IEEE
Sensors Letters Volume 2(3) (2017)
Article Subject (Sensor Application)

IEEE Sensors Council

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

Daniela Micucci[1]*, Marco Mobilio[1], and Paolo Napoletano[1]
[1] Department of Informatics, Systems and Communication, University of Milano - Bicocca, Milan, ITALY
* Member, IEEE

Abstract— Smartphones, smartwatches, fitness trackers, and ad-hoc wearable devices are being increasingly used to monitor human activities because are widespread and commonly used by people. Usually, machine-learning-based algorithms process data acquired by their sensors to classify human activities. The success of those algorithms mostly depends on the availability of training (labeled) data, which, if made publicly available, would allow researchers to make objective comparisons between techniques. Unfortunately, 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, such as the age, the gender, and so on.
In this letter, we present a new smartphone accelerometer dataset designed for activity recognition. The dataset includes 7,013 activities performed by 30 subjects, mostly females, 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, benchmarked with two different classifiers, thanks to its unique features will be of interest to the scientific community.

Index Terms—Sensor Application, Smartphone accelerometers, Dataset, Human Activity recognition, Fall detection

I. INTRODUCTION

In recent years, research on techniques able to recognize and monitor human activities is very active. Such techniques can be applied to several application domains spanning from leisure to health care. Dealing with health care, many attention is being posed on techniques able to recognize both activities of daily living (ADLs) and falls.

The recognition of ADLs may allow to infer the amount of physical activity that a subject perform daily. It is recognized, indeed, 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 [16].

The promptly recognition of falls may help in reducing the consequence (even fatal) that a fall may cause mostly in elderly people, whose number is rapidly increasing [13]. 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 [12]. Moreover, fallers who are not able to get up more likely require hospitalization or, even worse, die [11].

Techniques for human activity recognition usually rely on data acquired by sensors, which can be physically deployed in the ambient (ambient sensors) or worn by people (wearable sensors). Recently, a lot of attention has been payed to wearable sensors because they are less intrusive, also working outside own home, 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 publicly available the registered data [1, 7, 10]. The lack of a common source of data makes difficult to compare in an objective way the several newly proposed techniques and implementations [6,10,14].

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. In the former set we identified 11 datasets dated from 2008 to 2015, in the latter only 8 datasets recorded in the period 2012 to 2016. This trend enforces our conviction that smartphones are being more and more considered suitable devices for accelerometric data acquisition. Among the smartphones-based datasets, we identified 4 datasets that include ADLs only, and 4 including both ADLs and falls. Finally, among this last group MobiFall [14] and MobiAct [15] partly overlap because the more recent one extended the older.

Corresponding author: D. Micucci (daniela.micucci@unimib.it). All the authors contributed equally.

Digital Object Identifier: 10.1109/LSEN.XXXX.XXXXXXX (inserted by IEEE).

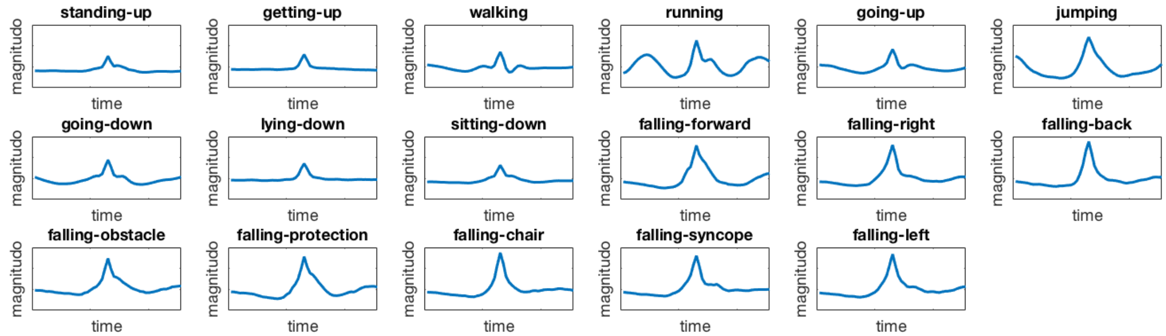

Fig. 1. Samples of acceleration shapes.

In smartphones-based datasets, the mean number of subjects for dataset is 24±16 (mean±standard deviation). For those datasets that specify the number of men and women (which are only 4), 7±5 are women and 18±15 are men. The mean age of the subjects spans from 21 to 43. Finally, information about the type of ADL, the type of fall, the subject's physical characteristics are missing in most of the datasets excepts for MobiFall and MobiAct. Thus, it is not possible to select data according to specific dimensions, such as the age, the sex, the weight of the subjects, or the type of ADL and fall.

To further contribute to the worldwide collection of accelerometer patterns, in this letter 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 and rich dataset of acceleration pattern captured by smartphones to be used as a common benchmark for the objective evaluation of human activity recognition techniques. Thus, such a dataset would have to contain many subjects, with different physical characteristics, performing a wide number of both ADLs and falls. 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 fulfill 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 7,013 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. This way it is possible to make training by selecting specific subject: by age, sex, weight and height. 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).

The closest dataset in terms of richness with respect to UniMiB SHAR is MobiAct: it has a greater number of subjects (57), but only 26% are women, and the subjects involved are younger (max 47 years old). The other datasets from smartphones have a common flaw: the inability to attribute the acceleration patterns to specific subjects, thus making impossible to reason in terms of people's physical characteristics. Finally, some datasets do not provide even the type of ADL or fall.

We benchmarked the dataset by performing several experiments. We evaluated two classifiers: the k-Nearest Neighbour (k-NN) and Support Vector Machines (SVM). Results achieved show how much the proposed dataset is challenging with respect to several classification tasks.

II. DATASET DESCRIPTION

The accelerometer data have been acquired by using a mobile application specially designed and implemented by the authors, and running on a Samsung Galaxy Nexus smartphone equipped with the Android OS version 5.1.1. The application records data through the built-in triaxial accelerometer with a sample frequency of 50 Hz. In addition, the application records audio signals with a sample frequency of 8,000 Hz. 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.

The voluntary subjects involved in the experimentation are 30 and their characteristics are presented in Table 1. Subjects performed 17 different types of activities: 9 activities of daily living (ADLs) and 8 simulated falls (FALLs). Activity types have been selected considering they are common in real-life scenarios according to previous studies on this topic [3-5,8]. Subjects performed the following ADLs: *walking*, *running*, *going upstairs* (going-up), *going downstairs* (going-down), *jumping*, *sitting down on a chair* (sitting-down), *standing up from the chair* (standing-up), *lying down on a bed* (lying-down), *getting up from the bed* (getting-up).

The FALLs performed were: *falling forward* (falling-forward), *falling backward* (falling-back), *falling left* (falling-left), *falling right* (falling-right), *falling with contact to an obstacle* (falling-obstacle), *syncope* (falling-syncope), *sitting down on empty chair* (falling-chair), *falling using compensation strategies to prevent the impact* (falling-protection).

Fig. 1 shows samples of acceleration shapes. For each activity, we displayed the average magnitudo shape obtained by averaging all the subjects' shapes.

Subjects wore the smartphone in their two trouser pockets: half of the times in the left one and the remaining times in the right one. Each activity was repeated at least 5 times. We asked each subject to clap their hands early before and after they performed the activity to be recorded. The presence of claps in the audio signals have helped us during the accelerometer data annotation process.

From the annotated accelerometer data, we extracted only signal windows of 50 Hz width (51 samples for each window) that were centered around a peak of magnitudo higher then *1.5g*, with *g* being the gravitational acceleration. For each activity, the accelerometer data vector is made of 3 vectors of 51 samples (a vector of size 1x153), one for each acceleration direction. The dataset is thus composed of 7,013 activity samples (5,097 ADLs and 1,699 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. The activity distribution is plotted in Fig. 2.

Table 1. The characteristics of the subjects.

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

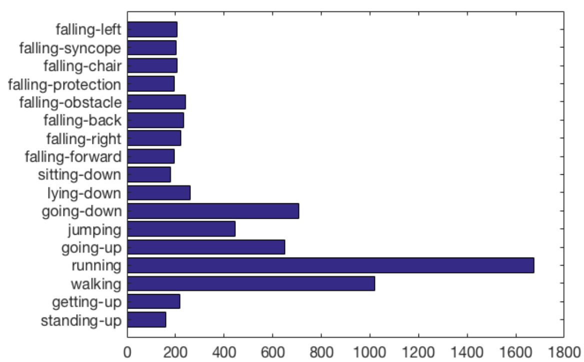

Fig. 2. Activity samples distribution.

III. DATASET EVALUATION

As activity recognizer, we experimented both the k-Nearest Neighbor (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;
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. 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 153-dimensional patterns obtained by concatenating the 51 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, frequency, or energy [3,9].

IV. EVALUATION METRICS

As shown in Fig. 2, each of the 17 sets of activities is different in size. To cope with the class imbalance problem of the dataset, we jointly used two assessment metrics: the *standard accuracy* and the *macro average accuracy* [2].

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:

- *Standard Accuracy* (*SA*)

$$SA = \frac{\sum_{a=1}^{|E|} TP_a}{\sum_{a=1}^{|E|} NP_a}$$

  SA is a global measure that does not consider the accuracy achieved on a single activity.

- *Macro Average Accuracy* (*MAA*)

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

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

V. RESULTS AND DISCUSSION

Results of the 5-fold and 30-fold evaluations with both k-NN and SVM are showed in Table 2. The AF-2 recognition task is very easy for both evaluations and classifiers. The MAA achieved by SVM is about 94% and 93% in the case of 5-fold and 30-fold evaluations respectively.

These results are close to those obtained by previous researchers on a similar classification task performed on different datasets [4,7]. The A-9 and F-8 recognition tasks are quite difficult, especially when we performed the 30-fold evaluation: we achieved a MAA of about 63% and 51% respectively. Therefore, the AF-17 recognition task is quite challenging: the MAA is about 61% and 54% in the case of 5-fold and 30-fold evaluations respectively.

Fig. 3 shows the confusion matrix corresponding to the experiment AF-17 (5-fold with SVM). Numbers represent the accuracy percentage (ranging from 0 to 100). The most misclassified activities are standing-up, getting-up which are confused with sitting-down and lying-down respectively.

This 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 his smartphone, the accelerometer data starts to be recorded. Data are then pre-processed and salient windows of accelerometric 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 letter. Obviously, the first periods the "test user" uses the application, there are no personal accelerometer data. The capability of the classifier to recognize activities relies only on training examples.

Table 2. SA and MAA for each configuration.

	5-fold evaluation				30-fold evaluation			
	KNN		SVM		KNN		SVM	
data	**SA**	**MAA**	**SA**	**MAA**	**SA**	**MAA**	**SA**	**MAA**
AF-17	78.77	61.81	78.83	61.54	78.77	61.81	78.83	61.54
AF-2	95.53	91.37	97.03	94.87	95.53	91.37	97.03	94.87
A-9	86.89	68.89	86.47	68.06	86.89	68.89	86.47	68.06
A-8	60.12	60.38	59.17	58.82	60.12	60.38	59.17	58.82

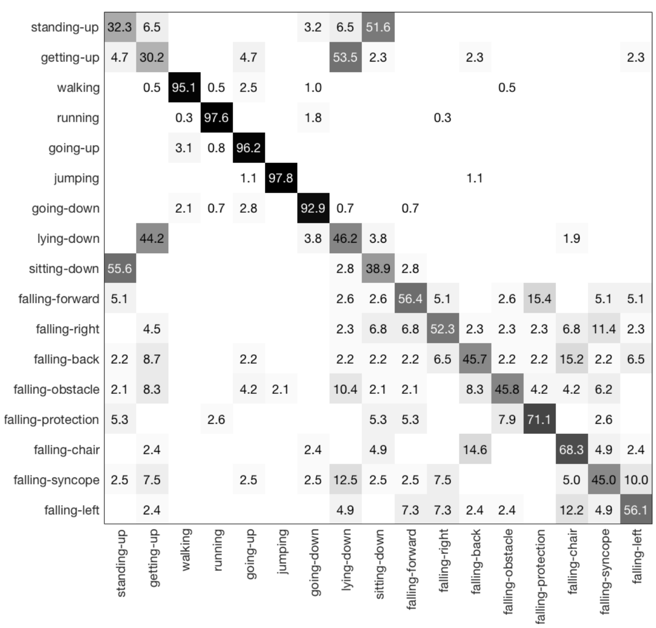

Fig. 3. Confusion matrix of the experiment AF-17 (5-fold with SVM).

The dataset, the Matlab scripts to repeat the experiments, and additional materials are available at the following address: http://www.sal.disco.unimib.it/technologies/unimib-shar/

VI. CONCLUSION

Many of the existing publicly available datasets include activities performed by a lower number of human subjects, mostly are male, with a narrow range of ages, and in most cases, fine grained ADLs and FALLs are not annotated.

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 AF-17 recognition task, as well as the A-9 and F-8 ones, performs quite well.

We are carrying 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.

ACKNOWLEDGMENT

The authors would like to thank Silvia Preda and Lisa Raimondi for their invaluable help in the collection of the accelerometer data during the acquisition campaign.

REFERENCES

[1] Chen J, Kwong K, Chang D, Luk J, Bajcsy R (2006) "Wearable sensors for reliable fall detection," Annual International Conference of the Engineering in Medicine and Biology Society.

[2] He H, Ma Y (2013) *Imbalanced learning: foundations, algorithms, and applications*. John Wiley & Sons.

[3] Igual R, Medrano C, Plaza I (2015) "A comparison of public datasets for acceleration-based fall detection," *Medical engineering & physics*, vol. 37, no. 9, pp. 870-878.

[4] Kangas K, Vikman I, Nyberg L, Korpelainen R, Lindblom J, Jämsä T (2012) "Comparison of real-life accidental falls in older people with experimental falls in middle-aged test subjects," *Gait & posture*, vol. 35, no. 3, pp. 500–505.

[5] Klenk J, Becker C, Lieken F, Nicolai S, Maetzler W, Alt W, Zijlstra W, Hausdorff J, Van Lummel R C, Chiari L, Lindemann U (2011) "Comparison of acceleration signals of simulated and real-world backward falls," *Medical engineering & physics*, vol. 33, no. 3, pp. 368-373.

[6] Leutheuser H, Schuldhaus D, Eskoer B M (2013) "Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset," *PLoS ONE*, vol. 8, no. 10.

[7] Li Q, Stankovic JA, Hanson MA, Barth AT, Lach J, Zhou G (2009) "Accurate, fast fall detection using gyroscopes and accelerometer derived posture information," International Workshop on Wearable and Implantable Body Sensor Networks, IEEE Computer Society.

[8] Medrano C, Igual R, Plaza I, Castro M, "Detecting falls as novelties in acceleration patterns acquired with smartphones," *PLoS ONE*, vol. 9, no. 4, 2014.

[9] Micucci D, Mobilio M, Napoletano P, Tisato F (2017) "Falls as anomalies? an experimental evaluation using smartphone accelerometer data," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 1, pp. 87-99.

[10] Ojetola O, Gaura E, Brusey J (2015) "Data set for fall events and daily activities from inertial sensors," ACM Multimedia Systems Conference.

[11] Tinetti M E, Liu W L, Claus E B (1993) "Predictors and prognosis of inability to get up after falls among elderly persons," *Journal of the American Medical Association*, vol. 269, no. 1, pp. 65-70.

[12] Tromp A M, Pluijm S M, Smit J H, Deeg D J, Bouter L M, Lips P (2001) "Fall-risk screening test: a prospective study on predictors for falls in community-dwelling elderly," *Journal of clinical epidiemology*, vol. 54, no. 8, pp. 837-844.

[13] United Nations (2013) "World Populatin Ageing Report 2013," http://goo.gl/z410Ju.

[14] Vavoulas G, Pediaditis M, Chatzaki C, Spanakis E G, Manolis T (2014) "The Mobifall dataset: fall detection and classification with a smartphone," *International Journal of Monitoring and Surveillance Technologies Research*, pp. 44–56.

[15] Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M (2016) "The Mobiact dataset: Recognition of activities of daily living using smartphones," International Conference on Information and Communication Technologies for Ageing Well and e-Health.

[16] World Health Organization (2010) "Global recommendations on physical activity for health," http://www.who.int/dietphysicalactivity/publications/9789241599979/.