Article

UniMiB SHAR: A Dataset for Human Activity Recognition Using Acceleration Data from Smartphones

Daniela Micucci *, Marco Mobilio and Paolo Napoletano

Department of Informatics, System and Communication, University of Milano Bicocca, Viale Sarca 336, 20126 Milan, Italy

*   Correspondence: daniela.micucci@unimib.it; Tel.: +39-02-6448-7866

Abstract: Smartphones, smartwatches, fitness trackers, and ad-hoc wearable devices are being increasingly used to monitor human activities. Data acquired by the hosted sensors are usually processed by machine-learning-based algorithms to classify human activities. The success of those algorithms mostly depends on the availability of training (labeled) data that, if made publicly available, would allow researchers to make objective comparisons between techniques. Nowadays, publicly available data sets are few, often contain samples from subjects with too similar characteristics, and very often lack of specific information so that is not possible to select subsets of samples according to specific criteria. In this article, we present a new dataset of acceleration samples acquired with an Android smartphone designed for human activity recognition and fall detection. The dataset includes 11,771 samples of both human activities and falls performed by 30 subjects of ages ranging from 18 to 60 years. Samples are divided in 17 fine grained classes grouped in two coarse grained classes: one containing samples of 9 types of activities of daily living (ADL) and the other containing samples of 8 types of falls. The dataset has been stored to include all the information useful to select samples according to different criteria, such as the type of ADL performed, the age, the gender, and so on. Finally, the dataset has been benchmarked with four different classifiers and with two different feature vectors. We evaluated four different classification tasks: fall vs no fall, 9 activities, 8 falls, 17 activities and falls. For each classification task we performed a subject-dependent (5-fold cross validation) and a subject-independent (leave-subject-out) evaluation. The major findings of the evaluation are the following: i) it is more difficult to distinguish between types of falls than types of activities; ii) subject-dependent evaluation outperforms the subject-independent one. The database will allow researchers to work to more robust features and classification schemes that permit to deal with different types of falls and with effect of data personalization.

Keywords: Smartphone accelerometers, Dataset, Activity of Daily Living recognition, Human Activity recognition, Fall detection

---

1. Introduction

Nowadays, many people lead a *sedentary life* due to the facilities that the increasingly pervasive technologies offer. Unfortunately, it is recognized that insufficient physical activity is one of the 10 leading risk factors for global mortality: people with poor physical activity is subjected to a risk of all-cause mortality that is 20% to 30% higher then people performing at least 150 minutes of moderate intensity physical activity per week [1]. Another important global phenomenon actually affecting our society is *population aging*: the decline or even decrease of the natural population growth rates due to a rise in life expectancy [2] and to a long-term downtrend in fertility (expecially in Europe [3]). Falls are a major health risk that impacts the quality of life of elderly people. Indeed, among elderly people, accidental falls occur frequently: the 30% of the over 65 population falls at least once per year; the proportion increases rapidly with age [4]. Moreover, fallers who are not able to get up more likely require hospitalization or, even worse, die [5].

Thus, research on techniques able to recognize activities of daily living (ADLs), also known as human activities (HA), and to detect falls is very active in recent years: the recognition of ADLs may allow to infer the amount of physical activity that a subject perform daily, while a prompt detection of falls may help in reducing the consequence (even fatal) that a fall may cause mostly in elderly people.

ADLs recognition and fall detection techniques usually accomplish their task by analizing samples from sensors, which can be physically deployed in the ambient (ambient sensors, e.g., cameras, vibration sensors, and microphones) or worn by people (wearable sensors, e.g., accelerometers, gyroscopes, and barometers) [6]. To train and evaluate their techniques, researchers usually build their own dataset of samples and rarely make it publicly available [7–9]. This practice makes difficult to compare in an objective way the several newly proposed techniques and implementations due to a lack of a common source of data [9–11]. Only very recently, Janidarmian et al. combined 14 publicly available datasets focusing on acceleration patterns in order to conduct an analysis on feature representations and classification techniques for human activity recognition [12]. Unfortunately, they do not make the resulting dataset available for downloading.

The few publicly available datasets can been primary divided into three main sets: acquired by ambient sensors, acquired by wearable devices, and a combination of the two. Recently, a lot of attention has been paid to wearable sensors because they are less intrusive, work outdoors, and often cheaper than the ambient ones. This is confirmed by the increasing number of techniques that are based on wearable sensors (see for example the survey by Luque et al. related to fall detection techniques relying on data from smartphones [13]).

Wearable sensors are divided in two main groups: ad-hoc wearable devices (e.g., SHIMMER sensor nodes), and smartphones (e.g., Android). For what concerns fall detection, several studies concluded that, in order to be used, fall detection devices must not stigmatize people nor disturb their daily life [14–16]. Unfortunately, devices such as ad-hoc wearable devices and ambient sensors are not well accepted by elderly people because mostly of their intrusiveness. On the contrary, smartphones are good candidate devices for hosting fall detection systems: they are are widespread and daily used by a very large number of person, included elderly people. This, on the one hand, reduces costs, and on the other, eliminates the problem of having to learn how to use a new device. Moreover, studies demonstrated that samples from smartphones sensors (e.g., accelerometer and gyroscope) are accurate enough to be used in clinical domain, such as ADLs recognition [17]. This is also confirmed by the amount of publications that rely on the use of smartphone as acquisition devices for fall detection systems [13,18] and ADLs recognition.

For these reasons we concentrate our attention to smartphones as acquisition devices both for ADL recognition and fall detection. Thus, we searched the publicly available datasets acquired with smartphones in order to identify their strengths and weaknesses so as to outline an effective method for carrying out a new acquisition campaign. We searched the most common repository (IEEE, ACM, Google, and Google Scholar) by using in our query the terms *ADL dataset* and *Fall dataset* in combination with the following words *smartphone*, *acceleration*, *accelerometer*, *inertial*, *IMU*, *sensor*, and *wearable*. We selected the first 100 results for each query. Removing duplicate entries, we obtained less then 200 different references. Then we manually examined the title, the abstract, and the introduction to eliminate references unrelated to ADL recognition and fall detection, and references that were based on ambient sensors such as camera, microphones, or RFID tags. We then read carefully the remaining references and discarded those that do not make publicly available the dataset used in the experimentation. Finally, we added the relevant references that we missed with our searches but were cited in the papers we selected. At the end of the process, we individuated 13 datasets with data from smartphones and 19 with data from wearable ad-hoc sensors. We then included only those datasets

that have been recorded starting from 2012[1] mostly because the oldest dataset including samples from smartphones is dated 2012. This choice makes the datasets homogeneous with respect to the sensors technologies related to acquisition sensors which rapidly evolves year by year.

At the end of the process, we individuated 13 datasets with data from smartphones and 13 with data from wearable ad-hoc sensors. In the following, we will detail some relevant characteristics of the 13 datasets from smartphones since our aim was to build a new dataset containing acceleration patters from smartphone able to complement the existing ones. As it will presented in Section 2, the datasets from ad-hoc wearable devices have be examined with the aim of identifying the most common ADLs and falls.

Table 1 shows the publicly available datasets recorded by means of smartphones and their characteristics. Table 1 also includes the dataset we realized in the last row, in order to ease the comparison.

Table 1. The publicly available datasets containing samples from smartphones sensors

Dataset	Year	ADLs	Falls	Nr. of subjects	Gender Female	Gender Male	Age (years)	Height (cm)	Weight (Kg)
DMPSBFD [19]	2015	yes	yes	5	-	-	-	-	-
Gravity [18]	2016	yes	yes	2	-	-	26 - 32 29 ± 4.2	170 - 185 178 ± 10.6	63 - 80 71.5 ± 12
MobiFall [11]	2014	yes	yes	24	7	17	22 - 47 27 ± 5	160 - 189 175 ± 7	50 - 103 76.4 ± 14.5
MobiAct [20]	2016	yes	yes	57	15	42	20 - 47 25 ± 4	160 - 193 175 ± 4	50 - 120 76.6 ± 14.4
RealWorld (HAR) [21]	2016	yes	no	16	7	8	16 - 62 32 ± 12	163 - 183 173 ± 7	48 - 95 74.1 ± 13.3
Shoaib PA [22]	2013	yes	no	4	0	4	25 - 30 -	-	-
Shoaib SA [23]	2014	yes	no	10	0	10	25 - 30 -	-	-
tFall [24]	2013	yes	yes	10	7	3	20 - 42 31 ± 9	161 - 184 173 ± 1	54 - 98 69.2 ± 13.1
UCI HAR [25]	2012	yes	no	30	-	-	19 - 48 -	-	-
UCI HAPT [26]	2015	yes	no	30	-	-	19 - 48 -	-	-
UCI UIWADS [27]	2013	yes	no	22	-	-	- -	-	-
UMA Fall [28]	2016	yes	yes	17	6	11	14 - 55 27 ± 10	155 - 195 172 ± 9	50 - 93 69.9 ± 12.3
WISDM [29]	2012	yes	no	29	-	-	-	-	-
UniMiB SHAR	2016	yes	yes	30	24	6	18 - 60 27 ± 11	160 - 190 169 ± 7	50 - 82 64.4 ± 9.7

The total number of datasets decreases to 11 because MobiAct and UCI HAPT are updated versions of MobiFall and UCI HAR respectively. Thus, in the following we will refer to 11 datasets overall, discarding MobiFall and UCI HAR.

---

[1] We considered the year in which the dataset has been made available. This year do not necessarily coincide with the year in which the related article has been published.

The 11 datasets have been recorded in the period 2012 to 2016 (column *Year*). Only 5 datasets out of 11 contain both falls (column *Falls*) and ADLs (column *ADLs*).

The average number of subjects for dataset is 18 (column *Nr. of subjects*). The datasets that specify the gender of the subjects (which are MobiAct, RealWorld (HAR), Shoaib PA, Shoaib SA, tFall, and UMA Fall) contain in mean 6 women and 13 men (columns *Gender - Female* and *Gender - Male* respectively).

DMPSBFD, UCI UIWADS, and WISDM do not specify the age of the subjects (column *Age*). In the remaining 8 datasets, subjects are aged between 21 and 43 on average with a standard deviation of 4 and11respectively.

Finally, only Gravity, MobiAct, RealWorld (HAR), tFall, and UMA Fall datasets provide detailed information about the height and the weight of the subjects (columns *Height* and *Weight* respectively).

The detailed information reported in Table 1 have been collected from the web site hosting the dataset, the readme files of each dataset, and the related papers. It is remarkable to notice that in many cases such information get lost in the downloaded dataset. Grey cells in Table 1 indicate that samples are stored so that they can be filtered according to the information contained in the cell. For instance, in all the datasets, with the exception of tFall, it is possible to select subsets of samples according to the specific ADL (column *ADLs*). For example, it is possible to select all the samples that have been labeled *walking*. tFall is an exception because the samples are simply labeled as generic ADL, thus not specifying which specific kind of ADL are.

For what concerns falls (column *Falls*), all the datasets have organized samples maintaining the information related to the specific type of fall they are related to (e.g., forward).

As specified in column *Nr. of subjects*, the samples are linked to the subjects that performed the related activities and, where provided, falls. This means that in all the datasets (with the exception of Shoaib PA) it is possible to select samples related to a specific subject. However, this information is unhelpful if there is no information on the physical characteristics of the subject. Looking at the double column *Gender*, only MobiAct, RealWorld (HAR), Shoaib PA, Shoaib SA, and UMA Fall maintain information related to the gender of the subject. Finally, it is surprising that only Gravity, MobiAct, RealWorld (HAR), and UMA Fall allow to select samples according to age, height, and/or weight of the subjects (columns *Age*, *Height*, and *Weight*).

In view of this analysis, only MobiAct, RealWorld (HAR), and UMA Fall allow to select samples according to several dimensions, such as the age, the sex, the weight of the subjects, or the type of ADL. MobiAct and UMA Fall allow to select samples also according to the type of fall. Unfortunately, the other datasets are not suitable in some experimental evaluations. For example, the evaluation of the effects of personalization in classification techniques [30] taking into account the physical characteristics of the subjects, that is, operating leave-one-subject-out cross-validation [31].

To further contribute to the worldwide collection of accelerometer patterns, in this paper we present a new dataset of smartphone accelerometer samples, named UniMiB SHAR (University of Milano Bicocca Smartphone-based Human Activity Recognition). The dataset was created with the aim of providing the scientific community with a new dataset of acceleration patterns captured by smartphones to be used as a common benchmark for the objective evaluation of both ADLs recognition and fall detection techniques.

The dataset has been designed keeping in mind on one side the limitations of the actual publicly available datasets, and on the other the characteristics of MobiAct, RealWorld (HAR), and UMA Fall, so to create a new dataset that juxtaposes and complements MobiAct, RealWorld (HAR), and UMA Fall with respect to the data that is missing. Thus, such a dataset would have to contain a large number of subjects (more than the 18 in average), with a large number of women (to compensate MobiAct,

RealWorld (HAR), and UMA Fall), with subjects over the age of 55 (to extend the range of UMA Fall[2]), with different physical characteristics (to maintain heterogeneity), performing a wide number of both ADLs and falls (to be suitable in several contexts). Moreover, the dataset would have to contain all the information required to select subjects or ADLs and falls according to different criteria, such as for example, all the female whose height is in the range 160-168 cm, all the men whose weight is in the range 80-110 Kg, all the walking activities of the subjects whose age is in the range 45-60 years.

To fulfil those requirements, we built a dataset including 9 different types of ADLs and 8 different types of falls. The dataset contains a total of 11,771 samples describing both activities of daily living (7,579) and falls (4,192) performed by 30 subjects, mostly females (24), of ages ranging from 18 to 60 years.. Each sample is a vector of 151 accelerometer values for each axis. Each accelerometer entry in the dataset maintains the information about the subject that generated it. Moreover, each accelerometer entry has been labeled by specifying the type of ADL (e.g., *walking*, *sitting*, or *standing*) or the type of fall (e.g., *forward*, *syncope*, or *backward*).

We benchmarked the dataset by performing several experiments. We evaluated four classifiers: k-Nearest Neighbour (k-NN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest (RF). Raw data and magnitudo have been considered as feature vectors. Finally, for each classification we performed a subject-dependent (5-fold cross validation) and a subject-independent (leave-subject-out) evaluation. Results show how much the proposed dataset is challenging with respect to a set of classification tasks.

The article is organized as follows. Section 2 describes the method used to build the datasets. Section 3 presents the dataset evaluation and Section 4 discusses the results of the evaluation. Finally, Section 5 provides final remarks.

2. Dataset Description

This section describes the method used to acquire and pre-process samples in order to produce the UniMiB SHAR dataset.

2.1. Data acquisition

The smartphone used in the experiments was a Samsung Galaxy Nexus I9250 with the Android OS version 5.1.1 and equipped with a Bosh BMA220 acceleration sensor. This sensor is a triaxial *low-g* acceleration sensor. It allows measurements of acceleration in three perpendicular axes, and allows acceleration ranges from $\pm 2g$ to $\pm 16g$ and sampling rates from 1KHz to 32Hz. The Android OS both limits to $\pm 2g$ with a resolution of 0.004g the acceleration range, and takes samples at a maximum frequency of 50Hz. However, the Android OS does not guarantee any consistency between the requested and the effective sampling rate. Indeed, the acquisition rate usually fluctuates during the acquisition. For the experiments presented in this paper, we resampled the signal in order to have a constant sampling rate of 50 Hz, which is commonly used in literature for activity recognition from data acquired through smartphones [23–25]. The accelerometer signal is for each time instant made of a triplet of numbers (x, y, z) that represents the accelerations along each of the 3 Cartesian axes.

We used also the smartphone built-in microphone to record audio signals with a sample frequency of 8,000 Hz, which are used during the data annotation process.

The subjects were asked to place the smartphone in their front trouser pockets: half of the time in the left one and the remaining time in the right one.

Acceleration triples and corresponding audio signals have been recorded using a mobile application specially designed and implemented by the authors, which stores data into two separated files inside the memory of the smartphone.

---

2   We do not consider RealWorld (HAR) since it contains ADLs only. Indeed, it is most difficult to recruit elderly subjects performing falls.

2.2. ADLs and Falls

In order to select both the ADLs and the falls, we analyzed the datasets in Table 1 and the most recent publicly available datasets recorded with wearable ad-hoc devices. As discussed in Section 1, we considered the datasets acquired from 2012 to be compliant with the year of the older smartphone-based dataset. This set includes, sorted by year of creation from the oldest the most recent, the following datasets: DLR v2 [32], Ugulino [33], USC HAD [34], DaLiAc [10], EvAAL [35], MHEALTH [36], UCI ARSA [27], BaSA [37], UR Fall Detection [38], MMsys [9], SisFall [39], UMA Fall³ [18], and REALDISP [40].

For what concerns ADLs, Figure 1 shows the most common ones in the overall 24 datasets we analyzed (11 with samples from smartphones and sketched in Table 1 and 13 with samples from wearable ad-hoc devices listed above). The y axis represents the number of datasets that include the specified ADL. ADLs are grouped by category. The following categories have been identified by analizyng the datasets: *Context-related*, which includes activities that someway deal with the context (e.g., *Stepping in a car*), *Motion-related*, which includes activities that imply some kind of physical movement (e.g., *Walking*), *Posture-related*, which includes activities in which the person maintains the position for a certain amount of time (e.g., *Standing*), *Sport-related*, which includes any kind of activity that requires a physical effort (e.g., *Jumping*), and *Others*, which includes activities that are presented in one dataset only (e.g., *Vacuuming* in category *Housekeeping-related*). The *Jogging* and *Running* activities deserve a clarification. In all the datasets we analyzed, they are mutually exclusive, that is, datasets that contain *Running*, do not contain *Jogging* and vice versa. The datasets REALDISP and MHEALT are an exception because they include both the activities. These datasets, besides being realized by the same institution, are primarily oriented towards the recognition of physical activities (warm up, cool down and fitness exercises). Moreover, none of the datasets analyzed exactly specify what the *Jogging* and *Running* activities are related to. Thus, even though they may be considered very similar activities, we have decided to keep them separated in oder to do not loose their specificity. We classify *Jogging* as a *sport-related* activity (in the sense, for instance, of jogging in the park), and *Running* as a *motion-related* activity (in the sense, for instance, of running for the bus). For each category, the x axis shows all the ADLs we found and that are present in at least 2 datasets. Under the label *Others* fall all the ADLs for the corresponding category that have been included in one dataset only (e.g., *Walking left-circle* in category *Motion-related*).

Tables 2 shows the 9 ADLs we have selected among the most popular included in the analyzed publicly available datasets. UniMiB SHAR includes the top 5 most popular *Motion-related* activities (i.e., *Walking*, *Going upstairs*, *Going downstairs*, *Sitting down*, and *Running*). Moreover, we detailed the generic *Standing up*, by including the *Standing up from sitting* and *Standing up from laying* activities. Finally, we included also the *Lying down from standing*.

In the *Sport-related* category, we did not included *Jogging* even if it is the most popular activity in its category because we included the *Running* activity in the *Motion-related* category. In *Sport-related* category, we chose the *Jumping* activity being the second most popular one.

Our dataset does not include *Postural-related* activities. Indeed, we were interested in acquiring acceleration data from activities related to movements both because from them it is possible to estimate the overall physical activity performed by a person, and because people are more likely to fall during movements [41].

We do not include ADLs belonging to categories such as *Housekeeping-*, *Cooking-*, or *Personal care-related* (those fall in the *Others* category in Table 1, because we are interested in **low order** activities of daily living, which include simple activities such as, *Standing*, *Sitting down*, *Walking*, rather than **high order** activities of daily living, which include complex activities such as, *Washing dishes*, *Combing hair*, *Preparing a sandwich*. The same holds for *contex-realted* activities that are intended as **high order**

---

³   UMA Fall contains samples from both smartphones and ad-hoc wearable devices.

.
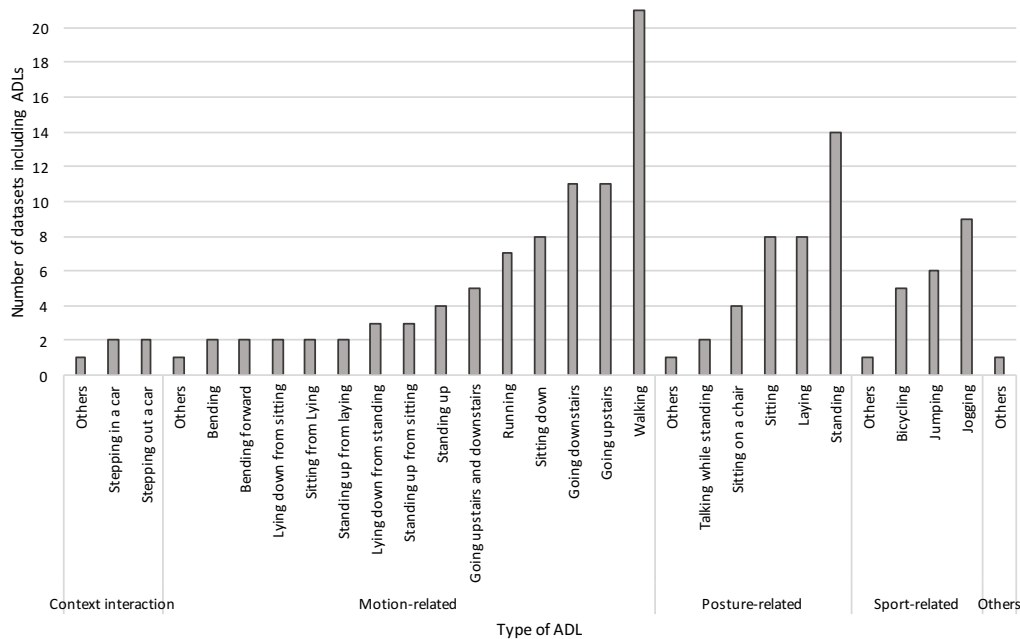
Figure 1. ADLs and theirs occurrence in the publicly available datasets analysed grouped by category

activities. This choice was also motivated by the fact that these activities are scarcely present in the analyzed datasets (in particular, each activity belonging in the above mentioned categories is present in only one of the 24 analyzed datasets).

Finally, among the ADLs related to movements, we selected the ADLs most used in literature as demonstrated by the analysis we performed, which is also confirmed by Pannurat et al. in [42].
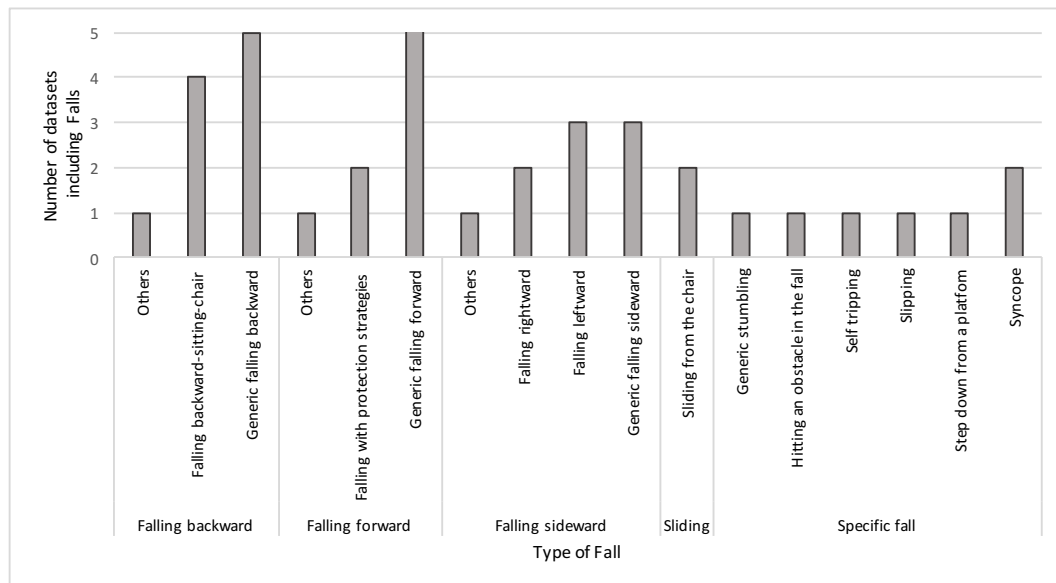
Table 2. ADLs performed by the subjects In the UniMiB SHAR dataset

Category	Name	Description	Label
Motion-related	Standing up from laying	From laying on the bed to standing	StandingUpFL
	Lying down from standing	From standing to lying on a bed	LyingDownFS
	Standing up from sitting	From standing to sitting on a chair	StandingUpFS
	Running	Moderate running	Running
	Sitting down	From standing to sitting on a chair	SittingDown
	Going downstairs	Climb the stairs moderately	GoingDownS
	Going upstairs	Down the stairs moderately	GoingUpS
	Walking	Normal walking	Walking
Sport-related	Jumping	Continuos jumping	Jumping

For what concerns falls, we analized DMPSBFD, Gravity, MobiAct, tFall, and UMAFall datasets from smartphones (see Table 1), and DLR v2, EvAAL, MMsys, SISFall, UMA Fall, and UR Fall Detection datasets from wearable ad-hoc devices, since they are the only datasets that contain falls. From this set, we excluded DLR v2, EvAAL because they do not specify the type of fall.

Figure 2 shows the most common falls in the resulting 9 datasets we analyzed. The y axis represents the number of datasets that include the specified fall. Likewise ADLs, falls are grouped by category. *Falling backward*, *Falling forward*, and *Falling sideward* include back-, front-, and side-ward falls respectively. *Sliding* category can be further specialized so that to include *Sliding from a chair*, *Sliding form a bed*, and *Generic sliding* that not specifies details about the type of sliding. Finally, the category *Specific fall* includes different type of falls that have not been further specialized.

For each category, the x axis shows all the types of falls we found. Under the label *Others* fall all the falls for the corresponding category that have been included in one dataset only. The *Specific fall* category is an exception since it includes falls types not particularly present in the analyzed datasets.

.

Figure 2. Falls and theirs occurrence in the publicly available datasets analysed grouped by category

Table 3. Falls performed by the subjects in teh UniMiB SHAR dataset

Category	Name	Description	Label
Falling backward	Falling backward-sitting-chair *Generic* falling backward	Fall backward while trying to sit on a chair Generic fall backward from standing	FallingBackSC FallingBack
Falling forward	Falling with protection strategies *Generic* falling forward	Falls using compensation strategies to prevent the impact Fall forward from standing, use of hands to dampen fall	FallingWithPS FallingForw
Falling sideward	Falling rightward Falling leftward	Fall right from standing Fall right from standing	FallingLeft FallingRight
Specific fall	Hitting an obstacle in the fall Syncope	Falls with contact to an obstacle before hitting the ground Getting unconscious	HittingObstacle Syncope

Choosing which falls to include in the dataset was driven by the following considerations: the number of falls should have been comparable to that of the other datasets, and the dataset should have included a set of representative types of falls. Thus, having four categories (not considering *Sliding*, which includes only one type of fall that has been considered by two datasets only), we selected two falls from each of them. In each category, we selected the first two most popular falls. The category *Falling sideward* is an exception since we preferred to choose the two most specific falls instead of including the too generic *Generic falling sideward*. Table 3 shows the 8 falls that we selected according the adopted criterion.

Finally, studies on this topic confirm the falls we selected are common in real-life [24,43–45].

2.3. Subjects

30 healthy subjects have been involved in the experiments: 24 were women and 6 men. The subjects, whose data are shown in Table 4, are aged between 18 and 60 years (27 $\pm$12 years), have a body mass between 50 and 82 kg (64.4 $\pm$ 9.7 kg), and a height between 160 and 190 cm (169 $\pm$ 7 cm). Note that we included more women and older ages to compensate for the lacks of MobiAct.

Table 4. The characteristics of the subjects

	Total	Female	Male	
subjects	30	24	6	
age	18 - 60	18 - 55	20 - 60	min - max
	27 ± 11	24 ± 9	36 ± 15	mean $\pm$ std
height	160 -190	160 - 172	170 - 190	min - max
	169 ± 7	166 ± 4	179 ± 6	mean $\pm$ std
weight	50 - 82	50 - 78	55 - 82	min - max
	64.4 ± 9.7	61.9 ± 7.8	74.7 ± 9.7	mean $\pm$ std

All the subjects performed both ADLs and Falls. The subjects gave written informed consent and the study was conducted in accordance with the WMA Declaration of Helsinki [46].

2.4. Protocols

To simplify the data annotation process, we asked each subject to clap her hands early before and after she performed the activity/fall to be recorded. Moreover, to reduce background noise, we asked each subject to wear gym trousers with front pockets.

Concerning ADLs, in order to avoid mistakes by the subjects due to too long sequences of activities, registrations have been subdivided in the three protocols showed in Table 5. Each protocol has been performed by each subject twice, the first one with the smartphone in the right pocket and the second in the left. Those smartphone positions were chosen because both they are the most natural ones and they are exactly the positions used in the analyzed references dealing with smartphones.

Protocol 1 includes *Walking* and *Running* activities. We opted for moderate walking and running so as to include even older people. *Protocol 2* includes activities related to both climbing and descending stairs, and jumps. In our registration, we selected straight stairs ramps, and asked each volunteer to perform jumps with a moderate elevation, with little effort, and spaced each other about 2 seconds. *Protocol 3* includes ascending and descending activities. The *Sitting down* and *Standing up from sitting* activities have been performed with a chair without armrests; the *Lying down from standing* and *Standing up from laying* have been performed on a sofa. The duration of the actives are in average with those reported in [47] that reviews a set of public datasets for wearable fall detection systems.

Falls have been recorded individually, always following the pattern of making a start and end clap (see Table 6). In cases where the volunteer ended in a prone position, the clap has been performed by an external subject to avoid as far as possible any movements that might lead to recording events outside the study. To carry out the simulation safely, a mattress of about 15 centimeters in height was used. Each fall was repeated six times, the first three with the smartphone in the right pocket, the others in the left. Finally, falls have been simulated, started from a standing straight up position, and self-started.

2.5. Segmentation and preprocessing

The audio files helped in the identification of the start and stop time instants for each recorded activity. From the labelled recorded accelerometer data, we extracted a signal window of 3 *sec* each time a peak was found, that is, when the following conditions were verified:

1. the magnitude of the signal m_t at time t was higher than $1.5g$, with g being the gravitational acceleration;
2. the magnitude m_{t-1} at the previous time instant $t - 1$ was lower than 0.

Each signal window of 3 *sec* was centered around each peak and it is likely that several overlap between subsequent windows may happen. We adopted this segmentation technique instead of

Table 5. The protocols for ADLs aquisition

Protocol	Action	Iteration
Protocol 1	Start the registration Put the smartphone in the pocket *clap* Walking for 30 seconds *clap* Running for 30 seconds *clap* Pull the smartphone from the pocket Stop the registration	1 time
Protocol 2	Start the registration Put the smartphone in the pocket *clap* Climb 15 steps *clap* Go down 15 steps *clap* Wait 2 seconds *clap* Jump 5 times *clap* Pull the smartphone from the pocket Stop the registration	1 time
Protocol 3	Start the registration Put the smartphone in the pocket	1 time
	clap Sitting down *clap* Standing up from sitting *clap* *Wait 2 seconds*	5 times
	clap Lying down from standing *clap* Standing up from laying *clap* *Wait 2 seconds*	5 times
	Pull the smartphone from the pocket Stop the registration	1 time

Table 6. Protocol for each fall

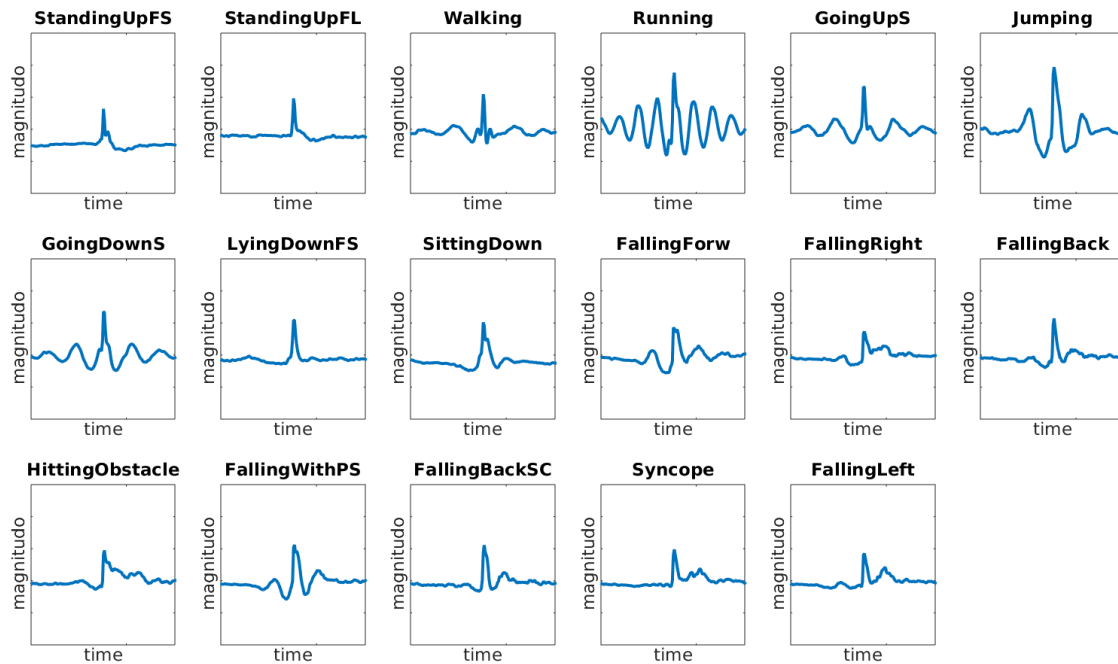| Action	Iteration
Start the registration Put the smartphone in the pocket *clap* fall *clap* Pull the smartphone from the pocket Stop the registration	6 times

Figure 3. Samples of acceleration shapes

selecting overlapped sliding windows because our dataset is mostly focused on motion-related recognition of ADLs and falls. The choice of taking 3 *sec* window has been motivated by: i) the cadence of an average person walking is within [90, 130] steps/min [48,49]; ii) at least a full walking cycle (two steps) is preferred on each window sample.

Figure 3 shows samples of acceleration shapes. For each activity, we displayed the average magnitude shape obtained by averaging all the subjects' shapes.

Since the device used for data acquisition records accelerometer data with a sample frequency of 50 Hz, for each activity, the accelerometer data vector is made of 3 vectors of 151 values (a vector of size 1x453), one for each acceleration direction. The dataset is thus composed of 11,771 samples describing both ADLs (7,759) and falls (4,192) not equally distributed across activity types. This is because the activity of running and walking were performed by subjects for a time longer than the time spent for other activities. Originally, 6,000 time windows of the running activity were found. In order to make the final dataset as much as balanced, we have deleted about 4,000 samples related to running activities. The resulting samples distribution is plotted in Figure 4, where the samples related to running activities are about 2,000. On our web site we release both datasets, the one balanced and the original one.

We preprocessed the acceleration signal $s(t)$ in order to remove the gravitational component $g(t)$. Since the gravitational force is assumed to have only low frequency components, we applied a Butterworth (BW) low-pass filter with a cut off frequency of 0.3 Hz [48]: $g(t) = BW(s(t), 0.3)$. The accelerometer data without gravitational component is then obtained as: $s(t) = s(t) - g(t)$.

3. Dataset Evaluation

We organized the accelerometer samples in order to evaluate four classification tasks:

1. *AF-17* contains 17 classes obtained by grouping all the 9 classes of ADLs and 8 classes of FALLs. This subset permits to evaluate the capability of the classifier to distinguish among different types of ADLs and FALLs;
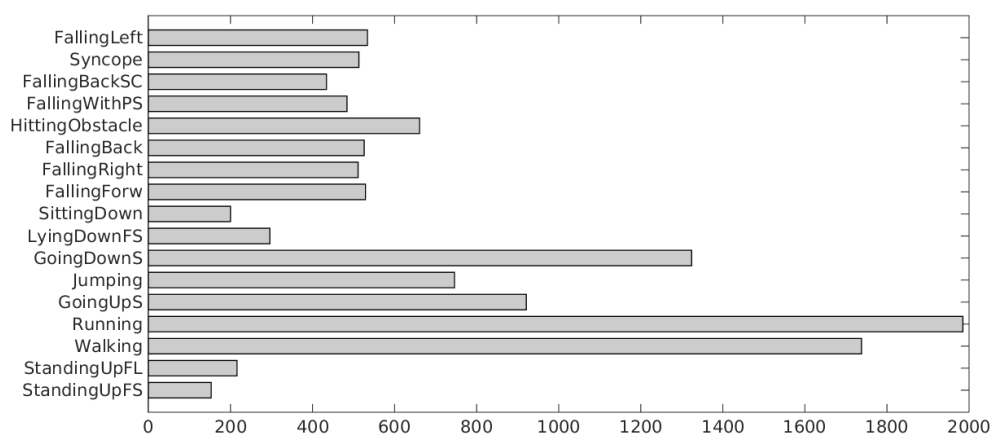
Figure 4. Activity samples distribution

2. *AF-2* contains 2 classes obtained by considering all the ADLs as one class and all the FALLs as one class. This subset permits to evaluate, whatever is the type of ADL or FALL, the classifier robustness in distinguishing between ADLs and FALLs;
3. *A-9* contains 9 classes obtained by considering all the 9 classes of ADLs. This subset permits to evaluate how much the classifier is capable to distinguish among different types of ADLs;
4. *F-8* contains 8 classes obtained by considering all the 8 classes of FALLs. This subset permits to evaluate how much the classifier is capable to distinguish among different types of FALLs.

We initially evaluated the classifiers by performing a traditional *5-fold* cross-validation. It means that all the data have been randomly split in 5 folds. Each fold has been considered as test data and the remaining ones as training data. Results are computed by averaging the result obtained on each test fold. The folds have been obtained by applying the stratified random sampling that ensures samples of the same subject in both the test and the training folds.

To make the dataset evaluation independent from the effect of personalization, we conducted another evaluation by performing a *leave-subject-out* cross-validation. Each test fold is made of accelerometer samples of one user only, namely the *test user*, while the training folds contain accelerometer samples of all the other users except the samples of the *test user*.

Previous studies demonstrated that classifiers trained on raw data perform better with respect to classifiers trained on other types of feature vector representations, such as magnitude of the signal, frequency, or energy [43,50]. However, in order to make the experiments comparable with others experiments presented in the state of the art, we considered two feature vectors:

1. *raw data*: the 453-dimensional patterns obtained by concatenating the 151 acceleration values recorded along each Cartesian direction;
2. *magnitude* of the accelerometer signal, that is a feature vector of 151 values.

We experimented four different classifiers:

1. k-Nearest Neighbour (k-NN) with $k = 1$;
2. Support Vector Machines (SVM) with a radial basis kernel;
3. Artificial Neural Networks (ANN). We set up a three-layers feed forward network with back propagation. The network architecture includes an input layer, a layer of hidden neurons and an output layer that includes a softmax function for class prediction. The number of hidden neurons n has been set in way that $n = \sqrt{m \times k}$, where m is the number of neurons in the input layer and k is the number of neurons in the output layer, namely the number of classes [51].

4. Random Forest (RF): bootstrap-aggregated decision trees with 300 bagged classification trees.

All the classifiers have been implemented exploiting the MATLAB Statistics and Machine Learning Toolbox and the Neural Network Toolbox.

3.1. Evaluation metrics

As shown in Figure 4, each of the 17 sets containing samples related to a specific activity is different in size. To cope with the class imbalance problem of the dataset we used as metric the *macro average accuracy* [52].

Given E the set of all the activities types, $a \in E$, NP_a the number of times a occurs in the dataset, and TP_a the number of times the activity a is recognized, MAA (*Macro Average Accuracy*) is defined by Equation 1.

$$MAA = \frac{1}{|E|} \sum_{a=1}^{|E|} Acc_a = \frac{1}{|E|} \sum_{a=1}^{|E|} \frac{TP_a}{NP_a}. \tag{1}$$

MAA is the arithmetic average of the accuracy Acc_a of each activity. It allows each partial accuracy to contribute equally to the evaluation.

4. Results and Discussion

In the following, we discuss separately the results achieved with the traditional 5-fold cross-validation and the leave-subject-out cross-validation.

4.1. Subject-dependent evaluation (5-fold evaluation)

The k-fold evaluation is the most employed evaluation scheme in literature [53]. This evaluation considers a training set and a test set made of activity samples performed by all the human subjects. The resulting classifier is subject-dependent and usually exhibits a very high performance. Results of the k-fold evaluation (here we used k=5) scheme are showed in Table 7 for raw data and magnitude. Overall, the performances achieved using raw data are better than the ones obtained using magnitude as feature vector. This confirm a result already achieved in previous works [43,50].

The AF-17 recognition task is quite challenging with a MAA of about 83% in the case of raw data with KNN, and a MAA of about 66% in the case of magnitude with RF. This means that is quite difficult to distinguish among types of activities especially in the case when magnitude is adopted as feature vector. Figure 5 shows the confusion matrix of the k-NN experiment in the case of raw data.

The A-9 classification task is quite easy, the MAA obtained by raw data with RF is about 88% while the MAA obtained by magnitude with SVM is about 79%. This means that is quite easy to distinguish between types of activities. Looking at the confusion matrix in Figure 5, the most misclassified pairs of activities are *Standing up from laying* and *Standing up from sitting*, *Lying down from standing* and *Sitting down*, *Going upstairs* and *Walking*, *Going downstairs* and *Walking*, *Jumping* and *Going downstairs*.

The F-8 recognition task is quite challenging: the MAA is about 78% and 57% in the case of raw data with KNN and magnitude with RF respectively. This result suggests that distinguish among falls is very complicated The most misclassified pairs of falls are *Falling with protection strategies* and *Generic falling forward*, *Syncope* and *Falling leftward*, *Generic falling backward* and *Falling backward-sitting-chair*, *Falling rightward* and *Falling with protection strategies*, *Falling rightward* and *Syncope*.

In contrast, the AF-2 recognition task is very easy for all the classifiers and for both raw data and magnitude with a MAA of about 99% achieved with raw data and SVM. These results are similar to those obtained by previous researchers on a similar classification task performed on different datasets [24,50]. This means that it is very easy to distinguish between falls and no falls.

To summarize, the F-8 and AF-17 are quite challenging classification tasks. The use of this dataset for those tasks will permit researchers:

Table 7. 5-fold evaluation. Mean Average Accuracy for each classification task using raw data and magnitude as feature vectors. In bold the best results for each classification task and feature vector employed.

	5-fold							
	raw data				magnitude			
data	KNN	SVM	ANN	RF	KNN	SVM	ANN	RF
AF-17	**82.86**	78.75	56.06	81.48	65.30	65.71	41.95	**65.96**
AF-2	97.78	**98.71**	98.57	98.09	95.56	**97.42**	96.71	95.74
A-9	87.77	81.62	72.13	**88.41**	77.37	**78.94**	62.81	75.14
F-8	**78.55**	75.63	55.07	78.27	53.31	56.34	37.66	**57.26**

Table 8. Leave-subject-out. Mean Average Accuracy for each classification task using raw data and magnitude of the signal as feature vectors. In bold the best results for each classification task and feature vector employed.

	leave-subject-out							
	raw data				magnitude			
data	KNN	SVM	ANN	RF	KNN	SVM	ANN	RF
AF-17	52.14	55.15	48.00	**56.53**	52.14	55.09	48.00	**56.58**
AF-2	92.90	**97.57**	95.41	97.02	92.90	**97.57**	96.07	97.05
A-9	63.79	63.32	63.63	**73.17**	63.79	63.36	63.63	72.67
F-8	43.66	**48.84**	38.50	45.88	43.66	**49.35**	38.50	45.26

- to design and evaluate more robust feature representations as well as more robust classification schemes for human activity recognition.
- to study more robust features to deal with accelerometer samples of different types of falls.

4.2. Subject-independent evaluation (leave-subject-out evaluation)

Table 8 shows the results obtained by performing the leave-subject-out evaluation. In this case the training set is made of activity samples of subjects not included in the test set. This evaluation is also known as subject independent evaluation and shows the feasibility of a real smartphone application for human activity recognition [54–56] where data of a given subject are usually not included in the training set of the classifier.

From the results it is almost evident the drop of performances with respect to the case of 5-fold evaluation. Human subject performs activities in a different way and this influences the recognition accuracy especially when it is necessary to distinguish between fine grained types of activities, that is in the case of AF-17, A-9 and F8 recognition tasks. In particular, in the case of AF-17 the best MAA is 56.58% using RF and magnitude. In the case of A-9 the best MAA is 73.17% using RF and raw data. In the case of F-8 the best MAA is 49.35% using SVM and magnitude. In contrast, distinguishing between coarse grained activities, such as falls vs no falls, is quite easy with a MAA of 97.57% with SVM for both raw data and magnitude. Overall the magnitude feature vector performs slightly better than the case of raw data. This suggests that using the magnitude as feature vector in the case of subject-independent evaluations could be more reliable than raw data.

The low performance achieved in the case of subject-independent evaluation permits researcher to investigate the following issues:

- the study of a more robust feature vector that is able to reduce as much as possible the performance gap between the subject-dependent and subject-independent evaluation;
- the study of on-line learning classification schemes that permit, with the use of a few subject-dependent data, to improve as much as possible the performance.

	StandingUpFS	StandingUpFL	Walking	Running	GoingUpS	Jumping	GoingDownS	LyingDownFS	SittingDown	FallingForw	FallingRight	FallingBack	HittingObstacle	FallingWithPS	FallingBackSC	Syncope	FallingLeft
StandingUpFS	93.3	6.7															
StandingUpFL	14.0	83.7						2.3									
Walking			99.1		0.3		0.6										
Running			3.5	96.0			0.5										
GoingUpS			15.2		83.2		1.6										
Jumping			3.4		0.7	87.9	8.1										
GoingDownS			22.0		2.7	0.8	74.6										
LyingDownFS	1.7	3.4						84.7	10.2								
SittingDown									100.0								
FallingForw	1.0								1.0	81.0	1.9	1.0	12.4			1.0	1.0
FallingRight	1.0							2.0	2.9	5.9	72.5	2.9		2.0	1.0	9.8	
FallingBack		1.0			1.0			1.9		5.7	1.0	71.4		1.9	11.4	1.9	2.9
HittingObstacle										1.5		1.5	87.1	1.5	0.8	4.5	3.0
FallingWithPS	1.0	1.0					1.0		1.0	4.2	14.6	1.0	1.0	68.8		3.1	3.1
FallingBackSC					1.2			3.5	2.3		4.7	7.0			77.9	1.2	2.3
Syncope		1.0							4.9	7.8	7.8	1.0		3.9	2.9	60.8	9.8
FallingLeft	0.9	2.8							3.8	3.8	2.8	0.9	3.8	0.9	5.7	8.5	66.0

Figure 5. Confusion matrix of the AF-17 classification achieved with k-NN

5. Conclusion

Almost all publicly available datasets from smartphones do not allow the selection of samples based on specific criteria related to the physical characteristics of subjects and the activities (and/or falls) they performed. Of the 11 datasets containing smartphone measurements, only MobiAct, RealWorld (HAR), and UMA Fall are the exception. These three datasets include more men than women. Considering only datasets that include falls (MobiAct and UMA Fall), the maximum age of the subjects is 47 years. Our goal was therefore to create a new dastaset that would be complementary to the more complete ones and that also include falls. The result is UniMiB SHAR dataset that includes 9 ADLs and 8 falls performed by 30 humans, mostly female, with a huge range of ages, from 18 to 60 years.

The classification results obtained on the proposed dataset showed that the raw data performs quite better than magnitude as feature vector in the case of subject-dependent evaluation, and, on the opposite, the magnitude performs quite better than raw data in the case of subject-independent evaluation. The classification of different types of activities is simpler than the classification of different types of falls. It is very easy to distinguish between falls and no falls for both raw data and magnitude. The subject-independent evaluation showed that recognition performance strongly depends on the subject data.

UniMiB SHAR dataset will permit researchers to study several issues, such as: i) robust features to deal with falls; ii) robust features and classification schemes to deal with personalization issues.

We are planning to carry out an evaluation of the state-of-the-art techniques for ADLs recognition on both UniMiB SHAR and all the publicly available datasets of accelerometer data from smartphone to have and objective comparison. Moreover, we have planned to make experimentation on personalization by using those datasets that include information about the characteristics of the subjects. We want to investigate whether the training set containing samples acquired by subjects with similar characteristics to the testing subject may result in a more effective classifier. Finally, we are planning to check if and how data from smartwatches and smartphones can jointly improve the

performances of the classifiers. To this end, we are improving the data acquisition application used for UniMiB SHAR.

Supplementary Materials: The dataset, the Matlab scripts to repeat the experiments, the app used to acquire samples, and additional materials (e.g., images with samples of acceleration shapes) are available at the following address: http://www.sal.disco.unimib.it/technologies/unimib-shar/.

Author Contributions: All authors contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   World Health Organization. Global recommendations on physical activity for health, 2010.
2.   United Nations. World Populatin Ageing Report, 2013.
3.   Grant, J.; Hoorens, S.; Sivadasan, S.; van het Loo, M.; DaVanzo, J.; Hale, L.; Gibson, S.; Butz, W. Low Fertility and Population Ageing. http://www.rand.org/pubs/monographs/MG206.html, 2004.
4.   Tromp, A.M.; Pluijm, S.M.; Smit, J.H.; Deeg, D.J.; Bouter, L.M.; Lips, P. Fall-risk screening test: a prospective study on predictors for falls in community-dwelling elderly. *Journal of Clinical Epidemiology* **2001**, *54*, 837–844.
5.   Tinetti, M.E.; Liu, W.L.; Claus, E.B. Predictors and prognosis of inability to get up after falls among elderly persons. *JAMA* **1993**, *269*, 65–70.
6.   Mubashir, M.; Shao, L.; Seed, L. A Survey on Fall Detection: Principles and Approaches. *Neurocomput.* **2013**, *100*, 144–152.
7.   Chen, J.; Kwong, K.; Chang, D.; Luk, J.; Bajcsy, R. Wearable sensors for reliable fall detection. Proceedings of the Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS), 2006.
8.   Li, Q.; Stankovic, J.A.; Hanson, M.A.; Barth, A.T.; Lach, J.; Zhou, G. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN), 2009.
9.   Ojetola, O.; Gaura, E.; Brusey, J. Data Set for Fall Events and Daily Activities from Inertial Sensors. Proceedings of the ACM Multimedia Systems Conference (MMSys), 2015.
10.   Leutheuser, H.; Schuldhaus, D.; Eskofier, B.M. Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset. *PLoS ONE* **2013**, *8*, 1–11.
11.   Vavoulas, G.; Pediaditis, M.; Chatzaki, C.; Spanakis, E.G.; Manolis, T. The MobiFall Dataset: Fall Detection and Classification with a Smartphone. *International Journal of Monitoring and Surveillance Technologies Research* **2014**, pp. 44–56.
12.   Janidarmian, M.; Fekr, A.R.; Radecka, K.; Zilic, Z. Acceleration Sensors in Human Activity Recognition. *Sensors* **2017**, *17*, 1–26.
13.   Luque, R.; Casilari, E.; Morón, M.J.; Redondo, G. Comparison and Characterization of Android-Based Fall Detection Systems. *Sensors* **2014**, *14*, 18543–18574.
14.   Steele, R.; Lo, A.; Secombe, C.; Wong, Y.K. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. *International Journal of Medical Informatics* **2009**, *78*, 788 – 801.
15.   Feldwieser, F.; Marchollek, M.; Meis, M.; Gietzelt, M.; Steinhagen-Thiessen, E. Acceptance of seniors towards automatic in home fall detection devices. *Journal of Assistive Technologies* **2016**, *10*, 178 – 186.
16.   Jeffs, E.; Vollam, S.; Young, J.D.; Horsington, L.; Lynch, B.; Watkinson, P.J. Wearable monitors for patients following discharge from an intensive care unit: practical lessons learnt from an observational study. *Journal of Advanced Nursing* **2016**, *72*, 1851 – 1862.
17.   Mourcou, Q.; Fleury, A.; Franco, C.; Klopcic, F.; Vuillerme, N. Performance Evaluation of Smartphone Inertial Sensors Measurement for Range of Motion. *Sensors* **2015**, *15*, 23168–23187.
18.   Vilarinho, T.; Farshchian, B.; Bajer, D.G.; Dahl, O.H.; Egge, I.; Hegdal, S.S.; Lones, A.; Slettevold, J.N.; Weggersen, S.M. A Combined Smartphone and Smartwatch Fall Detection System. Proceedings of the IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and

Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015.

19.   Dataset for Mobile Phone Sensing Based Fall Detection (DMPSBFD). https://figshare.com/articles/Dataset_for_Mobile_Phone_Sensing_Based_Fall_Detection/1444405, 2015. [Online; accessed 1-Oct-2016].

20.   Vavoulas, G.; Chatzaki, C.; Malliotakis, T.; Pediaditis, M.; Tsiknakis, M. The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones. Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE), 2016.

21.   Sztyler, T.; Stuckenschmidt, H. On-body Localization of Wearable Devices: An Investigation of Position-Aware Activity Recognition. Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom), 2016.

22.   Shoaib, M.; Scholten, H.; Havinga, P.J.M. Towards Physical Activity Recognition Using Smartphone Sensors. Proceedings of the IEEE International Conference on Ubiquitous Intelligence and Computing and International Conference on Autonomic and Trusted Computing (UIC/ATC), 2013.

23.   Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. Fusion of Smartphone Motion Sensors for Physical Activity Recognition. *Sensors* **2014**, *14*, 10146–10176.

24.   Medrano, C.; Igual, R.; Plaza, I.; Castro, M. Detecting falls as novelties in acceleration patterns acquired with smartphones. *PLoS ONE* **2014**, *9*.

25.   Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition Using Smartphones. Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2013.

26.   Reyes-Ortiz, J.L.; Oneto, L.; Sama, A.; Parra, X.; Anguita, D. Transition-Aware Human Activity Recognition Using Smartphones. *Neurocomputing* **2016**, *171*, 754 – 767.

27.   Casale, P.; Pujol, O.; Radeva, P. Personalization and User Verification in Wearable Systems Using Biometric Walking Patterns. *Personal Ubiquitous Comput.* **2012**, *16*, 563–580.

28.   Casilari, E.; Santoyo-Ramón, J.A.; Cano-García, J.M. Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection. *PLOS ONE* **2016**, *11*, 1–17.

29.   Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explor. Newsl.* **2011**, *12*, 74–82.

30.   Medrano, C.; Plaza, I.; Igual, R.; Sánchez, A.; Castro, M. The Effect of Personalization on Smartphone-Based Fall Detectors. *Sensors* **2016**, *16*.

31.   Khan, S.S.; Karg, M.E.; Kulia, D.; Hoey, J. Detecting falls with X-Factor Hidden Markov Models. *Applied Soft Computing* **2017**, *55*, 168 – 177.

32.   Frank, K.; Vera Nadales, M.J.; Robertson, P.; Pfeifer, T. Bayesian Recognition of Motion Related Activities with Inertial Sensors. Proceedings of the ACM International Conference Adjunct Papers on Ubiquitous Computing - Adjunct (Ubicomp), 2010.

33.   Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidiú, R.; Fuks, H., Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. In *Advances in Artificial Intelligence - SBIA 2012: 21th Brazilian Symposium on Artificial Intelligence, Curitiba, Brazil, October 20-25, 2012. Proceedings*; Barros, L.N.; Finger, M.; Pozo, A.T.; Gimenénez-Lugo, G.A.; Castilho, M., Eds.; Springer Berlin Heidelberg, 2012; pp. 52–61.

34.   Zhang, M.; Sawchuk, A.A. USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors. Proceedings of the ACM Workshop on Situation, Activity and Goal Awareness (SAGAware), colocated with the International Conference on Ubiquitous Computing (Ubicomp), 2012.

35.   Kozina, S.; Gjoreski, H.; Gams, M.; Luštrek, M. Three-layer Activity Recognition Combining Domain Knowledge and Meta- classification Author list. *Journal of Medical and Biological Engineering* **2013**, *33*, 406–414.

36.   Banos, O.; Villalonga, C.; Garcia, R.; Saez, A.; Damas, M.; Holgado-Terriza, J.A.; Lee, S.; Pomares, H.; Rojas, I. Design, implementation and validation of a novel open framework for agile development of mobile health applications. *BioMedical Engineering OnLine* **2015**, *14*.

37.   Leutheuser, H.; Doelfel, S.; Schuldhaus, D.; Reinfelder, S.; Eskofier, B.M. Performance Comparison of Two Step Segmentation Algorithms Using Different Step Activities. International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2014.

38. Kwolek, B.; Kepski, M. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer Methods and Programs in Biomedicine* **2014**, *117*, 489 – 501.

39. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A Fall and Movement Dataset. *Sensors* **2017**, *17*.

40. Banos, O.; Damas, M.; Pomares, H.; Rojas, I.; Tóth, M.A.; Amft, O. A Benchmark Dataset to Evaluate Sensor Displacement in Activity Recognition. Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp), 2012.

41. Robinovitch, S.N.; Feldman, F.; Yang, Y.; Schonnop, R.; Lueng, P.M.; Sarraf, T.; Sims-Gould, J.; Loughin, M. Video Capture of the Circumstances of Falls in Elderly People Residing in Long-Term Care: An Observational Study. *The Lancet* **2013**, *381*, 47–54.

42. Natthapon, P.; Thiemjarus, S.; Nantajeewarawat, E. Automatic Fall Monitoring: A Review. *Sensors* **2014**, *14*, 12900–12936.

43. Igual, R.; Medrano, C.; Plaza, I. A comparison of public datasets for acceleration-based fall detection. *Medical Engineering & Physics* **2015**, *37*, 870 – 878.

44. Kangas, M.; Vikman, I.; Nyberg, L.; Korpelainen, R.; Lindblom, J.; Jämsä, T. Comparison of real-life accidental falls in older people with experimental falls in middle-aged test subjects. *Gait & Posture* **2012**, *35*, 500 – 505.

45. Klenk, J.; Becker, C.; Lieken, F.; Nicolai, S.; Maetzler, W.; Alt, W.; Zijlstra, W.; Hausdorff, J.; van Lummel, R.; Chiari, L.; Lindemann, U. Comparison of acceleration signals of simulated and real-world backward falls. *Medical Engineering & Physics* **2011**, *33*, 368 – 373.

46. World Medical Association. WMA Declaration of Helsinki - Ethical principles for medical research involving human subjects, 2013.

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

48. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition using Smartphones. Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2013.

49. BenAbdelkader, C.; Cutler, R.; Davis, L. Stride and cadence as a biometric in automatic person identification and verification. Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition (FG), 2002.

50. Micucci, D.; Mobilio, M.; Napoletano, P.; Tisato, F. Falls as anomalies? An experimental evaluation using smartphone accelerometer data. *Journal of Ambient Intelligence and Humanized Computing* **2017**, *8*, 87–99.

51. Xu, H.; Liu, J.; Hu, H.; Zhang, Y. Wearable Sensor-Based Human Activity Recognition Method with Multi-Features Extracted from Hilbert-Huang Transform. *Sensors* **2016**, *16*, 2048.

52. He, H.; Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*; John Wiley & Sons, 2013.

53. Kittler, J.; others. Pattern recognition. A statistical approach **1982**.

54. Medrano, C.; Plaza, I.; Igual, R.; Sánchez, Á.; Castro, M. The effect of personalization on smartphone-based fall detectors. *Sensors* **2016**, *16*, 117.

55. Sztyler, T.; Stuckenschmidt, H. Online personalization of cross-subjects based activity recognition models on wearable devices. Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on. IEEE, 2017, pp. 180–189.

56. Weiss, G.M.; Lockhart, J.W. The impact of personalization on smartphone-based activity recognition. AAAI Workshop on Activity Context Representation: Techniques and Languages, 2012, pp. 98–104.