Human activity recognition using multisensor data fusion based on Reservoir Computing

4 authors, including:

Filippo Palumbo
Italian National Research Council

44 PUBLICATIONS   **371** CITATIONS

SEE PROFILE

Claudio Gallicchio
Università di Pisa

53 PUBLICATIONS   **380** CITATIONS

SEE PROFILE

Rita Pucci
The University of Edinburgh

13 PUBLICATIONS   **25** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Development, Analysis and Applications of Machine Learning Models for Structured Domains View project

FP7 GiraffPlus: Combing social interaction and long term monitoring for promoting independent living View project

Human activity recognition using multisensor data fusion based on Reservoir Computing

Filippo Palumbo [a,b,*], Claudio Gallicchio [b], Rita Pucci [b] and Alessio Micheli [b]

[a] *Institute of Information Science and Technologies "Alessandro Faedo", National Research Council, Pisa, Italy*
[b] *Department of Computer Science, University of Pisa, Pisa, Italy*

Abstract. Activity recognition plays a key role in providing activity assistance and care for users in smart homes. In this work, we present an activity recognition system that classifies in the near real-time a set of common daily activities exploiting both the data sampled by sensors embedded in a smartphone carried out by the user and the reciprocal Received Signal Strength (RSS) values coming from worn wireless sensor devices and from sensors deployed in the environment. In order to achieve an effective and responsive classification, a decision tree based on multisensor data-stream is applied fusing data coming from embedded sensors on the smartphone and environmental sensors before processing the RSS stream. To this end, we model the RSS stream, obtained from a Wireless Sensor Network (WSN), using Recurrent Neural Networks (RNNs) implemented as efficient Echo State Networks (ESNs), within the Reservoir Computing (RC) paradigm. We targeted the system for the EvAAL scenario, an international competition that aims at establishing benchmarks and evaluation metrics for comparing Ambient Assisted Living (AAL) solutions. In this paper, the performance of the proposed activity recognition system is assessed on a purposely collected real-world dataset, taking also into account a competitive neural network approach for performance comparison. Our results show that, with an appropriate configuration of the information fusion chain, the proposed system reaches a very good accuracy with a low deployment cost.

Keywords: AAL, activity recognition, neural networks, Reservoir Computing, sensor data fusion, WSN

1. Introduction

Activity Recognition (AR) is an emerging research topic, which is founded on established research fields such as ubiquitous computing, context-aware computing and multimedia, and machine learning for pattern recognition. Recognizing everyday life activities is a challenging application in pervasive computing, with a lot of interesting developments in the health care domain, the human behavior modeling domain and the human-machine interaction domain [20]. Inferring the activity of the users in their own domestic environments becomes even more useful in the Ambient Assisted Living (AAL) scenario, where facilities provide assistance and care for the elderlies and the knowledge of their daily activities can ensure safety and a successful aging.

From the point of view of the deployment of activity recognition solutions, we recognize two main approaches depending on whether the solution adopts wearable devices or not. The first kind of solutions generally use sensors (embedding accelerometers, or transducers for physiological measures) that make direct measures about the user movements. For example, a sensor placed on the user ankle may detect the number of steps based on the response of an embedded accelerometer that is shaken with a specific pattern every time the user makes a step. On the other hand, the disadvantage of this approach is that wearable devices can be intrusive on the user, even if, with recent advances in technologies of embedded systems, sensors tend to be smaller and smaller. Solutions that avoid the use of wearable devices instead, are motivated by the need for a less intrusive activity recognition systems. Among these solutions, those based on cameras are probably the most common [38]. However, even though this ap-

*Corresponding author. E-mail: filippo.palumbo@isti.cnr.it.

proach is physically less intrusive for the user, it suffers from several issues: low image resolution, target occlusion and time-consuming processing, which is still a challenge for real-time activity recognition systems. Furthermore, user privacy is also an important issue, especially if cameras are used to continuously monitor the user itself. More recently, a new generation of non wearable solution is emerging. These solution exploits the implicit alteration of the wireless channel due to the movements of the user, which is measured by devices placed in the environment and that measure the Received Signal Strength (RSS) of the beacon packets they exchange among themselves [14,60].

In this paper we propose an Activity Recognition system based on Multisensor data fusion (AReM) that makes use of both the approaches. Specifically, we use wearable and environmental sensors and we base the recognition of the user activity both on accelerometers embedded on the wearable sensors and on the RSS of the beacon packets exchanged between all the sensors (both wearable and environmental).

From the algorithmic point of view, the need for knowledge discovery of temporal patterns from potentially large amounts of raw sensor data has recently led to an increasing use of Machine Learning models in activity recognition systems (see e.g. [46] for a recent survey). In this context, the class of neural networks [27] provides an interesting range of models and architectures that are able to effectively learn in sequential data domains, with characteristics of robustness to noise and heterogeneity in the input sources considered. Neural networks for learning in sequential domains are particularly appropriate to deal with the characteristics of the tasks and the information involved in activity recognition applications, including methods based on windowing strategies, such as Time Delay Neural Networks (TDNNs) and Input Delay Neural Networks (IDNNs) [83,84], and methods based on explicit feedback connections in the class of Recurrent Neural Networks (RNNs) [39,78]. In particular, RNNs are widely used to process and capture dynamic knowledge from sequential information, allowing to provide predictions which depends on the temporal history of the input signals. Based on the short-term history of the signals received in input, RNNs are used in the context of human activity recognition to produce output values that represent classification labels/values corresponding to a target human activity. In this paper, within the class of RNNs, we take into consideration the efficient Reservoir Computing (RC) [52,82] paradigm in general, and the Echo

State Network (ESN) [33,34] model in particular. The ESN model is supported by theoretical investigations (see e.g. [25,77]) and by hundreds of relevant successful experimental studies reported in literature (see e.g. [53,82]). Recently, RC models have proved to be particularly suitable for processing noisy information streams originated by networks of heterogeneous sensors, resulting in successful real-world applications in tasks related to AAL and human activity recognition (see e.g. [2,5,9,19,61]), and allowing to successfully build intelligent sensor networks tailored to specific activity recognition contexts, as testified by the recent successful results of the RUBICON Project [3,6] and detailed in Section 2.

The proposed activity recognition process is realized by the means of two layers of classification. The first classification layer, represented by a custom decision tree algorithm, detects a subset of activities and directs the flow of collected data to different Machine Learning modules. These modules represent the second layer of the processing chain, and are implemented as ESNs. An important aspect of our work consists in the use of common wireless sensor devices deployed in the environment in combination with wearable sensors embedding accelerometers, in order to increase the performance of the activity recognition system.

In order to test the performance of our activity recognition system, we presented a preliminary version of AReM to the third edition of the EvAAL (standing for Evaluating AAL Systems through Competitive Benchmarking) international competition [15], where state-of-the-art systems are presented in the activity recognition track. We used the EvAAL opportunity not only to carry out a comparative study with other novel solutions in the activity recognition field presented by the competitors, but also to improve the AReM system through an iterative refinement process that, starting from the preliminary result obtained in [61], uses the outcomes of the competition experience to arrive to the improved system presented in this paper.[1] A more in-depth performance assessment of the proposed activity recognition system is presented

---

[1]This paper, based on ex-novo conducted experiments, extends the preliminary investigations presented in the conference paper [61], by proposing a more structured classification process and a complete performance analysis on a larger set of user activities, evaluated under different settings of the activity recognition system, also considering a competitive Machine Learning approach for performance comparison. Furthermore, an improved communication infrastructure is also proposed.

in this paper, using an activity recognition dataset collected through measurement campaigns targeted on the EvAAL activities, taking into account different possible settings based on the availability of sensor data sources and also considering the IDNN as a competitive approach for performance comparison.

The rest of the paper is organized as follows. In Section 2 current trends and related works are presented, Section 3 describes the overall architecture of the AReM activity recognition system, Section 4 describes the steps of the refinement process of the system through the real test site experiences, and Section 5 goes into the details of the refined system's performance focusing on ESNs outcomes. Finally, Section 6 draws the conclusions.

2. Related work

One of the main goals of AAL systems is the detection of human activity. This can be achieved by means of different tools and technologies spanning from smart homes (regular homes augmented with various type of sensors, including cameras, and actuators), mobile devices, and wearable sensors [72]. The activity itself can be represented and recognized at different resolutions, such as a single movement, action, activity, group activity, and crowd activity. The degree of the resolution should be chosen properly, according to the particular tool to be deployed. In the case of tools assisting the elderly doing rehabilitation tasks, a fine-grained resolution is needed in order to recognize the particular movement performed. Such systems, even though present an higher level of obtrusiveness (robotic tools [41]) or privacy issues (cameras [18,44]), are normally accepted since they are seen by the user as medical equipment with a precise and temporary scope. Within the broader context of a long-term monitoring AAL system, it is expected to be up and running all day long and over long periods of time in order to prevent cognitive or physical deterioration of the user. In this scenario, a system able to recognize simple activities (e.g. lying, sitting, standing, walking, bending, cycling, falling), instead of a single movement or action performed, is a good compromise between the possibility of inferring high-level activities (i.e. to infer activities of daily living) and the low level of obtrusiveness required together with the lower amount of data generated by the system. Furthermore, it enables short-term tasks dealing with in-home emergencies, like fall detection. In this field, solutions that

make use of wearable and mobile sensors are the more established and studied [10,54,73]. Data are mostly gathered from accelerometer and gyroscope sensors in the form of time series. Simple actions such as walking, jogging, and running can be represented in the form of periodic time-series patterns. Important issues in these kinds of solution are the number of the devices to be deployed and their position on the user's body. In [55], authors perform a study on how the position of the used device (embedding a triaxial accelerometer and a light sensor) influences the performance. They demonstrate that some positions are better for particular activities but confirm that there is not a generally valid position for all the possible activities to be recognized. Furthermore, in [80] authors suggest that for some activities more sensors improve the recognition. Also the use of ambient sensors has been deeply analyzed in the literature (e.g. [49,68,88]) in order to recognize more complex activities. In these solutions, a network of ambient sensors is usually used to model resident activities in the environment, as a sequence of sensor events. The main issue here is that the assumption of consistent predefined activities as a sequence of events does not hold in reality. Due to several physical, cognitive, cultural, and lifestyle differences, not all individuals perform the same set of tasks in similar sequences [72]. In order to overcome the deployment-related issues, we chose to use as primary data source the RSS of the beacon packets exchanged among a small set of wearable sensors (see Section 3.1 for the setting used in the experimentation) and some environmental devices placed in specific point of interests in the house related to the particular activity to be recognized (e.g. a device embedded in a stationary bike for the cycling activity). When used to infer close proximity and movements between near devices, the RSS is not so much affected by the particular orientation of the sensor's antenna. Furthermore, the use of ambient sensors keeps the number of worn sensors low and it removes the constraint of predefined sequences of activations since they are only used to infer proximity to relevant point of interests.

Recently, Machine Learning techniques have found wide applications in building human activity recognition systems based on data generated from sensors. Depending on the nature of the treated data, of the specific scenario considered and of the admissible trade-off among efficiency, flexibility and performance, different Machine Learning methods have been applied in this application area [46]. From a learning perspective, human activity recognition problems often in-

volve computational learning tasks characterized by a sequential nature. In this sense, the various estimations to be provided in relation to specific activities can be considered to be discernible basing on specific patterns of activations/values from a typically heterogeneous set of possibly noisy sensor sources (with potentially both continuous and discrete values), and based on the temporal order of such series. Focusing on supervised learning models for human activity recognition applications, a first example is represented by probabilistic-generative methods, including the family of Naïve Bayes classifiers and Hidden Markov Models (HMMs) [70] based classifiers, which have been applied to problems of daily-life human activity recognition (see e.g. [51,85,90]). However, the class of probabilistic-generative models often results to be severely limited in dealing with large amounts of low-level heterogeneous and noisy sensor data (see e.g. [57]). In this sense, discriminative methods often provide interesting approaches in terms of computational effort and predictive accuracy. For instance, the works in [37,71] make use of feed-forward neural networks for recognition of human activities from an accelerometer sensor placed on the waist of the user. Other works in the area of human activity recognition adopt Support Vector Machines (SVMs) [81]. For example, in [42] SVMs are used to approach an activity recognition problem considering a user in a smart environment with a variegate set of sensors. In [4,28,29] SVMs are used for problems of activity recognition using input from tri-axial accelerometer data. Some approaches, instead, are based on the use of multiple classification systems within a hierarchical activity recognition model (e.g. [23,47]). However, all these aforementioned discriminative approaches are based on learning models suitable for flat data domains, and therefore they are often restricted in the processing of sequential/temporal information. Other approaches which are also limited in the processing of dynamical information are based on decision tree models (e.g. [10,45,55]), instance based learning (e.g. [17]) or linear discriminant analysis (e.g. [67]).

On the other hand, in the neurocomputing area, neural network models for sequential domains processing represent good candidates for applications in human activity recognition problems, as they are characterized by the ability to effectively learn input-output temporal relations from a potentially huge set of noisy and imprecise heterogeneous streams of sensed data. In this context, delay neural networks are popular class of models [39], which represent temporal context using windowing/buffering techniques in conjunction with feed-forward neural architectures. In this concern, the IDNN model [83,84] represents a paradigmatic approach, and it is therefore considered in this paper for performance comparison with the proposed activity recognition system. Several works in literature report examples of applications of delay neural networks to problems in the field of human activity recognition, taking into consideration temporal windows with length typically in the range of 2–10 seconds. For example, in [43,65] delay networks are used to classify daily-life activities from accelerometer data, whereas a similar approach is used in [56] for human gesture recognition, and in [50] for recognition of workers activities. In [87] a delay network based approach is used to recognize human postures and activities from data collected by a smart-shoe device. More recently, in [74], delay neural networks are used for daily activity monitoring using accelerometer and gyroscope data generated by a smartphone. The versatility of IDNN based approaches in this context is also testified by applications in the area of animal activity recognition, as reported e.g. in [11,58,59]. Note, however, that windowing strategies adopted in delay neural networks may often imply several drawbacks, mainly related to the fact that the window length is a fixed a-priori defined parameter of the model, determining its memory length, i.e. the limit of the length of the longest sequences that can be successfully discriminated. An alternative approach is represented by RNNs architectures [39,78] with explicit recurrent connections, which are capable of dealing with sequential/temporal data through recursive processing based on the presence of feedback delay connections. As introduced in Section 1, a particular efficient approach to RNN training is represented by RC [52,82] networks, and in particular by ESNs [33,34]. Recently, promising results have been reported in applications of RC networks in the fields of AAL and human activity recognition. In this concern, the RC approach has been introduced and successfully experimentally assessed in tasks related to robot localization [19] and indoor user context localization in real-world environments from RSS data streams [5,7,8,26]. A further application of RC models for real-time user localization in indoor environments is presented in [12], in which by adopting a hybrid approach, RC networks have been showed to provide significant benefits to the accuracy of RSS-based localization systems. It is also worth mentioning the recent results of the European FP7 Project RUBICON [3,6], whose goal was to design and de-
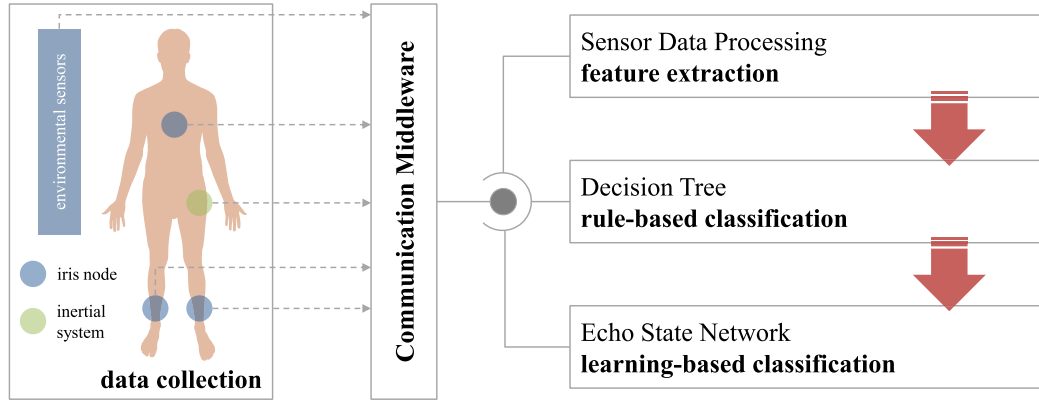
Fig. 1. The Activity Recognition multisensor processing chain.

velop a self-learning robotic ecology made up of sensors, actuators and robotic devices. Within the aims of the RUBICON project, ESNs distributed on the low-powerful nodes of a Wireless Sensor Network (WSN) have been used to approach supervised computational tasks in the field of AAL, and pertaining to the classification of human activities, using input data streams coming from sensorized environments. In this context, the work in [2] describes the application of ESNs to a set of supervised tasks pertaining to the recognition of user daily-life activities, within an AAL real-life test bed scenario, including more than 50 sensors of different types. Moreover, an application of ESNs to adaptive planning for context-aware robot navigation is presented in [9]. Further applications of ESNs in the area of AAL are reported in [75], in which ESNs are used to classify physical activities in equestrian sports from a 3-axis accelerometer, a gyroscope and a magnetic sensor stripped to a horse rider wrist, and in [31], in which standard ESNs are used to estimate people count from multiple PIRs in indoor corridors at the Fraunhofer IAIS facilities. Overall, the aforementioned applications of the RC approach to problems in the area of AAL and human activity recognition show the potentiality of this methodology, still deserving further investigations and deep experimental assessment, in particular in combination with already established techniques in this specific application domain.

3. Overall architecture

Following the recommendations of the technical annex of the EvAAL competition and the considerations made in Section 2, the main objective of the proposed system is to implement an activity recognition system

(ARS) that identifies the following activities: Lying, Sitting, Standing, Walking, Bending (both keeping the legs straight and keeping the legs folded), Falling and Cycling (using a stationary bike).

The proposed AReM activity recognition system can be described as a chain of processing steps, where each step is responsible for carrying out a particular task. The first task is to collect the raw data coming from the deployed wireless sensor network (WSN) and the inertial system embedded in the smartphone (Section 3.1). The data collected are forwarded to the next computational step by means of dedicated gateways exploiting the functionalities of the communication middleware (Section 3.2). Then the data are processed to extract significant features (Section 3.3), a first classification layer based on a custom decision tree is applied (Section 3.4), and finally the filtered data and features are given as input to dedicated learning modules, implemented by means of ESNs, for the final classification layer to provide the output activity (Section 3.5). The activity recognition chain presented in this paper is shown in Fig. 1.

We planned to use a sensor network that is partly worn by the user and partly deployed in the environment together with a smartphone carried out by the user. Specifically, we used four wearable (three part of the WSN and a smartphone) and one environmental sensors. The sensors composing the WSN are capable of measuring the signal strength of the incoming packets. The wearable sensors are placed on the chest and on the ankles of the user, while the smartphone was placed in the user's trouser pocket. Furthermore, the WSN wearable sensors, which are connected in a clique, exchange among themselves beacon packets, with the purpose of measuring the respective RSS among themselves. The environmental sensor filters

the beacon packets emitted by the wearable sensors to receive only to the beacons emitted by the sensor on the left ankle. The smartphone carried out by the user is used to exploit the data coming from the embedded inertial measurement system. All the data acquired by the wearable sensors are collected by the gateway that uses an instance of the communication middleware to transmit data to the other modules of the distributed system. The smartphone, instead, runs a mobile middleware instance.

The first classification layer is based on a custom decision tree, this classification technique has been successfully applied to activity recognition in numerous previous works [24,36,64,65]. It uses state-of-the-art techniques in order to discriminate a falling activity [16,48,69], to recognize the proximity of the user to a meaningful furniture in the environment [66], and to understand if the user is in a horizontal or vertical position [43].

The final activity classification is performed by ESN learning modules, which are dynamical neural networks able of directly processing temporal streams of sensed data generated by the nodes of the WSN (see Section 2). The internal parameters of the neural networks are adapted in a supervised fashion to the specific application by the means of learning on a sufficiently sampled training set of data, including both input data from the deployed sensors and ground-truth information, resulting in the definition of a *learning task*. Details on the ESN model can be found in Section 3.5.

3.1. Data collection

Recognizing human activities depends directly on the features extracted for motion analysis. In our activity recognition system we fuse information coming from inertial sensor embedded in a smartphone carried out by the user and the implicit alteration of the wireless channel due to the movements of the user, which is given by wireless sensors placed in the environment and on the user himself. These devices measure the RSS of the beacon packets they exchange among themselves in the WSN [14].

We collect RSS data using IRIS nodes embedding a Chipcon AT86RF230 radio subsystem that implements the IEEE 802.15.4 standard and programmed with a TinyOS firmware. Three of them are placed on the user's chest and ankles, another one is placed on a furniture in the environment representing a meaningful place for a particular activity (i.e. a stationary

bike for cycling activity recognition). For the purpose of communications, the beacon packets are exchanged by using a simple virtual token protocol that completes its execution in a time slot of 50 milliseconds [13]. A modified version of the Spin [89] token-passing protocol is used to schedule node transmission, in order to prevent packet collisions and maintain high data collection rate. When an anchor is transmitting, all other anchors receive the packet and perform the RSS measurements. The payload of the transmitting packet is the set of RSS values between the transmitting node and the other sensors sampled during the previous cycle. This packet has been received also by a sink node along with the node's unique ID. The sink collects all the payloads for storage and later processing. The RSS values are acquired for a given channel c for all the nodes $n = 1 \ldots N$ in the network, i.e., when the last node of the network has transmitted by using the channel c, the first sensor node starts with a new cycle by using a new channel. The data collected from each sensor pair (a_i, a_j), in the following called link, are formatted as a string with the following fields: the identifier of the receiver (ID), the RSS values measured between the receiver and the others transmitting sensors, the timestamp at which the string was acquired, and finally the channel used for the acquisition.

We used the inertial system embedded in a Huawei G510 smartphone to collect accelerometer data. The smartphone inertial system worn by the user was placed along its sagittal plane and the sampling rate was set according to the *SENSOR_DELAY_FASTEST* setting of the Android operating system that for the particular model used has proved to be ≈70 Hz.

3.2. Communication middleware

The data collected through the WSN and the smartphone come from different hardware and communication protocols (i.e. the WSN sink running TinyOS, the desktop gateway running Linux operating system, the mobile gateway running Android system). In order to hide this hardware and software heterogeneity, we exploited the GiraffPlus [21] middleware capabilities in terms of service discovery and communication.

In GiraffPlus, the service discovery and communication functionalities are realized by intelligent buses, namely the context, service and control buses. All communications between services happen in a roundabout way via one of them, even if, physically, the services are located on the same hardware node. Each of the buses handles a specific type of message/request
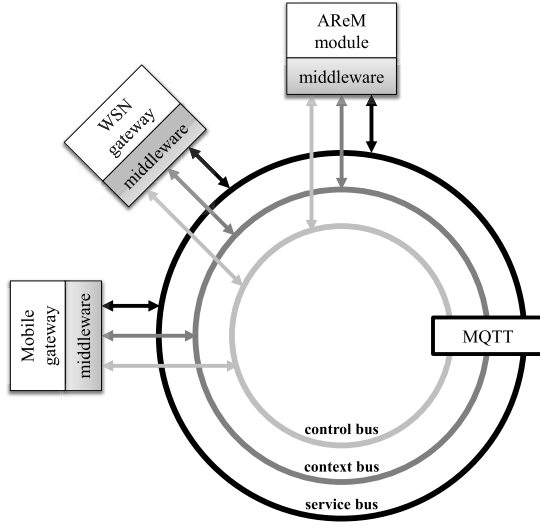
Fig. 2. The bus-based communication middleware integration.

```
_____ Sample Data _____
% -------
%    No. WSN nodes: 4    Time Slot [ms]: 50
%    No. Inertial:  1    Time Slot [ms]: 14
%    Step Time [ms]: 250
%    M:Magnitude, Ax,Ay:Accelerations
%    1:Chest, 2:Left Ankle, 3:Right Ankle, 4:Bicycle
% -------
%
%timestamp,minM,maxM,Ax,Ay,P14,std14,P12,std12,P13,
%std13,P23,std23
%
1368548213503,0.97,0.98,2.33,3.70,25.75,23.50,0.50,18.00,
1.41,31.25,0.43
1368548213753,0.98,1.10,2.41,3.66,24.68,23.10,0.40,18.50,
0.40,30.80,0.50
1368548214003,0.95,1.20,2.44,3.55,23.52,22.50,0.50,18.50,
0.00,30.55,0.20
1368548214253,1.10,1.15,2.34,3.79,24.05,22.50,0.00,17.50,
0.50,31.25,0.80
1368548214503,0.96,1.05,2.34,3.76,23.89,23.50,0.50,18.00,
0.20,31.50,0.10
%...
```

Fig. 3. A sample output sequence of the sensor data processing block.

and is realized by different kinds of topics. Topics are realized exploiting the communication connector interface based on the MQTT[2] protocol, a machine-to-machine (M2M) connectivity protocol designed as an extremely lightweight publish/subscribe messaging transport layer [30]. Figure 2 shows the main components of the proposed system interacting with the middleware buses in our distributed scenario. We exploited the adaptivity of the GiraffPlus middleware to integrate new sensors and communication protocols (like in the case of the 802.15.4-based TinyOS MAC protocol).

We integrated the inertial measurement system embedded in the smartphone using the mobile version of the GiraffPlus middleware running on Android devices [62], while the WSN sink was integrated in the system by means of the desktop version of the middleware [63] through an OSGi TinyOS wrapper.

Finally, we developed the AReM module as a GiraffPlus service that collects data exploiting the middleware context bus, processes the data extracting the requested features for the next steps of computation (Section 3.3), applies the decision tree (Section 3.4), and activates the selected ESN (Section 3.5).

3.3. Sensor data processing

As a result of the previous step we obtain data sampled at different frequencies with high dimensionality. Furthermore, consecutive values of a time series are

usually highly correlated, thus there is a lot of redundancy. In this step of the activity recognition chain we extract the needed time-domain features to compress the time series and slightly remove noise and correlations. We choose an epoch time of 250 milliseconds according to the EvAAL technical annex. In such a time slot we elaborate 5 samples of RSS (sampled at 20 Hz) for each of the four couples of WSN nodes (i.e. Chest-Right Ankle, Chest-Left Ankle, Right Ankle-Left Ankle, Bicycle-Left Ankle) and approximately 17 readings of accelerometer values (sampled at ≈60 Hz). The time-domain features extracted from these samples are shown in the Aggregate Values section of Fig. 3. They include the mean value and standard deviation for each reciprocal RSS reading from worn WSN sensors, the mean and standard deviation of the RSS between environmental node and left ankle worn sensor, the mean value of x and y axis accelerometer data and its max and min value magnitude over the 250 milliseconds time-slot. We consider the magnitude M as root-sum-of-squares of the three signals from the embedded tri-axial accelerometer sensor streaming expressed in g (9.80665 m/s^2):

$$M = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

Based on these features, the decision tree system and the ESN makes a prediction about the user activity.

3.4. Decision tree

Starting from the output given by the sensor data processing block of the proposed activity recognition
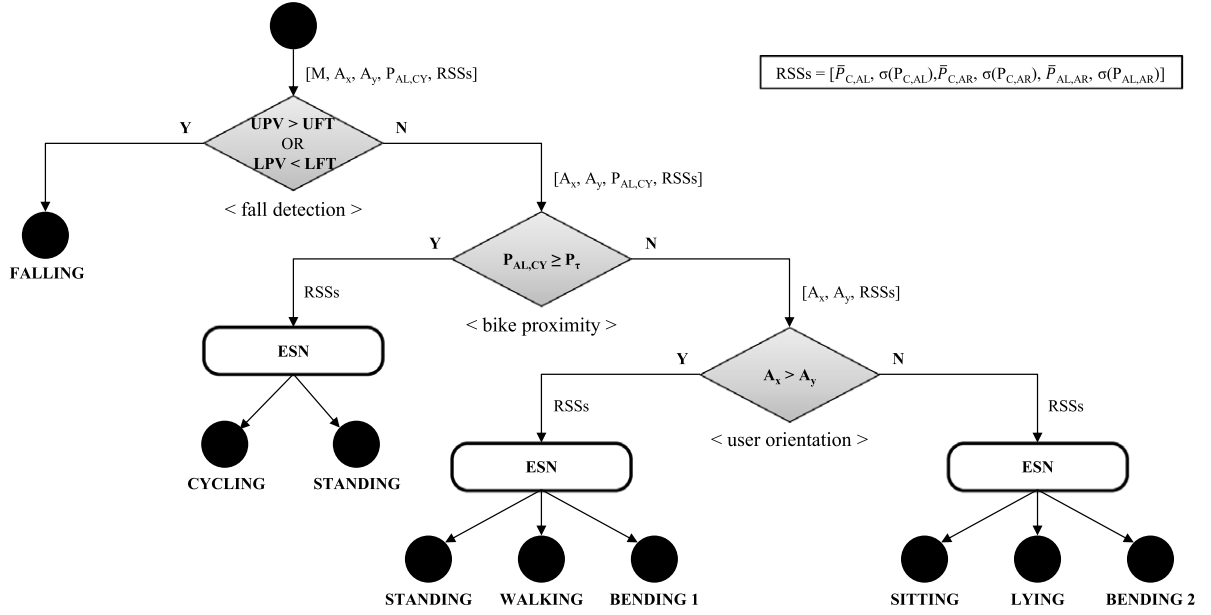
Fig. 4. Structure of the decision tree fusion mechanism.

chain, we arrive to the decision of the user activity output applying to each time-slot data a first layer of classification using a custom decision tree classification algorithm. The decision tree is used to discriminate the sub-group of activities to be recognized by the specific ESN module (see Section 3.5) in the second layer of classification. The advantages of custom decision trees include low computation requirements, a simple implementation, and a good understanding of the classifier's structure. The structure of the custom decision tree constructed is depicted in Fig. 4. The tree has three binary decision nodes and four leaf nodes, the latter representing the ESN to use except for the first one representing the falling activity.

The first decision node uses the magnitude M of the tri-axial accelerometer data to decide if a falling activity is happening. As proved in [16] the magnitude is a good measure to identify thresholds for falls detection. We define upper and lower fall thresholds as follows:

– Upper Fall Threshold (UFT): set at the level of the smallest magnitude Upper Peak Value (UPV) of 100 falls recorded. The UFT is related to the peak impact force experienced by the body segment (thigh) during the impact phase of the fall.
– Lower Fall Threshold (LFT): set at the level of the smallest magnitude Lower Peak Value (LPV) of 100 falls recorded. The LFT is related to the acceleration of the thigh at or before the initial contact of the body segment with the ground.
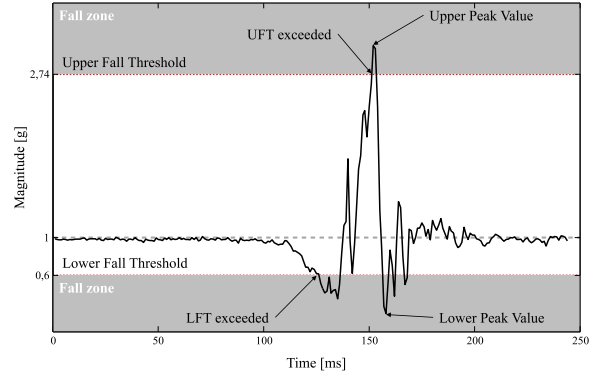

Fig. 5. Magnitude plot of the tri-axial embedded accelerometer, with thresholds used for falling detection. When stationary, the root–sum-of-squares signal from the tri-axial accelerometers is a constant +1 g.

Figure 5 shows the magnitude plot of one of the test made in the laboratory. Setting up a UFT of 2.74 g and a LFT of 0.6 g according to the results shown in [16], all the tests made were correctly recognized as falling activity. If the input data $LPV = \min(M)$ is greater than LFT, or $UPV = \max(M)$ is less than UFT, the second decision node is applied to the received time-slot data, otherwise a falling activity is detected.

The second decision node estimates the proximity [66] of the device placed on the user's left ankle to one of the WSN node placed to the stationary bike. We define a device i to be in-range of device j if the
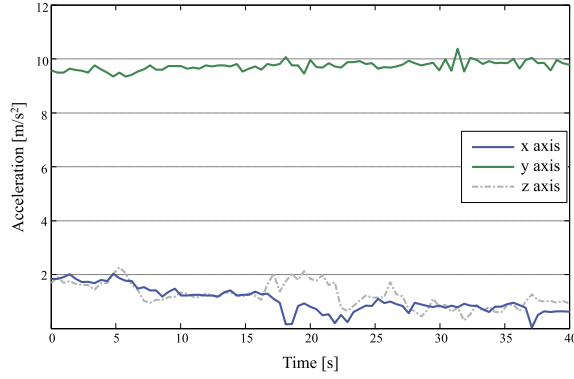
Fig. 6. Acceleration plot for the vertical position.

Fig. 7. Acceleration plot for the horizontal position.

received signal strength at j of the packet transmitted by i, $P_{i,j}$, falls below a power threshold P_τ. Thus, the proximity $Q_{i,j}$ is defined as:

$$Q_{i,j} = \begin{cases} 1, & P_{i,j} \geqslant P_\tau \\ 0, & P_{i,j} < P_\tau \end{cases} \tag{2}$$

If the $P_{AL,CY}$ ($i =$ ankle left AL, $j =$ stationary bicycle CY) is greater than the threshold set to $P_\tau = 20$ dBm we infer that the user is near the stationary bike assuming that he can be in the cycling or standing state. In this case, the RSSs vector is given as input to the ESN trained to recognize those activities. If not, the third decision node is applied to the received time-slot data. The RSSs vector is composed of the means and standard deviations of the reciprocal received signal strengths of the motes in the body area WSN. Specifically, the RSS between sensors placed at chest and left ankle ($\bar{P}_{C,AL}, \sigma(P_{C,AL})$), chest and right ankle ($\bar{P}_{C,AR}, \sigma(P_{C,AR})$), and left and right ankles ($\bar{P}_{AL,AR}, \sigma(P_{AL,AR})$).

The last decision node uses the x and y axis accelerometer mean values to determine if the user is in a vertical or horizontal position. As demonstrated in [43], it is clear that if the user is in a vertical or horizontal position, even if it exhibits periodic behavior related to the particular activity, he has distinctive patterns based on the relative magnitudes of the x, y, and z values. Figures 6 and 7 show that for vertical position activities like standing, walking, and bending keeping the legs straight (bending 1), the x relative magnitude A_x is usually less than the y relative magnitude A_y express in the figures as m/s^2. Conversely, if the user performs activities with his legs in a horizontal position like sitting, lying, and bending keeping the legs folded (bending 2), the x relative magnitude A_x is usu-
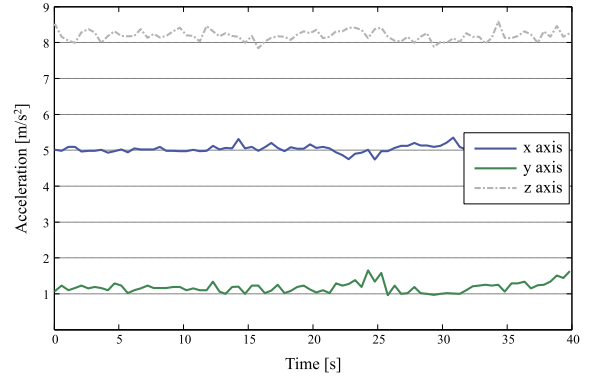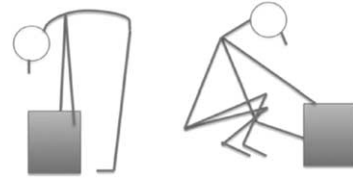

Fig. 8. The two types of bending activity. Bending 1 on the left and bending 2 on the right.

ally greater than the y relative magnitude A_y (Fig. 8 shows the two types of bending activity). Based on this decision we choose which trained ESN to use.

3.5. ESN

In this paper, the activity recognition process is realized by the means of RNNs [39,78], which are dynamical neural networks models particularly suitable for processing temporal sequences of noisy data, such is the case of human activity recognition. In this context, RNNs are used to produce a classification output at each time step, based on the history of the RSS signals received in input. Based on a training set of data samples, learning is used to adapt the network parameters to the specific context and input patterns. As mentioned in Section 1, for RNN modeling we take into consideration the efficient RC paradigm [52]. This results in an approach which, on the one hand strongly reduces the computational cost of training, and on the other hand has been experimentally proved to be particularly effective in real-world problems pertaining to application contexts of interest for this paper, such are AAL and user/robot localization (see Section 2).

In this paper, within the RC/ESN paradigm, we focus on the Leaky Integrator ESN (LI-ESN) [35], which

is particularly suitable for treating input data with the peculiarity of the RSS originated from a WSN [5,26].

The architecture of a LI-ESN is composed of an input layer with N_U units, a large and sparsely connected reservoir layer with N_R recurrent non-linear units, and a readout layer with N_Y feed-forward linear units. The reservoir encodes the input history of the driving input signal into a network state. The readout computes the output of the model by linearly combining the activation of the reservoir units.

At each time step t, the reservoir of the LI-ESN computes a state $\mathbf{x}(t) \in \mathbb{R}^{N_R}$ according to a state transition function:

$$\mathbf{x}(t) = (1-a)\mathbf{x}(t-1)$$
$$+ af\big(\mathbf{W}_{in}\mathbf{u}(t) + \hat{\mathbf{W}}\mathbf{x}(t-1)\big) \quad (3)$$

where $\mathbf{u}(t) \in \mathbb{R}^{N_U}$ is the input of the LI-ESN at time step t, $\mathbf{W}_{in} \in \mathbb{R}^{N_R \times N_U}$ is the input-to-reservoir weight matrix (possibly including a bias term), $\hat{\mathbf{W}} \in \mathbb{R}^{N_R \times N_R}$ is the recurrent reservoir weight matrix, f is the element-wise applied activation function, which typically is a non-linearity of a sigmoidal type (in this paper we use *tanh*), and $a \in [0, 1]$ is the leaking rate parameter, used to control the speed of the reservoir dynamics [35,52] (when $a = 1$ the standard ESN state transition function is obtained).

At each time step t, the readout computes the output of the model $\mathbf{y}(t) \in \mathbb{R}^{N_Y}$ through a linear combination of the elements in the state $\mathbf{x}(t)$, i.e.:

$$\mathbf{y}(t) = \mathbf{W}_{out}\mathbf{x}(t) \quad (4)$$

where $\mathbf{W}_{out} \in \mathbb{R}^{N_Y \times N_R}$ is the readout-to-reservoir weight matrix (possibly including a bias term).

In particular, in this paper we consider the case of multi-classification learning tasks among K classes, where each class corresponds to one of the activities. In this case, every element in the output vector at time step t, i.e. $y_i(t)$, corresponds to one of the considered activities, which means that $N_Y = K$ (in our paper, the maximum number of activities for one task is 7, see details in Section 5). In general, at time step t, the i-th activity is considered as identified if the corresponding element in the readout has a positive activation, i.e. whenever $y_i(t) > 0$. When it is required to identify one single activity at each time step, as in our application, the identified activity is the one corresponding to the readout unit with the largest activation, i.e. activity

i is recognized at time step t whenever

$$y_i(t) = \max_{j=1}^{N_Y}\big(y_j(t)\big) \quad (5)$$

The readout is the only part of the LI-ESN which undergoes a process of training, typically by using efficient linear methods, e.g. pseudo-inversion and ridge regression [52]. The reservoir parameters are left untrained after a proper initialization under the constraints of the *echo state property* (ESP) [25,34]. A sufficient and a necessary condition for the ESP to hold are provided in literature [33,35]. In practical applications the sufficient condition is often considered too restrictive, and the necessary condition is generally considered for the initialization of the reservoir parameters. Such necessary condition states that the system governing the reservoir dynamics (Eq. (3)) is locally asymptotically stable around the zero-state. For the case of LI-ESN, the necessary condition for the ESP can be expressed as:

$$\rho(\tilde{\mathbf{W}}) < 1 \quad (6)$$

where $\rho(\tilde{\mathbf{W}})$ is the spectral radius of the matrix

$$\tilde{\mathbf{W}} = (1-a)\mathbf{I} + a\hat{\mathbf{W}} \quad (7)$$

In general, in order to meet the condition in Eq. (6), a simple process consists in randomly initialize the weight values in the matrices $\mathbf{W}_{in}$ and $\hat{\mathbf{W}}$, and then rescale the matrix $\hat{\mathbf{W}}$ to obtained the desired value of the spectral radius (see e.g. [25,33,52] for details). In particular, in this paper we adopt the efficient weight encoding strategy (see details in [5,7]) consisting in randomly selecting the weight values in matrices $\mathbf{W}_{in}$ and $\hat{\mathbf{W}}$ from a small alphabet of possible values. Such approach has the advantage of greatly reducing the memory requirements for the storage of the network parameters, while not decreasing the performances of the RC models in applications, as experimentally shown in [5,7].

4. Fielding the system in a real smart environment

The proposed activity recognition system aims at becoming a robust solution for automatic and unobtrusive identification of user's activities in AAL scenarios. Real-time monitoring of human movements could be a useful tool for many purposes and fu-

ture applications such as life-log, health care or entertainment. To this purpose, we have performed a "deployment-and-evaluation" approach that includes different testbeds for each phase of development of the technology.

Specifically, we have conceived an approach that is based on the deployment and iterative refinement of the technology in three main steps. Once a prototype has been developed, it was first evaluated dealing with a subset of activities (i.e. standing up and down) in a laboratory settings. The results obtained in this phase are described in [61] and have been used to present the system to the EvAAL competition. After peer review, our team was accepted to the competition together with the teams: IJS (from the Jožef Stefan Institute of Ljubljana, Slovenia) [40], AmevaActivity (from the University of Seville, Spain) [22], and CUJ (from the University of Chiba, Japan) [32]. Competitors were invited to install and run their Activity Recognition System (ARS) during a predefined time slot. An actor performed a physical activity trip across the smart home with a reference ARS used to obtain the ground truth data. In order to get approximately the same ground truth for all the contestants, audio signals were used to synchronize the actor movements [1].

Starting from the encouraging results obtained in laboratory settings on the restricted set of activities, we performed further experiments in order to address the requirements of the EvAAL competition. With the aim of properly training the LI-ESNs for the activity recognition phase, a new measurement campaign was conducted collecting training data from the body area sensor network that would have been used in the competition. Three IRIS motes were used for the campaign, placed on the chest, right and left ankle of the test subject. Data from such a small WSN was recorded while an actor performed the activities: bending keeping the legs straight (i.e. bending 1), bending keeping the legs folded (i.e. bending 2), cycling, lying, sitting, standing, walking. Ground-truth was obtained manually, labeling each sequence of activities. Data gathered during this measurement campaign was organized into an *Activity Recognition dataset*, containing the RSS measured among the IRIS nodes during the aforementioned activities, at the frequency of 4 Hz (i.e. 4 samples per second). Each sequence in the dataset has the duration of 2 minutes (480 time steps) and corresponds to the actor performing a specific activity. The dataset contains a total number of 88 sequences, pertaining to the different activities, Table 1 shows the number of available sequences for each activity.

Table 1
Number of sequences for each activity in the Activity Recognition dataset

Activity	N. of sequences
Bending1	7
Bending2	6
Cycling, Lying, Sitting, Standing, Walking	15

Fig. 9. The sensors setup during the competition.

Figure 9 shows the sensors used for the measurement campaign, installed in the CIAMI living lab in Valencia, venue of the competition. In addition to the body area sensor network, as illustrated in Section 3.1, an environmental IRIS mote was installed on the stationary bike and a smartphone was placed in the actor's trousers pocket. The presented system placed in a top position confirming the effectiveness of the chosen approach. The results obtained in the competition are freely available on the official website[3] together with the details of all the competitors systems and the metrics used for the evaluation.

After the competition, the enhanced system presented in this paper was tested again under laboratory settings in order to perform a complete experimental assessment of the reservoir computing networks used for the activity recognition system, and to compare the obtained results with alternative methods (IDNNs) considered as baseline reference. The obtained results are illustrated in the following section.

---

[3]http://evaal.aaloa.org/.

5. Results

Based on the overall architecture of the AReM system described in Section 3, here we describe the results obtained by the learning modules performing the activity recognition process. As described in Section 3.4, the fall detection mechanism in the AReM system is implemented by exploiting state of the art solutions. Therefore, for the purposes of assessing the performance of our RC-based activity recognition process, here we restrict the attention to the remaining activities: Bending1, Bending2, Cycling, Lying, Sitting, Standing and Walking. Moreover, the performance achieved with the proposed RC-based approach is compared with the one obtained using IDNNs, representing a popular paradigmatic approach in the field of neurocomputing models for application in human activity recognition problems (see Section 2). Through such a comparison it is also possible to investigate whether the use of a recurrent approach for learning in sequence domains (such is the case of RC), not limited to the use of finite-size windows for treating streams of input data (as instead is the case of IDNN), allows to effectively take advantage on the considered tasks.

For our experiments, we considered both the settings in which only the RSS data from the wearable IRIS sensors (see Fig. 1) are available to the ARS, and the case in which also data from additional sensors (additional environmental RSS and smartphone accelerometer) are available (see details in Section 3). In the former setting, referred in the following as *RSS-based AReM setting*, or simply RSS-based, all the aforementioned activities are considered, and classification of the user actions is performed by a single learning model. In the latter case, referred in the following as *Heterogeneous AReM setting*, or simply Heterogeneous, the augmented set of available data allows to apply the rule-based classification process described in Section 3, through the application of the decision tree in Fig. 4. According to the structure of the considered decision tree, the classification of the user activity is performed in this case by resorting to 3 learning models, each of which specialized by training on subgroups of activities, and allowing a more accurate activity discrimination. Through a comparison of performances achieved under the RSS-based and the Heterogeneous settings, it is possible to experimentally assess the advantages brought by the availability of an augmented set of input data sources and by the design of the data fusion process, exploiting the specialization of the learning models trained on sub-sets of activities.

Table 2

Organization of the computational learning tasks and activities involved in the different ARS settings

ARS Setting		Activities
RSS-based AReM		Bending1, Bending2, Cycling, Lying, Sitting, Standing, Walking
Heterogeneous AReM	*task 1*	Cycling, Standing
	task 2	Bending1, Standing, Walking
	task 3	Bending2, Lying, Sitting

According to these settings, the collected Activity Recognition dataset described in Section 4 was used for the definition of different multi-classification learning tasks. In particular, 4 multi-classification learning tasks were arranged, the first one for the RSS-based ARS setting, and the remaining 3 for the different subsystems of the Heterogeneous setting. Table 2 provides the information about the organization of the learning tasks based on the different settings considered.

For each learning task in Table 2, we split the available data into a training set and an external separate test set (for performance evaluation only). The number of sequences in the training set and in the test sets is 64 and 24, respectively, for the learning task in the RSS-based setting. For the learning tasks in the Heterogeneous setting, the number of sequences in the training and test sets is 22 and 8, respectively, for task 1, 27 and 10, respectively, for task 2, and 26 and 10 respectively, for task 3 (note that the Standing activity is present in both tasks 1 and 2). For each multi-classification learning task, the values of the hyper-parameters of LI-ESNs and IDNNs were selected (in order to maximize the K-class classification accuracy, see Eq. (11)) on a validation set, according to a holdout model selection scheme over the training set of data. For every task in Table 2, the number of sequences in the validation set was ≈25% of the number of sequences in the training set, i.e. 14 for the learning task in the RSS-based ARS setting, and 4, 6 and 6, respectively, for the three learning tasks in the Heterogeneous ARS setting.

For model selection purposes (on the validation sets), in our experiments with LI-ESNs we considered, for every learning task, networks with reservoir dimension $N_R \in \{10, 50, 100, 300, 500\}$, 10% of connectivity, leaking rate $a \in \{0.1, 0.3, 0.5, 0.7, 1\}$. According to the weight encoding scheme described in [5,7] (see Section 3.5), the weights values in matrices $\mathbf{W}_{in}$ and $\hat{\mathbf{W}}$ were randomly chosen from a weight alphabet uniformly sampled in the range $[-0.4, 0.4]$, leading to a spectral radius (Eq. (6))

of approximately 0.9 for every setting. A number of 5 reservoir guesses were independently generated for each reservoir hyper-parametrization, and results were averaged over such guesses. The readout of LI-ESNs was trained using pseudo-inversion and ridge regression with regularization parameter $\lambda_r \in$ {0.0001, 0.001, 0.01, 0.1, 0.5, 1, 5, 10, 100, 1000}.

For what concerns the comparative experiments carried out using IDNNs, we considered networks trained using Back-propagation with learning rate $\eta \in$ {0.0001, 0.002, 0.0025, 0.003, 0.005, 0.01, 0.1} and momentum $\alpha \in$ {0, 0.001, 0.0001}. The values of the weight decay parameter and the number of hidden units were set to $\lambda_{wd} = 0.00001$ and $N_H = 100$, respectively, based on the results achieved on the validation set during preliminary experiments involving the ranges of values $\lambda_{wd} \in$ {0, 0.01, 0.001, 0.0001, 0.00001} and $N_H \in$ {10, 50, 100, 500}. The length of the input window (i.e. the input order of the delay network) was set to 10 (corresponding to 2.5 seconds), which is in line with typical settings adopted in literature (see Section 2). IDNNs were trained for a maximum number of 5000 epochs, stopping the training process whenever the error on the training set was stable for 10 consecutive epochs.

The predictive performance of the ARS systems proposed was assessed by computing accuracy and F1 score obtained for each learning task in Table 2, in accordance to the evaluating criteria of the EvAAL competition (see e.g. [76,79,86]). In particular, the performance of the models on each activity was evaluated by computing per-class (i.e. per-activity) measures of accuracy, recall, precision and F1 score, according to:

$$accuracy_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}$$

$$recall_i = \frac{tp_i}{tp_i + fn_i}$$

$$precision_i = \frac{tp_i}{tp_i + fp_i} \qquad (8)$$

$$F1_i = 2\frac{precision_i \, recall_i}{precision_i + recall_i}$$

where N_i is the number of samples available for the classification task involving the i-th activity, tp_i, tn_i, fp_i, fn_i are respectively the number of true positive, true negative, false positive and false negative classifications for the i-th activity. The overall performance of the ARS systems (multi-classification case) were com-

puted by averaging per-class accuracy, recall and precision over the activities:

$$accuracy_{av} = \frac{1}{K} \sum_{i=1}^{K} accuracy_i$$

$$recall_{av} = \frac{1}{K} \sum_{i=1}^{K} recall_i \qquad (9)$$

$$precision_{av} = \frac{1}{K} \sum_{i=1}^{K} precision_i$$

Based on the averaged recall and precision, the F1 measure for the case of multi-classification is computed as a macro-F1 score:

$$F1_{macro} = 2\frac{(precision_{av} recall_{av})}{precision_{av} + recall_{av}} \qquad (10)$$

A further measure of the performance of the models in the multi-classification case is given by the K-class classification accuracy, i.e. the accuracy computed over the prediction matrix for the complete set of activities (confusion matrix), defined as:

$$accuracy_K = \frac{\sum_{i=1}^{K} tp_i}{\sum_{i=1}^{K} N_i} \qquad (11)$$

where the maximum value of K used in our paper is 7.

In the results reported in the following, performances (accuracy and F1 scores) were averaged, and standard deviations were computed, over the 5 reservoir guesses considered for each LI-ESN hyper-parametrization. Analogously, for the experiments with IDNNs the performance corresponding to each network setting was averaged over 5 different runs, corresponding to different initializations of the weight values from a uniform distribution in $[-0.001, 0.001]$. In particular, it should be noted that the obtained standard deviations are very small in correspondence of every experimental setting considered (in the order of 0.1 percentage points for the accuracy and of 0.1–1 percentage points for the F1 score), not affecting the performance evaluation nor the comparison among the different cases. As such, for the sake of compactness of results presentation, standard deviations are not explicitly shown in the Tables reported in this section.

The performance achieved by LI-ESNs on the multi-classification learning task corresponding to the RSS-based ARS setting is shown in Table 3, which re-

Table 3

Test set per-class accuracy and F1 score achieved by LI-ESNs and IDNNs on the activities in the RSS-based ARS setting. The last row reports the accuracy averaged over the different activities and the macro-F1 score

Activity	Accuracy		F1	
	LI-ESN	*IDNN*	*LI-ESN*	*IDNN*
Bending1	0.981	0.968	0.897	0.833
Bending2	0.977	0.947	0.842	0.620
Cycling	0.999	0.961	0.997	0.888
Lying	0.976	0.900	0.931	0.632
Sitting	0.825	0.850	0.464	0.297
Standing	0.855	0.862	0.552	0.657
Walking	0.999	0.972	0.997	0.913
Overall	0.944	0.923	0.816	0.723

ports the test per-class accuracy and F1 score for each activity, along with the averaged accuracy (Eq. (9)) and macro-F1 (Eq. (10)) of the resulting ARS system. For performance comparison, Table 3 also reports the performance achieved by IDNNs in the same experimental conditions. Moreover, the confusion matrices for the two cases of LI-ESNs and IDNNs are graphically illustrated in Fig. 10, considering an enumeration of the activities according to the order in Table 3. Figure 10 shows for each case a compact graphical representation where each row of the matrix corresponds to the class assigned by the classifier (Output), while each column represents the activity actually performed (Target). A classification method with ideal performance will only have bars (Frequency) on the main diagonal of the matrix. The more bars on the non-diagonal cells are high, the worst the classification performance. There results show that the performance obtained by LI-ESNs in the RSS-based setting is already good, with averaged test accuracy of 0.944, macro F1 score of 0.816 and K-class classification accuracy of 0.805. As can be seen from Table 3 and Fig. 10, in general, single activities are well recognized, with per-activity accuracy between 0.825 and 0.999. However, due to the nature of the RSS input signals used, the activities Sitting and Standing results to be hardly distinguishable between each other (see Fig. 10a), with F1 scores close to 0.5. Specifically, the activities Sitting and Standing, as described in the EvAAL technical annex, do not include the phases of sitting down and standing up, in which the proposed system presents good results [1,61]. Indeed, when the user is seated, even though the sensors are closer than the standing still position, the most part of the body along with the chair is interposed between
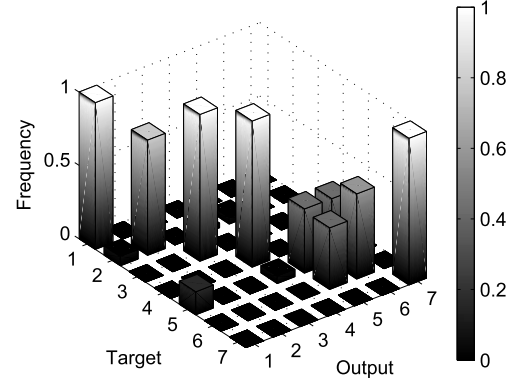
(a) RSS-based LI-ESN (AReM)
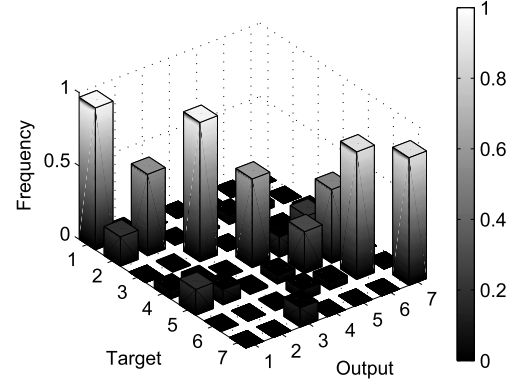
(b) RSS-based IDNN

Fig. 10. Confusion matrices of the performance in the multi-classification learning task obtained by the proposed LI-ESN based (AReM) system (a) and the IDNN based one (b), respectively. The two axes on the base of each graph represent the target activity class and the output class predicted by the system, respectively. The height of the bars represent the number of instances normalized to the total number of target elements in the corresponding class (Frequency). The smaller the bars outside of the main diagonal, the better is the performance. Activities are enumerated according to the same order as in Table 3: 1 for Bending1, 2 for Bending2, 3 for Cycling, 4 for Lying, 5 for Sitting, 6 for Standing, 7 for Walking.

the sensors, compensating the increment of RSS due to the closer positions. This makes the resulting RSS streams for Sitting practically very similar to the ones collected for Standing. Furthermore, it is worth to note that both such activities, involving the user in a still position, do not significantly perturb the streams of gathered RSS data. Table 3 also shows that the performance achieved by LI-ESNs is also generally higher than the one achieved by IDNNs. Indeed, in the RSS-based setting, IDNNs resulted in an averaged test accuracy of 0.923, macro F1 score of 0.723 and K-class classification accuracy of 0.675, which are respectively

Table 4

Test set per-class accuracy and F1 score achieved by LI-ESNs and IDNNs on the activities pertaining to the Heterogeneous ARS setting, task 1 (involving the activities Cycling and Standing). The last row reports the accuracy averaged over the different activities and the macro-F1 score

Activity	Accuracy		F1	
	LI-ESN	*IDNN*	*LI-ESN*	*IDNN*
Cycling	1.000	0.995	1.000	0.995
Standing	1.000	0.995	1.000	0.995
Overall	1.000	0.995	1.000	0.995

Table 5

Test set per-class accuracy and F1 score achieved by LI-ESNs and IDNNs on the activities pertaining to the Heterogeneous ARS setting, task 2 (involving the activities Bending1, Standing and Walking). The last row reports the accuracy averaged over the different activities and the macro-F1 score

Activity	Accuracy		F1	
	LI-ESN	*IDNN*	*LI-ESN*	*IDNN*
Bending1	0.996	0.987	0.990	0.969
Standing	0.995	0.996	0.994	0.995
Walking	1.000	1.000	1.000	1.000
Overall	0.997	0.995	0.995	0.988

2.1, 9.3 and 13 percentage points lower than the results obtained with LI-ESNs. Moreover, the proposed ARS setting based on LI-ESNs outperforms the analogous case with IDNNs also in terms of per-class accuracy for all the considered activities, with the exception of the aforementioned Sitting and Standing. In such cases, indeed, the per-class accuracy achieved by IDNNs is slightly higher than the one obtained by LI-ESNs, reflecting the fact that on specific tasks the windowing approach can be effective, and that in principle it is possible to obtain a good performance on single specific tasks using the LI-ESN recurrent approach or the IDNN windowing one. However, as a global result of an activity recognition system over a whole set of activities, a recurrent approach (such is the LI-ESN based one), which does not make use of an input window of a-priori determined length, results in a system of a more general applicability with better average performance, in particular in cases in which no specific information is available for ad hoc tuning of the window length.

The performance of the LI-ESNs on the 3 multi-classification learning tasks corresponding to the Heterogeneous ARS setting are reported in Tables 4, 5 and 6, which also present the results of IDNNs in the same experimental settings, for the sake of comparison.

Table 6

Test set per-class accuracy and F1 score achieved by LI-ESNs and IDNNs on the activities pertaining to the Heterogeneous ARS setting, task 3 (involving the activities Bending2, Sitting, Lying). The last row reports the accuracy averaged over the different activities and the macro-F1 score

Activity	Accuracy		F1	
	LI-ESN	*IDNN*	*LI-ESN*	*IDNN*
Bending2	0.941	0.889	0.832	0.728
Sitting	0.933	0.783	0.914	0.753
Lying	0.930	0.830	0.920	0.769
Overall	0.935	0.834	0.896	0.754

Results shown in such tables are computed by considering for each learning task only the sequences pertaining to the involved activities. In general, LI-ESNs showed very good performances on these 3 learning tasks, with average accuracy and macro F1 score above 0.93 and 0.89, respectively, for every task. In particular, for task 1, involving the activities Cycling and Standing (see Table 4), average accuracy and macro F1 score are both equal to 1. For task 2, involving the activities Bending1, Standing and Walking (see Table 5), average accuracy and macro F1 score are 0.997 and 0.995, respectively. For task 3, involving the activities Bending2, Sitting and Lying (see Table 6), average accuracy and macro F1 score are 0.935 and 0.896, respectively. It is also worth to note that Tables 4, 5 and 6 show that LI-ESNs generally outperform IDNNs on the three tasks considered in the Heterogeneous ARS setting. Indeed, the performance achieved by LI-ESNs improves the one obtained by IDNNs up to 10.1 and 14.2 percentage points, respectively, in terms of averaged accuracy and macro F1. The performance improvement obtained by the LI-ESN approach is particularly evident in correspondence of the activities in task 3, i.e. Bending2, Sitting and Lying (see Table 6).

The performance evaluation of the Heterogeneous ARS system as a whole is computed by aggregating the results of the learning models for tasks 1, 2 and 3, according to the decision tree in Fig. 4. It is worth to recall that in this case the classification output of the activity recognition system is the activity corresponding to the readout unit with the highest activation (see Eq. (5)) in the LI-ESN activated by the decision tree in Fig. 4 (analogously for the case of IDNNs).

Table 7 reports the test performance achieved by LI-ESNs and IDNNs for the Heterogeneous setting, considering all the sequences for all the activities. Note that results for the activity Standing are averaged be-

Table 7

Test set per-class accuracy and F1 score achieved by LI-ESNs and IDNNs on the activities in the Heterogeneous ARS setting. The last row reports the accuracy averaged over the different activities and the macro-F1 score

Activity	Accuracy		F1	
	LI-ESN	*IDNN*	*LI-ESN*	*IDNN*
Bending1	1.000	0.995	1.000	0.972
Bending2	0.976	0.950	0.832	0.728
Cycling	1.000	0.998	1.000	0.995
Lying	0.971	0.930	0.920	0.722
Sitting	0.972	0.911	0.914	0.771
Standing	1.000	0.997	1.000	0.992
Walking	1.000	1.000	1.000	1.000
Overall	0.988	0.969	0.956	0.885

tween the two possible cases represented in task 1 and task 2 (i.e. proximity and non-proximity to the bicycle). Graphical representations of the confusion matrices for the results of LI-ESNs and IDNNs under the Heterogeneous ARS setting are shown in Fig. 11. Figure 12 reports the output of the LI-ESN (Fig. 12a) and of the IDNN (Fig. 12b) models in correspondence of the Heterogeneous ARS setting. For the sake of clarity of representation, data pertaining to the different activities are grouped in Fig. 12 based on the ground-truth and concatenated according to the same order as in Tables 3 and 7.

As it can be observed in Table 7 and in Figs 11 and 12, results obtained by LI-ESN on single activities are extremely good, with test average accuracy between 0.971 and 1.000 and F1 scores between 0.832 and 1.000. In particular, the performance obtained for the two activities Standing and Sitting is greatly improved with respect to the RSS-based setting, showing the great advantage of the availability of the augmented set of sensor data to the ARS, which is also appreciable through a comparison between the confusion matrices in Figs 10 and 11. Averaged accuracy and macro F1 score for the Heterogeneous setting with LI-ESNs are 0.988 and 0.956, respectively, which represents an improvement of 4.4 and 14 percentage points with respect to the RSS-based case. The K-class classification accuracy for the Heterogeneous setting with LI-ESNs is 0.959, which overcomes the K-class classification in the corresponding RSS-based case of more that 15 percentage points. Moreover, the comparison between the results obtained by LI-ESN and by IDNN in the Heterogeneous ARS setting, reported in Table 7 and Figs 11 and 12, confirms the general superiority of the LI-ESN approach

in this application context. Indeed, the performance of LI-ESNs is always higher than or equal to the performance of IDNNs on the single activities considered, with an improvement up to approximately 6 and 20 percentage points in terms of per-class accuracy and F1, respectively. In this regard, it is also worth to observe that the advantage of IDNNs on the accuracy for the activities Sitting and Standing, noted for the RSS-based setting, does not hold anymore in the Heterogeneous setting. IDNNs in the Heterogeneous setting resulted in an average test accuracy of 0.969, macro F1 score of 0.885 and K-class classification accuracy of 0.897, which are respectively approximately 2, 7 and 6 percentage points lower than the results obtained with LI-ESNs. It is also interesting to observe that the results achieved with IDNNs in the two ARS settings considered, although generally inferior to the corresponding ones obtained with the proposed LI-ESN based ARS system, show also in this case a substantial improvement of the performance obtained in the Heterogeneous setting with respect to the RSS-based one. This aspect represents a further confirmation of the effectiveness of the use of the augmented set of available sensors within the decision tree fusion mechanism shown in Fig. 4 and described in Section 3.

6. Conclusion

In this work, we presented AReM, an activity recognition system that exploits the strengths of the RC approach for modeling RNNs, which resulted to be particularly appropriate for the identification and classification of temporal pattern in time series. Data collected from both environmental and wearable sensors were used as input for the classification process, realized by means of two classification layers. The first layer is based on a custom decision tree that exploits consolidated signal processing techniques in order to discern and group static features extracted from the data in terms of usefulness for the following steps of classification. The second layer is based on a set of RC networks implemented as LI-ESNs, realizing the final activity recognition step.

Our activity recognition system has been experimentally validated on a real-world activity recognition dataset, collected with iterative rounds of measurements campaigns. Extremely good performance results were obtained on an external test set, both in terms of average accuracy, up to 0.988, and in terms
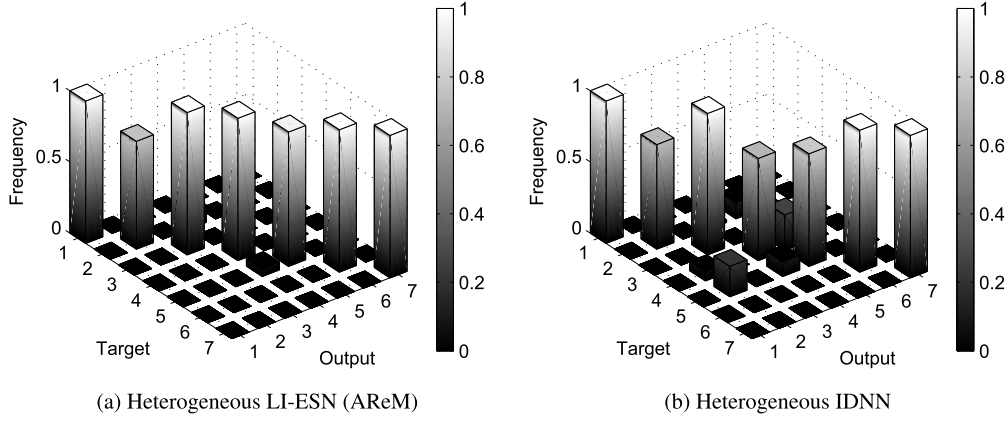
(a) Heterogeneous LI-ESN (AReM) (b) Heterogeneous IDNN

Fig. 11. Confusion matrices of the performance obtained under the Heterogeneous ARS settings by the proposed LI-ESN based (AReM) system (a) and the IDNN based one (b), respectively. Activities are enumerated according to the same order as in Tables 3 and 7: 1 for Bending1, 2 for Bending2, 3 for Cycling, 4 for Lying, 5 for Sitting, 6 for Standing, 7 for Walking.
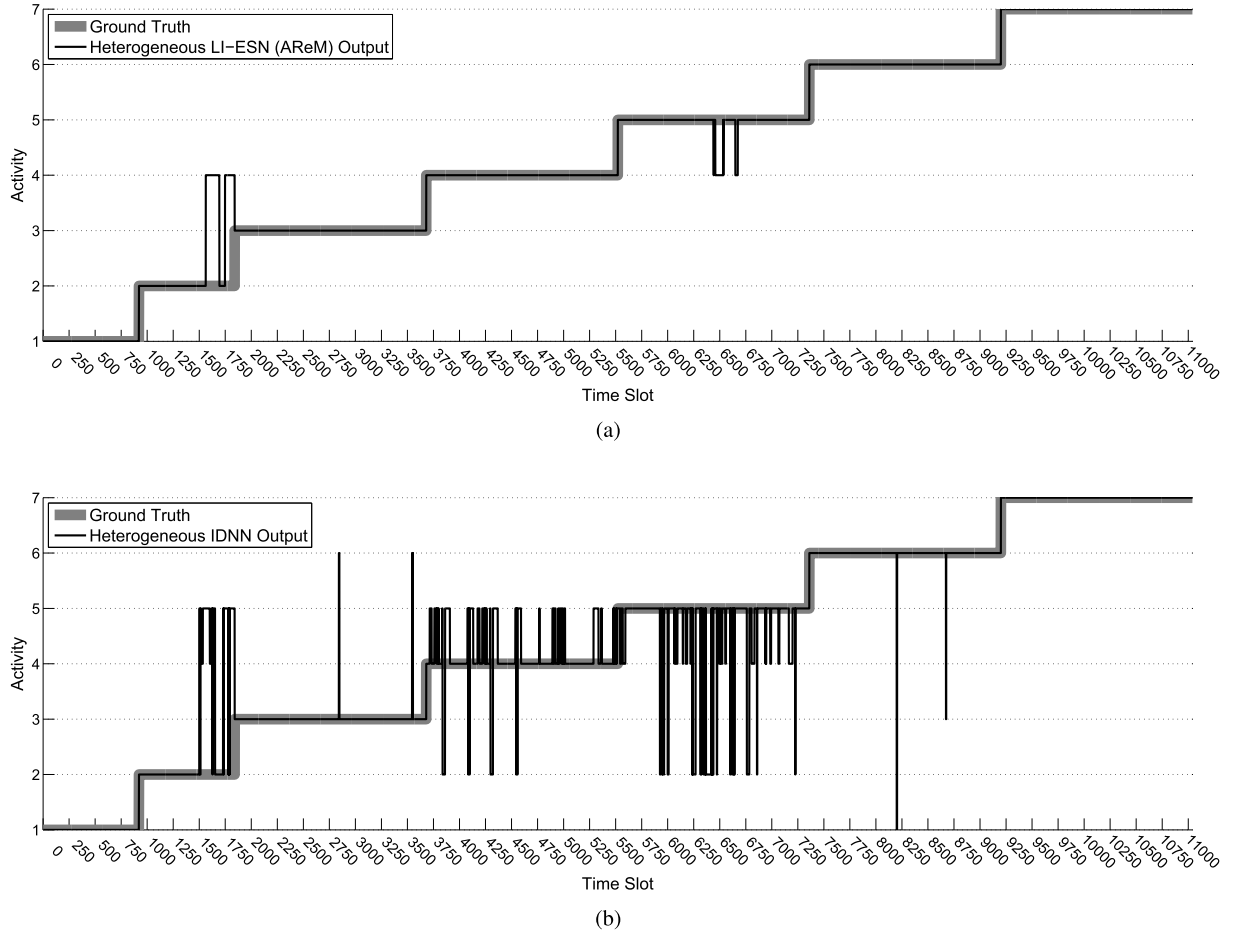

(a)

(b)

Fig. 12. Graphical representation of the results obtained by LI-ESN (AReM) (a) and by IDNN (b) under the Heterogeneous ARS setting on the test set of the activity recognition dataset. X-axis represents the progressive time-slots of 250 milliseconds. Y-axis represents the activity: 1 for Bending1, 2 for Bending2, 3 for Cycling, 4 for Lying, 5 for Sitting, 6 for Standing, 7 for Walking.

of macro F1 score, up to 0.956. The effectiveness of the refined multi-sensor processing chain has been assessed through a comparison of the performances achieved under different settings of the activity recognition system. In particular we took into consideration the case in which only RSS data from wearable sensors are available, and the case in which additional (environmental and wearable) sensors data can be used. The latter setting, allowing to better focus the training on specific sub-sets of activities, resulted in a significant performance enhancement. Moreover, for the sake of performance comparison with our proposed model, we took into consideration the competitive IDNN approach as a paradigmatic representative of the class of delay networks, a popular choice for neural networks learning in human activity recognition problems. This comparison on the one hand showed that the results obtained by the proposed LI-ESN approach generally outperform those achieved by IDNN in every setting considered, and on the other hand further confirmed the effectiveness of the proposed multi-sensor processing chain with the augmented set of available sensors.

Overall, the activity recognition system proposed in this paper intends to provide a flexible and robust solution for the identification of user's activities. The system is particularly interesting since it can constitute the premise not only for user monitoring applications, but can be part of a more general solution for ambient assisted living. To this purpose, as future work, we plan to use the proposed system in real test sites coming from EU funded project DOREMI[4] once they will become available for the experimentation.

Acknowledgements

References

[1] J.A. Álvarez-García, P. Barsocchi, S. Chessa and D. Salvi, Evaluation of localization and activity recognition systems for ambient assisted living: The experience of the 2012 evaal competition, *Journal of Ambient Intelligence and Smart Environments* **5**(1) (2013), 119–132.

[4] http://www.doremi-fp7.eu.

[2] G. Amato, D. Bacciu, M. Broxvall, S. Chessa, S. Coleman, M. Di Rocco, M. Dragone, C. Gallicchio, C. Gennaro, H. Lozano, T. McGinnity, A. Micheli, A. Ray, A. Renteria, A. Saffiotti, D. Swords, C. Vairo and P. Vance, Robotic Ubiquitous Cognitive Ecology for Smart Homes, *Journal of Intelligent & Robotic Systems* (2015), 1–25.

[3] G. Amato, M. Broxvall, S. Chessa, M. Dragone, C. Gennaro, R. Lopez, L. Maguire, T. McGinnity, A. Micheli, R. Renteria, G.O. Hare and F. Pecora, Robotic ubiquitous cognitive network, in: *Ambient Intelligence – Software and Applications*, P. Novais, K. Hallenborg, D.I. Tapia and J.M.C. Rodriguez, eds, Advances in Intelligent and Soft Computing, Vol. 153, Springer, Berlin, Heidelberg, 2012, pp. 191–195.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra and J.L. Reyes-Ortiz, Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine, in: *Ambient Assisted Living and Home Care*, Springer, 2012, pp. 216–223.

[5] D. Bacciu, P. Barsocchi, S. Chessa, C. Gallicchio and A. Micheli, An experimental characterization of reservoir computing in ambient assisted living applications, *Neural Computing and Applications* **24**(6) (2014), 1451–1464.

[6] D. Bacciu, S. Chessa, C. Gallicchio, A. Lenzi, A. Micheli and S. Pelagatti, A general purpose distributed learning model for robotic ecologies, in: *Proc. of the 10th International IFAC Symposium on Robot Control (SYROCO)*, Robot Control, Vol. 10, no. 1, 2012, pp. 435–440.

[7] D. Bacciu, S. Chessa, C. Gallicchio, A. Micheli and P. Barsocchi, An experimental evaluation of reservoir computation for ambient assisted living, in: *Neural Nets and Surroundings*, B. Apolloni, S. Bassis, A. Esposito and F.C. Morabito, eds, Smart Innovation, Systems and Technologies, Vol. 19, Springer, Berlin, Heidelberg, 2013, pp. 41–50.

[8] D. Bacciu, C. Gallicchio, A. Micheli, S. Chessa and P. Barsocchi, Predicting user movements in heterogeneous indoor environments by reservoir computing, in: *Proc. of the IJCAI Workshop on Space, Time and Ambient Intelligence (STAMI) 2011*, M. Bhatt, H.W. Guesgen and J.C. Augusto, eds, 2011, pp. 1–6.

[9] D. Bacciu, C. Gallicchio, A. Micheli, M. Di Rocco and A. Saffiotti, Learning context-aware mobile robot navigation in home environments, in: *Proc. of the 5th International Conference on Information, Intelligence, Systems and Applications, IISA 2014*, IEEE, 2014, pp. 57–62.

[10] L. Bao and S.S. Intille, Activity recognition from user-annotated acceleration data, in: *Pervasive Computing*, Springer, 2004, pp. 1–17.

[11] R. Barbuti, S. Chessa, A. Micheli and R. Pucci, Identification of nesting phase in tortoise populations by neural networks, Extended abstract, *The 50th Anniversary Convention of the AISB*, selected papers, 2013, pp. 62–65.

[12] P. Barsocchi, S. Chessa, A. Micheli and C. Gallicchio, Forecast-driven enhancement of received signal strength (RSS)-based localization systems, *ISPRS International Journal of Geo-Information* **2**(4) (2013), 978–995.

[13] P. Barsocchi, F. Potortì and P. Nepa, Device-free indoor localization for AAL applications, in: *Wireless Mobile Communication and Healthcare*, Springer, 2013, pp. 361–368.

[14] M. Bocca, O. Kaltiokallio and N. Patwari, Radio tomographic imaging for ambient assisted living, in: *Evaluating AAL Systems Through Competitive Benchmarking*, Springer, 2013, pp. 108–130.

[15] J.A. Botía, J.A.Á. García, K. Fujinami, P. Barsocchi and T. Riedel, *Evaluating AAL Systems Through Competitive Benchmarking*, Springer, 2013.

[16] A. Bourke, J. O'Brien and G. Lyons, Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm, *Gait & Posture* **26**(2) (2007), 194–199.

[17] T. Brezmes, J.-L. Gorricho and J. Cotrina, Activity recognition from accelerometer data on a mobile phone, in: *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, Springer, 2009, pp. 796–799.

[18] C.-Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A.A. Sawchuk and A.A. Rizzo, Towards pervasive physical rehabilitation using Microsoft kinect, in: *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, IEEE, 2012, pp. 159–162.

[19] S. Chessa, C. Gallicchio, R. Guzman and A. Micheli, Robot localization by echo state networks using RSS, in: *Recent Advances of Neural Network Models and Applications*, S. Bassis, A. Esposito and F.C. Morabito, eds, Smart Innovation, Systems and Technologies, Vol. 26, Springer International Publishing, 2014, pp. 147–154.

[20] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. Legrand, A. Rahimi, A. Rea, G. Bordello, B. Hemingway, P. Klasnja, K. Koscher, J. Landay, J. Lester, D. Wyatt and D. Haehnel, The mobile sensing platform: An embedded activity recognition system, *IEEE Pervasive Computing* **7**(2) (2008), 32–41.

[21] S. Coradeschi, A. Cesta, G. Cortellessa, L. Coraci, J. Gonzalez, L. Karlsson, F. Furfari, A. Loutfi, A. Orlandini, F. Palumbo et al., Giraffplus: Combining social interaction and long term monitoring for promoting independent living, in: *2013 The 6th International Conference on Human System Interaction (HSI)*, IEEE, 2013, pp. 578–585.

[22] M.Á.Á. de la Concepción, L.M.S. Morillo, L.G. Abril and J.A.O. Ramírez, Activity recognition system using non-intrusive devices through a complementary technique based on discrete methods, in: *Evaluating AAL Systems Through Competitive Benchmarking*, Springer, 2013, pp. 36–47.

[23] K. Ellis, S. Godbole, J. Kerr and G. Lanckriet, Multi-sensor physical activity recognition in free-living, in: *Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM, 2014, pp. 431–440.

[24] M. Ermes, J. Parkka, J. Mantyjarvi and I. Korhonen, Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions, *IEEE Transactions on Information Technology in Biomedicine* **12**(1) (2008), 20–26.

[25] C. Gallicchio and A. Micheli, Architectural and Markovian factors of echo state networks, *Neural Networks* **24**(5) (2011), 440–456.

[26] C. Gallicchio, A. Micheli, P. Barsocchi and S. Chessa, User movements forecasting by reservoir computing using signal streams produced by mote-class sensors, in: *Mobile Lightweight Wireless Systems (Mobilight 2011)*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 81, Springer, Berlin, Heidelberg, 2012, pp. 151–168.

[27] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edn, Prentice Hall, 1999.

[28] Z. He and L. Jin, Activity recognition from acceleration data using AR model representation and SVM, in: *Proc. of the International Conference on Machine Learning and Cybernetics*, Vol. 4, IEEE, 2008, pp. 2245–2250.

[29] Z. He and L. Jin, Activity recognition from acceleration data based on discrete consine transform and SVM, in: *IEEE International Conference on Systems, Man and Cybernetics, 2009, SMC 2009*, IEEE, 2009, pp. 5041–5044.

[30] U. Hunkeler, H.L. Truong and A. Stanford-Clark, MQTT-S a publish/subscribe protocol for wireless sensor networks, in: *3rd International Conference on Communication Systems Software and Middleware and Workshops, 2008, COMSWARE 2008*, IEEE, 2008, pp. 791–798.

[31] D.P. Hunt, D. Parry and S. Schliebs, Exploring the applicability of reservoir methods for classifying punctual sports activities using on-body sensors, in: *Proc. of the Thirty-Seventh Australasian Computer Science Conference*, Vol. 147, Australian Computer Society, Inc., 2014, pp. 67–73.

[32] N. Imamoglu, Z. Wei, H. Shi, Y. Yoshida, M. Nergui, J. Gonzalez, D. Gu, W. Chen, K. Nonami and W. Yu, An improved saliency for RGB-D visual tracking and control strategies for a bio-monitoring mobile robot, in: *Evaluating AAL Systems Through Competitive Benchmarking*, Springer, 2013, pp. 1–12.

[33] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks, Tech. rep., GMD – German National Research Institute for Computer Science, 2001.

[34] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**(5667) (2004), 78–80.

[35] H. Jaeger, M. Lukoševičius, D. Popovici and U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, *Neural Networks* **20**(3) (2007), 335–352.

[36] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell and B.G. Celler, Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, *IEEE Transactions on Information Technology in Biomedicine* **10**(1) (2006), 156–167.

[37] A.M. Khan, Y. Lee, S. Lee and T.-S. Kim, A tri-axial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer, *IEEE Transactions on Information Technology in Biomedicine* **14**(5) (2010), 1166–1172.

[38] S. Khawandi, B. Daya and P. Chauvet, Automated monitoring system for fall detection in the elderly, *International Journal of Image Processing* **4**(5) (2010).

[39] J. Kolen and S. Kremer (eds), *A Field Guide to Dynamical Recurrent Networks*, IEEE Press, 2001.

[40] S. Kozina, H. Gjoreski, M. Gams and M. Luštrek, Efficient activity recognition and fall detection using accelerometers, in: *Evaluating AAL Systems Through Competitive Benchmarking*, Springer, 2013, pp. 13–23.

[41] H.I. Krebs, J.J. Palazzolo, L. Dipietro, M. Ferraro, J. Krol, K. Rannekleiv, B.T. Volpe and N. Hogan, Rehabilitation robotics: Performance-based progressive robot-assisted therapy, *Autonomous Robots* **15**(1) (2003), 7–20.

[42] N.C. Krishnan and D.J. Cook, Activity recognition on streaming sensor data, *Pervasive and Mobile Computing* **10** (2014), 138–154.

[43] J.R. Kwapisz, G.M. Weiss and S.A. Moore, Activity recognition using cell phone accelerometers, *ACM SigKDD Explorations Newsletter* **12**(2) (2011), 74–82.

[44] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A.S. Rizzo and M. Bolas, Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft kinect sensor, in: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, IEEE, 2011, pp. 1831–1834.

[45] O.D. Lara and M.A. Labrador, A mobile platform for real-time human activity recognition, in: *Proc. of the Consumer Communications and Networking Conference (CCNC)*, IEEE, 2012, pp. 667–671.

[46] O.D. Lara and M.A. Labrador, A survey on human activity recognition using wearable sensors, *IEEE Communications Surveys & Tutorials* **15**(3) (2013), 1192–1209.

[47] H. Leutheuser, D. Schuldhaus and B.M. Eskofier, Hierarchical, multi-sensor based classification of daily life activities: Comparison with state-of-the-art algorithms using a benchmark dataset, *PloS One* **8**(10) (2013), e75196.

[48] Q. Li, J.A. Stankovic, M.A. Hanson, A.T. Barth, J. Lach and G. Zhou, Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information, in: *Sixth International Workshop on Wearable and Implantable Body Sensor Networks, 2009, BSN 2009*, IEEE, 2009, pp. 138–143.

[49] L. Liao, D. Fox and H. Kautz, Location-based activity recognition using relational Markov networks, in: *Proc. of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05*, Morgan Kaufmann Publishers Inc., 2005, pp. 773–778.

[50] J. Liju and V. Koshy, Accelerometer-based activity recognition in construction, *Journal of Computing in Civil Engineering* **25**(5) (2011), 370–379.

[51] X. Long, B. Yin and R.M. Aarts, Single-accelerometer-based daily physical activity classification, in: *Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, EMBC 2009*, IEEE, 2009, pp. 6107–6110.

[52] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, *Computer Science Review* **3**(3) (2009), 127–149.

[53] M. Lukoševičius, H. Jaeger and B. Schrauwen, Reservoir computing trends, *KI-Künstliche Intelligenz* **26**(4) (2012), 365–371.

[54] A. Mannini and A.M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, *Sensors* **10**(2) (2010), 1154–1175.

[55] U. Maurer, A. Smailagic, D. Siewiorek and M. Deisher, Activity recognition and monitoring using multiple sensors on different body positions, in: *Proc. of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN)*, IEEE, 2006, pp. 116–119.

[56] F. Moiz, P. Natoo, R. Derakhshani and W.D. Leon-Salas, A comparative study of classification methods for gesture recognition using a 3-axis accelerometer, in: *2011 International Joint Conference on Neural Networks, (IJCNN)*, IEEE, 2011, pp. 2479–2486.

[57] D.N. Monekosso and P. Remagnino, Behavior analysis for assisted living, *IEEE Transactions on Automation Science and Engineering* **7**(4) (2010), 879–886.

[58] R. Nathan, O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski and W.M. Getz, Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: General concepts and tools illustrated for griffon vultures, *The Journal of Experimental Biology* **215**(6) (2012), 986–996.

[59] M. Oczak, S. Viazzi, G. Ismayilova, L.T. Sonoda, N. Roulston, M. Fels, C. Bahr, J. Hartung, M. Guarino, D. Berckmans et al., Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network, *Biosystems Engineering* **119** (2014), 89–97.

[60] F. Palumbo, P. Barsocchi, F. Furfari and E. Ferro, AAL middleware infrastructure for green bed activity monitoring, *Journal of Sensors* **2013** (2013).

[61] F. Palumbo, P. Barsocchi, C. Gallicchio, S. Chessa and A. Micheli, Multisensor data fusion for activity recognition based on reservoir computing, in: *Evaluating AAL Systems Through Competitive Benchmarking*, J. Botía, J. Alvarez-Garcia, K. Fujinami, P. Barsocchi and T. Riedel, eds, Communications in Computer and Information Science, Vol. 386, Springer, Berlin, Heidelberg, 2013, pp. 24–35.

[62] F. Palumbo, D. La Rosa and S. Chessa, GP-m: Mobile middleware infrastructure for ambient assisted living, in: *2014 IEEE Symposium on Computers and Communication (ISCC)*, IEEE, 2014, pp. 1–6.

[63] F. Palumbo, J. Ullberg, A. Štimec, F. Furfari, L. Karlsson and S. Coradeschi, Sensor network infrastructure for a home care monitoring system, *Sensors* **14**(3) (2014), 3833–3860.

[64] J. Parkka, L. Cluitmans and M. Ermes, Personalization algorithm for real-time activity recognition using PDA, wireless motion bands, and binary decision tree, *IEEE Transactions on Information Technology in Biomedicine* **14**(5) (2010), 1211–1215.

[65] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola and I. Korhonen, Activity classification using realistic data from wearable sensors, *IEEE Transactions on Information Technology in Biomedicine* **10**(1) (2006), 119–128.

[66] N. Patwari and A.O. Hero III, Using proximity and quantized RSS for sensor localization in wireless networks, in: *Proc. of the 2nd ACM International Conference on Wireless Sensor Networks and Applications*, ACM, 2003, pp. 20–29.

[67] T. Peterek, M. Penhaker, P. Gajdoš and P. Dohnálek, Comparison of classification algorithms for physical activity recognition, in: *Innovations in Bio-Inspired Computing and Applications*, Springer, 2014, pp. 123–131.

[68] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, H. Kautz and D. Hahnel, Inferring activities from interactions with objects, *IEEE Pervasive Computing* **3**(4) (Oct. 2004), 50–57.

[69] A. Purwar, D. un Jeong and W.Y. Chung, Activity monitoring from real-time triaxial accelerometer data using sensor network, in: *International Conference on Control, Automation and Systems, 2007, ICCAS'07*, IEEE, 2007, pp. 2402–2406.

[70] L. Rabiner and B.-H. Juang, An introduction to hidden Markov models, *IEEE ASSP Magazine* **3**(1) (1986), 4–16.

[71] C. Randell and H. Muller, Context awareness by analysing accelerometer data, in: *The Fourth International Symposium on Wearable Computers*, IEEE, 2000, pp. 175–176.

[72] P. Rashidi and A. Mihailidis, A survey on ambient-assisted living tools for older adults, *IEEE Journal of Biomedical and Health Informatics* **17**(3) (2013), 579–590.

[73] N. Ravi, N. Dandekar, P. Mysore and M.L. Littman, Activity recognition from accelerometer data, in: *Proc. of the Seven-*

teenth Conference on Innovative Applications of Artificial Intelligence (IAAI), AAAI Press, 2005, pp. 1541–1546.

[74] V.N.T. Sang, N.D. Thang, V. Van Toi, N.D. Hoang and T.Q.D. Khoa, Human activity recognition and monitoring using smartphones, in: *5th International Conference on Biomedical Engineering in Vietnam*, Springer, 2015, pp. 481–485.

[75] M.D. Skowronski and J.G. Harris, Minimum mean squared error time series classification using an echo state network prediction model, in: *Proc. of the 2006 IEEE International Symposium on Circuits and Systems, ISCAS*, IEEE, 2006, p. 4.

[76] M. Sokolova and G. Lapalme, A systematic analysis of performance measures for classification tasks, *Information Processing & Management* **45**(4) (2009), 427–437.

[77] P. Tiňo, B. Hammer and M. Bodén, Markovian bias of neural-based architectures with feedback connections, in: *Perspectives of Neural-Symbolic Integration*, Springer, 2007, pp. 95–133.

[78] A.C. Tsoi and A.D. Back, Discrete time recurrent neural network architectures: A unifying review, *Neurocomputing* **15**(3–4) (1997), 183–223.

[79] T. van Kasteren, H. Alemdar and C. Ersoy, Effective performance metrics for evaluating activity recognition methods, in: *Proc. of the 24th International Conference on Architecture of Computing Systems (ARCS)*, 2011, pp. 301–310.

[80] K. Van Laerhoven, A. Schmidt and H.-W. Gellersen, Multi-sensor context aware clothing, in: *Proc. of the Sixth International Symposium on Wearable Computers, 2002, (ISWC 2002)*, IEEE, 2002, pp. 49–56.

[81] V.N. Vapnik, An overview of statistical learning theory, *IEEE Transactions on Neural Networks* **10**(5) (1999), 988–999.

[82] D. Verstraeten, B. Schrauwen, M. D'Haene and D. Stroobandt, An experimental unification of reservoir computing methods, *Neural Networks* **20**(3) (2007), 391–403.

[83] A. Waibel, Modular construction of time-delay neural networks for speech recognition, *Neural Computation* **1**(1) (1989), 39–46.

[84] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. Lang, Phoneme recognition using time-delay neural networks, *Acoustics, Speech and Signal Processing* **37**(3) (1989), 328–339.

[85] L. Wang, T. Gu, X. Tao, H. Chen and J. Lu, Recognizing multi-user activities using wearable sensors in a smart home, *Pervasive and Mobile Computing* **7**(3) (2011), 287–298.

[86] J. Ward, P. Lukowicz and H. Gellersen, Performance metrics for activity recognition, *ACM Trans. Intell. Syst. Technol.* **2**(1) (Jan. 2011), 1–23.

[87] T. Wenlong and E. Sazonov, Highly accurate recognition of human postures and activities through classification with rejection, *IEEE Journal of Biomedical and Health Informatics* **18**(1) (2014), 309–315.

[88] D.H. Wilson and C. Atkeson, Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors, in: *Pervasive Computing*, Springer, 2005, pp. 62–79.

[89] J. Wilson and N. Patwari, Spin: A token ring protocol for RSS collection, 2011.

[90] C. Zhu and W. Sheng, Human daily activity recognition in robot-assisted living using multi-sensor fusion, in: *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2009, pp. 2154–2159.