Article

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

Daniela Micucci *, Marco Mobilio and Paolo Napoletano

Department of Informatics, Systemic and Communication, University of Milano Bicocca, Viale Sarca 336, 20126 Milan, Italy; marco.mobilio@disco.unimib.it (M.M.); paolo.napoletano@disco.unimib.it (P.N.)
* Correspondence: daniela.micucci@unimib.it; Tel.: +39-02-6448-7866

Academic Editor: name
Received: date; Accepted: date; Published: date

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 smartphone accelerometer dataset designed for activity recognition. The dataset includes 11,771 activities performed by 30 subjects of ages ranging from 18 to 60 years. Activities are divided in 17 fine grained classes grouped in two coarse grained classes: 9 types of activities of daily living (ADL) and 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 two different classifiers and with different configurations. The best results are achieved with k-NN classifying ADLs only, considering personalization, and with both windows of 51 and 151 samples.

Keywords: smartphone accelerometers; dataset; 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 activity of daily living (ADLs) 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2017 doi:10.20944/preprints201706.0033.v1

Techniques for human activity recognition usually rely on data acquired by 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]. Recently, a lot of attention paid to wearable sensors because they are less intrusive, work outdoors, sometimes cheaper than the ambient ones, and, in the case of smartphones, widespread even in the elderly population.

Researchers usually rely on their own samples specifically recorded from sensor(s) to evaluate their new technique, and almost never make the registered data publicly available [7–9]. The lack of a common source of data makes difficult to compare in an objective way the several newly proposed techniques and implementations [9–11].

However, very recently, a few set of accelerometer datasets for human activity recognition have been collected by researchers worldwide and made publicly available. The datasets can be primarily divided in two main groups: those acquired by ad-hoc wearable devices (e.g., SHIMMER sensor nodes), and the other from Android-based smartphones.

We have thoroughly examined the datasets acquired with Android mobile devices in order to identify their strengths and weaknesses so as to outline an effective method for carrying out a new acquisition campaign. Table 1 shows the publicly available datasets that have been recorded by means of Android 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 [12]	2015	yes	yes	5	-	-	-	-	-
MobiFall [11]	2014	yes	yes	24	7	17	22 - 47 27 ± 5	160 - 189 175 ± 7	50 - 103 76.4 ± 14.5
MobiAct [13]	2016	yes	yes	57	15	42	20 - 47 25 ± 4	160 - 193 175 ± 4	50 - 120 76.6 ± 14.4
Shoaib PA [14]	2013	yes	no	4	0	4	25 - 30 -	-	-
Shoaib SA [15]	2014	yes	no	10	0	10	25 - 30 -	-	-
tFall [16]	2013	yes	yes	10	7	3	20 - 42 31 ± 9	161 - 184 173 ± 1	54 - 98 69.2 ± 13.1
UCI HAR [17]	2012	yes	no	30	-	-	19 - 48 -	-	-
UCI HAPT [18]	2015	yes	no	30	-	-	19 - 48 -	-	-
WISDM [19]	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

Our research results in 9 datasets recorded in the period 2012 to 2016 (column *Year*). From the description that authors provided, the total number of datasets decreases to 7 because MobiAct and UCI HAPT are updated versions of MobiFall and UCI HAR respectively. Thus, in the following we will refer to 7 datasets overall, discarding MobiFall and UCI HAR.

Only 3 datasets out of 7 contain both falls (column *Falls*) and ADLs (column *ADLs*).

The average number of subjects for dataset is 24 (column *Nr. of subjects*). The datasets that specify the gender of the subjects (which are only 2) contain in mean 8 women and 26 men (columns *Gender - Female* and *Gender - Male* respectively).

 doi:10.20944/preprints201706.0033.v1

Nor DMPSBFD, neither WISDM specifies the age of the subjects (column *Age*). In the remaining 5 datasets, subjects are aged between 22 and 39 on average with a standard deviation of 3 and 8 respectively.

Finally, only MobiAct and tFall 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 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 (ADLs 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. Perhaps, this information is unhelpful if there is no information on the physical characteristics of the subject. Looking at the double column *Gender*, only MobiAct, Shoaib PA, and Shoaib SA maintain information related to the gender of the subject. Finally, it is surprising that only MobiAct allows 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 allows to select data according to several dimensions, such as the age, the sex, the weight of the subjects, or the type of ADL and fall. Unfortunately, the other datasets are not suitable in some experimental evaluations. For example, the evaluation of the effects of personalization in classification techniques [20] taking into account the physical characteristics of the subjects, that is, operating leave-one-subject-out cross-validation [21].

To further contribute to the worldwide collection of accelerometer patterns, in this paper we present a new dataset of smartphone accelerometer data, 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 human activity recognition 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, so to create a new dataset that juxtaposes and complements MobiAct with regard to the data it is missing. Thus, such a dataset would have to contain a large number of subjects (more than the 24 in average), with a large number of women (to compensate MobiAct), with subjects over the age of 47 (to extend the range of MobiAct), 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 human activities 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 containing 17 fine grained classes of human activities grouped in two coarse grained activities classes: ADLs and falls. The dataset contains a total of 11,771 (7,759 ADLs and 4,192 falls) activities performed by 30 subjects, mostly females (24), of ages ranging from 18 to 60 years. Each accelerometric entry in the dataset maintains the information about the subject that generated it. Moreover, each accelerometric 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 two classifiers: the k-Nearest Neighbour (k-NN) and Support Vector Machines (SVM). Results show how much the proposed dataset is challenging with respect to several classification tasks.

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 with digital output. 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. For the experiments presented in this paper, the sampling rate was about 50 Hz, which is commonly used in literature for activity recognition from data acquired through smartphones [15–17]. 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 times in the left one and the remaining times 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 smartphones.

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. This set includes, sorted by year of creation from the oldest the most recent, the following datasets: DLR v2 [22], USC HAD [23], DaLiAc [10], EvAAL [24], UCI ARSA [25], BaSA [26], MMsys [9], and SisFall [27].

For what concerns the ADLs, Figure 1 shows the most common ones in the 17 datasets we analyzed. 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. Moreover, none of the dataset specifies exactly what the activities is 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. Moreover, we detailed the generic *Standing up*, by including the *Standing up from laying* and *Standing*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2017 doi:10.20944/preprints201706.0033.v1

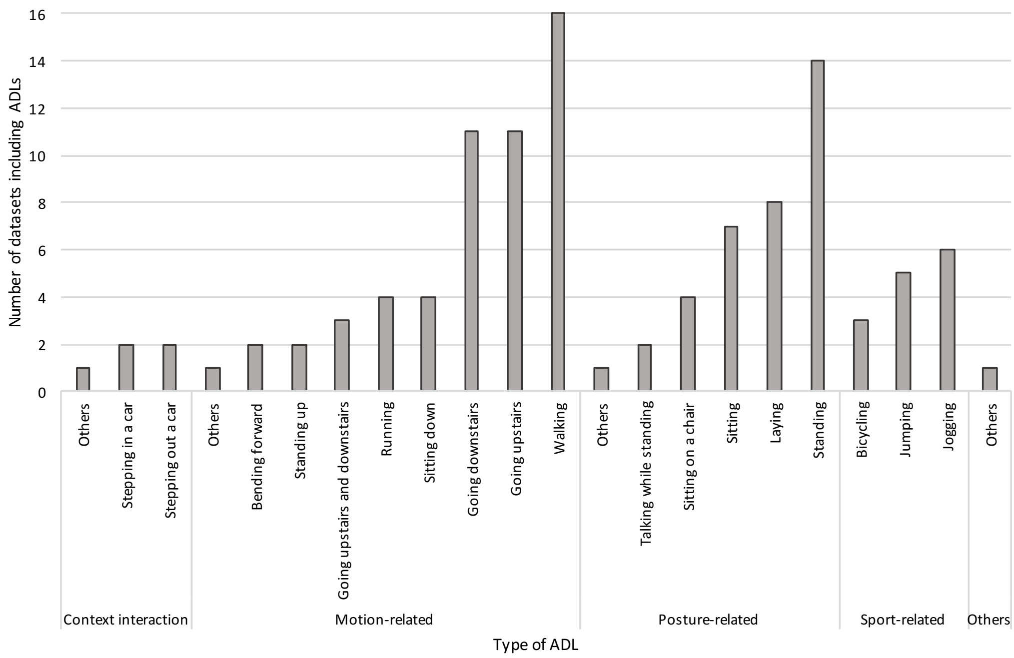


Figure 1. ADLs and theirs occurrence in the publicly available datasets analysed grouped by category

up from sitting activities. Finally, we included also the *Lying down from standing*, which was included by one dataset only.

In the *Sport-related* category, we do not included *Jogging* even if it is the most popular activity in its category. We prefer to consider *Running* because *Jogging* frequency is lower than *Running* being a *Sport-related* activity, especially in elderly. In *Sport-related* category, we have chosen 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 [28].

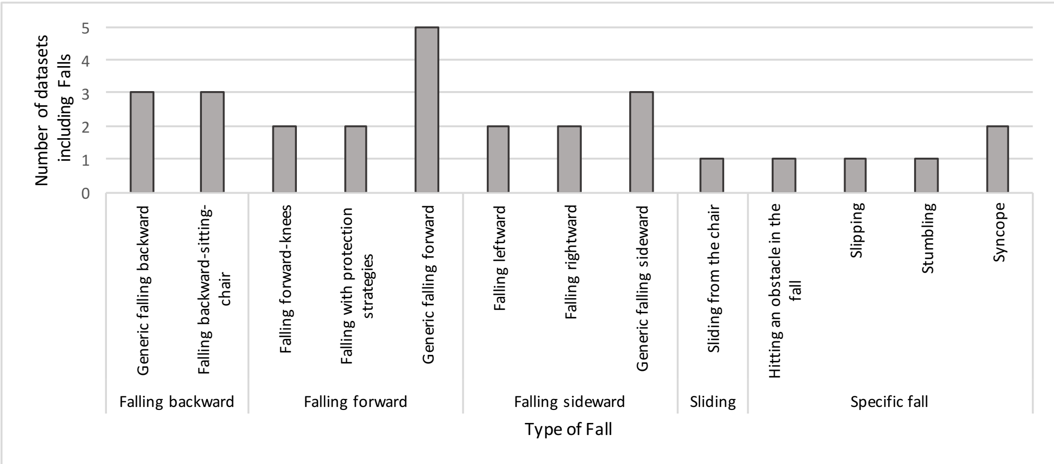
Finally, we do not include ADLs belonging to categories such as *Housekeeping-*, *Cooking-*, or *Personal care-related* (those fall in the *Others* category in Table 1, because we are interested in **low order** activities of daily living, which include simple activities such as, *Standing*, *Sitting 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** activities.

Table 2. ADLs performed by the subjects In the UniMiB SHAR dataset

Category	Name	Description	Label
Motion-related	Lying down from standing	From standing to lying on a bed	LyingDownFS
	Standing up from laying	From laying on the bed to standing	StandingUpFL
	Standing up from sitting	From standing to sitting on a chair	StandingUpFS
	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
	Running	Moderate running	Running

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2017 doi:10.20944/preprints201706.0033.v1

For what concerns falls, Figure 2 shows the most common ones in the datasets containing falls that we analyzed. This set has been derived from the 5 datasets in Table 1 that includes falls, and from MMsys and SisFall datasets, which are the only datasets among the four from wearable devices that both include falls and specify also the type of fall. 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 the total number of falls we identified. We reported also falls which appear in one dataset only because the datasets including falls are less that those including ADLs.



.

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	*Generic* falling backward Falling backward-sitting-chair	Generic fall backward from standing Fall backward while trying to sit on a chair	FallingBack FallingBackSC
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 leftward Falling rightward	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

For what concerns falls, their choice was suggested by the following considerations: the number of falls should have been comparable to that of 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 one dataset only), we selected two falls from each of them. In each category, we selected the first two most popular falls for each category. The category *Falling sideward* is an exception since we preferred to choose the two most specific falls instead of keeping the too generic *Generic falling sideward*. For falls in categories that are numerically distributed homogeneously among datasets, we have chosen falls not included in MobiAct to complement it. For example, in the *Specific forward* category we included *Falling with protection strategies* because *Falling forward-knees* has been included in MobiAct. Table 3 shows the 8 simulated falls that we selected according the adopted criteria.

Finally, studies on this topic confirm the activities we selected are common in real-life [16,29–31].

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 $\pm$ 11	24 $\pm$ 9	36 $\pm$ 15	mean $\pm$ std
height		160 -190	160 - 172	170 - 190	min - max
		169 $\pm$ 7	166 $\pm$ 4	179 $\pm$ 6	mean $\pm$ std
weight		50 - 82	50 - 78	55 - 82	min - max
		64.4 $\pm$ 9.7	61.9 $\pm$ 7.8	74.7 $\pm$ 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 [32].

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 for too long sequences of activities, registrations have been subdivided in the three protocols showed in Table 6. 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.

Table 5 shows the three protocols designed for ADLs acquisition. *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 has been performed with a chair without armrests; the *Lying down from standing* and *Standing up from laying* have been performed on a sofa.

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.

2.5. 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:

 doi:10.20944/preprints201706.0033.v1

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2017 doi:10.20944/preprints201706.0033.v1

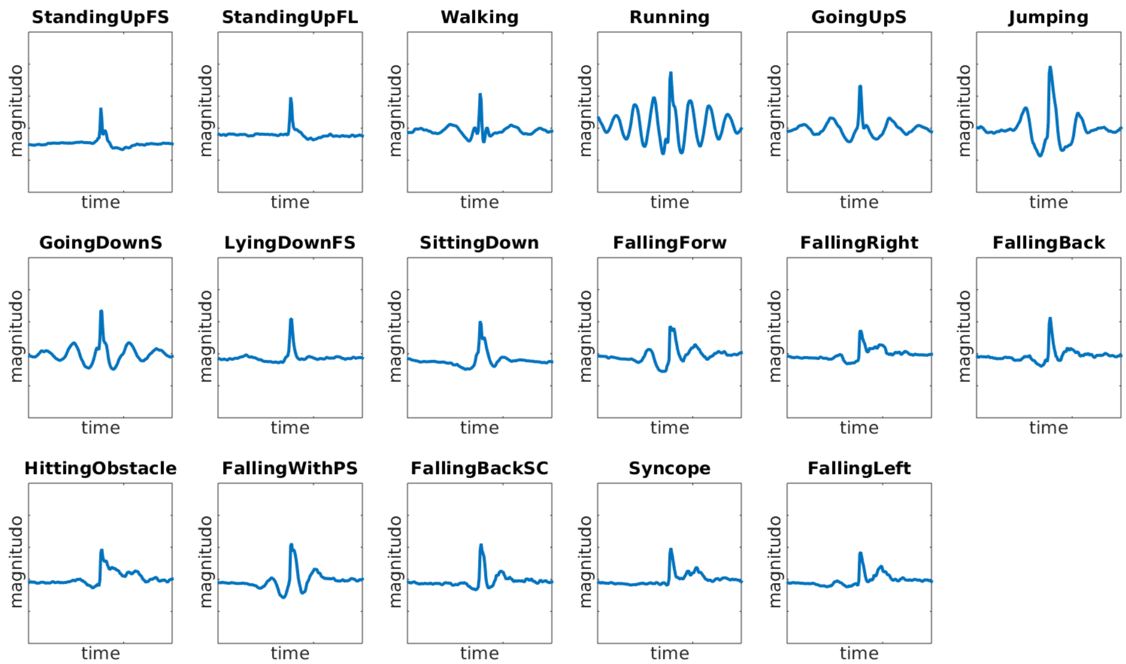


Figure 3. Samples of acceleration shapes

1. the magnitudo of the signal m_t at time t was higher than $1.5g$, with g being the gravitational acceleration;
2. the magnitudo 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. 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 [33,34]; 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 magnitudo 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 samples (a vector of size 1x453), one for each acceleration direction. The dataset is thus composed of 11,771 activity samples (7,579 ADLs and 4,192 FALLs) 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 running activities. The resulting activity distribution is plotted in Figure 4, where the running activities are about 2,000. On our web site we release both datasets, the one balanced and the original one.

3. Dataset Evaluation

As activity recognizer we experimented both the k-Nearest Neighbour (k-NN) with $k = 1$ and the Support Vector Machines (SVM) with a radial basis kernel on the following four subsets:

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;
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;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2017 doi:10.20944/preprints201706.0033.v1

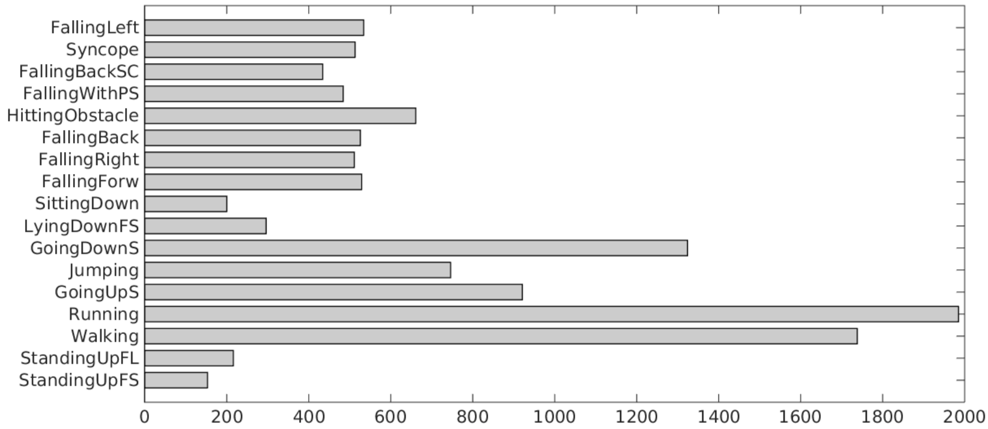


Figure 4. Activity samples distribution

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. This kind of evaluation with high probability leads to have data 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 30-fold cross-validation, also known as leave-one-out cross-validation. Each test fold is made of accelerometer data of one user only, namely the *test user*, while the training folds contain accelerometer data of all the other users except those of the *test user*.

The feature vectors used in the classification tasks are the raw data, that is, the 453-dimensional patterns obtained by concatenating the 151 acceleration samples recorded along each Cartesian direction. 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 magnitudo of the signal, frequency, or energy [29,35]. To make the experiments comparable with others experiments presented in the state of the art, we have also performed experiments considering a sub window of 51 samples taken from the original raw signals.

3.1. Evaluation metrics

As shown in Figure 4, each of the 17 sets of activities is different in size. To cope with the class imbalance problem of the dataset we used as metric the *macro average accuracy* [36].

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

Results of the 5-fold and 30-fold evaluations with both k-NN and SVM are showed in Table 7. The AF-2 recognition task is very easy for all the evaluation strategies, classifiers, and window sizes. The MAA achieved by SVM is about 98% and 97% in the case of 5-fold and 30-fold evaluations respectively. These results are similar to those obtained by previous researchers on a similar classification task performed on different datasets [16,35]. This means that it is very easy to distinguish between falls and no falls even in the case when no-subject dependent data are used for training.

In contrast, the AF-17 recognition task is quite challenging: the MAA is about 83% and 54% in the case of 5-fold and 30-fold evaluations respectively. This means that is quite difficult to distinguish among types of activities and falls especially in the case when no-subject dependent data are used for training.

The A-9 classification task is quite easy in the case of 5-fold, the MAA obtained by k-NN is about 88%, while it is difficult in the case of 30-fold evaluation, here the MAA obtained by SVM is about 65%. This last result depends mostly by the fact that each human subject performs activities in a different way. This effect of personalization changes accelerometer raw signals and so it influences the classification accuracy.

The F-8 recognition task is quite challenging: the MAA is about 78% and 49% in the case of 5-fold and 30-fold evaluations respectively. This result suggests that distinguish among falls is very complicated especially in the case when no-subject dependent data are used for training.

Figure 5 shows the confusion matrix of the k-NN experiment in the case of 151 samples and 5-fold cross-validation . Among all the ADLs, 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”, “Walking” and “Going downstairs”, “Going downstairs” and “Jumping”. Among all the falls the most misclassified pairs of falls are “Falling with protection strategies” and “Generic falling forward”, “Syncope” and “Falling rightward”, “Falling backward-sitting-chair” and “Generic falling backward”, “Falling rightward” and “Falling with protection strategies”, “Falling leftward” and “Syncope”.

These evaluation shows the feasibility of a real smartphone application for human activity recognition, which might be composed of a recording, a pre-processing, and a pre-trained classifier components. As soon as the “test user” starts the application on her smartphone, the accelerometer data starts to be recorded. Data are then pre-processed and salient windows of accelerometer signals (with a magnitudo peak higher than 1.5g) are selected. Those windows are then classified using a pre-trained classifier, which could be one of the classifiers that we trained using the dataset presented in this paper.

Table 7. Mean Average Accuracy for each classification task and window of 151 samples (W_1) and 51 (W_2) respectively. In bold the best result for each classification task and k-fold evaluation strategy.

	5-fold evaluation				30-fold evaluation			
	KNN		SVM		KNN		SVM	
data	W_1	W_2	W_1	W_2	W_1	W_2	W_1	W_2
AF-17	**83.02**	78.32	78.15	72.74	51.15	46.98	**54.70**	51.14
AF-2	97.65	93.20	**98.82**	96.67	92.67	86.30	**97.22**	92.68
A-9	**88.51**	88.34	82.68	81.75	62.31	62.08	62.47	**65.43**
A-8	**78.80**	71.70	76.09	66.72	43.50	37.73	**49.44**	39.71

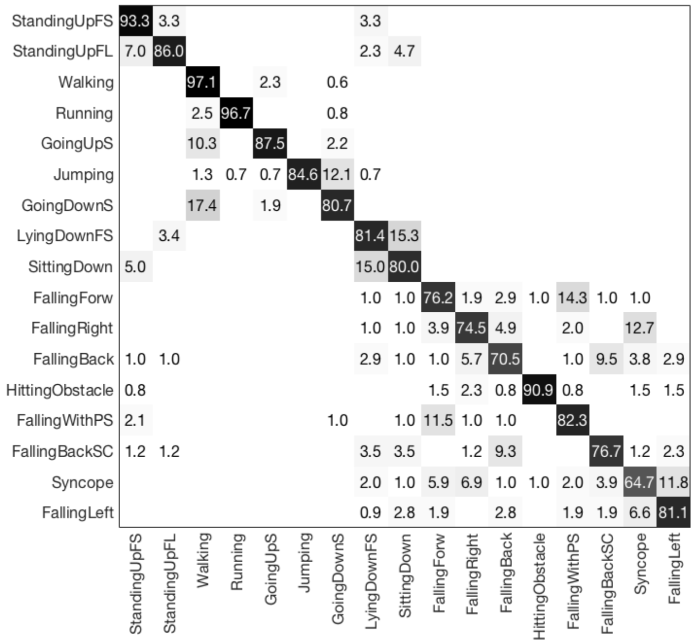


Figure 5. Confusion matrix of the AF-17 classification achieved with k-NN

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. MobiAct is the only dataset to make an exception. It is characterized by a high number of male subjects up to 47 years of age. Our goal was to create a new dataset containing also samples from subjects not included in MobiAct. The UniMiB SHAR dataset includes fine grained ADLs and falls performed by 30 humans, mostly female, with a huge range of ages, from 18 to 60 years. Classification results obtained on the proposed dataset showed that the window of 151 samples performed quite better that the 51 one. Moreover, we identified that personalization can lead to better results. Finally, we revealed that falls are quite complex to classify with respect to ADLs.

We are planning to carry out an evaluation of the state-of-the-art techniques for ADLs recognition on both UniMiB SHAR and all publicly available datasets of accelerometric 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.

 doi:10.20944/preprints201706.0033.v1

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. 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].
13. 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.
14. 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.
15. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. Fusion of Smartphone Motion Sensors for Physical Activity Recognition. Sensors. *Sensors* **2014**, *14*, 10146–10176.
16. Medrano, C.; Igual, R.; Plaza, I.; Castro, M. Detecting falls as novelties in acceleration patterns acquired with smartphones. *PLoS ONE* **2014**, *9*.
17. 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.
18. Reyes-Ortiz, J.L.; Oneto, L.; Samà, A.; Parra, X.; Anguita, D. Transition-Aware Human Activity Recognition Using Smartphones. *Neurocomputing* **2016**, *171*, 754 – 767.
19. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explor. Newsl.* **2011**, *12*, 74–82.
20. Medrano, C.; Plaza, I.; Igual, R.; Sánchez, A.; Castro, M. The Effect of Personalization on Smartphone-Based Fall Detectors. *Sensors* **2016**, *16*.
21. Khan, S.S.; Karg, M.E.; Kulić, D.; Hoey, J. Detecting falls with X-Factor Hidden Markov Models. *Applied Soft Computing* **2017**, *55*, 168 – 177.
22. 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.

23. 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.
24. 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.
25. Casale, P.; Pujol, O.; Radeva, P. Personalization and User Verification in Wearable Systems Using Biometric Walking Patterns. *Personal Ubiquitous Comput.* **2012**, *16*, 563–580.
26. 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.
27. Sucerquia, A.; López, J.D.; Vargas-Bonilla, J.F. SisFall: A Fall and Movement Dataset. *Sensors* **2017**, *17*.
28. 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.
29. Igual, R.; Medrano, C.; Plaza, I. A comparison of public datasets for acceleration-based fall detection. *Medical Engineering & Physics* **2015**, *37*, 870 – 878.
30. 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.
31. 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.
32. World Medical Association. WMA Declaration of Helsinki - Ethical principles for medical research involving human subjects, 2013.
33. 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.
34. 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.
35. 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.
36. He, H.; Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*; John Wiley & Sons, 2013.