International Journal of Medical Informatics 157 (2022) 104625

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

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

Francisco M. Garcia-Moreno [a], Maria Bermudez-Edo [a,*], Estefanía Rodríguez-García [b], José Manuel Pérez-Mármol [b], José Luis Garrido [a], María José Rodríguez-Fórtiz [a]

[a] Department of Software Engineering, Computer Sciences School, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18014 Granada, Spain
[b] Department of Physiology, Faculty of Health Sciences, University of Granada, Av. de la Ilustración, 60, 18016 Granada, Spain

ARTICLE INFO

Keywords:
Dependence assessment
IADL
Older adults
Machine learning
Wearable sensors
E-health
Prediction

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.

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

Abbreviations: AI, Artificial Intelligence; ML, Machine Learning; ADL, Activity of Daily Living; IADL, Instrumental Activity of Daily Living; BADL, Basic Activity of Daily Living; SS, Shopping Stages; TS, Trial Supervisor; k-NN, k-Nearest Neighbors; RF, Random Forest; SVM, Support Vector Machine; LBS, Lawton Brody Scale; BAN, Body Area Network; MBU, Mobile Base Unit; mSS, m-Health Service System; SDK, Software Development Kit; IJMEDI, International Journal of Medical Informatics; EDA, ElectroDermal Activity; SMOTE, Synthetic Minority Oversampling Technique; RFE, Recursive Feature Elimination; CV, Cross Validation; 5-FSCV, 5-Fold Stratified CV; RWCV, Record-Wise CV; SWCV, Subject-Wise CV.

* Corresponding author: Department of Software Engineering, Computer Sciences School, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n, 18014 Granada, Spain.

E-mail addresses: fmgarmor@ugr.es (F.M. Garcia-Moreno), mbe@ugr.es (M. Bermudez-Edo), rodrigueze@ugr.es (E. Rodríguez-García), josemapm@ugr.es (J.M. Pérez-Mármol), jgarrido@ugr.es (J.L. Garrido), mjfortiz@ugr.es (M.J. Rodríguez-Fórtiz).

https://doi.org/10.1016/j.ijmedinf.2021.104625
Received 26 March 2021; Received in revised form 20 October 2021; Accepted 22 October 2021
Available online 3 November 2021
1386-5056/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

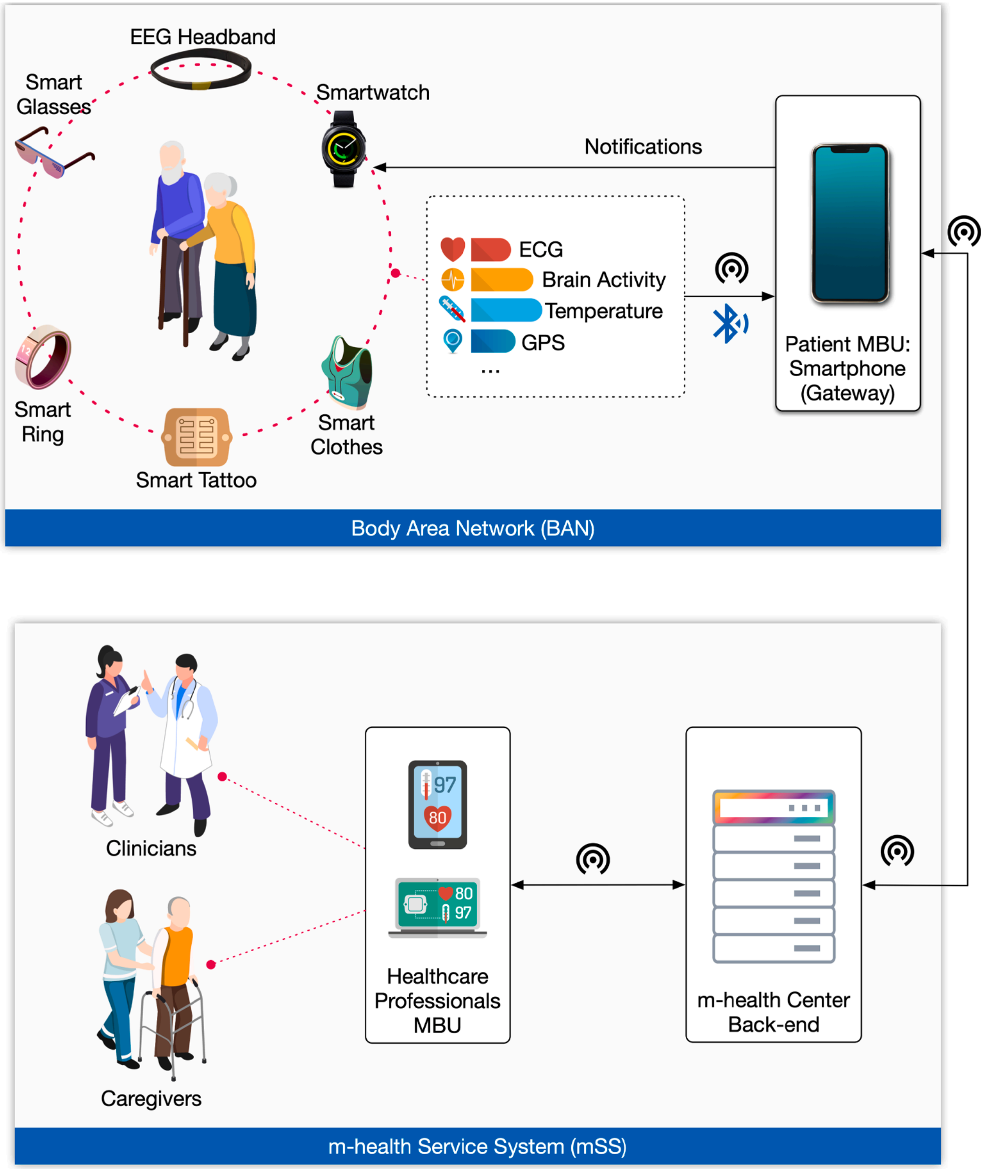

Fig. 1. A generic m-health scenario overview.

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

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.

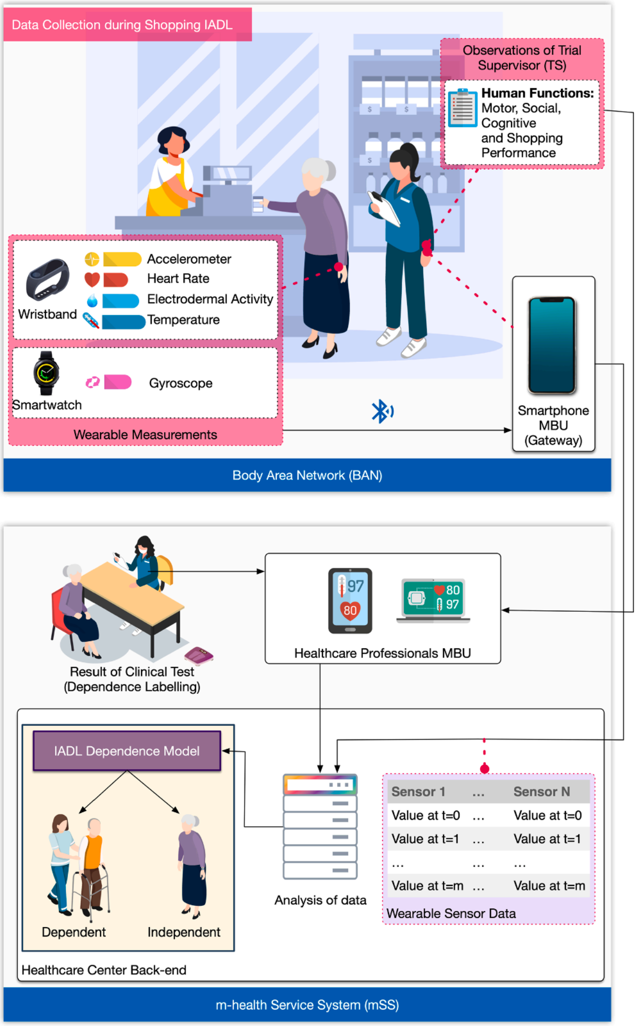

Fig. 2. Overview of the proposed m-health scenario.

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

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].

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.

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

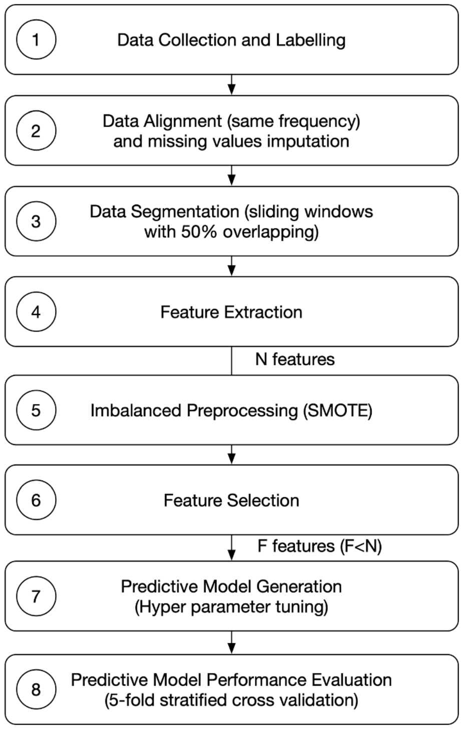

Fig. 3. Machine learning pipeline of the IADL dependence classification process. N features are the total features extracted from the Data Segmentation step. The feature selection step selects the F most relevant features for this study out of the N total features collected ($F < N$).

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.

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.

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

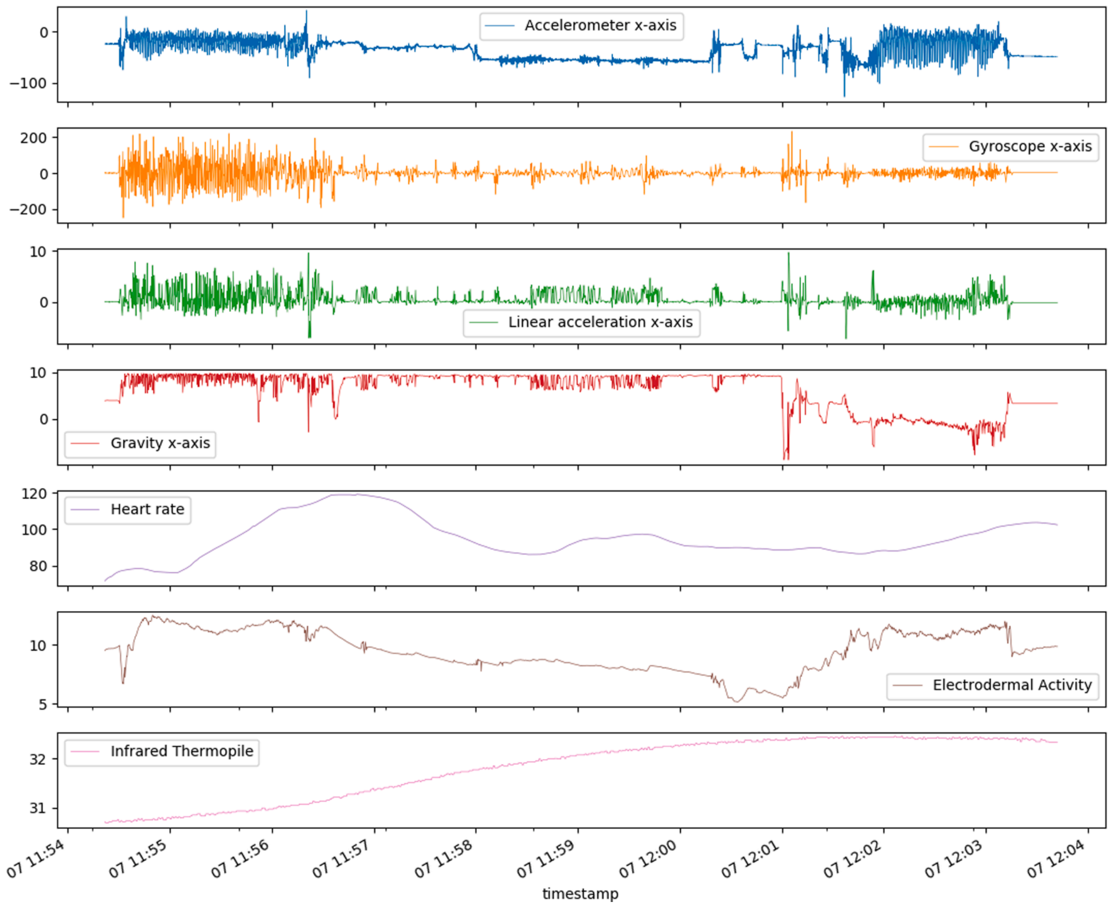

Fig. 4. Some aligned wearable signals from an anonymous subject at 25 Hz during the performance of the shopping IADL.

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.

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.

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).

$$\text{maximum} = 220 - \text{subject age} \tag{1}$$

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.

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

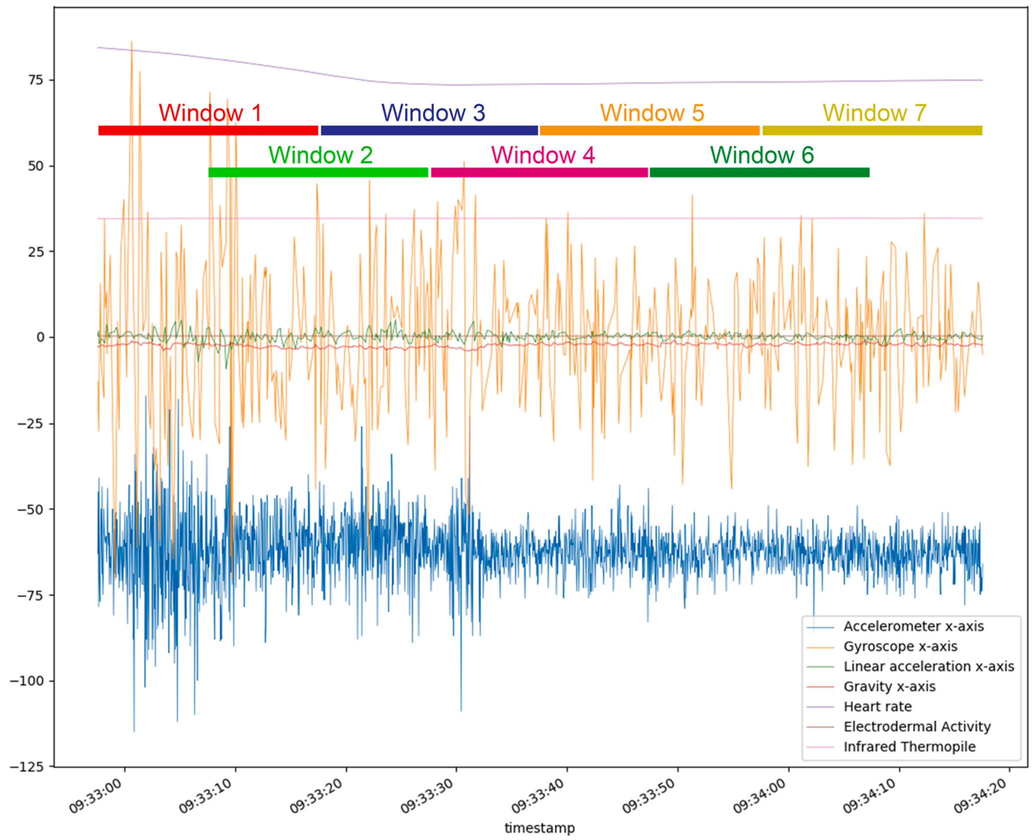

Fig. 5. Sliding windows approach using 50% overlapping.

raw data with different window sizes measured in seconds (0.5 s, 1 s, 1.5 s, 2 s, 2.5 s and 3 s). 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.

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.

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.

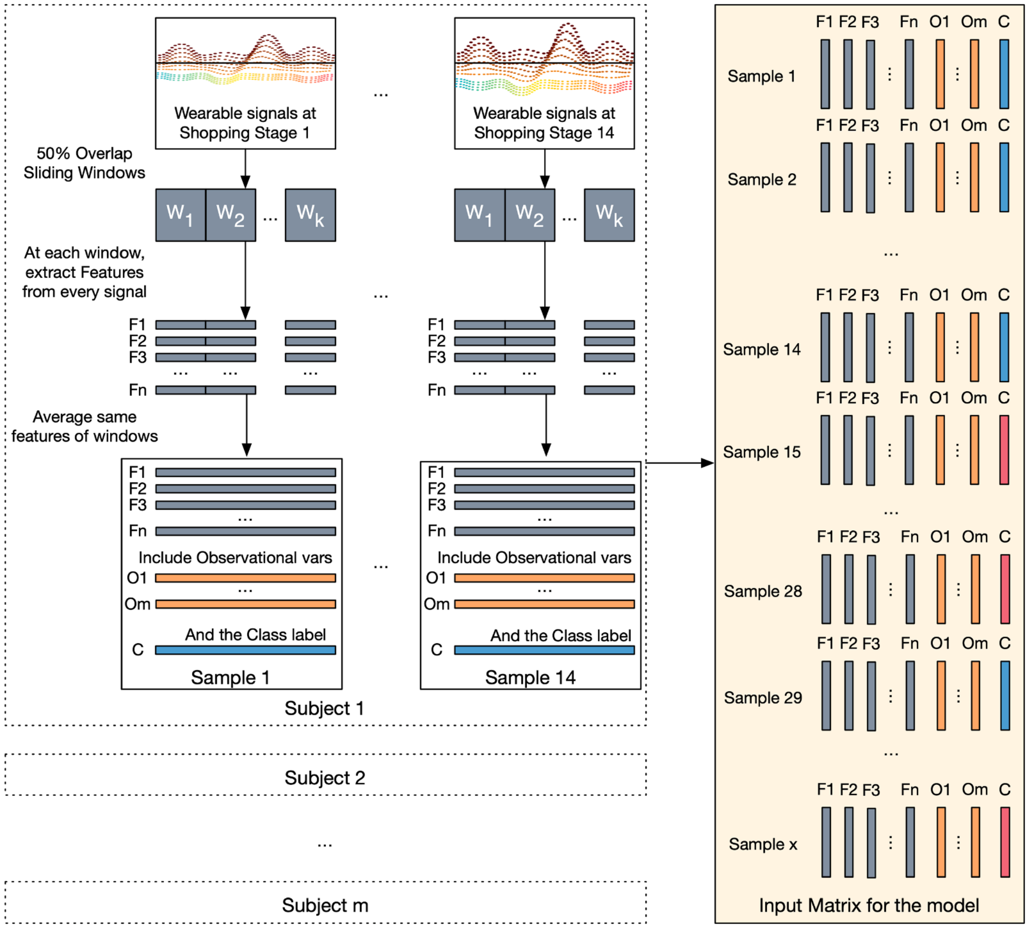

Fig. 6. Creation of the input matrix for each model (k-NN, RF or SVM). The blue color of the field C (label of the sample) represents a dependent subject, while the red color represents an independent subject. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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).
[a]Triaxial sensor: accelerometer and gyroscope.
[b]Uniaxial sensor: heart rate, EDA and IT.
Total features: $(8{\cdot}3){\cdot}2 + (8{\cdot}3) = 72$ wearable features.

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 Section 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.

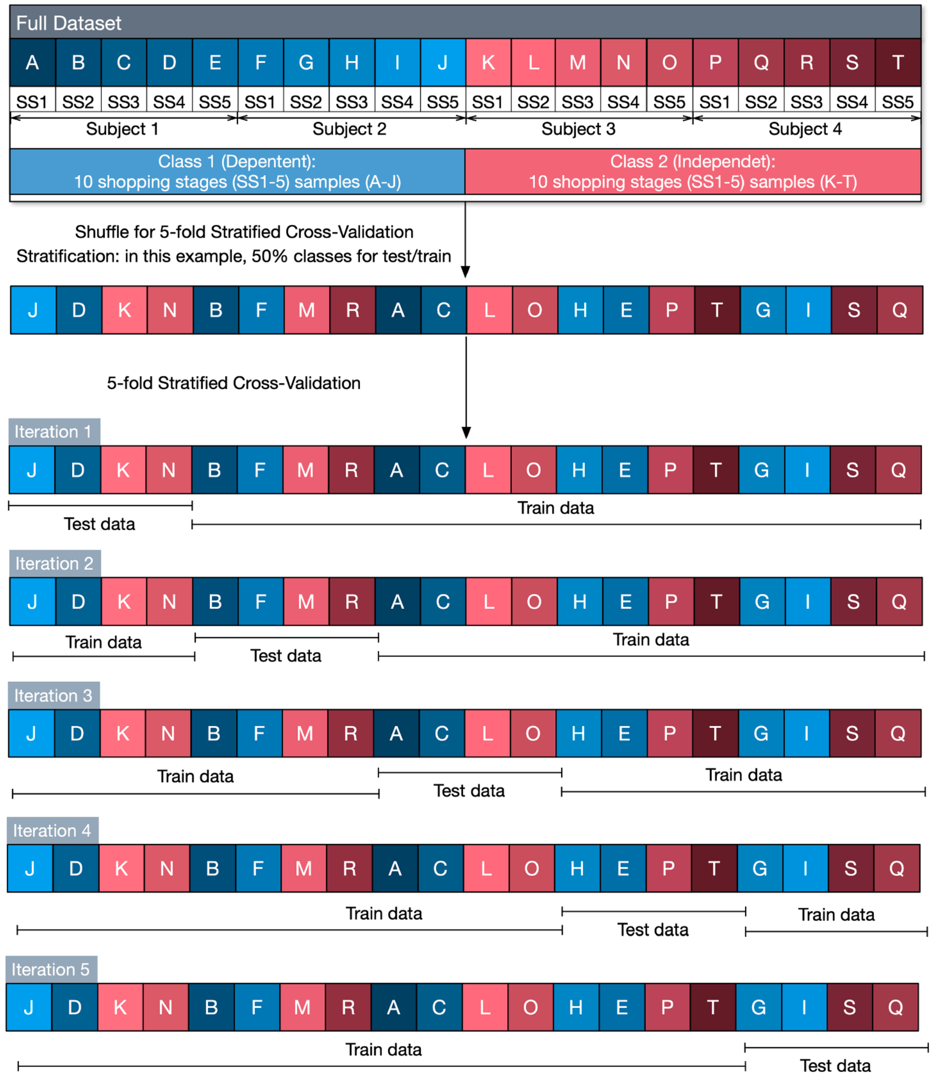

Fig. 7. Example of 5-Fold Stratified Cross Validation with 20 samples.

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).

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).

The results were stabilized when we took over 10 features and we achieved the best performance for windows of 0.5 s (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

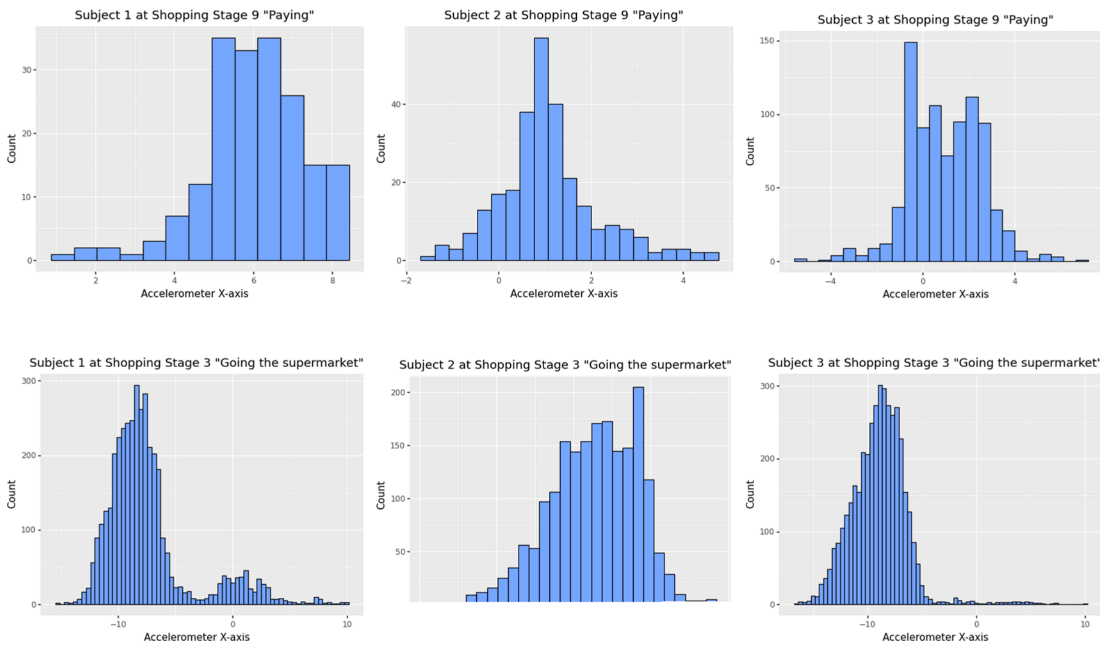

Fig. 8. Example of three subjects' data distributions where Accelerometer x-axis feature is different per subject and per split.

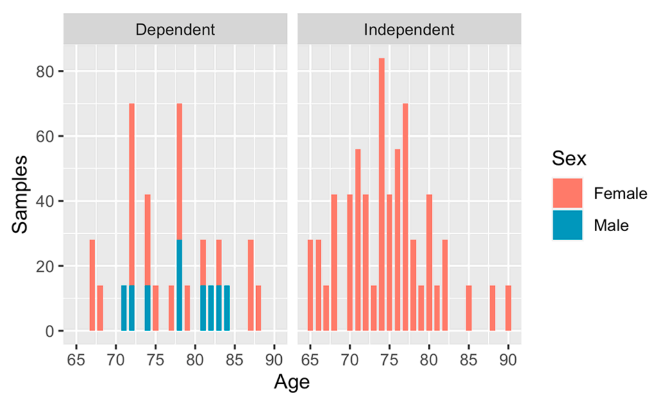

Fig. 9. Ex2: Distribution of subject's age and sex grouped by the result of Lawton & Brody scale.

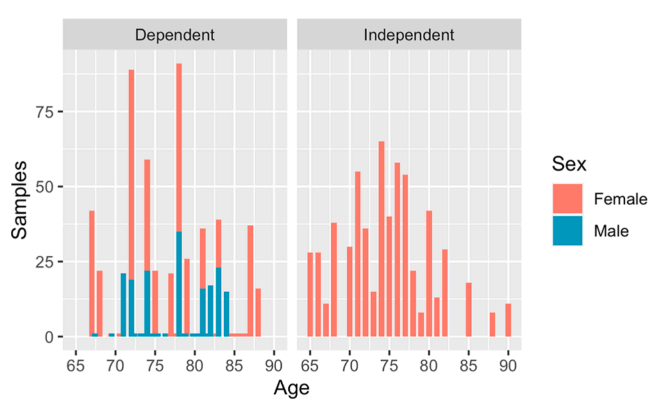

Fig. 10. Ex2: Distribution of subject's age and sex after SMOTE and grouped by the result of Lawton & Brody scale.

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).

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.

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).

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

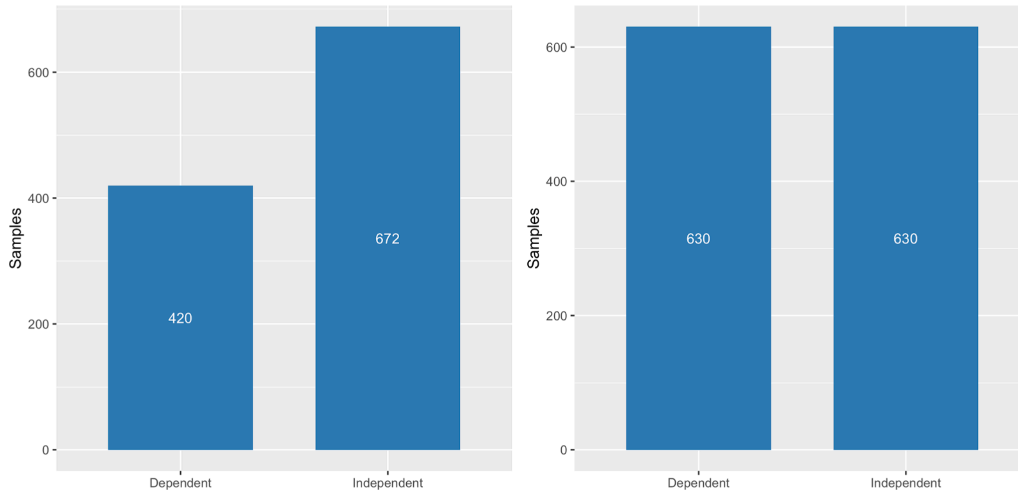

Fig. 11. Ex2: Class proportion comparison between original imbalance data (left) and preprocessed data with Smote to balance the original data (right).

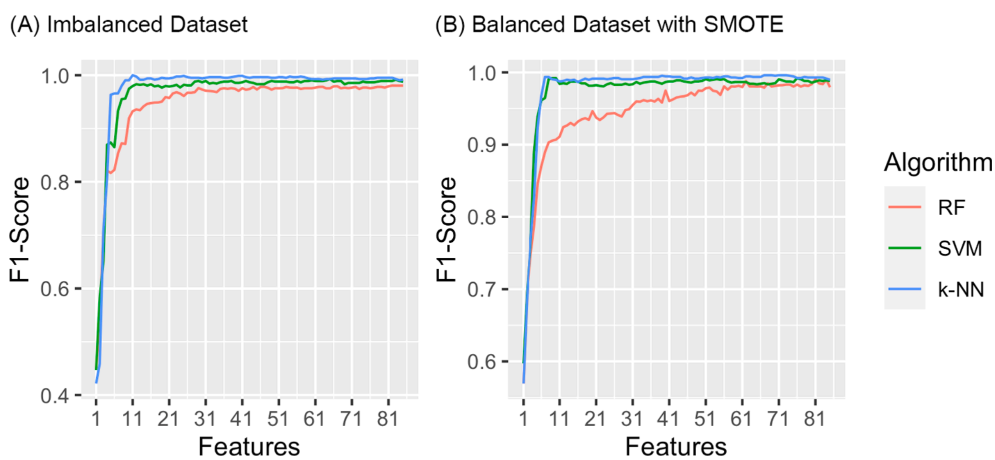

Fig. 12. Ex2: F1-score over all features with the original Dataset (A) and after SMOTE (B).

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

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).

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

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
39	0.9951802	0.9963303	1.0000000	0.9941279

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.

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).

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).

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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research conducted in this publication was funded by the Spanish Ministry of Economy and Competitiveness (Agencia Estatal de Investigación) under grant number: PID2019-109644RB-I00/AEI/10.13039/501100011033. Funding for open access charge:

Universidad de Granada / CBUA.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2021.104625.

References

[1] J. Devi, The scales of functional assessment of Activities of Daily Living in geriatrics, Age 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 Med. J. 3 (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 (3 Part 1) (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, Med. J. 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, Int. J. Med. Health Sci. 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: Translat. Res. Clin. Intervent. 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, Am. J. Occup. 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 Programs Biomed. 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. Zhu, L. Cilliers, C.-P. Wong, P. Bamidis, Y.-C. Li, Li, Mobile and wearable technologies in healthcare for the ageing population, Computer Methods Programs Biomed. 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 Lett. 29 (16) (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 Trans. Biomed. Eng.. 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 Trans. Biomed. Circuits Syst. 5 (4) (2011) 320–329, https://doi.org/10.1109/TBCAS.2011.2160540.
[14] T. Broens, A.V. Halteren, M.V. Sinderen, K. Wac, Towards an application framework for context-aware m-health applications, Int. J. Internet Protocol Technol. 2 (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 Health Technol. Inform. 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, Med. Eng. Phys. 37 (7) (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 J. Biomedical Health Inform. 23 (4) (2019) 1585–1594, https://doi.org/10.1109/JBHI.622102010.1109/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 (IADL), StatPearls Publishing, StatPearls., 2020.
[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 Quality Life Outcomes. 10 (1) (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, Comput. Methods Programs Biomed. 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 J. Biomed. Health Inform. 23 (4) (2019) 1618–1630, https://doi.org/10.1109/JBHI.622102010.1109/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 Med. 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 (3) (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, Front. 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, Clin. Interventions 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 (4) (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, 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 (4) (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, Int. J. Medical Inform. 153 (2021) 104510, 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, Can. Med. Assoc. J. 187 (2015) E442–E449, https://doi.org/10.1037/1082-989X.7.1.19.
[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 J. Health Phys. 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, Psychol. 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 Constr. 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, 16th International Conference on Intelligent Environments (2020), https://doi.org/10.1109/IE49459.2020.9155016.
[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 Inform. 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 Med. Assoc. J. 187 (15) (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, J. Artificial Intelligence Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[57] T.N. Lal, O. Chapelle, J. Weston, A. Elisseeff, in: Studies in Fuzziness and Soft ComputingFeature Extraction, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 137–165.
[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 Sci. 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 Med. 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 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 Trans. Inform. Technol. Biomed. 14 (6) (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, Proc. Comput. Sci. 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 Technol. Lett. 2 (1) (2015) 2–5, https://doi.org/10.1049/htl2.v2.110.1049/htl.2014.0104.
[66] S.-T. Liaw, A. Georgiou, H. Marin, Evaluation of digital health & information technology in primary care, Int. J. Med. Inform. 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, Appl. Clin. Inform. 09 (03) (2018) 604–634, https://doi.org/10.1055/s-0038-1668091.
[68] S.J. Czaja, Current findings and issues in technology and aging, J. Appl. Gerontol. 40 (5) (2021) 463–465, https://doi.org/10.1177/0733464821998579.
[69] T.d.S. Alexandre, L.P. Corona, D.P. Nunes, J.L.F. Santos, Y.A.d.O. Duarte, M. L. Lebrão, Disability in instrumental activities of daily living among older adults: gender differences, Revista de Saúde Pública. 48 (3) (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), J. Gerontol. Series A: Biol. Sci. Med. Sci. 56 (7) (2001) M448–M453, https://doi.org/10.1093/gerona/56.7.M448.