On the Personalization of Classification Models for Human Activity Recognition

ANNA FERRARI [ID], DANIELA MICUCCI [ID], MARCO MOBILIO [ID], AND PAOLO NAPOLETANO [ID]

Department of Computer Science, Systems and Communications, University of Milano-Bicocca, 20126 Milan, Italy

Corresponding author: Daniela Micucci (daniela.micucci@unimib.it)

ABSTRACT Recently, a significant amount of literature concerning machine learning techniques has focused on automatic recognition of activities performed by people. The main reason for this considerable interest is the increasing availability of devices able to acquire signals which, if properly processed, can provide information about human activities of daily living (ADL). The recognition of human activities is generally performed by machine learning techniques that process signals from wearable sensors and/or cameras appropriately arranged in the environment. Whatever the type of sensor, activities performed by human beings have a strong subjective characteristic that is related to different factors, such as age, gender, weight, height, physical abilities, and lifestyle. Personalization models have been studied to take into account these subjective factors and it has been demonstrated that using these models, the accuracy of machine learning algorithms can be improved. In this work we focus on the recognition of human activities using signals acquired by the accelerometer embedded in a smartphone. The contributions of this research are mainly three. A first contribution is the definition of a clear validation model that takes into account the problem of personalization and which thus makes it possible to objectively evaluate the performances of machine learning algorithms. A second contribution is the evaluation, on three different public datasets, of a personalization model which considers two aspects: the similarity between people related to physical aspects (age, weight, and height) and similarity related to intrinsic characteristics of the signals produced by these people when performing activities. A third and last contribution is the development of a personalization model that considers both the physical and signal similarities. The experiments show that the employment of personalization models improves, on average, the accuracy, thus confirming the soundness of the approach and paving the way for future investigations on this topic.

INDEX TERMS Personalization, human activity recognition, ADL, similarity, machine learning, smartphone.

I. INTRODUCTION

Human activity recognition (HAR) is a field of research that aims at defining and experimenting new techniques able to automatically recognize human activities exploiting signals recorded by wearable and/or environmental devices [1]. In the majority of cases, environmental devices require an installation in the home environment and devices such as cameras are perceived as intrusive devices, especially by elderly people [2]. For these reasons, in recent years the focus has shifted to the use of wearable devices. Among them, special attention is currently being paid to smartphones, smartwatches, and fitness devices [3]–[7]. This is mainly

The associate editor coordinating the review of this manuscript and approving it for publication was Jihwan P. Choi [ID].

due to their wide diffusion among the population and to the presence of various types of sensors integrated in the devices (e.g., accelerometer, gyroscope, orientation, and GPS). HAR techniques based on signals from sensors of wearable devices have been proposed for several application domains: sport tracking uses signals such as GPS and accelerometer for evaluating the activity of the user [8], detection of falls exploits wearable devices preventing mortality of elder people [5], [7], behavior analysis, based on physical measurements, prevents dementia diseases [9], and many others. Most of these HAR techniques rely on accelerometers because they have low power consumption and permit continuous sensing over a complete day [10].

During the last decade, plenty of traditional machine learning as well as deep learning methods for HAR that

use accelerometers has been proposed in literature [1], [6], [11]–[13]. However, real-world HAR systems may achieve non satisfying recognition accuracy in real world applications because HAR techniques may struggle to generalize to new users and/or new environments [14], [15].

Several factors may affect the accuracy of activity recognition methods: i) position of the device (e.g., pocket, hand, or bag); ii) differences between different brands of sensors, in terms of sensitivity range and sampling frequency; iii) human characteristics, such as age, gender, weight, height, lifestyle, and physical abilities. While factors related to the position and the characteristics of the devices have been largely investigated, few works have explored the effects of human characteristics on recognition accuracy [5], [14], [16]–[18]. Lane *et al.* [17] proposed a new method to take into account human factors. This method exploits the similarity between users to weight training data and thus to improve the recognition accuracy. Unfortunately, results achieved by these researchers are not reproducible because the dataset used for the experimentation is not publicly available and moreover, the authors mainly focused on the automatic annotation of inertial signals and not classification of activities of subjects. The approach proposed by Lane *et al.* [17] deserves further investigation and thus it has been the starting point of the research we performed and whose results are presented in this paper.

Our research question was: *can personalization methods be employed for increasing the accuracy of HAR techniques based on accelerometer signals?*

To reply to this research question, we have:

- experimented several personalization methods on three public datasets in order to make the results reproducible and thus allowing future research on this topic;
- defined a suitable procedure to split the experimental data into training, validation, and test sets;
- defined a new personalization method that includes and generalizes the different ideas discussed in the state-of-the-art.

The paper is structured as follows: Section II briefly introduces the literature on human activity recognition and then discusses recent papers that deal with the concept of personalization; Section III presents the proposed personalization methods; Section IV presents the datasets used for the experimentation and some descriptive and exploratory analysis of the data; Section V presents the results of the experiments and finally, Section VI summarizes the conclusions and outlines future work.

II. RELATED WORK
A. HAR USING MACHINE LEARNING

Literature on ADLs recognition is extensive and includes approaches that are designed on different assumptions. Therefore, several orthogonal criteria that can be used to classify the proposed approaches. Common used classification criteria are based on: i) the type of information used to infer the models, ii) the type of machine learning techniques used to process the signals, and iii) the type of sensors used to acquire the signals to be processed.

The type of information used to build the classifiers divides the approach into three main categories: data-driven, knowledge-driven, and hybrid.

Data-driven approaches use data mining and machine learning techniques to learn activity models. Data-driven approaches are able to handle uncertainty and temporal information. The flaw is that data-driven approaches require large datasets of labelled data to train classifiers [19]. Chen and Nugent [20] provide a recent survey of data-driven approaches.

Knowledge-driven approaches use a-priori contextual information to infer the activities performed. The prior knowledge may include for example, the implicit relationships between activities, the related temporal and spatial context, and the entities involved (objects and people) [21]. Knowledge-driven approaches are semantically clear and easy to get started. However, they suffer in handling uncertainty and temporal information [19]. Approaches may be further classified in logic-, ontology-, and mining-based. See Chen *et al.* [19] for a deep analysis. Examples of knowledge-driven approaches are those from Riboni and Bettini [22] and from Chen *et al.* [23]. More approaches are discussed by Chen *et al.* in reference [20].

Finally, hybrid approaches combine data- and knowledge-driven approaches to take the advantages from each of them. Examples of hybrid approaches are those from Azkune and Almeida [21], Stevenson and Dobson [24], and Civitarese *et al.* [25].

HAR activity is mainly carried out with the support of machine learning techniques. Initially, traditional machine learning techniques, such as Support Vector Machines, *k*-Nearest Neighbor, and Decision Tree were used. See for example the survey by Lara *et al.* [1]. In recent years, due to their successful use in the computer vision area [26], deep learning techniques are also increasingly being used in sensor-based human activity recognition. Wang *et al.* [13] discuss the recent advance of sensor-based deep learning approaches. Deep learning techniques present advantages over traditional approaches, such as the automatic high-level feature extraction [13]. However, deep learning techniques, unlike traditional approaches, require a large number of samples and an expensive hardware to estimate the model [27].

Two main kinds of sensors are used in HAR: the environmental and the wearable ones. Environmental sensors are immersed in the environment and mainly include cameras, RFID tags and readers, and microphones [28], [29]. On the opposite, wearable sensors are embedded in the devices that are worn by users. The mostly used sensors are accelerometers, gyroscopes, and magnetometers [30]–[32]. In the last years, smartphones and smartwatches have became increasingly powerful devices, embed several sensors including the inertial ones, and people have began to interact with them as part of their daily living [1]. Moreover, the on body-sensing
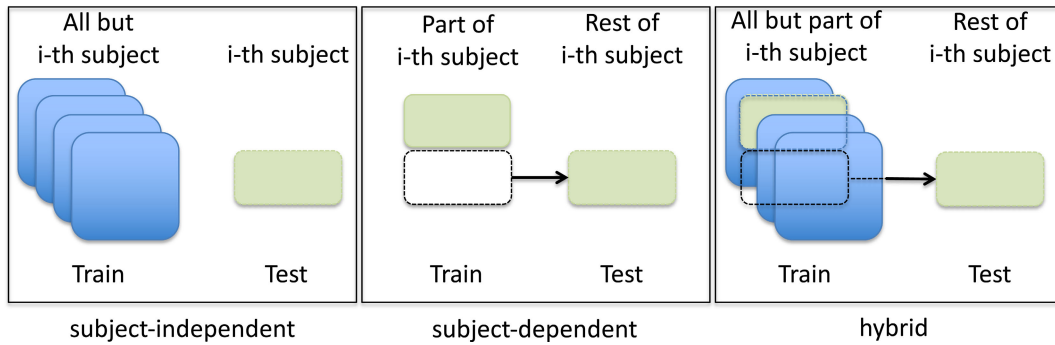
proves to be the most prevalent monitoring technology for activity recognition [33]. For this reason, in recent years, several HAR approaches have been designed to recognize ADLs by processing signals acquired from smartphones and smartwatches [34]–[37].

B. PERSONALIZATION IN HAR

Although research on activity recognition techniques from wearable devices is very active, the resulting systems are limited in their ability to generalize to new users and/or new environments, and require considerable effort and customization to achieve good performance in a real-context [14], [15].

One of the most relevant difficulty to face with new situations is due to the population diversity problem [17], that is, the natural differences between users' activity patterns, which implies that different executions of the same activity are different.

According to Zunino *et al.* [38], two factors influence why the same activity is carried out in a different way: the *inter-subject variability*, which either refers to anthropometric differences of body parts or to personal styles in accomplishing the activity (in other words, different subjects may differently perform the same activity); and the *intra-subject variability*, which represents the random nature of each class of activity due to pathological conditions or environmental factors (in other words, a subject never performs the same activity in the same exact way).

Thus, as users of mobile sensing applications increase in size, the differences between people cause the accuracy of classification to degrade quickly [17].

To face this problem, activity classification models should be able to generalize as much as possible with respect to the final user and the real execution context.

In order to achieve generalizable activity recognition models, three approaches are mainly adopted in literature: *subject-independent*, *subject-dependent*, and *hybrid*.

The *subject-independent* (also called *impersonal*) model does not use the end user data for the development of the activity recognition model. It is based on the definition of a single activity recognition model that must be flexible enough to be able to generalize the diversity between users and it

should be able to have good performance once a new user is to be classified.

The *subject-dependent* (also called *personal*) model only uses the end user data for the development of the activity recognition model. The specific model, being built with the data of the final user, is able to capture her/his peculiarities, thus it should well generalize in the real context. The flaw is that it must be implemented for each end user [39].

The *hybrid* model uses the end user data and the data of the other users for the development of the activity recognition model. In other words the classification model is trained both on the data of the users and on a part of the data of the final user. The idea is that the classifier should recognize easier the activity performed by the final user.

Figure 1 shows a graphical depiction of the three models to better clarify their differences.

Tapia *et al.* [40] introduced the subject-independent and subject-dependent models, and later Weiss and Lockhart [18] the hybrid model.

The models were compared by different researchers and also extended in order to achieve better performance.

Medrano *et al.* [5] demonstrated that the subject-dependent approach achieves higher performance then subject-independent approach for falls detection, called respectively *personal* and *generic fall detector*. Shen *et al.* [41] achieved similar results for activity recognition and come to the conclusion that the subject-dependent (termed *personalized*) model tends to perform better than the subject-independent (termed *generalized*) one because user training data carries her/his personalized activity information. Lara *et al.* [42] consider subject-independent approach more challenging because in practice, a real-time activity recognition system should be able to fit any individual and they consider not convenient in many cases to train the activity model for each subject.

Weiss and Lockhart [18] and Lockhart and Weiss [43] compared the subject-independent and the subject-dependent (termed *impersonal* and *personal* respectively) with the hybrid model. They concluded that the models built on the subject-dependent and the hybrid approaches achieve same

performance and outperform the performance of the model based on the subject-independent approach.

Similar conclusions are achieved by Lane *et al.* [17], who compare subject-dependent and subject-independent (respectively named *isolated* and *single*) models with another model called *multi-naive*. In this case, subject-dependent approach outperformed the other two approaches as the amount of the available data increases.

Chen and Shen [44] compared the subject-independent, subject-dependent, and hybrid (respectively called *rest-to-one*, *one-to-one*, and *all-to-one*) models, and once again the subject-dependent model outperforms the subject-independent model, whereas the hybrid model achieves the best performance. The authors also classify subject-independent and hybrid models as *generalized* models, while the subject-dependent model falls into the category of the *personalized* models.

Same results have been achieved by Vaizman *et al.* [45], who compared the subject-independent, subject-dependent, and hybrid (respectively called *universal*, *individual*, and *adapted*) models. Furthermore, they introduced context-based information by exploiting many sensors, such as, location, audio, and phone-state sensors.

Yu *et al.* [46] exploited the hybrid model and compare it to a new model called *incremental hybrid model*. The latter is trained first with the subject-independent approach and then it is incrementally updated based on personal data from a specific user. The difference from the hybrid is that the incremental hybrid model gives more weights to personal data during training. Similarly, Siirtola *et al.* [47] proposed an incremental learning method. The method initially uses a subject-independent model, which is updated with a 2-steps feature extraction method from the test subject data. Afterwards, the same authors proposed a 4 steps subject-dependent model [48]. The proposed method initially uses a subject-independent model, collects and labels the data from the user based on the subject-independent model, trains a subject-dependent model on the collected and labelled data, and classifies activity based on the subject-dependent model. Vo *et al.* [49] exploited a similar approach. The proposed approach first trains a subject-dependent model from data of subject *A*. The model of subject *A* is then transferred to subject *B*. Then, the unlabelled samples of subject *B* are classified to the model of subject *A*. These data are finally used to adjust model for the subject *B*.

The results described above confirm that the performance of the classifiers increases as the training data include the data of the final user. This means that the classifier achieves good performance both preserving the collected data of all users and also by introducing a small amount of data belonging to the user under test.

Alternative solutions consider the similarity between users as crucial factor for obtaining a classification model able to adapt to new situations.

In these direction, Sztyler and Stuckenschmidt [10] and Sztyler *et al.* [50] proposed a personalized variant of the hybrid model. The classification model is trained using the data of those users that are similar to the final user based on signal patterns similarity. They found that people with same fitness level also have similar acceleration patterns regarding the running activity, whereas gender and physique could characterize the walking activity. The heterogeneity of the data is not eliminated but it is managed in the classification procedure.

A similar approach is presented by Lane *et al.* [17]. The proposed approach consists in exploiting the similarity between users in order to weight the collected data. The similarities are calculated based on signal pattern data, or on physical data (e.g., age and height), or on lifestyle index. The value of similarity is used as weight. The higher the weight, the more similar two users are and the more that signals from those users is used for classification.

Garcia-Ceja and Brena [51], [52] exploited inter-class similarity instead of the similarity between subjects (called inter-user similarity) presented by Lane *et al.* [17]. The final model is trained using only the instances that are similar to the target user for each class.

Hong *et al.* [14] proposed a different solution where the generalization is obtained by a combination of activity recognition models (trained by a subject-dependent approach). This combination permits to achieve better activity recognition performance for the final user.

Reiss and Stricker [53] proposed a model that consists of a set of weighted classifiers (experts). Initially all the weights have the same values. The classifiers are adapted to a new user by considering a new set of suitable weights that better fit the labeled data of the new user.

In this paper we propose a model that extends and combines the different models above described. It is based on the similarity between users but it is more flexible because, by changing a single parameter, we can decide which approach will be implemented among subject-dependent, subject-independent, and hybrid. Furthermore, we choose Adaboost which is a more flexible classifier because it combines different weak classifiers, which can ideally recall the proposal of Hong *et al.* [14] and Reiss and Stricker [53].

III. PROPOSED PERSONALIZATION METHODS

In the previous section we discussed literature methods that exploit information about the user under test to improve the accuracy of recognition algorithms. More performing models include personalization prospective both in term of physical characteristics and in term of combination of different classifiers.

In this work we propose personalization models based on similarity between users in term of physical attributes and/or signals patterns. Personalization models are used to weight users training data of the classifier, that in our case is the Adaboost classifier. We demonstrate that a classifier trained on data personalized in this way is more powerful, in terms of recognition accuracy, with respect to a classifier trained without personalization.
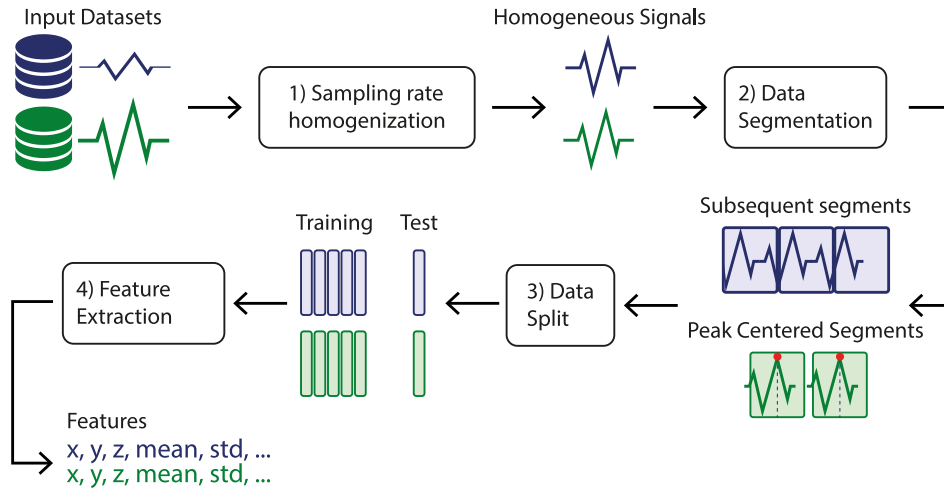
FIGURE 2. Data preparation and feature extraction pipeline.

The rationale behind similarity-based personalization is based on two intuitions:

1) Users with different physical characteristics, such as age or weight, walk or run in a different way. This results in a different accelerometer signal. Focusing the training data on those users that are more similar to the user under test may help to increase recognition capabilities of the classifier. We refer to this aspect as *physical-based similarity*.
2) Independently from similarities based on physical characteristics, accelerometer signals from two different users may be more similar with respect to other users performing the same activity. We refer to this aspect as *signal-based (or sensor-based) similarity*.

Before entering into the core of the methods we propose, we introduce a method of data split for training and testing data that we consider an indispensable step in order to reliably validate the methods of personalization under test. Figure 2 shows the steps of the method that are described in the following.

A. DATA PREPARATION

For validating personalization models we selected three different datasets: UniMiB-SHAR [54], MobiAct [55], and Motion Sense [56]. Each signal of the datasets is composed of three accelerometer components along the x, y, and z axis. Since public available datasets are usually acquired by different research groups using different devices and protocols, it is common to have non-homogeneous characteristics in terms of data collection, such as position of the sensor, sampling frequency of the sensor, number of participants, activities not performed by all users, etc. This inhomogeneity leads sometimes to not-comparable results among different datasets.

Machine learning methods take a segment made of N subsequent accelerometer samples as input. It means that the original accelerometer signals need to be segmented before being fed into a classifier.

Another important point concerning data preparation is represented by the split between training and test set. In literature there are two very common procedures, that are the k-cross and leave-one-subject-out validation. Those are mutually exclusive. As described later on, for our scope, we had to employ the k-cross validation on the top of the leave-one-subject-out one.

Defining a common protocol for data preparation is necessary in order to manage different datasets and it is important for the reproducibility of the experiments.

In the following we describe the steps of the data preparation protocol that are pictured in Figure 2.

1) SAMPLING RATE HOMOGENIZATION

Sensors acquire data at a given sampling frequency, with a given range of intensity values, and so on. Each manufacturer designs its own sensor with operating specifications that may be different from the typical ones. Moreover, sensors acquire data at not constant sampling frequency. The three datasets we considered have different sampling rates. Since machine learning methods require input data in a given format (e.g., number of samples per second and intensity range), that is, consistent over time, we had to pre-process data from the datasets in order to make them homogeneous in terms of sampling rate. To this end, we have chosen the lowest sampling rate among all the sampling rates of the datasets, then we have subsampled all the signal to fit that frequency. The sampling rate chosen is 50Hz. Literature suggests that about 50Hz is a suitable sampling rate that permits to model human activities [57], thus our choice does not negatively affect the results. This step is represented by the action number 1 in Figure 2.

2) DATA SEGMENTATION

It is a common practice to divide the original signal data in segments, or windows, which contain a certain number

of samples taken from the original accelerometer signal at a given frequency rate. The length of each segment is a parameter of the classifier and must be compatible with the temporal duration of the activities. In other words, the segment should contain at least one occurrence of any activities included in the dataset list. Otherwise, signals that contain incomplete portions of activity may be erroneously classified. Usually, the slowest activity is walking and it is common practice to consider 2.5 seconds as the minimum temporal interval to observe two human steps. It means that, a 2.5 seconds segment, at a sampling frequency of 50Hz, contains 125 samples.

Once the length of each segment has been fixed, there are two possible ways for segmenting data:

- *Subsequent (overlapped) segments*. Each signal is divided into subsequent windows of a given size and each of them can be overlapped with the previous and the next one. The overlapping percentage is a parameter of the algorithm. In this paper we considered windows of 5 seconds. We computed the analysis with no overlapping and with 50% of overlapping and compare the results [54].
- *Segments centered on the peak of the signal*. Each signal is divided into subsequent windows but only if the intensity of the recorded signal is higher than a given threshold. Usually the threshold is $2g$, where g is the gravitational constant. When a value v exceeds the threshold, a segment is taken around the temporal position of the peak value v.

Data segmentation allows to augment data and consequently to reduce the overfitting of machine learning algorithms that is most of the time caused by a low amount of training data. In Figure 2 we show the data segmentation (action number 2) with a dataset with subsequent segments (the blue dataset) and another dataset that exploits peak centered segments (the green dataset).

3) DATA SPLIT PROCEDURE

A human activity dataset is composed of inertial signals recorded through a smartphone worn by human subjects that during experimental sessions performed a certain number of activities. Ideally, there is a list of activities to be performed by all the subjects. One of the recurring problems with the literature datasets is that not all the subjects performed all the activities that are in the dataset list.

The k-fold cross validation separates the dataset into k groups which, alternately, compose the training and the test sets disregarding if the segments of a given subject is either in the training and in the test split (shown in Figure 2 as action number 3). This behavior is not suitable for investigating personalization models, because we have to be sure that data from the user under test are or not within the training set. To this end, the subject-out-validation strategy is a more suitable way to split training and test data. This strategy considers training data made of segments from all the subjects but the subject under test.

To better explore personalization methods, we also need to ensure that the training and test splits are composed by the same list of activities and more important that all the users of dataset performed the same list of activities. This is not always true, because, as discussed above, it may happen that not all the subjects performed all the activities of the list. To this scope, we removed those subjects from the dataset that did not perform all the activities included in the dataset list or alternatively, if this leads to a huge lost of subjects, we removed from the dataset those activities not performed by all subjects.

In this paper we explored personalization methods on the top of three different data splits: *subject-dependent*, *subject-independent*, and *hybrid* (see Figure 1). The subject-dependent split considers the training and test sets made of only signals from the subject under test. The subject-independent split considers a test set made of only signals from the subject i, and a training set made of signals from all the subjects but i. The hybrid split is a subject-independent split with the injection of a small amount of signals of the subject i within the training set. In order to realize the hybrid approach we need that some data from the subject under test is part of the training data. At the same time, to make the hybrid approach comparable with the others, we need the test set is always the same whatever is the data-split adopted. To ensure that, the data from each subject is divided using the k-fold cross validation strategy with the constraint that each fold contains the same number of activities. Given a subject under test, one of the k folds is taken as test set for all the three data splits: *subject-dependent*, *subject-independent*, and *hybrid*. The remaining folds are used as training data of the *subject-dependent* data split, and also in combination with the data from other subjects in the *hybrid* data split. The training data of the *subject-independent* is made of all the $k - 1$ folds of each subjects that are not included in the corresponding test sets.

4) FEATURE EXTRACTION

In literature it is shown that using a set of features instead of raw data improves the classification accuracy [27]. Furthermore, features extraction reduces the data dimensionality while extracting the most important peculiarity of the signal. Each accelerometer signal of the datasets is composed of three accelerometer components along the x, y, and z axis. An entire segment is as follows:

$$
\begin{aligned}
(\underbrace{acc_{x_1},\ acc_{x_2}\ acc_{x_3}\ \ldots\ acc_{x_n}}_{x-dimension\ acceleration}\ \cdots \\
\underbrace{acc_{y_1}\ acc_{y_2}\ acc_{y_3}\ \ldots\ acc_{y_n}}_{y-dimension\ acceleration}\ \cdots \\
\underbrace{acc_{z_1}\ acc_{z_2}\ acc_{z_3}\ \ldots\ acc_{z_n}}_{z-dimension\ acceleration})
\end{aligned}
\tag{1}
$$

where $n = s \cdot f$, and f is the sampling rate.

We considered a vector of hand-crafted time and frequency domain features calculated on each segment and for each

TABLE 1. 18 hand-crafted features.

Features	
Minimum	$min = \min_{j=1,...n}(x_j)$
Maximum	$max = \max_{j=1,...n}(x_j)$
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Median	$Me = x_{0.5} : F(x_{0.5}) = 0.5$
Standard Deviation (SD)	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Interquartile Difference	$ID = x_{0.75} - x_{0.25}$
Skewness	$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$
Kurtosis	$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{s^4}$
Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
Total Sum	$TS = \sum_{i=1}^{n} x_i$
Range	$R = max - min$
Mean of Peak's Distance	$m_p = \frac{1}{s^2}\sum_{j=1}^{s}\sum_{i=1}^{s} d(p_i, p_j)$
Entropy	$H(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$
Sum of the spectral power components	$SP = \frac{1}{n}\sum_{j=1}^{f} \|FFT_j\|^2$
Mean of the spectral components	$\mu_f = \frac{1}{n}\sum_{j=1}^{n} FFT_j$
Median of the spectral components	$Me_f = FFT_{0.5} : F(f_{0.5}) = 0.5$
First Cepstral Coefficient	$c(1) = \mathcal{F}^{-1}\{\log \|A(f)\|\}$, where $A(f)$ is the Fourier Transformation

accelerometer direction. Table 1 presents the features we considered [27], [58]–[61]. The resulting feature vector is obtained by concatenating the feature extracted from the component x, y, and z of the accelerometer signal. This is presented in Figure 2 in action number 4 as the last step of the data preparation protocol.

B. SUBJECT SIMILARITY

In this paper we define a new personalization model including and generalizing the different models presented in the-state-of-the-art. The general idea is that people with different physical aspects, life style, or habits may walk, run, etc., in a different way and that accelerometer signals related to the same activity, disregarding the physical similarities between subjects, may have common characteristics [17].

To take into account such a diversity, we introduce the concept of similarity between subjects and then we exploit the similarity to weight the training data in order to give more importance to data that are more similar to data of the user under test.

Each subject i can be described with a feature vector $\mathbf{g}_i = \{g_1, \ldots, g_K\}$. Similarity between two subjects i and j is defined as follows:

$$\text{sim}(i, j) = e^{-\gamma d(i,j)} \quad (2)$$

where γ is a scale parameter and $d(i, j)$ is the Euclidean distance between the feature vector of two subjects:

$$d(i, j) = \sqrt{\sum_{k=1}^{K}(g_{k,i} - g_{k,j})^2} \quad (3)$$

The resulting similarity value ranges from 0 to 1: 0 means that the two subjects are dissimilar, and 1 means that the two subjects are equal. The idea is to take advantage of the
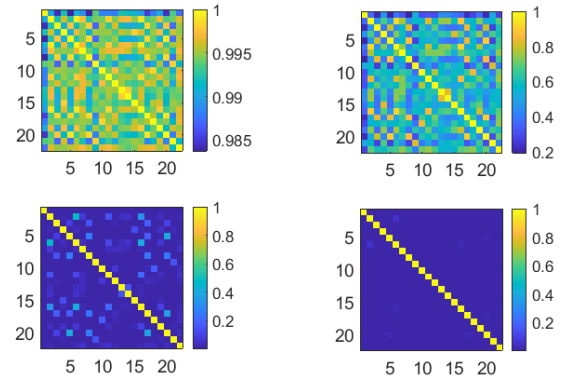

FIGURE 3. Physical similarity matrix for different values of $\gamma = 0.01, \ 1, \ 10, \ 40$ (clock-wise order).

similarity between subjects as follows. Given a subject i under test, all the training data are weighted by using the similarity between the user i and the rest of the users. We can define three types of similarity: *physical-based* ($\text{sim}^{physical}$), *sensor-based* (sim^{sensor}), and *physical combined with sensor-based similarity* ($\text{sim}^{physical+sensor}$).

Figure 3 shows the training data which are weighted according to the similarity. Each sample is so considered with different importance in the classification with respect to the similarity between the subject in the training and in the test set.

1) SIMILARITY BASED ON PHYSICAL CHARACTERISTICS

For each subject we define a feature vector that is made of three real values $\mathbf{g}^{physical} = (age, weight, height) = (g_1^p, g_2^p, g_3^p)$. Each component of the triplet ranges from 0 to 1 because all the ages, weights, and heights of the subjects have been normalized to fit the range of real number $[0 - 1]$. The choice of these characteristics is inspired by the literature and it is subject to the availability of the metadata of the public datasets. We decided to not consider lifestyle of the subjects as further subject characteristic because this information is usually not available in public datasets.

Figure 3 shows some examples of physical similarity between subjects from the dataset MotionSense. The examples have been obtained with different values of γ ranging from a low to a high value. Each figure is the visual representation of the similarity matrix between all the subjects of the dataset. Clearly the diagonal of the matrix is always 1 (each subject is similar to him/herself). The parameter γ plays a crucial role in the definition of personalization models. The parameter γ determines the shape of the exponential function: higher values of γ correspond to more separation between subjects. With $\gamma = 0$ all the subjects has similarity 1.

2) SIMILARITY BASED ON SIGNAL DISTANCE

For each subject we define a feature vector that is made of 18 real values described in Table 1: $\mathbf{g}^{sensor} = (g_1^s, \ldots, g_{18}^s)$. Each subject i has N_i segments. We calculate the similarity between 2 subjects i and j by summing up the similarity

between each segment of the subjects:

$$\text{sim}^{sensor}(i,j) = \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \text{sim}(x_{in}, x_{jm})$$

$$= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} e^{-\gamma d(x_{in}, x_{jm})} \quad (4)$$

3) SIMILARITY BASED ON PHYSICAL CHARACTERISTICS AND SIGNAL DISTANCE

For each subject we define a similarity with respect to the other subjects that is made of the weighted sum of physical and sensor similarity:

$$\text{sim}^{physical+sensor}(i,j)$$
$$= \alpha \cdot \text{sim}^{sensor}(i,j) + \beta \cdot \text{sim}^{physical}(i,j) \quad (5)$$

where α and β are such that $\alpha + \beta = 1$.

C. PERSONALIZATION MODELS

In order to evaluate the influence of the similarity-based personalization strategies, we have performed 2 group of experiments:

- *Experiments without similarity-based personalization*: This group of experiments ignores the similarity between subjects that is equal to consider for all the types of similarities $\gamma = 0$. For this first group we considered the following dataset splits:
  - subject-dependent model: the training set and the test set are composed only by the data of the test subject. For each subject under test, the procedure is repeated k times according to the k-cross-validation.
  - subject-independent model: we train the classifier using the data of all subjects but the test. For each subject under test, the procedure is repeated k times according to the k-cross-validation.
  - hybrid model: we train the classifier using the data of all subjects and a portion of the data of the test which are the $(k-1)$ splits not used for the test. For each subject under test, the procedure is repeated k times.
- *Experiments with similarity-based personalization*: This group of experiments considers the similarity between users by integrating the sample data with the similarity weights and so the classification is influenced by the similarity between users. For the second group of experiments, we considered the following dataset splits:
  - subject-independent-weighted model: we train the classifier using the data of all subjects but the test as in the first group of experiments. The segments of the training data are weighted with the similarity between the subjects in the training and the subject of the test. For each subject under test, the procedure is repeated k times according to the k-cross-validation.

FIGURE 4. Accuracy as the values of γ increases.

  - hybrid weighted model: we train the classifier using the data of all subjects and a portion of the data of the test (which are the $k-1$ splits not used for the test). The segments of the training data are weighted with the similarity between the subjects in the training and the subject of the test. For each subject under test, the procedure is repeated k times according to the k-cross-validation.
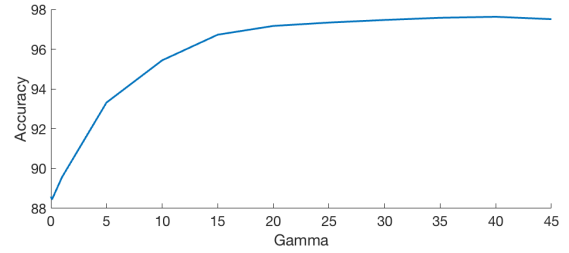
For this group of experiments, it is not possible to employ the subject-dependent split because by definition it does not contain samples from other users and then it is not possible to compute the similarity between different subjects.

D. CLASSIFIER

To evaluate the goodness of the personalization strategies we considered the Adaboost classifier which permits to weight training data before starting the training process [62]. We have also experimented Support Vector Machines and k-Nearest Neighbor. However for these classifiers the adoption of the similarity-based weighting procedure did not lead to remarkable accuracy modifications whatever was the value of γ.

The classification was repeated several times: for each model and for each value of γ. The choice of appropriate values of γ is not trivial. The value of γ is arbitrary in $(0, +\infty)$, zero and infinity excluded.

Whatever is the classifier, the performance are measured using accuracy. 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, accuracy is define as in Equation (6):

$$Acc = \frac{1}{|E|} \sum_{a=1}^{|E|} Acc_a = \frac{1}{|E|} \sum_{a=1}^{|E|} \frac{TP_a}{NP_a}. \quad (6)$$

Acc is the arithmetic average of the accuracy Acc_a of each activity.

A grid searching procedure permitted us to select the best value according with accuracy results. In Figure 4 we observe the accuracy behavior as γ increases. We notice that if the value of γ grows, also the value of the accuracy grows until the value 40, and then starts to decrease. This behavior is exactly what we expected because when $\gamma \rightarrow +\infty$ the number of training data decreases and so the accuracy.

TABLE 2. Statistics of subjects physical characteristics.

	Sex		Age	Weight	Height
Dataset	Male	Female	(years)	(kg)	(cm)
UniMiB-SHAR	6	22	18-60	50-82	160-190
			27 ± 11	$64.4 \pm 9, 7$	169 ± 7
MobiAct	43	14	20-47	50-120	158-193
			25.19 ± 4.45	76.8 ± 14.16	175.73 ± 7.77
Motion Sense	13	9	18-46	48-102	161-190
			28.8 ± 5.46	72.12 ± 16.21	174.2 ± 8.9

If $\gamma = 0$ we also have a decreasing of accuracy because it correspond to a unpersonalized model.

IV. DATASET DESCRIPTION

We experimented three public datasets containing accelerometer signals of Activities of Daily Living (ADLs) and Falls recorded by smartphones.

A. UNIMIB-SHAR

The dataset contains tri-axial acceleration data organized in 3 seconds windows around the peak. Signals of 17 different activities (ADLs and Falls) are collected and performed by 30 subjects. For each of them sex, age, weight, and height are known [54]. The original sampling rate is 50Hz. We have chosen segments of 3 seconds for this dataset.

The subjects placed the smartphone used for the acquisition (a Samsung Galaxy Nexus I9250) half of the times in the left trouser's pocket and the remaining times in the right one. They repeat each activity several times. After having applied the data split procedure described in section III-A3, the number of subjects is 27 and the number of the activity is 13.

B. MOBIACT

This dataset includes tri-axial acceleration data of 15 ADLs and Falls recorded with a Samsung Galaxy S3 and performed by 67 participants. The windows size we considered is of 5 seconds with a sample rate of 87Hz. Additional information on the subjects are sex, age, weight, and height. The smartphone is located with random orientation in a loose pocket chosen by the subject [55]. The original sampling rate is 87Hz. We have chosen segments of 5 seconds for this dataset.

After having applied the data split procedure described in section III-A3, we removed 10 subjects and 2 activities because of missing values.

C. MOTION SENSE

This dataset contains time-series data generated by the accelerometer sensors of an iPhone 6s worn by 24 participants. Each of the subjects performed 6 activities (only ADLs). The smartphone were kept in the participant's front pocket. After having applied the data split procedure described in section III-A3, we removed 2 subjects [56]. The original sampling rate is 50Hz. We have chosen segments of 5 seconds for this dataset.
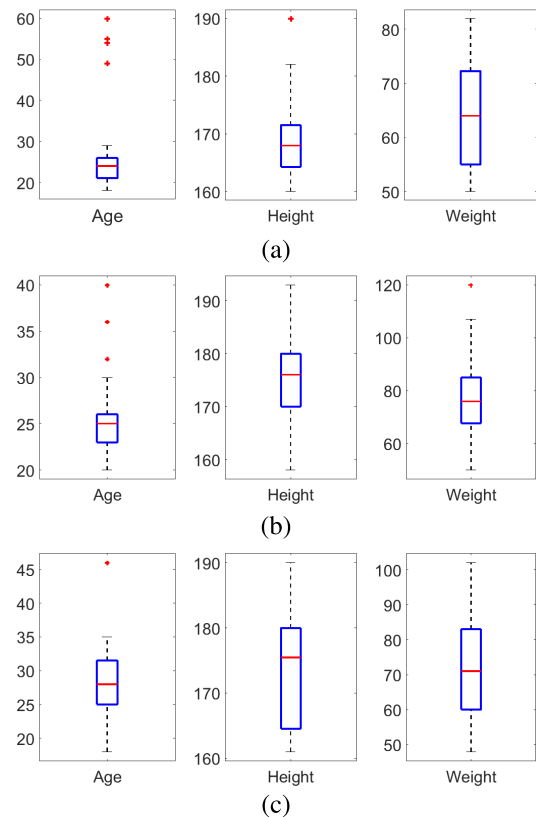

FIGURE 5. Boxplots over age, weight, and height: (a) UniMiB-SHAR; (b) MobiAct; (c) Motion sense.

D. DESCRIPTIVE AND EXPLORATORY ANALYSIS

Table 2 shows the statistics of the considered datasets. The datasets present almost the same distribution of subject characteristics. The box-plots in Figure 5 show more information about dispersion, position, and outliers of the subjects characteristics. There are many outliers for the variable Age which means that there are some person with age out of 1.5 interquartile difference. Despite of this fact, the variability is not low. For height and weight we observe more variability and few outliers.

An other descriptive statistics we performed is the frequency distribution of the activities. In Figure 6 we present activities distribution with respect to the three datasets. As we can see in the graphs, UniMiB-SHAR and MobiAct contain more activities than Motion Sense and, more important, some activities, such as walking, standing, running, and jogging,
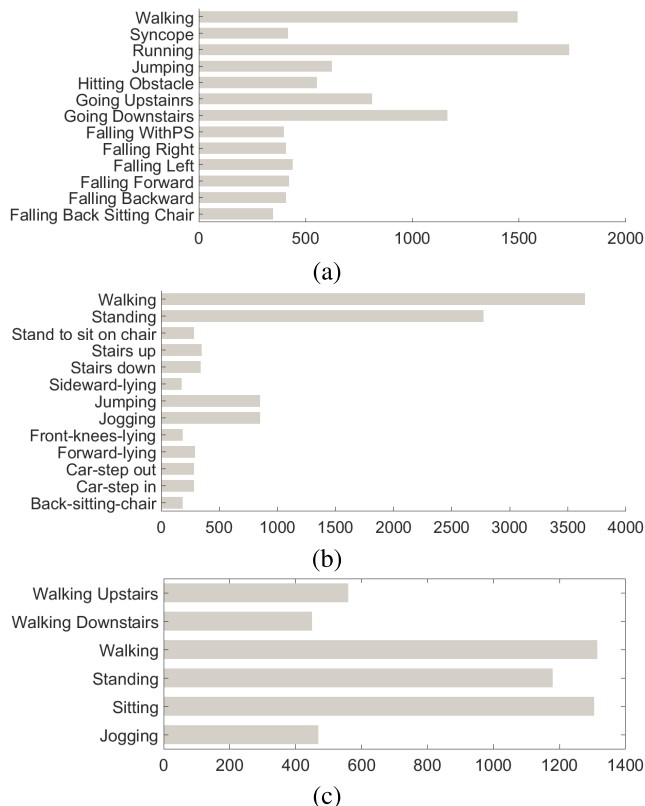
FIGURE 6. Activity sample distribution: (a) UniMiB-SHAR; (b) MobiAct; (c) Motion sense.

FIGURE 7. Multidimensional scaling over physical characteristics. (a) UniMiB-SHAR; (b) MobiAct; (c) Motion sense.

are more frequent than other activities such as stairs up, falling right, and others.

To further motivate our work we performed a multidimensional scaling analysis of the physical characteristics of subjects of each dataset [63]. This analysis may permit to highlight similarity between subjects that share similar physical characteristics. We applied the analysis to a matrix $D^{physical}$ of size $N \times N$ where N is the number of subjects of a given database. The matrix $D^{physical}$ is calculated by calculating the similarity between all the N subjects of the given database using equation (2) mentioned in section III-B.

Figure 7 shows the results of such an analysis mapped into 2 dimensions. Each row of the figure is related to a given dataset. Each point of the graph represents one subject of the dataset and each color shows a given cluster of subjects. For each dataset, in the figures on the left, red stands for young subjects and blue for older subjects, while in the figures on the right, green stands for normal Body Max Index (BMI) subjects and red stands for subjects with an abnormal BMI. The result of this analysis substantially confirms that a kind of implicit separability, on the basis of physical characteristics, between subjects exists.

The results of the multiscale analysis highlighted that age and physical characteristics are discriminator factors under subjects. In all the cases, age discriminates subjects, while BMI discriminates in two cases over three.
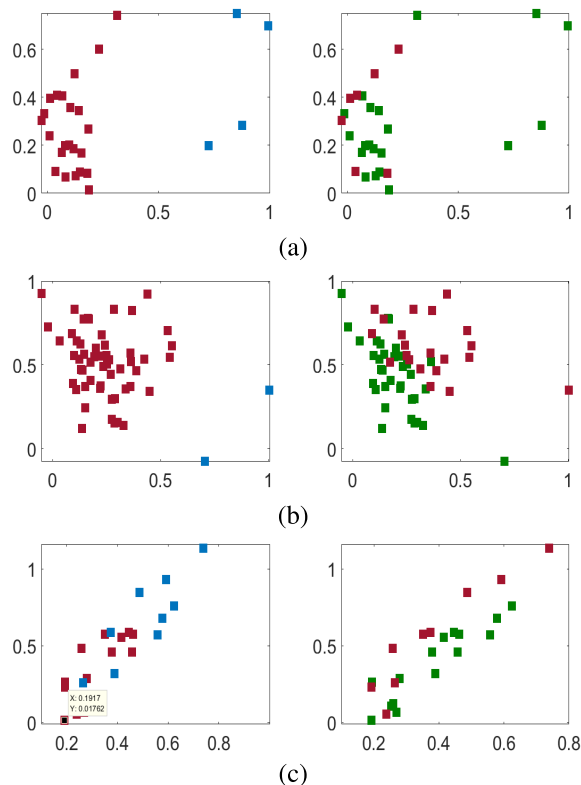
Until now we just explored datasets in terms of physical aspects. Here we want to investigate if signals present some discriminator tendency. For this reason we exploited the Principal Component Analysis (PCA) [64] over the features presented in Table 1. The first and second principal component have been computed for all datasets and are shown in Figure 8. The points on the graphs represent the subjects and the blue color stands for subjects with a $21 < BMI < 28$ and the red color stands for subjects with BMI values outside the interval considered above. PCA analysis shows evidently that subjects are separable on the basis of the features extracted from the segments of the original accelerometer signals.

V. RESULTS

As discussed in section III-C we have performed 2 groups of experiments: 1) *experiments without similarity-based personalization* and 2) *experiments with similarity-based personalization*.

For sake of comparison we have experimented two different classification strategies for each group of experiments: AdaBoost combined with hand crafted (AdaBoost-HC), see Table 1, and AdaBoost combined with deep features (AdaBoost-CNN). We have used the Matlab version of AdaBoost (MathWorks). To further evaluate the goodness of deep features, we experimented the use of Support Vector Machines (SVM) combined with deep features (SVM-CNN)
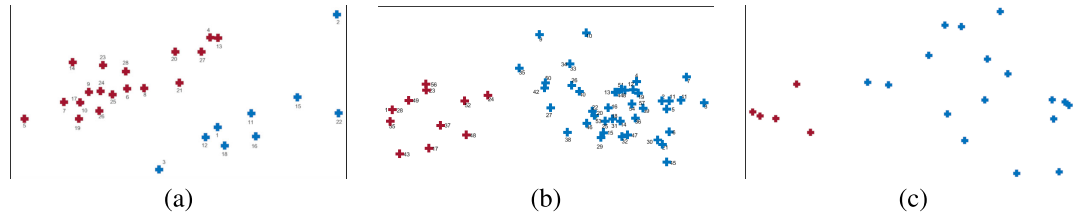
FIGURE 8. PCA using 18 features. (a) UniMiB-SHAR; (b) MobiAct; (c) Motion sense.

in the case of the first group of experiments. We have used the Matlab version of SVM (MathWorks).

Deep features have been achieved by using the Residual Network developed in [27]. Table 3 details the network architecture proposed for this study. The input size of the network is $1 \times 128 \times 3$, that corresponds to 3 segments along the three axes x, y, and z. The network architecture is made of an initial convolutional block, 3 residual stages, each containing a variable number n of residual blocks, average pooling layer, fully connected layer, and softmax layer. A convolutional block is made of three layers: convolutional, batch normalization, and ReLu. A residual block is made of 2 subsequent convolutional blocks and an addition operator that sums the input of the residual block with the output of the residual block itself. Each convolutional layer is $1 \times 3 \times f_{maps}$, where f_{maps} is the number of feature maps of the filter. The best values for n and f_{maps} have been found by following a grid search approach: n ranged between 3 and 21, while f_{maps} ranged between 10 and 200.

The network has been trained on the UCI-HAR [65] dataset and then used as feature extractor. In particular, we have taken the last pooling layer before the last fully connected layer thus obtaining a 1024-dimensional feature vector.

A. EXPERIMENTS WITHOUT SIMILARITY

For the first group of experiments, we have randomly initialized the weights of the algorithm while for the second group, we have weighted all the data belonging to a given subject with the corresponding weights given by the similarity between the given subject and the test subject.

Table 4 reports the results achieved by using the Adaboost classifier for both the groups of experiments. The second group of experiments is highlighted using a light gray color while the remaining are numbers related to the first group of experiments. On average, looking at the fourth column of the first group of experiments, it is quite clear that the hybrid strategy works better than the subject-independent one, the improvement is of about 3% (see also Table 5). This behavior was some how expected, because the hybrid strategy considers the training set containing a small amount of data that belongs to the subjects under test. The presence of data of subject under test permits the AdaBoost to better specialize on the subject under test.

Moreover, apart from UniMiB-SHAR, the subject-dependent strategy achieves an average accuracy of

TABLE 3. Residual network architecture.

Layer name	Shape
conv1	$\{1 \times 3\} \times n$
conv2_n	$\{1 \times 3 \times f_{maps}\} \times n$
conv3_n	$\{1 \times 3 \times 2f_{maps}\} \times n$
conv4_n	$\{1 \times 3 \times 4f_{maps}\} \times n$
avg_pool_x	1×32
fully conn.	$(1 \times 4f_{maps}) \times 15$
softmax	1×15

about 35% lower than both subject-independent and hybrid strategies. In the case of UniMiB-SHAR, this is not true because the dataset is made of segments taken around peaks of the accelerometer signal while the other datasets are made of segments taken subsequently with a zero or 50% of overlap. In case of walking, there is an high probability that in a segment of 3 seconds there are many peaks (higher than 2g). Let us suppose that the number of peaks is 5, it means that for a segment of 3 seconds we take 5 segments, that are quite similar, for classification. The resulting dataset contains a large amount of redundant segments. A subject-dependent strategy takes advantage of this redundancy and specializes very well the classifier especially when the training set is made of only data from the subject under test. In the case of MobiAct and Motion Sense datasets, this is not true because we take a segment of 5 seconds that shares 50% of its length with the previous and subsequent segments.

The comparison between AdaBoost-HC, AdaBoost-CNN, and SVM-CNN is showed in Table 5. Results are averaged across datasets. Numbers show that overall, the use of hand-crafted features outperform the use of CNN features. This behavior is explainable by the fact that the hand-crafted feature vector is of length 21 while the CNN is of length 1048. This behavior is also confirmed by the results of the SVM-CNN approach that, apart from the subject-dependent case, are quite similar to the results achieved by the AdaBoost-CNN approach.

B. EXPERIMENTS WITH SIMILARITY

The three similarity-based personalizations are then computed on the basis of physical characteristics, signal characteristics, and physical combined with signal characteristics. These personalizations have been applied to both subject-independent and hybrid strategies. The corresponding results

TABLE 4. % Accuracy of the first and second (light grey) of experiments achieved using AdaBoost-HC.

	UniMiB-SHAR	MobiAct	Motion Sense	AVG
subject-dependent	84.79	45.57	43.55	57.97

	UniMiB-SHAR Δ max % $\sim$ 24	MobiAct Δ max % $\sim$ 7	Motion Sense Δ max % $\sim$ 4	AVG Δ max % $\sim$ 14
subject-independent	56.80	81.29	72.48	70.19
hybrid	61.66	83.73	73.82	73.07
subject-independent- Physical Similarity	57.39	81.62	72.45	70.49
subject-independent- Sensor Similarity	57.00	82.45	74.03	71.16
subject-independent- Physical Sensor Similarity	56.93	82.64	73.85	71.14
hybrid- Physical Similarity	**85.44**	89.43	77.76	84.21
hybrid- Sensor Similarity	84.71	90.76	**78.06**	**85.51**
hybrid- Physical Sensor Similarity	84.87	**90.90**	77.86	84.53

TABLE 5. % Accuracy of the first and second (light grey) of experiments using AdaBoost-HC, AdaBoost-CNN and SVM-CNN.

	AdaBoost+HC Δ max % $\sim$ 14	AdaBoost+CNN Δ max % $\sim$ 30	SVM+CNN —	AVG Δ max % $\sim$ 24
subject-dependent	57.97	62.32	86.01	68.76
subject-independent	70.19	60.61	60.60	63.80
hybrid	73.07	72.36	70.23	71.89
subject-independent- Similarity	70.93	61.67	-	66.30
hybrid- Similarity	**84.75**	**90.03**	-	**87.39**

TABLE 6. % Average of accuracy achieved on the subject-dependent split and on subject-independent and hybrid combined with similarity-based personalization methods.

	no-similarity	similarity	Δ%
subject-dependent	60.15	-	-
subject-independent	65.40	66.30	0.9
hybrid	72.72	87.39	14.7

for AdaBoost-HC are highlighted in Table 4 with a light grey color. On average, looking at the fourth column of the group 2 of experiments, it is quite clear that the similarity-based personalizations lead to a considerable improvement only in the case of the hybrid strategy. In the case of UniMiB-SHAR, the maximum improvement that we achieved is of about 0.5% and 24% for subject-independent and hybrid strategy respectively. In the case of MobiAct, the maximum improvement is of about 1.5% and 7% for subject-independent and hybrid strategy respectively. In the case of Motion Sense, it is of about 1.5% and 4% for subject-independent and hybrid strategy respectively. In Table 5 results achieved by the AdaBoost-CNN approach confirm that the use of similarity increase the performance with a highest improvement of about 30%. Across datasets, on average, similarity-based personalizations lead to an improvement of performance of about 0.9% and 14.7% for subject-independent and hybrid strategy respectively (see also Table 6).

The fact that similarity-based personalizations combined with the hybrid strategy work better than personalizations combined with the subject-independent strategy is not surprising. Similarities between subjects are used to weight the data of the training set. In practice data belonging to more similar subjects are more important than data belonging to less similar subjects.

Among similarity-based personalizations, differences are few. It is clear from numbers that physical, signal, and physical + signal-based similarities lead to almost the same improvement of accuracy.

VI. CONCLUSION AND FUTURE WORK

Recently, a significant amount of literature concerning machine learning techniques has focused on automatic human activity recognition (HAR) by using accelerometer recorded by smartphones. Real-world HAR systems may achieve not satisfying recognition accuracy in real world applications because HAR techniques may struggle to generalize to new users and/or new environments. Several factors may affect the accuracy of activity recognition methods: i) position of the device; ii) differences between different brands of sensors; iii) human characteristics. While factors related to the position and the characteristics of the devices have been largely investigated, few works have explored the effects of human characteristics on recognition accuracy.

In this paper we have experimented several personalization methods on three public datasets in order to make the results reproducible and thus allowing future research on this topic. The personalization methods experimented are based on the concept of similarity between users. This means that users may have similar physical characteristics or have similar accelerometer signals and that, such a similarity can be employed to weight training data in a way that data belonging to more similar subjects to the subject under test count more than data of less similar subjects.

We have combined personalization methods with suitable splits of the data: subject-independent and hybrid. The first split considers training set made of data from all the subjects but the subject under test, while the second split considers a training set made of data from all the subjects but the user under test and a small amount of data of the user under test. Experiments, on average, prove that personalization methods improve accuracy of the classifier only if combined with a hybrid split. In this case the increment of accuracy, on average, is of about 11%.

This paper confirms that personalization methods can be effective especially if a small amount of subject-dependent data are included in the training set. The way we carried out the experimentation makes it possible to reproduce the results and more important it paves the way for future investigations on this topic.

As future work, we are considering to include also the gyroscope in the analysis and more important we are considering to prove personalization methods on other publicly available datasets in order to further confirm results achieved here. To this end we are trying to exploit the platform to collect, unify, and distribute inertial labeled signals for human activity recognition that we are building [66], [67].

REFERENCES

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors.," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[2] G. Grossi, R. Lanzarotti, P. Napoletano, N. Noceti, and F. Odone, "Positive technology for elderly well-being: A review," *Pattern Recognit. Lett.*, to be published.

[3] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, Mar. 2011.

[4] M. Habib, M. Mohktar, S. Kamaruzzaman, K. Lim, T. Pin, and F. Ibrahim, "Smartphone-based solutions for fall detection and prevention: Challenges and open issues," *Sensors*, vol. 14, no. 4, pp. 7181–7208, Apr. 2014.

[5] C. Medrano, R. Igual, I. Plaza, and M. Castro, "Detecting falls as novelties in acceleration patterns acquired with smartphones," *PLoS ONE*, vol. 9, no. 4, Apr. 2014, Art. no. e94811.

[6] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, "A survey of online activity recognition using mobile phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, Jan. 2015.

[7] D. Micucci, M. Mobilio, P. Napoletano, and F. Tisato, "Falls as anomalies? An experimental evaluation using smartphone accelerometer data," *J. Ambient Intell. Humanized Comput.*, vol. 8, no. 1, pp. 87–99, Feb. 2017.

[8] S. Kaghyan and H. Sarukhanyan, "Accelerometer and GPS sensor combination based system for human activity recognition," in *Proc. Comput. Sci. Inf. Technol. (CSIT)*, 2013, pp. 1–9.

[9] E. Rovini, C. Maremmani, and F. Cavallo, "How wearable sensors can support parkinson's disease diagnosis and treatment: A systematic review," *Frontiers Neurosci.*, vol. 11, p. 555, Oct. 2017.

[10] T. Sztyler and H. Stuckenschmidt, "Online personalization of cross-subjects based activity recognition models on wearable devices," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, Mar. 2017, pp. 180–189.

[11] J. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2015, pp. 1–7.

[12] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.

[13] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.

[14] J.-H. Hong, J. Ramos, and A. K. Dey, "Toward personalized activity recognition systems with a semipopulation approach," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 1, pp. 101–112, Feb. 2016.

[15] R. Igual, C. Medrano, and I. Plaza, "A comparison of public datasets for acceleration-based fall detection," *Med. Eng. Phys.*, vol. 37, no. 9, pp. 870–878, Sep. 2015.

[16] F. Bagala, C. Becker, A. Cappello, L. Chiari, K. Aminian, J. M. Hausdorff, W. Zijlstra, and J. Klenk, "Evaluation of accelerometer-based fall detection algorithms on real-world falls," *PLoS ONE*, vol. 7, no. 5, May 2012, Art. no. e37062.

[17] N. D. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A. T. Campbell, and F. Zhao, "Enabling large-scale human activity inference on smartphones using community similarity networks (CSN)," in *Proc. 13th Int. Conf. Ubiquitous Comput. (UbiComp)*, 2011, pp. 355–364.

[18] G. M. Weiss and J. W. Lockhart, "The impact of personalization on smartphone-based activity recognition," in *Proc. AAAI Workshop Activity Context Represent., Techn. Lang.*, 2012, pp. 1–7.

[19] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 790–808, Nov. 2012.

[20] L. Chen and C. D. Nugent, "Sensor-based activity recognition review," in *Human Activity Recognition and Behaviour Analysis: For Cyber-Physical Systems in Smart Environments*. Cham, Switzerland: Springer, 2019, pp. 23–47.

[21] G. Azkune and A. Almeida, "A scalable hybrid activity recognition approach for intelligent environments," *IEEE Access*, vol. 6, pp. 41745–41759, 2018.

[22] D. Riboni and C. Bettini, "COSAR: Hybrid reasoning for context-aware activity recognition," *Pers. Ubiquitous Comput.*, vol. 15, no. 3, pp. 271–289, Mar. 2011.

[23] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, Jun. 2012.

[24] J. Ye, G. Stevenson, and S. Dobson, "USMART: An unsupervised semantic mining activity recognition technique," *TiiSACM Trans. Interact. Intell. Syst.*, vol. 4, no. 4, pp. 1–27, Nov. 2014.

[25] G. Civitarese, C. Bettini, T. Sztyler, D. Riboni, and H. Stuckenschmidt, "NewNECTAR: Collaborative active learning for knowledge-based probabilistic activity recognition," *Pervas. Mobile Comput.*, vol. 56, pp. 88–105, Mar. 2019.

[26] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64270–64277, 2018.

[27] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Hand-crafted features vs residual networks for human activities recognition using accelerometer," in *Proc. IEEE 23rd Int. Symp. Consum. Technol. (ISCT)*, Jun. 2019, pp. 153–156.

[28] X. Tao, *Wearable Electronics and Photonics*. Amsterdam, The Netherlands: Elsevier, 2005.

[29] C. Ho, A. Robinson, D. Miller, and M. Davis, "Overview of sensors and needs for environmental monitoring," *Sensors*, vol. 5, no. 1, pp. 4–37, Oct. 2008.

[30] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Human activities recognition using accelerometer and gyroscope," in *Proc. Eur. Conf. Ambient Intell. (AMI)*, 2019, pp. 357–362.

[31] S. R. Shakya, C. Zhang, and Z. Zhou, "Comparative study of machine learning and deep learning architecture for human activity recognition using accelerometer data," *Int. J. Mach. Learn. Comput.*, vol. 8, pp. 577–582, Nov. 2018.

[32] C. Torres-Huitzil and M. Nuno-Maganda, "Robust smartphone-based human activity recognition using a tri-axial accelerometer," in *Proc. IEEE 6th Latin Amer. Symp. Circuits Syst. (LASCAS)*, Feb. 2015, pp. 1–4.

[33] R. Khusainov, D. Azzi, I. Achumba, and S. Bersch, "Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations," *Sensors*, vol. 13, no. 10, pp. 12852–12902, Sep. 2013.

[34] M. A. Labrador and O. D. L. Yejas, *Human Activity Recognition: Using Wearable Sensors Smartphones*. Boca Raton, FL, USA: CRC Press, 2013.

[35] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "Smartwatch-based sitting detection with human activity recognition for office workers syndrome," in *Proc. Int. ECTI Northern Sect. Conf. Electr., Electron., Comput. Telecommun. Eng. (ECTI-NCON)*, Feb. 2018, pp. 160–164.

[36] S. K. Polu, "Human activity recognition on smartphones using machine learning algorithms," *Int. J. Innov. Res. Sci. Technol.*, vol. 5, no. 6, pp. 31–37, 2018.

[37] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Gener. Comput. Syst.*, vol. 81, pp. 307–313, Apr. 2018.

[38] A. Zunino, J. Cavazza, and V. Murino, "Revisiting human action recognition: Personalization vs. generalization," in *Proc. Int. Conf. Image Anal. Process. (ICAR)*, 2017, pp. 469–480.

[39] M. Berchtold, M. Budde, H. R. Schmidtke, and M. Beigl, "An extensible modular recognition concept that makes activity recognition practical," in *Proc. Annu. Conf. Artif. Intell. (AAAI)*, 2010, pp. 400–409.

[40] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor," in *Proc. 11th IEEE Int. Symp. Wearable Comput.*, Oct. 2007, pp. 37–40.

[41] C. Shen, Y. Chen, and G. Yang, "On motion-sensor behavior analysis for human-activity recognition via smartphones," in *Proc. IEEE Int. Conf. Identity, Secur. Behavior Anal. (ISBA)*, Feb. 2016, pp. 1–6.

[42] D. Lara, A. J. Pérez, M. A. Labrador, and J. D. Posada, "Centinela: A human activity recognition system based on acceleration and vital sign data," *Pervasive Mobile Comput.*, vol. 8, no. 5, pp. 717–729, Oct. 2012.

[43] J. W. Lockhart and G. M. Weiss, "The benefits of personalized smartphone-based activity recognition models," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2014, pp. 614–622.

[44] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[45] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE Pervas. Comput.*, vol. 16, no. 4, pp. 62–74, Oct. 2017.

[46] T. Yu, Y. Zhuang, O. Mengshoel, and O. Yagan, "Hybridizing personal and impersonal machine learning models for activity recognition on mobile devices," in *Proc. 8th EAI Int. Conf. Mobile Comput., Appl. Services*, 2016, pp. 117–126.

[47] P. Siirtola, H. Koskimäki, and J. Röning, "Personalizing human activity recognition models using incremental learning," 2019, *arXiv:1905.12628*. [Online]. Available: http://arxiv.org/abs/1905.12628

[48] P. Siirtola, H. Koskimäki, and J. Röning, "From user-independent to personal human activity recognition models exploiting the sensors of a smartphone," 2019, *arXiv:1905.12285*. [Online]. Available: http://arxiv.org/abs/1905.12285

[49] Q. V. Vo, M. T. Hoang, and D. Choi, "Personalization in mobile activity recognition system using K-medoids clustering algorithm," *Int. J. Distrib. Sensor Netw.*, vol. 9, no. 7, Jul. 2013, Art. no. 315841.

[50] T. Sztyler, H. Stuckenschmidt, and W. Petrich, "Position-aware activity recognition with wearable devices," *Pervasive Mobile Comput.*, vol. 38, pp. 281–295, Jul. 2017.

[51] E. Garcia-Ceja and R. Brena, "Building personalized activity recognition models with scarce labeled data based on class similarities," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell. (UCAmI)*, 2015, pp. 265–276.

[52] E. Garcia-Ceja and R. Brena, "Activity recognition using community data to complement small amounts of labeled instances," *Sensors*, vol. 16, no. 6, p. 877, Jun. 2016.

[53] A. Reiss and D. Stricker, "Personalized mobile physical activity recognition," in *Proc. 17th Annu. Int. Symp. Int. Symp. Wearable Comput. (ISWC)*, 2013, pp. 25–28.

[54] D. Micucci, M. Mobilio, and P. Napoletano, "UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones," *Appl. Sci.*, vol. 7, no. 10, p. 1101, Oct. 2017.

[55] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, and M. Tsiknakis, "The MobiAct dataset: Recognition of activities of daily living using smartphones," in *Proc. Int. Conf. Inf. Commun. Technol. Ageing Well e-Health*, 2016, pp. 143–151.

[56] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Protecting sensory data against sensitive inferences," in *Proc. 1st Workshop Privacy Design Distrib. Syst. (W-P2DS)*, 2018, pp. 1–6.

[57] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. Conf. Innov. Appl. Artif. Intell. (IAAI)*, 2005, pp. 1541–1546.

[58] M. Janidarmian, A. R. Fekr, K. Radecka, and Z. Zilic, "A comprehensive analysis on wearable acceleration sensors in human activity recognition," *Sensors*, vol. 17, no. 3, p. 529, Mar. 2017.

[59] H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell, "The jigsaw continuous sensing engine for mobile phone applications," in *Proc. 8th ACM Conf. Embedded Networked Sensor Syst. (SenSys)*, 2010, pp. 71–84.

[60] S. R. Vanrell, D. H. Milone, and H. L. Rufiner, "Assessment of homomorphic analysis for human activity recognition from acceleration signals," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1001–1010, Jul. 2018.

[61] S. Bianco, P. Napoletano, and R. Schettini, "Multimodal car driver stress recognition," in *Proc. 13th EAI Int. Conf. Pervas. Comput. Technol. Healthcare (PervasiveHealth)*, 2019 pp. 302–307.

[62] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[63] T. F. Cox and M. A. Cox, *Multidimensional Scaling*. Boca Raton, FL, USA: CRC Press, 2000.

[64] I. T. Jolliffe, "Principal component analysis: A beginner's guide. Introduction and application," *Weather*, vol. 45, no. 10, pp. 375–382, 1990.

[65] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN13)*, 2013, pp. 1–6.

[66] A. Ferrari, D. Micucci, M. Marco, and P. Napoletano, "A framework for long-term data collection to support automatic human activity recognition," in *Proc. Intell. Environ. (IE), Workshop Reliable Intell. Environ. (WoRIE)*, 2019, p. 367.

[67] A. Ferrari, M. Mobilio, D. Micucci, and P. Napoletano, "On the homogenization of heterogeneous inertial-based databases for human activity recognition," in *Proc. IEEE World Congr. Services (SERVICES)*, Jul. 2019, pp. 295–300.

ANNA FERRARI received the bachelor's degree in statistics and the bachelor's degree in mathematics from the University of Milano-Bicocca, in 2009 and 2016, respectively, and the M.Sc. degree in statistics from the Humboldt University of Berlin, in 2011. She is currently pursuing the Ph.D. degree in computer science with the University of Milano-Bicocca.

Her research focuses on the application of machine learning and deep learning techniques in the field of human activity recognition. The improvement of such techniques is the core of her work, which finds its application on human healthcare, healthcare monitoring, and personal wellbeing.

DANIELA MICUCCI received the Ph.D. degree in mathematics, statistics, computational sciences, and computer science from the University of Milan, in 2004.

She is currently an Assistant Professor with the University of Milano-Bicocca, where she is also the Head of the Software Architecture Laboratory, Department of Computer Science. Her research interests include software engineering, in particular software architectures, real-time systems, and self-healing and self-repairing systems.

Dr. Micucci is currently active in several European and National projects. She is also regularly involved in the program committees of workshops and conferences in her areas of interest.

MARCO MOBILIO received the master's and Ph.D. degrees in computer science from the University of Milano-Bicocca. He is currently a Postdoctoral Researcher with the University of Milano-Bicocca. His previous works included the definition of software architectures for embedded systems in the ambient intelligence field. He is currently involved in the H2020 EU-funded Project NGPaaS, as a part of the 5G-PPP Phase 2 initiative. In this context, he focuses on the development monitoring and analytics solutions for cloud platforms. His research interests include software engineering, self-healing, and ambient assisted living, specifically in human activity recognition from inertial sensors.

PAOLO NAPOLETANO received the master's degree in telecommunications engineering from the University of Naples Federico II, with a thesis focused on transmission of electromagnetic fields in 2003, and the Ph.D. degree in information engineering from the University of Salerno, Italy, with a thesis focused on computational vision and pattern recognition in 2007. He is currently an Assistant Professor of computer science (tenure track-RTDB) with the Department of Informatics, Systems and Communication, University of Milano-Bicocca. His current research interests focus on signal, image, and video analysis and understanding, multimedia information processing and management, and machine learning for multimodal data classification and understanding.

• • •