Toward an Unsupervised Approach for Daily Gesture Recognition in Assisted Living Applications

Alessandra Moschetti, Laura Fiorini, Dario Esposito, Paolo Dario, *Member, IEEE*, and Filippo Cavallo, *Member, IEEE*

***Abstract*— Activity Recognition is important in assisted living applications to monitor people at home. Over the past, inertial sensors have been used to recognize different activities, spanning from physical activities to eating ones. Over the last years, supervised methods have been widely used, but they require an extensive labeled data set to train the algorithms and this may represent a limitation of concrete approaches. This paper presents a comparison of unsupervised and supervised methods in recognizing nine gestures by means of two inertial sensors placed on the index finger and on the wrist. Three supervised classification techniques, namely random forest, support vector machine, and multilayer perceptron, as well as three unsupervised classification techniques, namely k-Means, hierarchical clustering, and self-organized maps, were compared with the recognition of gestures made by 20 subjects. The obtained results show that the support vector machine classifier provided the best performances (0.94 accuracy) compared with the other supervised algorithms. However, the outcomes show that even in an unsupervised context, the system is able to recognize the gestures with an average accuracy of ∼0.81. The proposed system may be therefore involved in future telecare services that could monitor the activities of daily living, allowing an unsupervised approach that does not require labelled data.**

***Index Terms*— Gesture recognition, unsupervised analysis, wearable sensors, accelerometers.**

I. Introduction

THANKS to recent advances in medicine, life expectancy has grown in the last years and by 2075 people aged 65 and over are expected to account for 34% of the European population [1]. Aging can cause decreases in people's physical and cognitive abilities thus increasing the need for care from nurse practitioners (+94% in 2025) [2] and physician's assistants (+72% in 2025) [3]. Since the majority of elderly people would prefer to stay in their homes as long as possible, and considering the costs of the nursing home care, it is important to prevent the deterioration of health conditions, to support old persons in daily activities, and to delay entry into institutional care facilities [4]. ICT and Robotics technologies, among other vantages, can help to prevent, support, and maintain the independent living of the elderly population [5], by monitoring elderly people at their own place and supporting families and care staff in the delivery of assistance [6]. Particularly, the ability to recognize gestures and daily activities is useful to facilitate caregivers to better provide help in personal hygiene, hydration, etc. Additionally, recent studies correlate changes in the daily behaviors of older people to cognitive problems [7]. Recognizing eating and drinking activities, for example, would also help to check food habits, determine whether the person is still able to maintain his or her daily routine, and detect changes in it [8], so to facilitate prompt intervention by caregivers.

When dealing with elderly users, it is important to consider three important characteristics, i.e. ease-of-use (easy configuration and maintenance), coverage (no limited working area) and privacy [9]. As stated from literature evidence, three main approaches have been used in activity recognition based on different sensing technologies: vision, environmental and wearable sensors. Vision-based technologies raise issues linked to privacy, illumination variations, occlusion and background change [10]. The second approach relies on the interaction of the user with specific objects or appliances, assuming that the use of a certain object is strictly linked to a precise activity, but requires a large amount of sensors that need to be installed at the user's place [10]. On the other hand, thanks to the miniaturization and affordability of micro-electromechanical systems (MEMs) and in particular of Inertial Measurement Units (IMUs), the approach based on wearable sensors is gaining popularity. The use of wearable sensors makes it possible to collect data about users' movements without forcing them to stay in front of a camera or interacting with specific objects [11].

Several works have focused on the recognition of human activities using wearable sensors [12]. As detailed in the next paragraph, many of them adopted supervised machine learning, showing promising results, but conversely requiring labeled dataset, which can sometimes be difficult to generate and need to be updated each time a new activity is added [13].

Therefore, this paper proposed a step toward an unsupervised pattern-learning algorithm that can recognize gestures

Manuscript received July 26, 2017; revised October 12, 2017; accepted October 12, 2017. Date of publication October 20, 2017; date of current version November 22, 2017. This work was supported in part by DAPHNE Project (REGIONE TOSCANA PAR FAS 2007–2013, BANDO FAS SALUTE 2014, CUP J52I16000170002). The associate editor coordinating the review of this paper and approving it for publication was Prof. Aime Lay-Ekuakille. *(Corresponding author: Filippo Cavallo.)*

The authors are with the BioRobotics Institute, Scuola Superiore Sant'Anna, 56025, Pontedera, Italy (e-mail: alessandra.moschetti@santannapisa.it; laura.fiorini@santannapisa.it; dario.esposito@santannapisa.it; paolo.dario@santannapisa.it; filippo.cavallo@santannapisa.it).

Digital Object Identifier 10.1109/JSEN.2017.2764323

TABLE I

REVIEW OF STUDIES ON ACTIVITY RECOGNITION (DT = DECISION TREE, MLP = MULTI-LAYER PERCEPTRON, RANDOM FOREST = RF, SVM = SUPPORT VECTOR MACHINE)

Ref.	Used Sensors	Sensor Position	Activities	Machine Learning	Results
[11]	Accelerometer	Chest, right thigh and left ankle	Standing, stair descent, sitting, sitting down, sitting on the ground, from sitting to sitting on the ground, from lying to sitting on the ground, lying down, lying, walking, stair ascent and standing up.	**Supervised:** k-Nearest Neighbor, SVM, Supervised Learning Gaussian Mixture Models and RF **Unsupervised:** k-Means, Gaussian Mixture Models and Hidden Markov Model	Accuracy: **S:** > 0.850 **U:** >0.756 (min values in a 10fold Cross Validation)
[14]	IMU, magnetometer, GPS, light, pressure	Pocket position and wrist	Walking, running, cycling, standing, sitting, elevator ascents, elevator descents, stair ascents and stair descents	**Supervised:** C4.5, and CART based DT, Naïve Bayes, MLP and SVM	Accuracy: Up to 0.95 with Smartphone and 0.89 with Smartwatch
[15]	Accelerometer and gyroscope	Wrist	Sit-Stand, Stand-Sit, Sit-Lie, Lie-Sit, Stand-Lie, Lie-Stand, Standing, Sitting, Lying, Step Forward, Step Backward	**Supervised:** SVM	F-measure: 0.93 (leave-one-subject-out cross-validation)
[13]	Accelerometer and gyroscope	Pocket position	Walking, running, sitting, standing, and lying down	**Unsupervised:** k-means clustering, mixture of Gaussian (GMM), hierarchical agglomerative clustering (HIER), and DBSCAN	Accuracy: > 0.720 (min value)
[16]	Accelerometer and Location	Right thigh, on the waist and on the right hand	Sitting, standing, lying, walking, sit-to-stand, stand-to-sit, lie-to sit, and sit-to-lie categorized into stationary and motional activities. Five specific types of hand gestures: using mouse, typing on a keyboard, flipping a page while reading a book, stir-fry cooking, and dining using a spoon	**Supervised:** Three-level dynamic Bayesian Network	Accuracy: > 0.850
[17]	Accelerometer and gyroscope	Pocket position and wrist	Smoking, eating, typing, writing, drinking coffee, giving a talk, walking, jogging, biking, walking upstairs, walking downstairs, sitting, and standing.	**Supervised:** SVM, k-Nearest Neighbor and DT	Average accuracy >0.970 for simple activities and >0.895 for complex ones using wrist position and 10-fold stratified cross validation
[18]	Accelerometer	Front-right pocket and wrist	Walking, Jogging, Climbing Stairs, Sitting, Standing, Kicking Soccer Ball. Dribbling Basketball, Playing Catch with Tennis Ball (two people), Typing, Handwriting, Clapping, Brushing Teeth, Folding Clothes. Eating Pasta, Eating Soup, Eating Sandwich, Eating Chips, Drinking from a Cup	**Supervised:** RF, DT, Instance Based, Naïve Bayes and MLP	Accuracy = 0.703 with RF and impersonal model
[19]	Accelerometer	Wrist	Eating with chopstick, eating with the spoon and eating with the hand	**Supervised:** Naïve Bayes, BayesNet, Boosting, Bagging and DT	Accuracy = 0.680
[20]	Accelerometer and gyroscope	Wrist	Eating with the hand, with chopsticks or with spoon and other activities like smoking, drinking tea, washing one's face, shaving, applying makeup etc..	**Supervised:** ST	Accuracy = 0.920 in distinguishing between eating and non-eating gestures and = 0.86 for eating modes
Our work	Accelerometer	Index Finger and Wrist	Eat with the hand, eat with the fork, eat with the spoon, drink with the glass, drink with the cup, answer the phone, brush teeth, brush hair and use the hairdryer	**Supervised:** RF, MLP, and SVM **Unsupervised:** k-Means, Self-Organizing Map, and Hierarchical Clustering	Accuracy: **S:** > 0.908 **U:** >0.803 (min values in a leave-one-subject-out cross-validation)

without labeled data, favoring the use of data generated in real cases. In order to improve the gesture recognition capability, a ring placed on the index finger and a bracelet placed on the wrist are used.

II. RELATED WORKS

As can be seen from literature evidence (see Table I), many studies have focused on the recognition of human activity using wearable sensors. Much attention has been paid to the recognition of physical activities, which are very important to maintain a healthy lifestyle. Some works reached good results in terms of accuracy using only inertial sensors [11], [13], [15] or coupling these sensors with other ones [14].

Some works increased the number of activities to be recognized, involving also the use of the hand (Hand-Oriented-Activities). In particular, using inertial sensors [17], [18] and location [16] they were able to recognize hand- oriented activities mixed with not hand-oriented ones.

Finally, some studies focused on the recognition of eating activities and especially eating modes. In these cases,

the authors focused their attention on the identification of specific eating gesture in order to monitor also the food intake. In particular, using inertial sensors, different eating modes were distinguished among other activities [19], [20].

A summary of the related works in term of used sensors, recognized activities machine learning approach, and results can be found in Table I.

To promptly intervene in case of changes in daily behaviors, it is important to be able to discriminate among complex activities such as eating and drinking and performing acts of personal hygiene [7], [21]. As can be seen in the aforementioned works, when dealing with more complex actions, often the experimental data set includes very different activities, thus making easier to distinguish between eating, drinking and other activities. On the other hands, works that are interested in recognizing Hand-Oriented-Activities are not able to distinguish between different eating modes (Table I).

Assuming that people make specific gestures when they perform eating, drinking and other activities, it is possible to infer the activity from the performed gesture [22]. In our previous work, we focused on the recognition of daily gestures using wearable sensors placed on the wrist and on the fingers using two supervised machine learning algorithms, namely Decision Tree (DT) and a polynomial kernel Support Vector Machine (SVM) [23]. The selection of gestures to be recognized made it possible to discriminate among more complex activities that are very similar to each other, all including a movement of the hand to the head. While the best results were obtained by using three sensors, placed on the wrist, on the distal phalange of the thumb, and on the intermediate phalange of the index (F-measure equal to 0.91), we decided to focus our attention on the use of two sensors, wrist and index, that show good results while decreasing the invasiveness (F-measure equal to 0.88).

Considering the machine learning approach, for the best of our knowledge, few recent works use unsupervised approaches for activity recognition problems [11], [13]. For instance, Kwon et al. [13] proposed an unsupervised activity recognition approach and, using the accuracy and Normalized Mutual Information (NMI) as measures of the goodness of the algorithms, achieved high accuracy even when the number of activities, k, is unknown.

Therefore, the aim of this paper is twofold: firstly it investigates the use of unsupervised machine learning approaches, providing the basis to implement algorithms for daily gesture recognition in real conditions; secondly it demonstrates the advantages of extracting kinematics features of movement not only at the level of wrist, but also of finger by using a novel instrument based on a bracelet and ring. Indeed, the characterization of fingers' movement permits to identify aspects related to fine manipulation, digital grasping, etc. that in the end allow a deeper and effective recognition of gestures above all very similar gestures.

In particular, we compare the results obtained by three supervised learning algorithms with the ones obtained by three unsupervised ones, considering that unsupervised machine learning algorithms do not require labeled data, thus avoiding the need for a training dataset, and can therefore be more adaptable to real applications of human activity recognition.

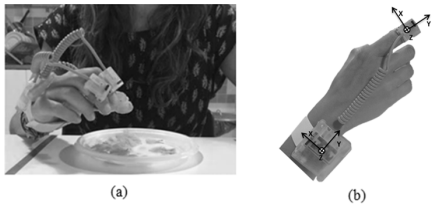

Fig. 1. Example of eating with the hand with the full configuration of sensors (a) and the considered configuration of sensors with the orientation of the axes (b). The z-axes are pointing toward the finger and the wrist.

III. Methodology

The main goal of this analysis is to demonstrate that our system is able to distinguish among nine daily gestures with an unsupervised approach. In fact, in order to realize the potential of this system in the real world, pattern learning algorithms should be able to operate without labeled data, as it is too resource intensive for a person to verify the large quantities of data that are generated by a gesture. An overview of the steps made to pursue the objectives is presented in Fig. 2.

In this section, we introduce the system architecture, the experimental settings and protocol and the feature extraction and selection.

A. Instrumentation

Data were acquired using five sensor units placed on the wrist and on the hand as described in our previous work [23]. In this work, we decided to consider in the analysis only the sensors placed on the index finger and on the wrist, which showed a good trade-off between recognition accuracy and obtrusiveness. In Fig.1 the two configurations are shown.

The sensor units consist of an INEMO-M1 board with a LSM303DLHC (six-axis geomagnetic module, STMicroelectronics), an L3G4200D (three-axis digital gyroscope, STMicroelectronics), an I2C digital output, and a dedicated microcontroller. From each unit, the acceleration and angular velocity are collected at 50 Hz.

A low-pass filter with a cut-off frequency of 5 Hz was implemented on board to filter data of the accelerometer and gyroscope in order to remove high-frequency noise and tremors.

B. Experimental Settings and Participants

Twenty young participants (11 females and 9 males, whose ages ranged from 21 to 34 (29.3 ± 3.4)) performed nine different gestures in the DomoCasa Laboratory, a 200 m^2 fully furnished apartment located in Peccioli, Pisa (Italy). This location was chosen to allow users to perform gestures in a natural way, reducing as far as possible the unnatural movement arising from the laboratory setting. The gestures were chosen in order to detect complex activities that consist of a similar movement, i.e. hand to head movement, and therefore can be easily confused. Moreover, these activities would make it possible to identify eating and drinking movements, which are important for the monitoring of elderly persons. In particular, the gestures performed are described in Table II.

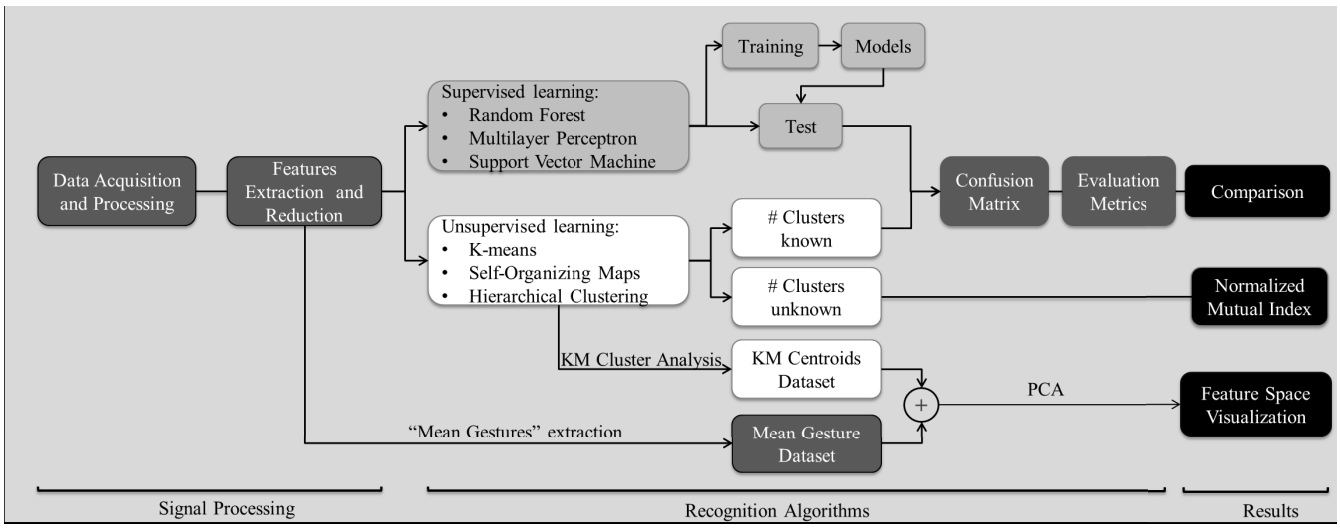

Fig. 2. Schematic representation of the methodological approach.

TABLE II
GESTURE DESCRIPTION

Gesture	Description
HA: Eating with the hand	Participants took the food with the hand and moved it to the mouth and back to the table.
GL: Drinking with a glass	Participants were asked to grasp the glass, move it to the mouth, and then leave it on the table.
FK: Eating with a fork	Participants had to take a piece of already cut fruit with the fork, eat it, and then move the hand back to the table without leaving the fork.
SP: Eating with a spoon	Participants had to use the spoon, load it with some yoghurt, and move it to the mouth and then back to the table without leaving the spoon.
CP: Drinking with a cup	Participants were asked to grasp the mug, move it to the mouth and then back on the table, leaving it
PH: Answering the telephone	Participants had to take the phone, move it to the head and back on the table after a few seconds
TB: Brushing the teeth with a toothbrush	Participants were asked to take the toothbrush from the sink, move it to the mouth to brush the teeth, and put it back on the sink
HB: Brushing the hair with a hairbrush	The gesture consisted in taking the hairbrush from the sink, moving it to the head, using it two or three times, and putting it back on the sink
HD: Drying the hair with a hair dryer	Participants were asked to take the hair dryer from the sink, move it to the head, and dry the hair

Users were asked to perform a sequence of 40 gestures for each kind of gesture without any constrictions in the way in which objects were picked up and the gestures were made. At the beginning of each session, users had to keep the hand and the forearm still on a plane surface in order to calibrate each session and compare the position of the sensors among the different sequences, referring to the first acquired gesture (HA). Users were observed during the acquisition and gestures were manually labeled in order to track the beginning and end of each gesture.

C. Feature Extraction and Selection

After the segmentation of the signal in single gestures according to the label, features were extracted. According to the state of the art [12], four features from the acceleration were evaluated along each direction. In particular, the mean (M), standard deviation (SD), mean absolute deviation (MAD) and the root mean square (RMS) were calculated. Our dataset thus included 7200 gestures with 24 features (3 axes $\times$ 4 features $\times$ 2 sensors) labeled with the corresponding gesture.

The linear correlation coefficient (Pearson) between each feature was computed to keep in the analysis only the features with a coefficient below 0.85 (absolute values) in order to reduce the noise due to the redundancy of data [11].

The final dataset was then composed of 15 uncorrelated features. An ANOVA test confirmed that the nine gestures were statistically different for all of these selected features ($p < 0.05$). In particular, as regards the index sensor unit, we selected the M and SD values along the three axes and the RMS values along the x- and y-axes. As regards the wrist sensor, we selected the M and SD values along the three axes and the RMS value along the x-axis. Once the dataset had been reduced, a Z-norm was computed to avoid distortion and have a zero mean and a unit standard deviation.

The obtained dataset was then used with supervised and unsupervised machine learning algorithms to compare the results.

D. Evaluation Measures

The results obtained with the unsupervised and supervised machine learning algorithms were presented as a confusion matrix.

Thus, the overall accuracy, F-measure, precision, and recall were computed as described in Eqs. from 1 to 4 [12] (where TP stands for True Positive, TN stands for True Negative,

TABLE III
COMPARISON OF SUPERVISED AND UNSUPERVISED ANALYSIS (K = 9) IN TERMS OF ACCURACY, F-MEASURE, PRECISION AND RECALL, COMPUTED AS THE MEAN VALUE

		Accuracy	F-Measure	Precision	Recall
Supervised					
	RF	0.932	0.936	0.941	0.932
	MLP	0.909	0.914	0.919	0.909
	SVM	0.938	0.942	0.947	0.938
Unsupervised					
	KM	0.818	0.818	0.818	0.818
	SOM	0.817	0.816	0.816	0.817
	HC	0.803	0.810	0.817	0.803

FP stands for False Positive and FN for False Negative).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

These evaluation metrics were used to compare and discuss the performances of these two approaches. With regard to the unsupervised analysis, the performances were evaluated with an external criterion [24] by comparing the output with our a priori knowledge.

IV. EXPERIMENTS

In this section, we describe the analysis of the reduced dataset with three supervised machine-learning techniques and with three unsupervised ones, both with k known and unknown. Finally, the visualization of the feature space to identify similarity and diversity between gestures is introduced.

A. Supervised Approach

The main goal of this work is to evaluate whether the proposed sensor configuration is able to distinguish among gestures even with the unsupervised approach. In this context, the supervised analysis is used as the "gold standard" for the comparison between the two approaches.

Particularly, in this work Multi-Layer Perceptron (MLP), Random Forest (RF), and SVM were applied as supervised algorithms. The performance of these algorithms was tested by a Leave-One-Subject-Out cross-validation (LOSO) technique, where 19 participants were used as training set, while the remaining one was used as a test set for the algorithms. All the participants were used as test set. In this phase, the analysis was performed by using the Weka Data Mining Suite [25]. In particular, MLP and RF were used in the default conditions, while for the SVM a radial basis function kernel was used.

The three techniques showed high results in terms of accuracy and the F-measure. In particular, as reported in Table III (supervised), the best results were obtained with the SVM with an accuracy of 0.938 and a F-measure of 0.942. The RF algorithm showed a lower accuracy and F-measure (0.932 and 0.936 respectively) with respect to SVM, but they were still higher than those obtained by MLP, which achieved an accuracy value of 0.909 and an F-measure of 0.914. RF, MLP, and SVM also showed good precision: 0.940, 0.919, and 0.947 respectively (Tab. III). These results confirm that the system is able to distinguish among the selected gestures when the system is trained, even with unknown subjects.

Considering the F-measure of the single gestures (Fig. 3a), the FK was one of the worst recognized (0.855 with MLP, 0.862 with RF, and 0.876 with SVM), often being confused with the SP. In addition, the HB reached a low value of F-measure, especially with the MLP approach (0.826 with MLP, 0.893 with RF, and 0.892 with SVM). This gesture was often confused with HD, which is justifiable considering the similarity of the gestures. In the supervised analysis, the highest values of F-measure were reached for PH and HA, which are higher than 0.97 and 0.95 respectively for all the algorithms.

B. Unsupervised Approach – K-Known

Three unsupervised machine learning clustering techniques were used to group the performed gestures into clusters. Particularly, in this work, the K-Mean (KM) algorithm, Self-Organizing Maps (SOMs), and Hierarchical Clustering (HC) were applied and compared. In particular, KM was applied considering the Euclidian distance with five replicates in order to avoid local minima. We chose these algorithms because they are widely used in similar applications [13], [26]. The Machine Learning and Pattern Recognition Matlab Toolboxes [27] were used for the unsupervised analysis.

In a first step, it was assumed that the number of performed activities is known. In this case, the three unsupervised algorithms show high and comparable results to the supervised ones concerning accuracy and the F-measure ($\sim$0.81 both). As regards precision and recall, the results are comparable for all the measures ($\sim$0.81) (Tab. III). Particularly, the worst recognized gestures are HB and HD, which are often mutually confused (see Fig. 4). Other gestures with a low F-measure are FK, TB, and CP (Fig. 3a). As depicted in Fig. 4, SP and FK are mutually confused due to the similarity of the two gestures. On the contrary, PH and HA are the most recognized gestures,

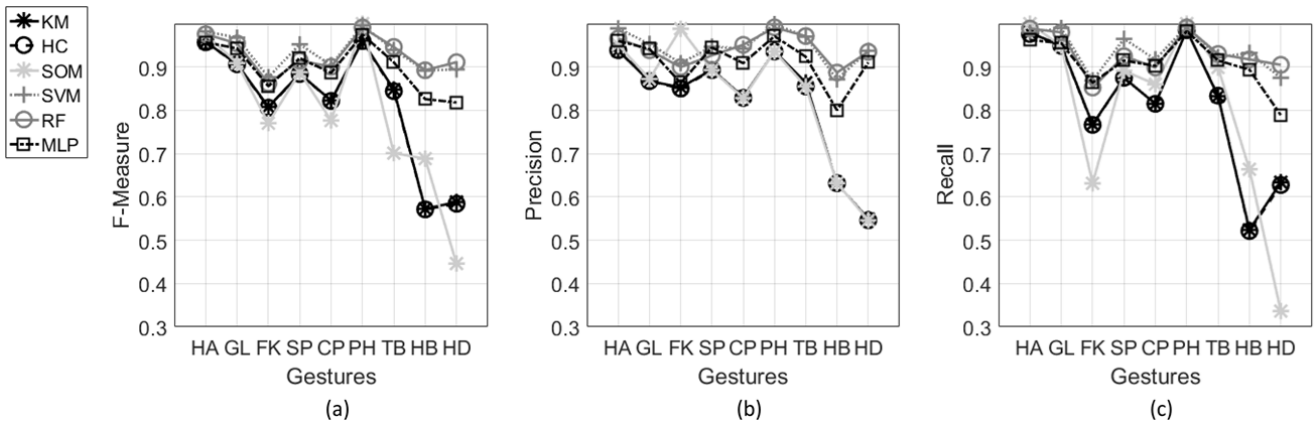

Fig. 3. Comparison of supervised and unsupervised analysis (k = 9): F-Measure (a) precision. (b) recall (c).

as confirmed by the high values of the F-measure (> 0.95 and > 0.96, respectively, for the three algorithms).

C. Unsupervised Approach – K-Unknown

In order to evaluate the performance of these unsupervised methods for activity recognition, the same algorithms were applied also considering the lack of knowledge of the number of activities performed (k) as proposed in [13]. In particular, this method emphasizes whether the unsupervised method could be useful even when there is an arbitrary number of activities. In this case, to evaluate the performance of the clustering algorithms we used NMI and accuracy as external criteria to compare the performances [13], [28]. In particular, the NMI index was evaluate as:

$$NMI = \frac{\sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij}\log\left(\frac{n\cdot n_{ij}}{n_i\cdot n_j}\right)}{\sqrt{\sum_{i=1}^{r} n_i \log\frac{n_i}{n}\sum_{j=1}^{s} n_j \log\frac{n_j}{n}}} \tag{5}$$

where r is the number of clusters, s is the number of classes, n_{ij} is the number of instances in cluster i and j, n_i is the number of instances in cluster i, n_j is the number of instances in cluster j and n is number of instances. It is important to notice that, this index does not require the condition that the number of activities is the same as the cluster.

As regard the accuracy measure, since the number of activity is different from the number of cluster we adapt the definition of accuracy as presented in [13]. For each trial, we consider, as correct-correspondence between cluster and activity (true positive), the cluster that has the largest portion of the true cluster. Fig. 5 shows the experiment results for the three approaches.

The analysis reveals that KM and SOM have the maximum NMI values for k = 9, whereas HC has the maximum value for k = 10 (Fig. 5). Accuracy results underline that the high values of this index is for k = 9 for all the approaches. Particularly, for k = 9 we obtained high and comparable accuracy performance for KM, HC and SOM (Tab. III), whereas, as regards the NMI index, KM and HC shows higher performance (0.760 and 0.765 respectively) rather than the SOM (0.759).

For k < 9, the indexes decrease significantly because different clusters merge into one. Some gestures are very similar (for instance HB and HD) and noisy, due to the inter- and intra- subject variability. A possible explanation is the fact that the gestures are "near" in the considered feature space (see Sect. IV.D), consequently they can be consider as one. Similarly for k > 9, the algorithms divide a normal cluster into several clusters, and thus the indexes value are lower.

The higher value of the NMI and accuracy indexes obtained with k = 9 suggests that the unsupervised approach could distinguish among the different gestures even when k is unknown.

D. Feature Space Visualization

Unsupervised learning has been used to reveal the implicit relationships in the dataset [29]. Firstly, we computed the "*Mean Gesture*" dataset, where all the features (F) were obtained as the mean value over the totality of a specific gesture. Thus, each gesture was described by a $1\times F$ row of the final $G\times F$ dataset (where F is the total number of features and G is the total number of gestures). Additionally, the standard deviation of the "*Mean Gesture*" had been computed to include the variability of a single gesture in the analysis. Similarly, also the "*centroids*" dataset has been built considering the centroid of the cluster obtained with the KM algorithm to corroborate the comparison between the two methods. Then, the two datasets were merged in one and analyzed to investigate similarity and differences among gestures.

Firstly, Principal Components Analysis (PCA) was applied in order to reduce the number of features thus to improve the visualization of the feature space. According to the Kaiser Rule [30], we consider only components with eigenvalues greater than 1. Hence, three components were selected to describe our feature space.

Then the Similarity Matrix (SM) was computed to explore the relationships between points and to quantify the "similar gestures" in the dataset. The equation of the i-th element of the SM is based on the normalized Euclidian distance ($\|ED\|$) computed as:

$$\forall p, q \in P \quad ED_{p,q} = \sqrt[2]{\sum_{i=1}^{m}(p_i - q_i)^2} \tag{6}$$

where p and q are two generic points of our space (P) and m in the dimension of our space.

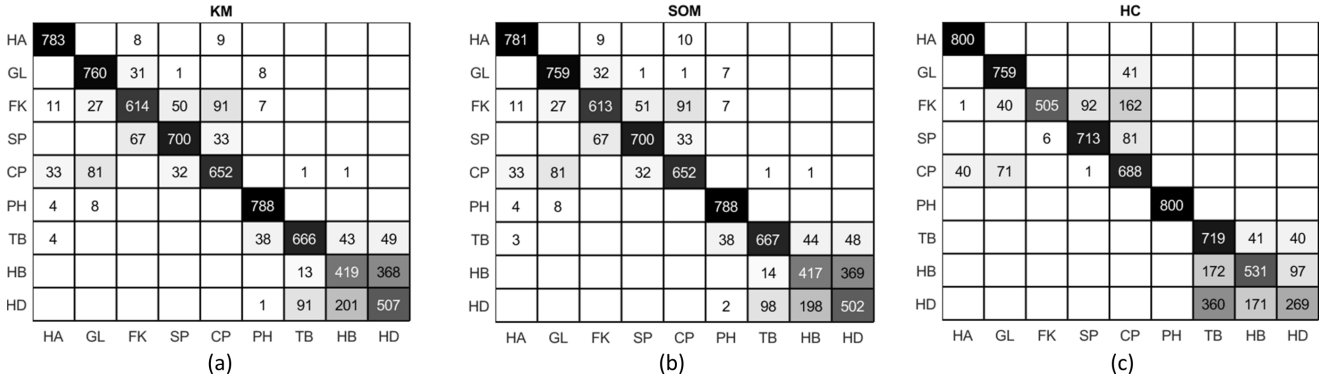

Fig. 4. Confusion matrix for k = 9 (a) K-means (b) self-organizing map (c) hierarchical clustering.

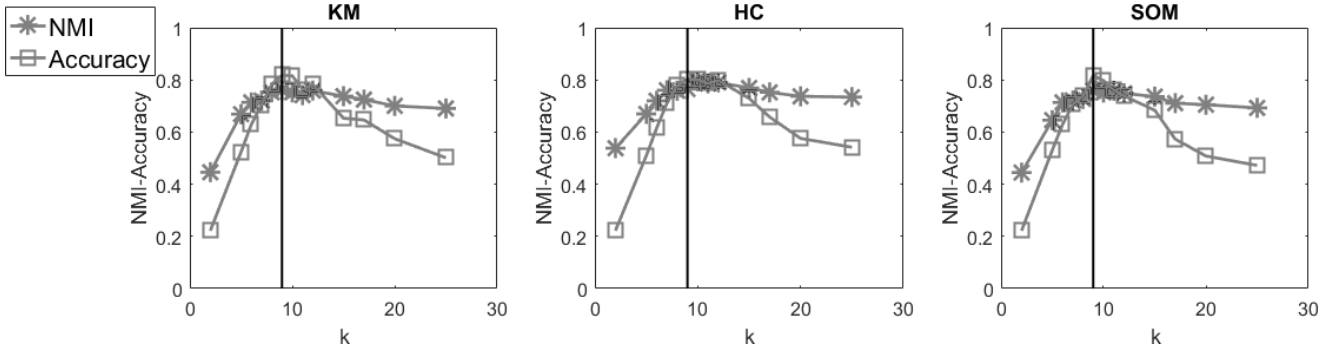

Fig. 5. Comparison of NMI index and accuracy when the number of clusters (k) is not known a-priori.

Thus the normalized Euclidian distance can be calculated as:

$$\|ED\|_i = \left| \frac{ED_i - \max(ED)}{\max(ED) - \min(ED)} \right| \tag{7}$$

Finally, the $\|ED\|$ (Eqs. 6 and 7) has been computed between each "*Mean Gestures*" and between two corresponding points (i.e. HA in the "mean gesture" dataset and "*centroids*" dataset) in the PCA space in order to investigate whether the cluster algorithms are able to correctly describe the gestures.

Fig. 6 reports the "*Mean Gestures*" (circle) and the *"Centroids"* (diamonds) in the feature space whereas the gray grid represents the inter-subject variation of each gestures. It is important to notice that some gestures are very similar and their variability around the mean gesture is partially overlapped. These results confirm the performance of unsupervised algorithms for k $\neq$ 9.

The visualization of the feature space confirms that the most similar gestures are HB and HD, which are the closest "mean gestures" in the space. Other close points are CP and HA ($\|ED\| = 0.265$) and FK and CP ($\|ED\| = 0.280$). On the contrary, SP and HB are the most distant points in the space ($\|ED\| = 1.000$). SP is also far from TB ($\|ED\| = 0.956$), PH ($\|ED\| = 0.879$), and HD ($\|ED\| = 0.868$). These results are aligned with the outputs of the confusion matrix (Fig. 4).

From the comparison of the "mean gestures" with the KM centroids, HB and HD show the highest values of $\|ED\|$ computed between these corresponding points in the 3D-PCA space (1.000 and 0.525 respectively). In effect, these gestures are mutually confused, as confirmed by the lower values of the

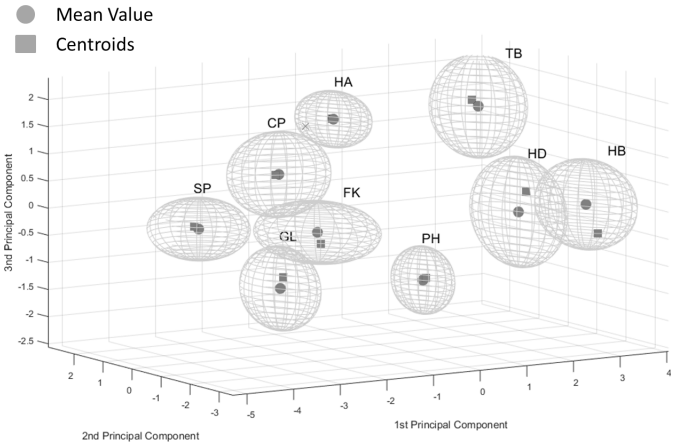

Fig. 6. Features space representation considering the mean values of each gesture (circle) and the centroids (diamond) on the plane of the first, second and third principal component. The gray grid represent the standard deviation.

F-measure (0.572 for HB and 0.588 for HD). In contrast, HA has the most similar points. Other similar points are TB (0.185) and SP (0.072). Nevertheless, as shown in Fig. 6, all the centroids are included in the range of the standard deviation.

V. Discussion

The aim of this work was to evaluate whether the selected sensors' configuration is able to discriminate among the gestures using an unsupervised approach compared to a supervised one. Starting from our previous work [23], where it was shown that the addition of the sensor on the index finger improved the recognition rate with respect to a single sensor on the wrist (the F-measure increases from 0.622 to 0.884 using a polynomial kernel SVM in a LOSO analysis), we compared

the outcomes of unsupervised and supervised approach using the same sensors' configuration.

The proposed analysis achieves high results for both approaches (complete results are reported in Table III), showing that our system is able to distinguish among the different gestures achieving good values of accuracy with respect to the state of the art. For instance, in [18], the recognition of eating activities with a wrist sensor was very low (accuracy equal to 0.527 and 0.627 for eating soup and drinking activities in the case of an impersonal analysis) with respect to our system (overall accuracy of $\sim$0.81 for the unsupervised analysis), suggesting that the addition of the index can improve the recognition rate. The explanation of that could lie in the fact that a ring sensor on finger allow to encompass kinematics features of movement more related to, for example, fine manipulation, digital grasping or pinch, that thus make possible the distinction of gestures that would be very similar if only tracked by wrist sensor.

Regarding the unsupervised analysis, the three methods have difficulty in showing a precise separation among the clusters. The collected gestures are not distributed in a spherical shape and they are noisy, as confirmed by the analysis of the feature space, the $\|ED\|$ values and the standard deviation of the mean value (Fig. 6) reported in the previous section.

As concern the k-unknown analysis, we evaluated how the system is able to manage the situation by comparing the NMI and the accuracy parameters (Fig. 5). The highest value of NMI correspond to $k = 9$ for KM and SOM, similarly, the high accuracy values are for $k = 9$. Therefore, this means that the proposed system configuration and the selected set of features are able to discriminate among the different gestures even in this "blind" condition. Nevertheless, the proposed approach presents some limitations that we would like to overcome in our future works. In particular, we force the algorithm to cluster all gestures in the k clusters without considering the option "new gesture", in other words, a specific gesture has been assigned to the nearest cluster. New sophisticated algorithm should be able to adapt and learn "new gesture" from streaming sensor data. In this context, we plan to investigate and develop a new algorithm to properly manage unseen gestures.

As regards the approach with $k = 9$, the supervised and unsupervised approaches present similar behavior in recognizing specific gestures in terms of precision, recall, and F-measure as depicted in Fig. 3. Both approaches show the best results in terms of F-measure in PH and HA, while among the worst recognized are HB and HD as shown in Fig. 3. These results are confirmed also by the analysis of the feature space (Fig. 6) and the unsupervised confusion matrix (Fig. 5).

It is worth to highlight that our activity dataset has been designed to include similar gestures all involving the movement of the hand to the head, thus making more difficult to recognize them. According to the state of the art (see Table I), only few works consider this set of gestures that can be easily confused but are important for the recognition of daily activities. Indeed, among the important activities to be recognized, there are feeding and personal hygiene, which are in fact included in the dataset proposed in this work. The recognition of these activities allows to monitor people at home and fosters the ability to detect changes in daily patterns in order to identify possible critical situations and check whether elderly persons are still able to live at their own home. Moreover, the possibility to recognize eating and drinking activities could allow to check on the diet of elderly persons, helping them to maintain a healthy lifestyle.

In the proposed work, the experimentation was carried out with young healthy people to evaluate whether this system could be used to recognize significant daily gestures. It is necessary, therefore, to test the system also with elderly people to check the performances of the same configuration of sensors. Hence, future experimentation will involve old persons that could have physical impairment also linked to neurodegenerative diseases like Parkinson's Disease.

To make the activity monitoring really part of daily life, it is important to have systems that require little training or configuration effort and that integrate easily in the person everyday life [21]. Nowadays, these aspects are a concrete technical challenge that the researchers need to address.

The use of unsupervised algorithms allows to overcome issues related to the need of labelled data, thus making easier to analyze large quantity of data and getting a step closer to real applications. The combination of sensors described in this work provides good recognition rate with unsupervised approaches, even if according to the state of the art, one of the drawbacks of the use of wearable sensors is the perceived obtrusiveness. However, in order to overcome this limitation, wearable sensors can become part of already used objects or accessorized, like jewelry [31]. In this way, the presented configuration can come a step closer to the application in daily life.

VI. Conclusion

Recognition of daily activities is crucial in the monitoring of elderly people at home, improving the ability of the caregivers to check on the conditions of the persons and increasing the possibility for old persons to stay longer at their own place.

Hence, in this paper, we proposed a comparison between unsupervised and supervised approaches for the recognition of nine daily gestures. We evaluated the performance and obtained high performances for both approaches (see Table III). This work shows, therefore, how a sensor on the wrist and one on the index finger can be used to recognize gestures associated with daily activities even with an unsupervised approach. These results highlight the possibility to get a step closer to real applications in the recognition of daily activities.

Moreover, the use of an inertial ring, such as the one developed in [32], together with the use of a sensor on the wrist could be further investigated to evaluate the recognition rate of other activities, such as physical activities. In this way, it could be possible to increase the number of recognized activities, increasing the ability of the system to detect changes in daily routine and analyze the lifestyle of people.

Future works will be, therefore, focused on the increase in the number of activities that can be recognized by using the proposed system and on the analysis of how the

unsupervised approach manages the addition of new items to be recognized.

REFERENCES

[1] (Sep. 29, 2016). *Nearly 27 Million People Aged 80 or Over in the European Union*. Accessed: Oct. 2016. [Online]. Available: http://ec.europa.eu/eurostat/documents/2995521/7672228/3-29092016-AP-EN.pdf/4b90f6bb-43c1-45ed-985b-dfbe9564157a

[2] D. I. Auerbach, "Will the NP workforce grow in the future? New forecasts and implications for healthcare delivery," *Med. Care*, vol. 50, no. 7, pp. 606–610, 2012.

[3] R. S. Hooker, J. F. Cawley, and C. M. Everett, "Predictive modeling the physician assistant supply: 2010–2025," *Public Health Rep.*, vol. 126, no. 5, pp. 708–716, 2011.

[4] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 579–590, May 2013.

[5] M. Aquilano *et al.*, "Ambient assisted living and ageing: Preliminary results of RITA project," in *Proc. IEEE EMBC*, Dec. 2012, pp. 5823–5826.

[6] G. Turchetti, S. Micera, F. Cavallo, L. Odetti, and P. Dario, "Technology and innovative services," *IEEE Pulse*, vol. 2, no. 2, pp. 27–35, Apr. 2011.

[7] D. J. Cook and N. C. Krishnan, *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior From Sensor Data*. Hoboken, NJ, USA: Wiley, 2012.

[8] N. K. Suryadevara, S. C. Mukhopadhyay, R. Wang, and R. K. Rayudu, "Forecasting the behavior of an elderly using wireless sensors data in a smart home," *Eng. Appl. Artif. Intell.*, vol. 26, no. 10, pp. 2641–2652, 2013.

[9] L. Wang, T. Gu, X. Tao, and J. Lu, "Toward a wearable RFID system for real-time activity recognition using radio patterns," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 228–242, Jan. 2017.

[10] A. Wang, G. Chen, J. Yang, S. Zhao, and C. Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4566–4578, Nov. 2016.

[11] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015.

[12] 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.

[13] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6067–6074, Oct. 2014.

[14] J. J. Guiry, P. van de Ven, and J. Nelson, "Multi-sensor fusion for enhanced contextual awareness of everyday activities with ubiquitous devices," *Sensors*, vol. 14, no. 3, pp. 5687–5701, 2014.

[15] B. Mortazavi *et al.*, "Can smartwatches replace smartphones for posture tracking?" *Sensors*, vol. 15, no. 10, pp. 26783–26800, 2015.

[16] C. Zhu and W. Sheng, "Realtime recognition of complex human daily activities using human motion and location data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2422–2430, Sep. 2012.

[17] M. Shoaib, S. Bosch, H. Scholten, P. J. Havinga, and O. D. Incel, "Towards detection of bad habits by fusing smartphone and smartwatch sensors," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2015, pp. 591–596.

[18] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inf. (BHI)*, Feb. 2016, pp. 426–429.

[19] H. J. Kim, M. Kim, S. J. Lee, and Y. S. Choi, "An analysis of eating activities for automatic food type recognition," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, 2012, pp. 1–5.

[20] S. Sen, V. Subbaraju, A. Misra, R. K. Balan, and Y. Lee, "The case for smartwatch-based diet monitoring," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2015, pp. 585–590.

[21] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 81–94, Mar. 2016.

[22] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, "A survey of diet monitoring technology," *IEEE Pervasive Comput.*, vol. 16, no. 1, pp. 57–65, Mar. 2017.

[23] A. Moschetti, L. Fiorini, D. Esposito, P. Dario, and F. Cavallo, "Recognition of daily gestures with wearable inertial rings and bracelets," *Sensors*, vol. 16, no. 8, p. 1341, 2016.

[24] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

[25] M. Hall *et al.*, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[26] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[27] (May 15, 2017). *Pattern Recognition and Machine Learning Toolbox*. [Online]. Available: https://it.mathworks.com/matlabcentral/fileexchange/55826-pattern-recognition-and-machine-learning-toolbox

[28] A. A. Strehl and J. Ghosh, "Cluster ensembles-a knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.

[29] F. Li and S. Dustdar, "Incorporating unsupervised learning in activity recognition," in *Proc. AAAI Workshop*, vol. WS-11-04. San Francisco, CA, USA, Aug. 2011, pp. 38–41.

[30] H. F. Kaiser, "The application of electronic computers to factor analysis," *Edu. Psychol. Meas.*, vol. 20, pp. 141–151, Apr. 1960.

[31] A. L. Ju and M. Spasojevic, "Smart jewelry: The future of mobile user interfaces," in *Proc. Workshop Future Mobile User Interfaces*, 2015, pp. 13–15.

[32] D. Esposito and F. Cavallo, "Preliminary design issues for inertial rings in Ambient Assisted Living applications," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, Pisa, Italy, May 2015, pp. 250–255.

Alessandra Moschetti received the master's (Hons.) degree in biomedical engineering from the University of Pisa, Italy, in 2012, and the Ph.D. *(cum laude)* degree in biorobotics from Scuola Superiore Sant'Anna, Pisa, in 2017. She currently holds a Post-Doctoral position with the Biorobotics Institute of Scuola Superiore Sant'Anna. Her research interests include activity recognition, biomedical signal processing, wearable sensors, and human robot interaction in assisted living applications.

Laura Fiorini received the master's (Hons.) degree in biomedical engineering from the University of Pisa in 2012 and the Ph.D. *(cum laude)* degree in biorobotics from the Scuola Superiore Sant'Anna in 2016. She currently holds a post-doctoral position at the BioRobotics Institute of Scuola Superiore Sant'Anna. Her research interests include ambient assisted living, cloud service robotics, ICT system for cognitive activation, pattern recognition, signal processing, and experimental protocol.

Dario Esposito received the master's degree in electronical engineering from the Federico II University of Naples, Italy, in 2011. Since 2011, he has been an Assistant Researcher with the BioRobotics Institute of Scuola Superiore Sant'Anna. His research interests are focused on assistive robotics, wearable devices based on sensor systems, and wireless sensor networks.

Paolo Dario has been a Visiting Professor at Brown University, Providence, RI, USA, the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, the École Normale Superieure de Cachan, France, the College de France, Paris, France, the Polytechnic University of Catalunya, Barcelona, Spain, Waseda University, Tokyo, Japan, Zhejiang University, Hangzhou, and Tianjin University, China. He is currently a Professor of Biomedical Robotics and the Director of The BioRobotics Institute of Scuola Superiore Sant'Anna (SSSA), Pisa, Italy. He is the Co-ordinator of the Ph.D. Program in BioRobotics at SSSA. He is the coordinator of many national and European projects and the author of more than 500 scientific papers, over 300 on ISI journals. His main research interests are in the fields of medical robotics, bio-robotics, biomechatronics and micro/nano engineering, and robotics.

Filippo Cavallo received the M.Sc. degree in electrical engineering, the Ph.D. degree in bioengineering from the BioRobotics Institute of Scuola Superiore Sant'Anna, Pisa, Italy. He is currently an Assistant Professor with the BioRobotics Institute of Scuola Superiore Sant'Anna, Pisa, Italy, focusing on cloud and social robotics, ambient assisted living, wireless and wearable sensor systems, biomedical processing, acceptability and ICT and AAL roadmapping. He is an author of various papers on conferences and ISI journals. He participated in various national and European projects and currently is Project Manager of Robot-Era, AALIANCE2, and Parkinson Project. From 2005 to 2007, he was a Visiting Researcher with the EndoCAS Center of Excellence, Pisa, Italy, Takanishi Lab, Waseda University, Tokyo, Japan (2007), the Tecnalia Research Center, Basque Country, Spain, working on wearable sensor system for AAL.