obtains a decent accuracy on detecting motor imagery. To that end, we performed several experiments with different session and window (splits within the sessions) sizes. Additionally, we improved our previous deep learning model: with different preprocessing techniques, such as averaging data from sliding windows and noise removal; and applying some tuning techniques to prevent overfitting, such as early stopping. Afterwards, we validate our findings by classifying the left/right motor imagery in sessions where the subject is continuously changing directions every few seconds.

The remainder of this paper is organized as follows. Section 2 describes the related work concerning BCI devices and ML algorithms applied to motor imagery, paying special attention to session size. Section 3 introduces the Material and Methods, and also presents the deep learning proposal for classifying two motor imagery tasks: left- and right-hand movement imagery. Section 4 registers and analyzes the results or our experiments. Finally, Section 5, concludes the paper and describes the future work.

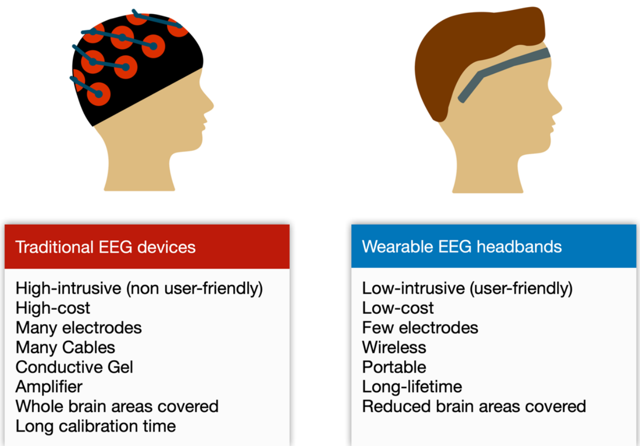

Figure 1. Differences between traditional EEG devices and wearable EEG headbands.

2. Related work

In recent years, deep learning (DL) —a subfield of machine learning— is gaining momentum. An enormous amount of papers using DL have been published in different areas [14–17]. Health is one of such fields [17] and even DL is commonly used to analyze EEG signals [18]; with obvious restrictions on mobility. During the execution of body part movements or their imagination (motor imagery), such as the hands or eyes, some signals originate in our brain, hence, we can use this correlation to detect movement or intention of movements. Brain activities have different frequencies and can be measured using EEG for driving a BCI system, because they can provide high temporal resolution, reflecting brain dynamic changes in milliseconds [10,19]. The signals measured can be stored and analyzed by means of features extraction to classify the movements with which they are related, inferring the intention of the user to move a specific part of the body, for example, the right or left hand. However, most of the research uses accurate and expensive devices, which are quite intrusive, as the users need to stick sensors all over his/her head, connected with cables to expensive machines [18]. Only a few researchers have experimented with low-cost and low-intrusive EEG devices, such as headbands [18]. For example, Bird et al. [20] used the DL technique long short-term memory (LSTM), which is a type of recurrent neural network (RNN), to learn the attentional emotion and sentiment of one user while providing visual stimuli.

As mentioned above, traditional EEG devices are high-cost, invasive (electrodes implanted in the brain), intrusive and require user-training and long calibration time [19,21]. Although there are non-invasive devices (such as EEG caps) which obtain the EEG data, placing electrodes to the scalp with a conductive gel, they are still high-intrusive because they need an awkward cap with cables linked to machines [10]. Thus, they are uncomfortable to wear in outdoors activities for the users. Nonetheless, there are new wearable devices, usually called BCI headbands. They are low-cost, non-invasive, and low-intrusive, which are more comfortable than intrusive EEG caps, because they do not need conductive gels or any cable. One example of these headbands is Muse, that we used in this research. The low-cost and low-intrusive Muse headband has been used to recognize users and a range of activities (reading, listening to music, playing computer games, watching movies, and relaxing) [22], movements (jaw-clenching and head and eye movements) [23], facial expressions [24], motor imagery [13,25,26], etc.

Most of the previous studies on motor imagery classification applied ML/DL algorithms use EEG recordings greater than 10 s. The BCI devices were both low-intrusive and intrusive. Both Zhang et al. [27] and Chen et al. [28] classified 5 intentions (eyes closed, opening-closing both feet, both fists, left fist, and right fist) with the LSTM technique achieving an accuracy close to 98%, but using 120 s-session size from MI-EEG dataset recorded with a high-intrusive EEG cap (Electro-Cap International USA). Rodriguez et al. [25] achieved 80% accuracy applying convolutional neural networks (CNN) and LSTMs layers to learn 4 motion intentions of a single user (hand, foot, mathematical activity, and relaxation state), but the response time was 30 s. Garcia-Moreno et al. [13] also used CNN+LSTM to classify 2 intentions (right and left hands), reaching 98.9% accuracy for 20 s response time. Li et al. [26] used SVM for classifying left and right hands motor imagery. The average accuracy of the eight participants was 95.1% accuracy (the best individual participant's accuracy was 98.9%), with a response time of 10 s.

Only few studies on motor imagery detection used EEG-recording durations lower than 10 s. However, either they used intrusive EEG devices or caps (such as Easycap, Germany or Electro-Cap International, USA) placed on the scalp (a conductive gel or paste is needed); or they used public datasets recorded with intrusive devices (such as MI EEG [29], Datasets 2A [30] and 2B [31] of BCI Competition IV [32–38]. For example, the studies [39,40] were one of the first approaches on motor imagery classification, where they applied logistic regression (LR) and reached an accuracy of 90.5% with recordings of 3 s using intrusive EEG devices. In 2011, Bhattacharyya et al. [32] used 9 s-recordings with an intrusive BCI and applied SVM for left and right motor imagery, achieving 82.14% accuracy, although they use unusual validation split (50–50%, instead of 90–10% or 80–20%). Recent studies [33–38] used DL techniques such as CNN and LSTM with recordings of 2–5 s (the best results reached an accuracy of 93.9% with recordings of 4 s [36]), using high-intrusive EEG caps and intra-subject protocol. Two of these recent works [34,37] used only a few electrodes; 3 out of the 22 of the original dataset [31]. In particular, the work of Ha and Jeong [37] achieved the best accuracy, which is 78.44%. They also obtained the results by averaging 9-subject's accuracies (intra-subject). However, both studies used data recorded with high-intrusive EEG devices. Only one of these recent work [38] reduced the response time to 2 s with a 70% accuracy for intra-subject (by averaging 9-subject's accuracies) and 40% for cross-subject. However, this work also use data from a high-intrusive EEG cap with 22 electrodes.

In conclusion, previous works on motor imagery focus on the accurate detection of the brain activity, rather than on the response time of a potential actuator. Hence, the session sizes are generally big (bigger than 10 s) to ensure that the activity of the brain is accurately captured (see Table 1, sorted by session size in descending order). Only a few works lower the session size to 4 or 5 s but using intrusive EEG devices. Although some of these works use only 3 or 6 channels of the intrusive device, and therefore, they could resemble a low-intrusive device with only few sensors, the accuracy is below 78% for 4 s session size. We will study in this paper how to lower the session size (lower than 4 s, if possible), but maintaining the accuracy above 80%. We chose 80% because most of the related work using session-sizes up to 4 s report an accuracy lower than 80%.

Table 1. Review of previous works about motor imagery using machine learning.

Work	Method	Channels	Intrusive	Own Dataset	Subjects	Classes	Session Size	Validation Split	Accuracy
[27]	CNN+LSTM	EEG: 64	Yes	No	Cross-Subject: 108	Five	120 s	75%–25%	98.3%
[28]	LSTM	EEG: 64	Yes	No	Intra-Subject: 109	Five	120 s	5 × 5-fold	97.8%
[25]	CNN+LSTM	Muse: 4	Low	Yes	Intra-Subject: 1	Four	30 s	90%–10%	80.13%
[13]	CNN+LSTM	Muse: 4	Low	Yes	Cross-Subject: 4	Binary	20 s	90%–10%	98.9%
[26]	SVM	Muse: 4	Low	Yes	Intra-Subject: 8	Binary	10 s	4-fold	95.1%
[32]	SVM	EEG: 2	Yes	No	Intra-Subject: 2	Binary	9 s	50%–50%	82.14%
[33]	CNN	EEG: 28	Yes	Yes	Intra-Subject: 2	Binary	5 s	80%–20%	86.41%

[36]	RLDA CNN	EEG: 22 EEG: 44	Yes Yes	No Yes	Intra-Subject: 9 Intra-Subject: 20	Four Four	4 s 4 s	ICV	73.7% 93.9%
[35]	LSTM	EEG: 6	Yes	No	Intra-Subject: 9	Binary	4 s	5x5-fold	79.6%
[37]	CNN	EEG: 3	Yes	No	Intra-Subject: 9	Binary	4 s	60%–40%	78.44%
[34]	CNN+SAE	EEG: 3	Yes	No	Intra-Subject: 9	Binary	4 s	10x10-fold	77.6%
[39,40]	LR	EEG: 128	Yes	Yes	Intra-Subject: 29	Three	3 s	50%–50%	90.5%
[38]	CNN	EEG: 22	Yes	No	Intra-Subject: 9 Cross-Subject: 9	Four	2 s	4-fold	~70% ~40%

SVM: support vector machine; RLDA: regularized linear discriminant analysis; ICV: inner cross validation.

3. Materials and methods

This section describes the experiment protocols that we have followed; the foundations of the algorithms we have used (deep learning); as well as the workflow proposed for our DL pipeline.

3.1 Experiment Protocols

We performed two sets of experiments. The first set explores the reduction of session-sizes and window-sizes, keeping a decent accuracy of the motor imagery detection. To that end, we performed several experiments with different window and session sizes for each motor imagery (right and left). The second set validates the minimum response time achieved in the first set. To that end, we performed several experiments in which the participants where continuously changing direction in motor imagery every few seconds.

Since the movement of eyeballs have an impact on brain waves [26], participants in every experiment rotated the eyes in the corresponding direction while imagining to pick up a bottle of water, but avoiding to move the hand, to touch the bottle or to blink the eyes.

3.1.1 Exploring Response Time Reduction

For the first set of the experiment, we recruited four healthy adults between the ages of 33 and 55 (3 females and 1 male).

We recorded forty EEG sessions per participant with a low-intrusive headband (see Figure 2), performing two different tasks: motor imagery for left and right hands (20 sessions per hand). Each session lasted 20 s and we labelled them (0-1) depending on the tasks (left/right hand). These sessions took place in a silent and distraction-free environment, as the lack of concentration could influence the results.

Figure 2. A participant wearing the low-intrusive EEG headband.

Summarizing, the experimental protocol consisted of these ordered steps:

1. The participant is sitting down on a chair with arms extended in parallel, resting on a table.
2. A bottle of water is on the table, approximately 5 cm to the left of the left hand.
3. The participant's head faces forward, while the eyes rotate to the left, looking to the bottle.
4. We asked the participant to imagine picking up the bottle with the left hand, but without moving the hand only thinking about it for 20 s.
5. Then, we asked the participant to relax and to close the eyes for 20 s.
6. We move the bottle on the right side of the table (approximately 5 cm to the right of the right hand).
7. The participant repeats steps 3–5, but for right hand motor imagery.
8. We repeated steps 2–7 for 20 times for each participant.

3.1.2 Validating the Response Time Reduction

To validate the results of the first set of experiments, we created a second set of experiments in which we tried to detect the changes in the motor imagery (right or left) in a continuous session. Due to the restrictions of the pandemic of 2020 and lockdowns, we could not use the same participants of the first experiment. In this second experiment, we recruited three new healthy adults —2 females and 1 male— aging 33–50.

The protocol for this set of experiments consisted of the following ordered steps:

1. The participant is sitting down on a chair with arms extended in parallel, resting on a table.
2. Two bottles of water are on the table. One of them is approximately 5 cm to the left of the left hand and the other bottle is 5 cm to the right of the right hand.
3. The participant's head face forward, while the eyes rotate to the left, looking to the bottle.
4. We asked the participant to imagine picking up the bottle with the right hand, but without moving the hand; only thinking about it for 6 s.
5. Then, we asked the participant to imagine picking up the bottle with the left hand, but without moving the hand, only thinking about it for 6 s.
6. We repeated steps 4–5 for 5 times for each participant: 1 min in total.
7. Then, we repeated steps 1–6 for 20 times.

3.2 Deep Learning Foundations

Deep learning has been traditionally used for computer vision, speech recognition, or natural language processing, and has been extended recently to several fields [15]. Additionally, in EEG signals, DL has been successfully used in the last years [18].

DL applies multiple iterative non-linear transformations of data, simulating the connections of the neurons in a brain. The parameters of the transformations are refined iteratively by minimizing a cost function (i.e., minimizing the error between the predicted and the real signal). DL means several layers of neural networks. However, there is not a consensus on how many layers make it deep. In practice, several DL approaches use just three layers. In a neural network, we have neurons or units organized in these layers: one input layer, one output layer, and one or more hidden layers.

The hidden layers of a deep neural network could be fully connected (FC, also known as the dense layer), recurrent neural network (RNN), or convolutional neural network (CNN). In a FC, all neurons received as input all the weighted outputs of the preceding layer. Generally, it is followed by a non-linear activation function (such as relu and sigmoid). In an RNN, one neuron receives the preceding output of the previous layers and its own output of the previous values.

In particular, one of the most used layers in an RNN is the long short-term memory (LSTM). As LSTM layers are recurrent, they retain memory of the previous time step. This feature makes these layers suitable for time series where the lags between events are uncertain. LSTM hidden layers (neurons) are called units. These units are inside a cell. Like RNN, the LSTM cell has a state and therefore can remember information of the previous timestep (see Figure 3). The common input of an LSTM is a triplet consisting of samples, timesteps, and features.

In a CNN, the outputs are convoluted, and one neuron only takes a subset of output signals (the closest outputs) of the previous layers (depending on the convolution of said outputs). CNN layers detect better the spatial component of the data, which means they can select the best features, and the RNN layers detect better the temporal component of the data [18].

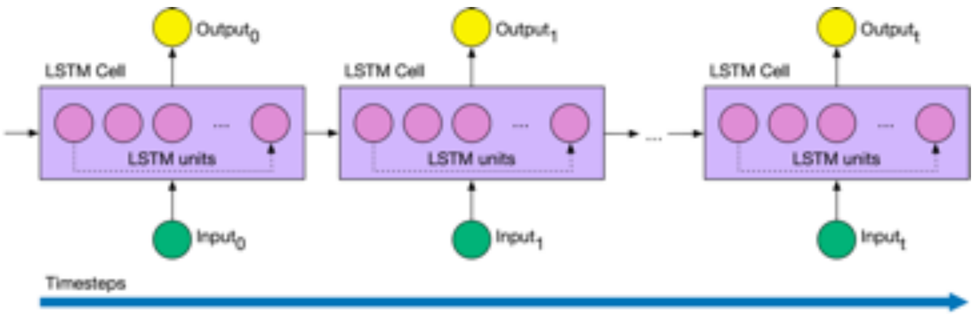

Figure 3. Long short-term memory (LSTM) network diagram.

3.3 Deep Learning Pipeline

This section describes the workflow of our DL pipeline which consists of three main steps: (1) EEG data acquisition; (2) Data preprocessing; and (3) DL architecture for building a model to classify two imagery movement intentions. Figure 4 presents the pipeline which will be explained in detail in the next subsections.

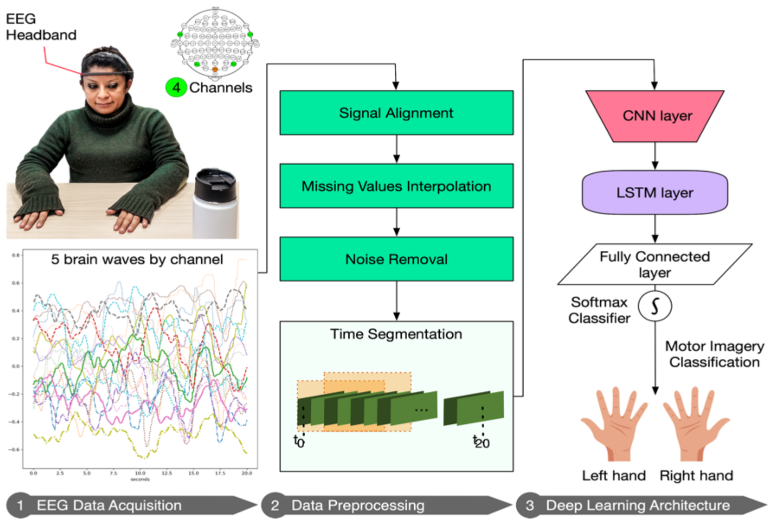

Figure 4. Deep learning pipeline: data acquisition, preprocessing, and deep learning architecture.

3.3.1 EEG Data Acquisition

Although previous studies commonly measure hand movements and motor imagery with electrodes located in the primary motor cortex (C3 and C4 electrodes), they use high-intrusive and high-cost EEG devices, which have a lot of cables and electrodes covering all of the lobes of the brain [25,27,28,32–36,39,40] (See Figure 5). Low-cost and low-intrusive EEG headbands have less sensors than the intrusive devices, and do not usually have electrodes in the central area of the brain —motor cortex— (Figure 5, red color). Usually, they have surrounding electrodes located on frontal (Figure 5, pink color) and temporal (Figure 5, blue color) brain areas. The frontal brain lobe is related to brain functions such as movement control, reasoning, emotions, speech, or problem solving [41]. In addition, the temporal lobe is related to auditory stimuli interpretation, meaning, memory, and processing [41]. Thus, the electrodes of these wearables could potentially measure motor imagery. In fact, previous studies [13,25,26] using low-intrusive headbands reached accurate results detecting motor imagery, although they use recordings longer than 10 s (as we presented in Table 1).

We used a Muse headband (version 2) by InteraXon [5], an EEG wearable device with 4 electrodes (channels) which can detect brain signals in a low-intrusive way. This device records EEG using 4 gold-plated cup bipolar electrodes, which are located on the frontal and temporal brain areas. Figure 5 shows (in light green) the location of the electrodes that follows the 10-20 international standard and the brain areas covered. These active electrodes (or channels) are: TP9 (left ear), TP10 (right ear), AF7 (left forehead), AF8 (right forehead) for average sized heads. The frontal electrodes could be placed in the location of F7 and F8 electrodes for small sized heads as Li et al. [26] mentioned. The reference electrode in Muse is Fpz. It does not capture brain signals, but measures the potential differences (voltage) between the active electrodes and Fpz.

Since Software Development Kits (SDK) by InteraXon were discontinued (preventing software developers or researchers to create their custom applications), we used the Mind Monitor application [18] to record the brain signals in real-time.

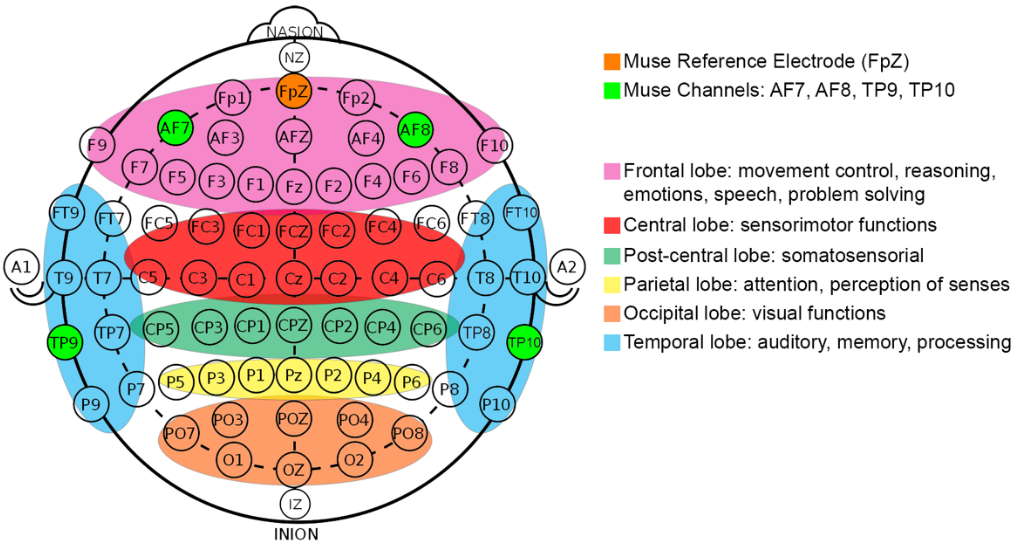

Figure 5. Muse electrodes position and brain functions.

In the headband, each channel (TP9, TP10, AF7, and AF8) captures a raw EEG signal, measured in microvolts (μV). Thus, we have four raw EEG signal data in total. These raw data consist of 4 signals coming from the 4 channels and they are ranging from 0 to 1682 μV approximately (see a signal sample in Figure 6, ranging from 600 to 800 μV). The Mind Monitor app supports collecting the five common brain waves, or frequency bands (α, β, δ, γ, and θ) and applies a notch filter of 50 Hz to delete electrical power interferences. This app automatically processes the raw data coming from each channel to obtain the brain waves, using the logarithm of the power spectral density (PSD). In ascending order, the frequency bands obtained are: 1) delta (< 4Hz, common in continuous-attention tasks), 2) theta (between 4 and 7 Hz, appears in repressing a response or action), 3) alpha (between 8 and 15 Hz, spikes when relaxing or closing eyes), 4) beta (between 16 and 31 Hz, displays active thinking, focus, high alert, or anxiety), and 5) gamma (> 32 Hz, reflects cross-modal sensory processing). Figure 7 shows these 5 waves per each of the 4 channels (in total 20 signals) for left- and right-hand motor imagery. Furthermore, the headband also records other signals such as accelerometer, heart rate, breath, and muscle movement sensors (allowing to record blinking and jaw clenching). Although, some of these sensors are neither supported for developers nor by the Mind Monitor app.

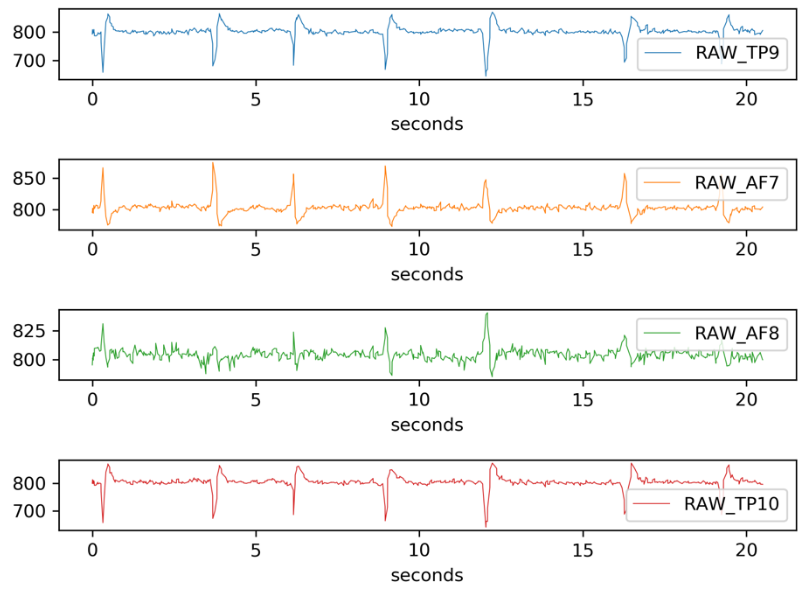

Figure 6. Sample of raw EEG signal data through the four Muse channels.

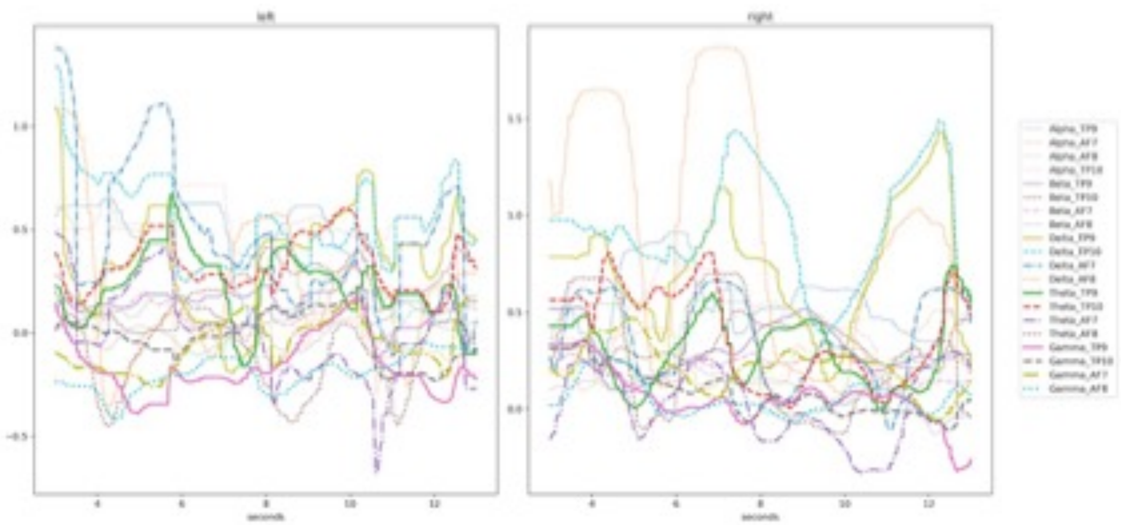

Figure 7. Sample of a participant's EEG signal for the five waves per channel. The left figure shows the brain waves for left-hand motor imagery and the right figure shows the brain waves for right hand motor imagery.

A signal vector can be represented as follows $[X^1_{t=0}, X^2_{t=0}, X^3_{t=0}, X^4_{t=0}]$, where X is the point of the signal at time 0 and for the channels 1, 2, 3, or 4 (i.e., TP9, TP10, AF7, and AF8). Therefore, in this way, we can represent the vectors of every wave type replacing "X" for the corresponding Greek letter (α, β, δ, γ, and θ).

3.3.2 Data Preprocessing

First, we need to preprocess a total amount of 20 brain wave signals recorded at a sampling rate of 256 Hz (default value). Since the Mind Monitor app could receive these data with a small delay, we have to align the signals. Additionally, sensors are prone to produce missing values. To avoid inconsistencies in the data, we interpolate the signals to fill up the missing values when possible (small chunks of missing values).

Second, we carried out a noise removal using an exploratory data analysis (EDA) in order to investigate the spatial representation. We could visualize all the signals for left and right hand imagery motions and discarded the noisy signals, mainly due to a lack of contact with headband electrodes on the participant's skin.

Third, we segmented the resulting data in different session sizes. As we mentioned in previous sections, we have data recorded during a session size of 20 s. Since our aim is to decrease this session size in order to reduce the response time of the prediction, we segmented the data in different session sizes. This preprocessing consisted in getting the signal data from the start of the recording and discarded the rest of the session data; simulating different sessions with different session size. In particular, we use sessions from 1 s to 20 s in intervals of 0.5 s (i.e. [1, 1.5, 2, 2.5, ..., 19.5, 20]). We also use small session sizes of [0.01, 0.025, 0.05, 0.1, 0.25, 0.5] seconds. In addition, we discarded the first two seconds of all recordings in order to prevent some noisy values, due to the setup of each experiment; and the subject needs to start the concentration process.

Fourth, we performed a new data segmentation again (in windows) to each session resulting from the previous segmentation, applying sliding windows with 50% overlapping (see Figure 8). We decided to use overlapping as a data augmentation technique, because this increases the sample size re-using the existing data, contributing to improve accuracy and stability [18]. We extracted the values X^i_t, (where i is the channel and t the time) of each signal corresponding to a window size. The number of values per signal depends on the window size and sampling rate (Equation 1). For example, for a 2.5 s window at a sampling rate of 256 Hz, the number of values of a signal is 640 ("points" in Figure 8). Since the timestep parameter of LSTM layers is the number of windows in which we can split the complete session, we applied the arithmetic mean to average those 640 points of each signal. Hence, we constructed one vector per window. This vector will be the neural network input. For example, if we have a window size of 2 s and sampling rate is 256 Hz, the number of windows (timesteps) will be 512. We tested with several windows sizes (0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1, 1.5, 2, 2.5, and 5), in order to get the best performance for every session and window (as in previous works [42,43]):

$$\text{signal values per window} = \frac{\text{window size (in seconds)}}{\text{sampling rate (in seconds)}} \qquad (1)$$

Finally, for every window, we can represent the vectors as a concatenation of the 5-wave values, where its components are the values of every wave signal (α, β, δ, γ, and θ) of each EEG channel (In Figure 8, the superindices 1, 2, 3, and 4, corresponding to TP9, TP10, AF7, and AF8). The vectors are represented as:

$[\alpha^1, \alpha^2, \alpha^3, \alpha^4, \beta^1, \beta^2, \beta^3, \beta^4, \delta^1, \delta^2, \delta^3, \delta^4, \gamma^1, \gamma^2, \gamma^3, \gamma^4, \theta^1, \theta^2, \theta^3, \theta^4]_{t=i}$ where i is the time, which increases in intervals of 256 Hz (0.00390625 s).

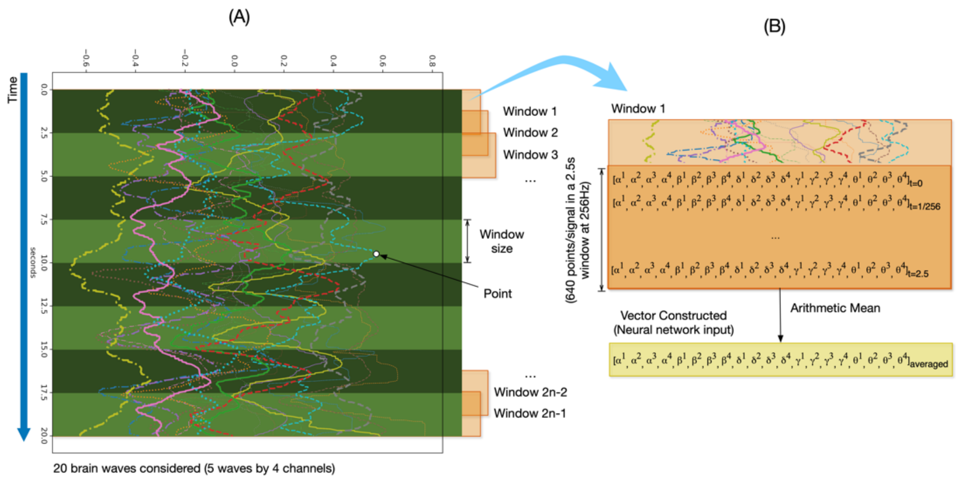

Figure 8. Construction of vector windowed input for a 20-session size applying 50% overlapping of sliding windows. (A) Data from one participant with the 20 brain waves (5 waves per each of the 4 channels —AF7, AF8, TP9 and TP10) before segmentation in windows. (B) Detailed vector values of a single window and the final vector, calculated as the arithmetic mean of the window vectors.

3.3.3 Deep Learning Architecture

In this work, we propose a DL architecture based on 1D-CNN and LSTM layers (Figure 9). The 1D-CNN layer is used to extract the most relevant signals, hereafter referred to as features, (out of the initial 20); and LSTM neural network is used to learn the sequence of the time series. In addition, we tuned the parameters of the layers (e.g., neurons) until we get the best performance. The architecture is the following:

- The input layer. This layer is three-dimensional following this triplet (samples: None, timesteps: number of windows, features: 20). In our case, samples are the number of observations, i.e., the total number of recordings, which the DL model does not know a priori and for that reason is "None". Since we used 50% overlapping of sliding windows, we defined the timesteps as the number of windows covering the full session (in total, 2n-1 windows). In addition, features, in our case, are the number of signals, i.e., twenty.
- A 1D-CNN layer with 32 filters of size 1; and kernel size equals 1, which means that each output is calculated based on the previous 1 timestep.
- A LSTM layer with 32 neurons. In order to prevent overfitting, we used 0.1 dropout, 0.0000001 regularizer and early stopping [44]. Early stopping ensures that the neural network stops the training at a certain epoch where further training would overfit the model.
- A fully connected layer with a softmax function activation, for performing the classification of the two motor intentions.

The implementation of the signal processing and the DL model architecture were coded with the Python and Keras library [45]. In Keras, the sample value of the input triplet is commonly set as "None", because we do not know the total amount of samples to use a priori in the training phase. Therefore, the algorithm will accept any number of samples. Afterwards, when we train the model, we specify the samples split for training and validation.

In the deep learning domain, it is common to split the sample size in three datasets: (1) training; (2) validation: unseen data by the training phase; (3) and testing, unused data in training and validation. Furthermore, for the training and validation, it is common to split the datasets into 80%–20% or 90%–10% partitions depending on the available dataset size. In our case, we split the data into 80% for building the model and 20% for testing and, finally, we split again this 80% into 80% for training and 20% for validation. Finally, we use accuracy (correct predictions/total predictions) as the performance metric to validate our DL model.

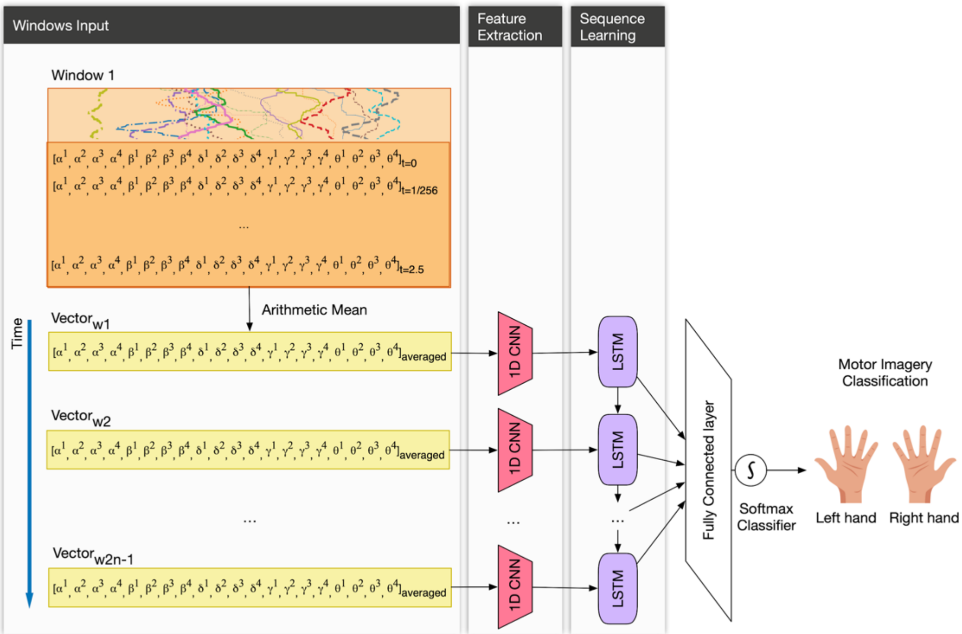

Figure 9. DL Architecture proposed based on 1-dimensional convolutional neural network (1D-CNN) and long short-term memory (LSTM) layers.

4. Results

4.1 Set of Experiments 1: Exploring Response Time Reduction

Before running the experiments, we discarded several recording sessions (104) that reported a lot of noise in several signals (see Figure 10). For example, Figure 10 in the left-hand size shows that the recordings do not show any variation of the signals. We believe this is due to the wrong contacts of the sensors with the skin of the participant. The right-hand side of Figure 10 shows some signals with outliers. Ultimately, the sample size consisted of 56 samples corresponding to two participants (28 recordings per participant) and with a total recording duration of 17 s (reduced from 20 s because some recordings did not reach 20 s after the preprocessing phase). Although the rest of the participants had some clean recordings, they were discarded because we set a minimum of 14 recordings to compare them, and they do not match this requirement.

In Figure 11, we present the results of the experiments. The x-axis represents the session size (measured in seconds) and y-axis is the test accuracy performance. Each subgraphic corresponds to different window size configurations as we explain below.

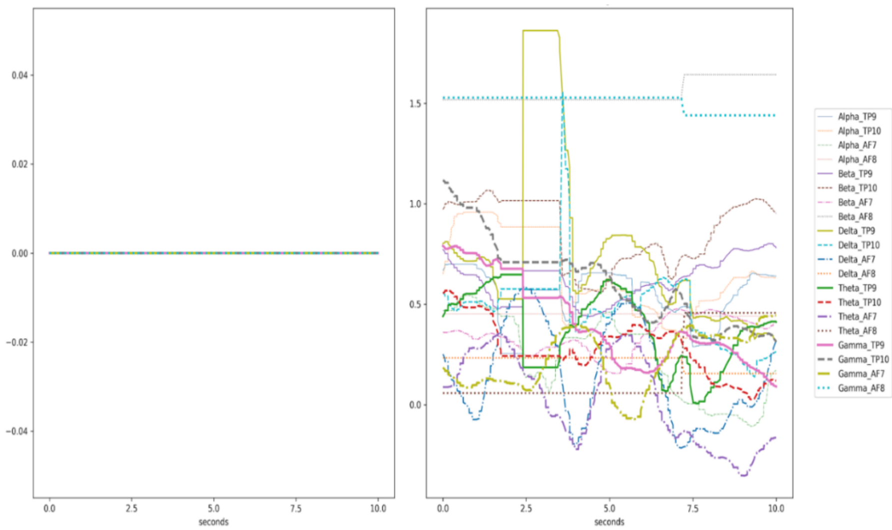

Figure 10. Examples of two noisy signals. (Left) Recordings do not show any variation of the signals, possibly due to wrong contact with the skin. (Right) Some signals with outliers in Delta TP9/TP10.

As shown in Figure 11 section A, little window sizes (0.01, 0.025, 0.05, 0.1;) reported the worst results, because they never had a stable accuracy over the different session sizes. Intermediate window sizes (0.25, 0.5, and 1; Figure 11 section B) reported good results from 6.5 s of session size onwards. This means that the response time of the prediction has been reduced from 20 s to 6.5. This is already an improvement. Figure 11 section C shows big window sizes (1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5). These window sizes report also good results from 6.5 s window sessions onwards, varying between 0.75 and 0.9 accuracy for session sizes between 4 and 14 s. Figure 11 section D shows window sizes of 1.5, 2, 3, 4.5, and 5 s that reach an accuracy between 0.83 and 0.92 for session sizes between 2 and 14 s. Summarizing, with small session sizes, we cannot discriminate between left- and right-hands intention; and with big session sizes a user will have to wait for a long response time.

Therefore, the best stable accuracies appear on windows sizes corresponding to 1.5, 2, 3, 4.5, and 5 s. Since, 1.5 s-window size is stable from 4 s of session size and 2 s-window size is stable from 2 s, we selected 2 s-window size as the best candidate for our DL model, which reduces to 2 s the time-response for the motor imagery classification. In particular, the accuracy of 2 s-window size, ranged from 0.83 and 0.92 approximately for session sizes between 2 and 13.5 s (Figure 12). After 14 s it reaches 0.99 accuracy. This configuration —2 s-window size— maintains a balance between the time-response and an acceptable accuracy performance, because it is able to make an accurate prediction of imagery motion from 2–2.5 s of session duration.

With respect to the related research, we used the brain waves α, β, δ, γ, and θ through TP9, TP10, AF7, and AF8 channels as in previous results [13], due to the accurate classification achieved. For that reason, we discarded raw data as in [25] or only the gamma waves of AF7 and AF8 channels used in [26]. We also prevented overfitting using dropout layers, regularizers, and early stopping. Moreover, we also tested different window and session sizes configurations to find the best setup. However, more importantly, our results go one step further, we achieved to reduce the session size from 20 s to 2 s getting an acceptable accuracy between 83% and 92% also with 2 s window size (1 single window). Thus, our DL model is available to predict the correct imagery motion in a time-response lower than the related work. Although 2 s of time-response is not a real-time (immediate) response, these results are promising overcoming all of the existing previous work (Table 1).

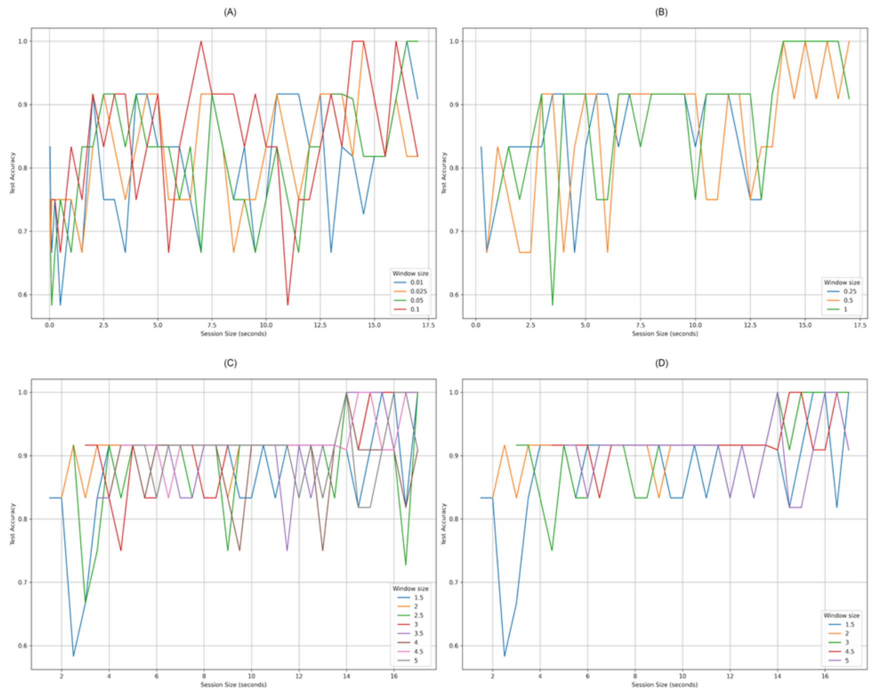

Figure 11. Test accuracy curves for different window sizes. (A) Unstable accuracies for very little window sizes. (B) Stable accuracies from 6.5 s session size. (C) Some stable window sizes from 4 s session size. (D) Best stable window sizes between 2 and 4 s of session size.

Furthermore, the participants reported a good usability of the system during the experiment. They felt very good wearing the headband, because it is comfortable, low-intrusive, and the data were collected transparently for them.

4.2 Experiment 2: Validating the Response Time Reduction

In order to validate the response time reduction achieved in the previous experiment, we obtained data from all of the participants recruited; 20 recordings per participant. The main idea behind this experiment is to validate the model for motor imagery classification for the 2 s session sizes, while a participant is imaging picking up bottles during a minute, changing the hand-side every 6 s. Since we recorded the data from the Mind Monitor app, we used labelling buttons for creating labels (marks) in the recordings. Thus, we know exactly when we asked the participant to change form right to left or vice versa. We selected 6 s to change sides, in order to cover the 2 s for classification and the possible delays. These delays involve asking the participant to change sides and to manually press labelling button in the app. This could imply a latency in these two actions (see Figure 12).

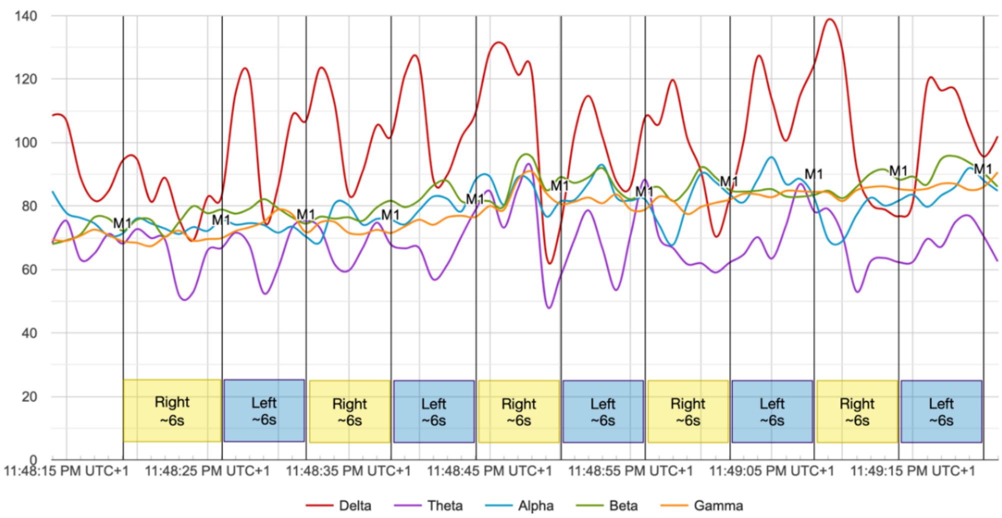

Figure 12. Recording example of experiment 2. M1 is the mark 1, labelled to know when the participant was imaging right/left side. Averaged channel values per wave type.

The number of trials consisted of 60 recordings in total (20 times 3), and every trial had 10 samples (5 for right; and 5 for left motor imagery). We created three different models, one per participant and reserved 20% of the recordings for testing (4 recordings per participant). Then, the resulting 16 recordings were split again into 80%–20% for training and validation, respectively (12 for training; 4 for validation). In terms of samples, we used in total 120 samples for training the model, 40 samples for validation, and 40 samples for testing. We repeated this model creation process over 5 iterations in order to randomize the selection of recordings, and finally averaged the results of the 5 iterations.

We tested the DL pipeline for the 2 s time-response with different window sizes. The results are presented in Table 2. The best test accuracy averaged over the three participants was 83.8% using 0.5 s window size. One participant achieved the best test accuracy of 93% (the only male), followed by the other participants with 80.5% and 78%.

Table 2. Validation of DL models for 2 s time-response.

WS	TrAcc1	VaAcc1	TeAcc1	TrAcc2	VaAcc2	TeAcc2	TrAcc3	VaAcc3	TeAcc3	TrAvg	VaAvg	TeAvg
2s	0.773846	0.760000	0.700000	0.795385	0.746667	0.705000	0.829231	0.786667	0.820000	0.799487	0.764444	0.741667
1.5s	0.747692	0.753333	0.720000	0.807692	0.806667	0.725000	0.847692	0.880000	0.850000	0.801025	0.813333	0.765000
1s	0.912308	0.840000	0.795000	0.829231	0.826667	0.740000	0.949231	0.920000	0.905000	0.896923	0.862222	0.813333
0.5s	0.920000	0.833333	0.805000	0.906154	0.860000	0.780000	0.940000	0.913334	0.930000	0.922051	0.868889	0.838333
0.25s	0.896923	0.873333	0.765000	0.832308	0.840000	0.745000	0.916923	0.873333	0.890000	0.882051	0.862222	0.800000
0.1s	0.841538	0.833333	0.760000	0.803077	0.773333	0.700000	0.840000	0.853333	0.855000	0.828205	0.820000	0.771667
0.05s	0.835384	0.866667	0.730000	0.692308	0.706667	0.720000	0.818462	0.853333	0.845000	0.782051	0.808889	0.765000

Averaged all accuracies from 5 iterations.
Best test accuracies in bold letters. Best individual test-accuracy underlined.
WS: window size; TrAccX: train accuracy of participant X; VaAccX: validation accuracy of participant X; TeAccX: test accuracy of participant X; TrAvg: train accuracy average; VaAvg: validation accuracy average; TeAvg: test accuracy average.

These results validate our previous set of experiments, confirming that in two seconds, we are able to classify the motor imagery with an accurate value. These results are not exactly the same as in the first set of experiments, partially due to the different participants. It is well-known that the brain activity is quite personal and can change over time [46]. Although we can have some general patterns, neither all the persons have the same sensibility in brain activity, nor the same concentration ability [46]. In any case, these results overcome the related works using the Muse headband (low-cost EEG device), reducing time-response and maintaining a decent accuracy. Furthermore, our accuracy results also overcome the related works which reduce the time-response in the range of 2–4 s (these studies used high-intrusive devices). The best work [36] reached 93.9% accuracy (but with 4 s time-response), followed by [39,40] with 90.5% accuracy (3 s time-response), and [36] with 70% accuracy (2 s time-response). The underlined differences on the motor imagery classification among the participants should be study further with the help of brain specialists, and study the innate skills on imaging the hand-movement of each participant.

5. Conclusions and Future Work

In this paper, we have explored what is the smallest response time that still obtains a competitive accuracy on the detection of motor imagery using a low-cost and low-intrusive EEG headband. We used a deep learning pipeline, based on 1D-CNN and LSTM layers, to explore the best reduction of the response time in motor imagery classification with a low-cost and low-intrusive BCI headband.

Our results showed that motor imagery classification for left and right hands, with low-cost and low-intrusive headband devices (with 4 electrodes), is feasible, accurate, and can reduce the time-response to 2 s. The proposed deep learning model architecture, based on 1D-CNN and LSTM layers, ensures an accuracy between 83% and 92% with unseen data for a time-response of 2 s or longer. These results overcome related works with intrusive and with non-intrusive EEG devices. High-intrusive EEG caps, with hundreds of electrodes, reach accuracies between 40% and 93.9% and provide a time-response of 4 s (two times slower than our architecture). Low-intrusive devices in the literature use session sizes greater than 10 s.

We have found that low-intrusive devices (headbands) can compete with high-intrusive devices in some tasks; in particular, in motor imagery for right and left hand. Although the electrodes of the headbands do not cover the primary cortex area (C3 and C4 electrodes), they cover the frontal lobe of the brain, which is related with movement control functions of the brain. Hence, the headbands could still detect movement (motor) imagery.

This proposal has the potential of increasing people's independence, allowing them to interact with computer systems and applications without using their bodies, but just thinking or imagining the movement.

In the future, we would like to increase the sample size to enhance the accuracy, and also to apply our algorithms to other participants comparing not only the participants, but also to further study the personal responses and how they affect the personalities in the classification. Additionally, we want to compare the effect of different percentages of overlapping sliding windows. We will also explore the possibility to further reduce the response time. This is not an irrelevant task, as in real life we have to offer an immediate response, and 2 s, which is reported as the best accuracy in our experiments, could slow down the user experience. Furthermore, although the headband used in this work is low-intrusive compared to similar EEG devices, it is still awkward to wear in outdoors activities. In the future, probably other non-intrusive devices will appear, which will be suitable for outdoors activities, and we would like to study them. These next-generation devices may also have more electrodes than the Muse and located on motor cortex (maintaining the low-intrusiveness); and could improve the accuracy and response time. Furthermore, the properties of the Muse headband such as low-cost and low-intrusive contribute to democratize the adoption of these BCI wearable technologies in the health domain. In this line, another of our objectives is the recognition of activities, mental state, and emotions of the users while they are performing daily life activities and the relationship with their health status. The headband will be part of our previous e-health system based on microservices and the cloud [42,47], which includes other wearable sensors and applications to monitor and intervene in the healthcare of the elderly. Thus, the integration of the proposal presented in this paper with our previously designed e-health system will be able to contribute to a practical solution for healthcare systems.

Author Contributions

"Conceptualization, All authors; methodology, F.M.G.M., and M.B.E.; software, F.M.G.M., and J.L.G.; validation, F.M.G.M., and M.J.R.F.; formal analysis, F.M.G.M, and M.B.E.; investigation, All authors; resources, All authors; data curation, F.M.G.M., and M.B.E.; visualization, F.M.G.M.; writing—original draft preparation, F.M.G.M.; writing—review and editing, M.B.E., J.L.G. and M.J.R.F.; supervision, M.J.R.F. and M.B.E.; project administration, J.L.G., and M.J.R.F.; funding acquisition, All authors. All authors have read and agreed to the published version of the manuscript."

Funding

This research work is partially funded by: the Spanish Ministry of Economy and Competitiveness (Agencia Estatal de Investigación-AEI) and the European Regional Development Funds (FEDER, UE) through the research project Ref. TIN2016-79484-R; and by the Spanish Ministry of Science and Innovation through the research project Ref. PID2019-109644RB-I00 / AEI / 10.13039/501100011033 and the Scholarship Program FPU Ref. FPU18/00287.

Acknowledgments

We would like to thank Ms Fernanda Lugmaña-Hidalgo for the photographs and time to participate in this work. We would also like to thank the participants from University of Granada (Spain) for their wiliness of taking part in this study.

Appendix A

AFX	position of an electrode in BCI, with number X
BCI	brain computer interfaces
CNN	convolutional neural network
DL	deep learning
EEG	electroencephalography

Fpz reference electrode in BCI

LR logistic regression

LSTM long short-term memory

ML machine learning

NN neural network

RNN recurrent neural network

SDK software development kits

SVM support vector machines

TPX position of an electrode in BCI, with number X

References

1. Dunn, J.; Runge, R.; Snyder, M. Wearables and the medical revolution. *Per. Med.* **2018**, *15*, 429–448.
2. Aggravi, M.; Salvietti, G.; Prattichizzo, D. Haptic assistive bracelets for blind skier guidance. In Proceedings of the 7th Augmented Human International Conference 2016 on–AH '16, Geneva, Switzerland, 25–27 February 2016; pp. 1–4.
3. Majumder, S.; Mondal, T.; Deen, M. Wearable sensors for remote health monitoring. *Sensors* **2017**, *17*, 130.
4. Elsaleh, T.; Enshaeifar, S.; Rezvani, R.; Acton, S.T.; Janeiko, V.; Bermudez-Edo, M. IoT-Stream: A lightweight ontology for internet of Things data streams and Its USE with data analytics and event Detection SERVICES. *Sensors* **2020**, *20*, 953.
5. Enshaeifar, S.; Zoha, A.; Markides, A.; Skillman, S.; Acton, S.T.; Elsaleh, T.; Hassanpour, M.; Ahrabian, A.; Kenny, M.; Klein, S.; et al. Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. *PLoS ONE* **2018**, *13*, e0195605.
6. Fico, G.; Montalva, J.-B.; Medrano, A.; Liappas, N.; Mata-Díaz, A.; Cea, G.; Arredondo, M.T. Co-creating with consumers and stakeholders to understand the benefit of internet of things in smart living environments for ageing well: The approach adopted in the Madrid Deployment Site of the ACTIVAGE Large Scale Pilot. In *IFMBE Proceedings*; Springer: Singapore, Singapore, 2018; pp. 1089–1092, ISBN 10.1007/9789811.
7. Greene, B.R.; McManus, K.; Redmond, S.J.; Caulfield, B.; Quinn, C.C. Digital assessment of falls risk, frailty, and mobility impairment using wearable sensors. *Npj. Digit. Med.* **2019**, *2*, 1–7.
8. Zhang, X.; Yao, L.; Huang, C.; Sheng, Q.Z.; Wang, X. Intent recognition in smart living Through DEEP recurrent neural networks. In Proceedings of the International Conference on Neural Information Processing; Guangzhou, China, 14–18 November 2017; pp. 748–758.
9. Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; Mcalpine, D.; Zhang, Y. A Survey on deep learning based brain Computer INTERFACE: Recent advances and New frontiers. *arXiv* **2019**, arXiv:physics/0402096.
10. Nicolas-Alonso, L.F.; Gomez-Gil, J. Brain computer interfaces–A review. *Sensors* **2012**, *12*, 1211–1279.
11. InteraXon Muse 2: Brain Sensing Headband–Technology Enhanced Meditatiomn. Available online: https://choosemuse.com/muse-2/(accessed on 27 January 2020).
12. Jurcak, V.; Tsuzuki, D.; Dan, I. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *Neuroimage* **2007**, *34*, 1600–1611.
13. Garcia-Moreno, F.M.; Bermudez-Edo, M.; Rodriguez-Fortiz, M.J.; Garrido, J.L. A CNN-LSTM deep Learning classifier for motor imagery EEG detection using a low-invasive and low-Cost BCI headband. In Proceedings of the 16th International Conference on Intelligent Environments (IE), Madrid, Spain, 20–23 July 2020; pp. 84–91.
14. Bermudez-Edo, M.; Barnaghi, P. Spatio-temporal analysis for smart city data. In Proceedings of the Companion of the The Web Conference 2018 on The Web Conference 2018–WWW '18, Lyon, France, 23–27 April 2018.
15. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
16. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A.Y. Multimodal deep learning. In Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, WA, USA, 28 June–2 July, 2011.
17. Ravi, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.-Z. Deep learning for health informatics. *IEEE J. Biomed. Heal. Inf.* **2017**, *21*, 4–21.

18. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001.
19. Ramadan, R.A.; Vasilakos, A.V. Brain computer interface: Control signals review. *Neurocomputing* **2017**, *223*, 26–44.
20. Bird, J.J.; Faria, D.R.; Manso, L.J.; Ekárt, A.; Buckingham, C.D. A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction. *Complexity* **2019**, *2019*, 1–14.
21. Lotte, F. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proc. IEEE* **2015**, *103*, 871–890.
22. Wiechert, G.; Triff, M.; Liu, Z.; Yin, Z.; Zhao, S.; Zhong, Z.; Zhaou, R.; Lingras, P. Identifying users and activities with cognitive signal processing from a wearable headband. In Proceedings of the IEEE 15th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Palo Alto, CA, USA, 8 August 2016; pp. 129–136.
23. Salehzadeh, A.; Calitz, A.P.; Greyling, J. Human activity recognition using deep electroencephalography learning. *Biomed. Signal Process. Control* **2020**, *62*, 102094.
24. Zhao, D.; MacDonald, S.; Gaudi, T.; Uribe-Quevedo, A.; Martin, M.V.; Kapralos, B. Facial expression detection employing a brain computer interface. In Proceedings of the 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece, 23–25 July 2018; pp. 1–2.
25. Rodriguez, P.I.; Mejia, J.; Mederos, B.; Moreno, N.E.; Mendoza, V.M. Acquisition, analysis and classification of EEG signals for control design. arXiv **2018**, arXiv:physics/0402096.
26. Li, Z.; Xu, J.; Zhu, T. Recognition of Brain Waves of Left and Right Hand Movement Imagery with Portable Electroencephalographs. *arXiv* **2015**, arXiv:physics/0402096.
27. Zhang, D.; Yao, L.; Zhang, X.; Wang, S.; Chen, W.; Boots, R. EEG-based intention recognition from spatio-temporal representations via Cascade and parallel convolutional Recurrent neural networks. *aRxiv* **2017**, arXiv:physics/0402096.
28. Chen, W.; Wang, S.; Zhang, X.; Yao, L.; Yue, L.; Qian, B.; Li, X. EEG-based motion intention recognition via multi-task RNNs. In Proceedings of the 2018 SIAM International Conference on Data Mining; Society for Industrial and Applied Mathematics: San Diego, CA, USA, 3-5 May 2018; pp. 279–287.
29. Kaya, M.; Binli, M.K.; Ozbay, E.; Yanar, H.; Mishchenko, Y. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Sci. Data* **2018**, *5*, 180211.
30. Brunner, C.; Leeb, R.; Müller-Putz, G.R.; Schlögl, A.; Pfurtscheller, G. BCI Competition IV 2008–Graz data set A. Available online: http://www.bbci.de/competition/iv/desc_2a.pdf (accessed on 24 November 2020).
31. Brunner, C.; Leeb, R.; Müller-Putz, G.R.; Schlögl, A.; Pfurtscheller, G. BCI Competition 2008–Graz data set B. Available online: http://www.bbci.de/competition/iv/desc_2b.pdf (accessed on 24 November 2020).
32. Bhattacharyya, S.; Khasnobish, A.; Konar, A.; Tibarewala, D.N.; Nagar, A.K. Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms. In Proceedings of the IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), Paris, France, 11–15 April 2011; pp. 1–8.
33. Tang, Z.; Li, C.; Sun, S. Single-trial EEG classification of motor imagery using deep convolutional neural networks. *Optik* **2017**, *130*, 11–18.
34. Tabar, Y.R.; Halici, U. A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* **2017**, *14*, 016003.
35. Wang, P.; Jiang, A.; Liu, X.; Shang, J.; Zhang, L. LSTM-based EEG classification in motor imagery tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2086–2095.
36. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420.
37. Ha, K.-W.; Jeong, J.-W. Motor imagery EEG classification using capsule networks. *Sensors* **2019**, *19*, 2854.
38. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013.
39. Blankertz, B.; Dornhege, G.; Krauledat, M.; Muller, K.-R.; Kunzmann, V.; Losch, F.; Curio, G. The Berlin brain–computer interface: EEG-based communication without subject training. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2006**, *14*, 147–152.
40. Tomioka, R.; Aihara, K.; Müller, K.-R. Logistic Regression for Single Trial EEG Classification. In Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 4–7 December 2006; 19, pp. 1377–1384.
41. Schmidt, R.F.; Thews, G. *Human Phys.*; Springer: Berlin, Heidelberg, 1989.
42. Garcia-Moreno, F.M.; Bermudez-Edo, M.; Garrido, J.L.; Rodríguez-García, E.; Pérez-Mármol, J.M.; Rodríguez-Fórtiz, M.J. A microservices e-Health system for ecological frailty assessment using wearables. *Sensors* **2020**, *20*, 3427.
43. Bermudez-Edo, M.; Barnaghi, P.; Moessner, K. Analysing real world data streams with spatio-temporal correlations: Entropy vs. Pearson correlation. *Autom. Constr.* **2018**, *88*, 87–100.

44. Pereyra, G.; Tucker, G.; Chorowski, J.; Kaiser, Ł.; Hinton, G. Regularizing neural networks by penalizing confident output distributions. In Proceedings of the 5th International Conference on Learning Representations (ICLR 2017), Toulon, France, 24–26 April 2017.
45. Chollet, F. Others Keras Available online: https://keras.io (accessed on 24 November 2020).
46. Lebedev, A.V.; Kaelen, M.; Lövdén, M.; Nilsson, J.; Feilding, A.; Nutt, D.J.; Carhart-Harris, R.L. LSD-induced entropic brain activity predicts subsequent personality change. *Hum. Brain Mapp.* **2016**, *37*, 3203–3213.
47. García-Moreno, F.M.; Rodríguez-García, E.; Rodríguez-Fórtizm M.; Garrido, J.L.; Bermúdez-Edo, M.; Villaverde-Gutiérrez, C.; Pérez-Mármol, J.M. Designing a smart mobile health system for ecological frailty assessment in elderly. *Proceedings* **2019**, *31*, 41.

UNIVERSIDAD
DE GRANADA