

Experimental Analysis of Deep Echo State Networks for Ambient Assisted Living

Claudio Gallicchio (✉) and Alessio Micheli

Department of Computer Science, University of Pisa,
Largo B. Pontecorvo 3, Pisa, Italy
gallicch@di.unipi.it, micheli@di.unipi.it

Abstract. The Reservoir Computing (RC) paradigm represents a state-of-the-art methodology for efficient building of recurrent neural networks, which in the last years has proved effective in learning real-world temporal tasks from streams of sensorial data in the Ambient Assisted Living (AAL) domain. Recently, the study of RC networks has been extended to the case of deep architectures, with the introduction of the deep Echo State Network (DeepESN) model. Featured by a layered composition of recurrent units, DeepESNs are inherently able to develop a hierarchically structured representation of temporal information, at the same time preserving the RC characterization of training efficiency.

In this paper, we discuss the introduction of the DeepESN approach in the field of AAL. To this aim, we perform a comparative experimental analysis on two real-world benchmark datasets related to inferring the user's behavior from data streams gathered from the nodes of a wireless sensor network. Results show that DeepESNs outperform standard RC networks with shallow architecture, suggesting a multiple-time scales nature of the involved temporal data and pointing out the great potentiality of the proposed approach in the AAL field.

Keywords: Deep Learning, Reservoir Computing, Deep Echo State Network, Ambient Assisted Living, Human Activity Recognition

1 Introduction

Being able to recognize the behavior of humans in their every-day environments is one of the key objectives of Ambient Assisted Living (AAL) applications. This ability can be indeed exploited in diverse applicative contexts aiming at improving the quality of life of older people, e.g. by monitoring the regularity of their activities, enhancing the degree of personalization of smart home services based on their habits, or anticipating their needs in the place where they live or work. Among the possible solutions, the use of wireless sensor networks (WSN) [10] as a mean to gather relevant data for the purpose of modeling the user's behavior turns out to be a reasonable trade-off between intrusiveness, user acceptance and quality of the data.

In typical AAL scenarios, vast amount of temporal data is generated through the interaction of humans with the sensors onboard the nodes of the deployed WSN. In this context, the interest in the adoption of Machine Learning methodologies to discover relevant patterns from streams of sensorial information is constantly increasing [34]. In particular, the class of Recurrent Neural Networks (RNNs) [33, 29] are recognized for their remarkable ability to effectively approach learning tasks characterized by a distinct sequential/temporal nature in presence of noisy and imprecise data, and it is therefore considered as particularly appropriate to approach the difficulties of the learning problems occurring in the AAL applicative domains [37]. The Reservoir Computing (RC) [40, 35] paradigm and the Echo State Network (ESN) [31, 30] model represent a theoretically grounded methodology [17] for efficiently modeling and train RNNs. Characterized by a huge popularity in many applicative domains involving temporal data processing (see e.g. [36, 40]), the ESN approach has recently gained a great success in real-world AAL-related tasks. Examples of relevant applications in this context include indoor user context localization [9, 11, 5, 21], robot localization [14, 12], adaptive planning in personalized robotic applications [8], human gesture recognition [20], human activity recognition [37, 2] and health care monitoring for medical applications [7, 23, 3]. In addition to this, ESNs have been adopted as core learning methodology in recent European initiatives, such as the FP7 RUBICON¹ (Robotic UBIquitous COgnitive Network) project [1, 13], and the FP7 DOREMI² (Decrease of cOgnitive decline, malnutRition and sedEntariness by elderly empowerment in lifestyle Management and social Inclusion) project [38, 6]. Moreover, the ESN approach has also recently been investigated in perspective of the realization of a learning service for the Internet of Things [4] and promises to greatly help in addressing the challenges involved by interaction between robotic devices and the IoT, in the so-called Internet of Robotic Things [39] framework.

Recently, with the introduction of the DeepESN model in [27, 18], the study of hierarchically organized RC architectures is arousing an increasing interest. Keeping the extreme efficiency of training algorithms as in standard RC networks, DeepESNs are capable of developing progressively more abstract representations of temporal information in the levels of the architecture, potentially allowing to naturally capture the structure of sequential data featured by multiple time-scales.

In this paper we investigate the introduction of the DeepESN approach in learning tasks from streams of sensorial data in the AAL domain. In particular, we provide an experimental analysis on two representative real-world benchmark datasets concerning the identification of user behavior in indoor environment based on data gathered from a small WSN. Specifically, a first dataset regards the prediction of user's spatial context in typical office environments, while a second dataset targets a problem of Human Activity Recognition (HAR). In both the cases the analysis is conducted in comparison to standard shallow

¹ EU FP7 RUBICON project (contract no. 269914), <http://fp7rubicon.eu/>

² EU FP7 DOREMI project (contract no. 611650), <http://www.doremi-fp7.eu/>

ESNs, in order to assess the impact of the proposed approach in this specific applicative context.

The rest of this paper is organized as follows. In Section 2 we describe the major characterizations of the DeepESN model. The results of the experimental assessment on the two real-world benchmark datasets are provided and discussed in Section 3. Finally, Section 4 concludes the paper.

2 Deep Echo State Networks

Within the randomized neural networks framework [16], the RC paradigm [40, 35] and the ESN model [31, 30, 17] represent the state-of-the-art RNN methodology for efficiently learn in temporal domains. ESNs (and in general all types of RC networks) are based on a conceptual and practical distinction between a dynamical component, called *reservoir*, and a feed-forward *readout* tool, which is the only trained part of the network’s architecture. The reservoir implements a randomized (temporal) filter which has the role of embedding the history of input signals received by the network into a state space representation that provides a contextual memory to the system for each new input. The idea is that if such temporal embedding is rich enough, then the representation of the problem at hand in the state space is likely to be solved by a trained linear readout tool, and adaptation of the dynamical reservoir is not necessary. The well known characterization of training efficiency of the RC approach naturally stems from such consideration.

Recently, the ESN approach has been extended in the direction of deep learning with the introduction of the DeepESN model [27, 26, 18], in which the dynamical reservoir part of the network architecture is hierarchically organized into layers. The study of DeepESNs has a twofold objective: on the one hand it aims at the development of efficiently trained deep neural network models for learning in temporal domains, and on the other hand, by putting aside the aspects related to learning of the recurrent connections, it allows to highlight and stress the intrinsic properties of deep recurrent architectures. The analysis conducted so far has shown that the layered composition of RNN layers is indeed able to develop structured and rich representations of temporal information featured by multiple time-scales dynamics [27, 18], as also pointed out through investigations on the short-term memory abilities [15], as well as by means of theoretical studies in the field of dynamical systems [26] and Lyapunov exponents [24, 25]. Under an even broader perspective, the study of DeepESNs also allows to open up interesting discussions about the true nature of deep learning for temporal data processing [28]. On the application side, DeepESNs proved effective in both synthetic and real-world cases, outperforming state-of-the-art results on several versions of the multiple superimposed oscillator task [28] and, recently, showing a very good predictive performance in a medical task related to diagnosis of Parkinson’s disease [22]. Additional details and an up-to-date overview on the advancements in the study of DeepESN can be found in [19].

The reservoir architecture of a DeepESN is composed of N_L layers of recurrent units, where here we assume that each layer has the same size, denoted by N_R . Figure 1 gives a graphical illustration of the hierarchical reservoir organization in a DeepESN. As it can be seen, the DeepESN state computation follows the order in the architectural composition such that the first reservoir layer is fed by the external input, while each successive layer receives in input the output of the previous one.

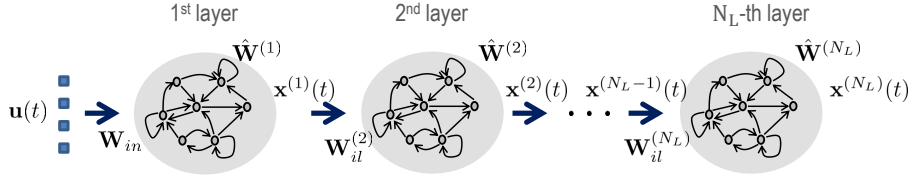


Fig. 1. Layered reservoir architecture of a DeepESN.

From a dynamical system point of view the reservoir of a DeepESN realizes a discrete-time non-linear dynamical system that is driven by the external input signal. More in detail, denoting by $\mathbf{x}^{(i)}(t) \in \mathbb{R}^{N_R}$ the reservoir state of layer i at time step t , we can represent the global state of the whole network at time t by $\mathbf{x}(t) = (\mathbf{x}^{(1)}(t), \mathbf{x}^{(2)}(t), \dots, \mathbf{x}^{(N_L)}(t)) \in \mathbb{R}^{N_L N_R}$, which varies within a global state space that is the product of all the individual reservoir spaces. Under this viewpoint, the entire architecture of the stacked reservoirs is a dynamical system that is ruled by global state transition function F , which expresses how the new state of the network depends on the current input and on its previous state. Such global function F can be decomposed into its layer-wise components, i.e. $F = (F^{(1)}, F^{(2)}, \dots, F^{(N_L)})$, where each $F^{(i)}$, for $i = 1, 2, \dots, N_L$, denotes the state transition function implemented at layer i .

In the following, we shall refer to the case of leaky integrator reservoir units [32] and we will omit the bias terms in equations for the ease of presentation. Denoting by $\mathbf{u}(t) \in \mathbb{R}^{N_U}$ the external input at time step t , the state transition function of the first reservoir layer, i.e. $F^{(1)}$, computes the state of the first reservoir layer as follows:

$$\begin{aligned} \mathbf{x}^{(1)}(t) &= F^{(1)}(\mathbf{u}(t), \mathbf{x}^{(1)}(t-1)) \\ &= (1 - a^{(1)})\mathbf{x}^{(1)}(t-1) + a^{(1)} \tanh(\mathbf{W}_{in}\mathbf{u}(t) + \hat{\mathbf{W}}^{(1)}\mathbf{x}^{(1)}(t-1)), \end{aligned} \quad (1)$$

where, $a^{(1)} \in [0, 1]$ is the leaking rate parameter of the first layer, $\mathbf{W}_{in} \in \mathbb{R}^{N_R \times N_U}$ denotes the input weight matrix, $\hat{\mathbf{W}}^{(1)} \in \mathbb{R}^{N_R \times N_R}$ is the recurrent reservoir weight matrix for the first layer and \tanh denotes the element-wise applied hyperbolic tangent activation function. Following the pipeline on the hierarchical reservoir architecture, each successive layer $i > 1$ is fed by the output of the previous layer at the same time step. Thereby, the state of the reservoir

layer i at time step t is computed by the state transition function $F^{(i)}$ as follows:

$$\begin{aligned} \mathbf{x}^{(i)}(t) &= F^{(i)}(\mathbf{x}^{(i-1)}(t), \mathbf{x}^{(i)}(t-1)) \\ &= (1 - a^{(i)})\mathbf{x}^{(i)}(t-1) + a^{(i)} \tanh(\mathbf{W}_{il}^{(i)}\mathbf{x}^{(i-1)}(t) + \hat{\mathbf{W}}^{(i)}\mathbf{x}^{(i)}(t-1)), \end{aligned} \quad (2)$$

where $a^{(i)} \in [0, 1]$ is the leaking rate parameter of the i -th layer, $\mathbf{W}_{il}^{(i)} \in \mathbb{R}^{N_R \times N_R}$ is the inter-layer reservoir weight matrix for layer i and $\hat{\mathbf{W}}^{(i)} \in \mathbb{R}^{N_R \times N_R}$ is the recurrent reservoir weight matrix for layer i . Note that in the above presented mathematical description of the model, whenever the reservoir comprises only one layer (i.e. for $N_L = 1$), then a standard *shallow* ESN is obtained.

As in the standard RC framework, the reservoir part is left untrained after being initialized under the conditions prescribed by the Echo State Property (ESP) [30, 41], which has been extended to the case of deep RC architectures in [26]. Typically, the reservoir is initialized according to the necessary condition for the ESP, which expresses a stability constraint on the developed network dynamics. Denoting by $\rho(\cdot)$ the spectral radius³ of its matrix argument, the necessary condition for the ESP [26] requires that:

$$\max_{i=1,2,\dots,N_L} \rho((1 - a^{(i)})\mathbf{I} + a^{(i)}\hat{\mathbf{W}}^{(i)}) < 1, \quad (3)$$

where \mathbf{I} is the identity matrix of size $N_R \times N_R$. Thereby, the initialization process of a DeepESN consists in a random initialization (e.g. from a uniform distribution in $[-1, 1]$) of the reservoir in each layer, followed by a re-scaling to ensure that the condition in equation 3 is satisfied. Moreover, matrices \mathbf{W}_{in} and $\mathbf{W}_{il}^{(2)}, \dots, \mathbf{W}_{il}^{(N_L)}$ are randomly initialized (e.g. from a uniform distribution in $[-1, 1]$) and then re-scaled to a desired value of their 2-norms, where $s_{in} = \|\mathbf{W}_{in}\|_2$ act as input scaling parameter and $s_{il}^{(i)} = \|\mathbf{W}_{il}^{(i)}\|_2$ is the inter-layer scaling parameter of the i -th layer.

The output of the DeepESN is computed by the readout tool by means of a linear combination of the reservoir states in the hierarchy. Although different alternatives are possible, here we consider the case in which the input for the readout layer is composed by the the global state of the network. Accordingly, at time step t the output of the DeepESN, denoted by $\mathbf{y}(t) \in \mathbb{R}^{N_Y}$ is computed as follows:

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

where $\mathbf{W}_{out} \in \mathbb{R}^{N_Y \times N_L N_R}$ is the readout weight matrix, which is adjusted on a training set, typically by direct methods such as Moore-Penrose pseudo-inversion or ridge regression [35].

For the purposes of our experimental analysis, we can distinguish two cases of learning tasks on temporal data. Specifically, in the case of sequence-to-sequence tasks an output element is required in correspondence of each input element, and equation 4 is applied to every time step of the computation. In the case of sequence-to-element tasks, the output is required only in correspondence of

³ The maximum among the eigenvalues in modulus.

the last element of the input sequence, and equation 4 is applied only to the last time step (i.e. for each input sequence in correspondence of the last global state of the DeepESN). In case of binary classification tasks, the output of the network is discretized in $\{-1, 1\}$ by applying the sign function to the output in equation 4, whereas in case of multi-class classification tasks the output class label is typically obtained by identifying the readout unit with the maximum activation.

Further details on the DeepESN model can be found in recent literature [27, 26].

3 Experiments

In this section we present the results of the experimental assessment of DeepESNs on two real-world benchmark datasets in the context of AAL applications. Specifically, the adopted datasets are described in Section 3.1, while the experimental settings and results achieved by DeepESNs are comparatively discussed in Section 3.2.

3.1 Datasets

In our experiments we took into consideration two real-world datasets related to the identification of human behavior in indoor environments. Both the datasets have been designed and developed within the activities and collaborations of our research group⁴ as benchmarks in the perspective of being adopted for evaluation purposes of methods to be used in AAL domains. To this aim, both the datasets have been made freely available for download on the prominent and well-known UCI Machine Learning Repository⁵. As regards the specific aims of this paper, the considered datasets are used as useful benchmarks for the assessment of the impact of deep RNN architectures for time-series processing on tasks from sensorial data.

The major features of the adopted datasets are summarized in the following. **Indoor Movement Forecasting.** The first dataset that we take into consideration pertains to scenario of anticipating user movements in a real-world indoor office environment [9]. The prototypical environmental setting that is taken into consideration consists in a couple of rooms separated by a corridor. The user is walking in one of the two rooms and the goal is to anticipate whether she/he will change room or not, once arrived in a marker position (symmetrically placed in both the rooms). A WSN is placed in the environment, comprising 5 IRIS nodes, 4 of which act as anchors fixed nearby the corners of the rooms and the last one is a mobile worn by the user. The input data consists in the 4-dimensional stream of Received Signal Strength (RSS), sampled at the frequency of 8 Hz, exchanged between the mobile and the anchors during the user's movements until the marker position is reached. The corresponding learning task is modeled

⁴ Computational Intelligence & Machine Learning (CIML) group, Department of Computer Science, University of Pisa. Website <http://www.di.unipi.it/groups/ciml/>

⁵ <https://archive.ics.uci.edu/ml/index.php>

as a sequence-to-element binary classification task, in which the target output associated to each input sequence is $+1$ for sequences that will lead to a room change, and -1 for those leading to a room preservation. The dataset contains information pertaining to 3 couples of rooms, gathered in a real-world setting as described in [9], for a total number of 314 sequences. Data pertaining to each couple of rooms is re-scaled in the range $[-1, 1]$, individually for each RSS trace. Figure 2 shows examples of noisy input signals corresponding to both the cases of room change and room preservation. The Indoor Movement Forecasting dataset is freely available at the address <https://archive.ics.uci.edu/ml/datasets/Indoor+User+Movement+Prediction+from+RSS+data>.

Human Activity Recognition. The second dataset considered in our experimental analysis is related to the recognition of human activities from RSS data [37]. The goal is to recognize the action performed by the user within a set of 7 daily-life activities, i.e. bending with legs straight, bending with legs folded, cycling, lying, sitting, standing and walking. Sensor information is gathered from a small WSN composed by 3 IRIS nodes, worn by the user and placed on the chest, on the left ankle and on the right ankle. Input data is obtained from the time series of RSS information exchanged among the 3 sensors, where the average and the standard deviation of each of the 3 RSS traces was computed over a time slot of 250 milliseconds. The dataset thereby comprises 6-dimensional input sequences, with a sampling frequency of 4 Hz, corresponding to the RSS-based setting of the dataset presented in [37]. The corresponding learning task is modeled as a sequence-to-sequence multi-class classification task, in which the (ground-truth) target output at each time step is represented as a $(+1/-1)$ 1-of-7 encoding of the activity that the user is correspondingly performing. In our experimental analysis, in order to obtain signals in a similar range of values for each input dimension, we re-scaled the RSS averages by a factor of 100 and the RSS standard deviations by a factor of 10. Examples of the re-scaled input signals in correspondence of some of the activities considered in the Human Activity Recognition dataset are illustrated in Figure 3, showing the high level of noise in the involved time-series and the difficulty of recognizing clear patterns by visual inspection. The Human Activity Recognition dataset comprises a

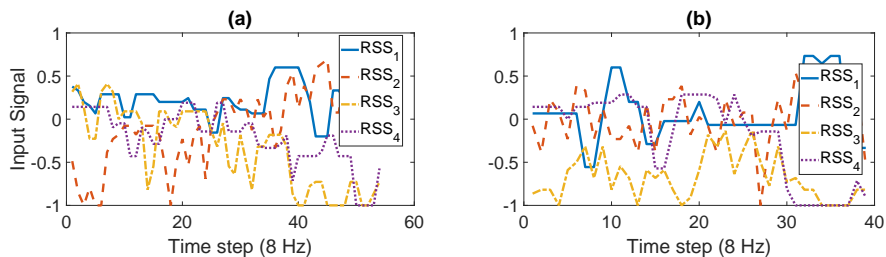


Fig. 2. Examples of RSS traces from the *Indoor Movement Forecasting* dataset. Input signals correspond to: **(a)** room change (target class $+1$); **(b)**: room preservation (target class -1).

total number of 88 sequences, and can be freely downloaded at the following address: [https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+\(AReM\)](https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+system+based+on+Multisensor+data+fusion+(AReM)).

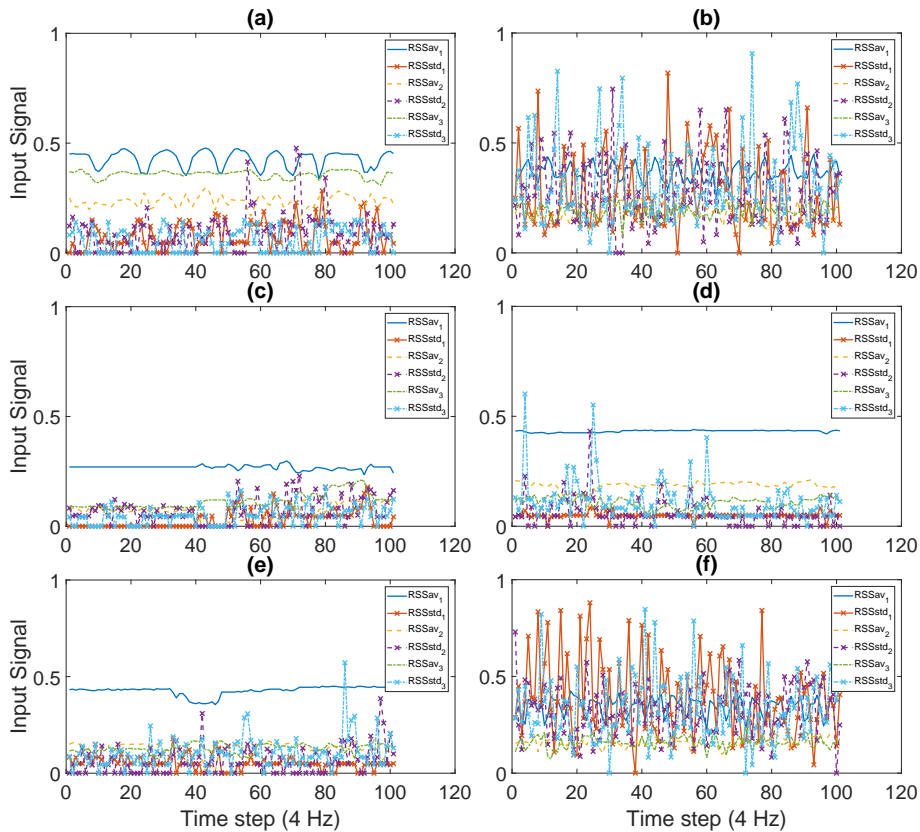


Fig. 3. Examples of input traces (100 time steps-long excerpts) from the *Human Activity Recognition* dataset. Input signals correspond to: (a): bending (with legs straight), (b): cycling, (c): lying, (d): sitting, (e): standing, (f): walking.

3.2 Results

In our experiments, we considered DeepESNs with $N_L = 10$ layers, each of which consisted in a fully-connected reservoir with $N_R = 10$ units, resulting in a total amount of 100 reservoir units. Note that, as pointed out also in [9], such a choice allowed us to focus our analysis on a reservoir size that practically allows the RC networks embedding into the nodes of a WSN. Moreover, we assumed that the same values of the spectral radius, inter-layer scaling and

leaking rate parameters are adopted in each reservoir layer, i.e. for each $i = 1, 2, \dots, N_L$ we set $\rho^{(i)} = \rho$, $s_{il}^{(i)} = s_{il}$ (for $i > 1$) and $a^{(i)} = a$. As regards the readout training, we used ridge regression with regularization parameter denoted by λ_r . For each task, we performed a model selection procedure to chose the values of the RC hyper-parameters on a validation set, varying their values in the ranges reported in Table 1, and according to the cross-validation schemes described in the following. To this end, for each reservoir hyper-parameterization we independently generated 20 networks guesses (with different random seeds), and averaged the achieved results on such guesses.

Hyper-parameter	Range of values
spectral radius ρ	0.7, 0.8, 0.9, 1
input scaling s_{in}	0.1, 0.5, 1, 2, 5
inter-layer scaling s_{il}	0.1, 0.5, 1, 2, 5
leaking rate a	0.1, 0.3, 0.5, 0.7, 1
readout regularization λ_r	$10^{-9}, 10^{-8}, \dots, 10^{-2}, 10^{-1}$

Table 1. Range of RC hyper-parameters values considered for model selection.

The predictive performance on the Indoor Movement Forecasting task was assessed in terms of (2 class) accuracy, while in the case of the Human Activity Recognition task we evaluated the models performance in terms of their 7-class accuracy, i.e. the rate of samples that are correctly assigned to their target class label among the 7 possibilities.

As regards the model selection scheme, for the Indoor Movement Prediction task we considered a setting close to the heterogeneous case in [9]. Specifically, data pertaining to the first two couples of rooms were used as training set, while data corresponding to the third couple of rooms represented an external test set. Training data was then split into 5 folds, for a stratified nested 5-fold cross-validation scheme designed for the purpose of model selection (on the validation set). For the case of the Human Activity Recognition task, we adopted a stratified 3-fold cross-validation, with a further internal level of stratified 4-folds cross-validation used for model selection. In this regard, it is also worth noticing that the model selection schemes adopted in this paper represent an even more rigorous evaluation assessment procedure than those adopted in the reference works for both the AAL tasks considered (see [9] and [37]).

For the sake of performance comparison, we ran experiments on the two AAL tasks also with standard shallow ESNs, following the same settings adopted for DeepESNs as described above. In particular, to the aim of comparison, it is important to stress that we used ESNs with the same number of total recurrent units (and hence trainable parameters) as in the case of DeepESNs, with 100 units organized in a non-layered fully-connected reservoir architecture. This allowed us to effectively and directly assess (by comparison) the effect of the hi-

model	Training	Validation	Test
shallow ESN	0.97(± 0.00)	0.91(± 0.03)	0.84(± 0.04)
DeepESN	0.98(± 0.01)	0.95(± 0.02)	0.90(± 0.03)

Table 2. Training, validation and test accuracy achieved by DeepESN and shallow ESN on the Indoor Movement task.

model	Training	Validation	Test
shallow ESN	0.81(± 0.01)	0.75(± 0.02)	0.74(± 0.01)
DeepESN	0.86(± 0.01)	0.77(± 0.03)	0.77(± 0.02)

Table 3. Training, validation and test 7-class accuracy achieved by DeepESN and shallow ESN on the Human Activity Recognition task.

erarchical organization of DeepESN reservoir state dynamics on the AAL tasks under a fair condition on the number of free parameters of the learner.

The averaged predictive performance achieved by DeepESNs and shallow ESNs on the Indoor Movement Forecasting task is reported in Table 2 (averaged results and standard deviations are computed on the different reservoir guesses). Results show that, in presence of a similar performance on the training set, DeepESN achieves a higher accuracy than shallow ESN on both validation and test sets, reaching respectively 95% and 90% of accuracy. This suggests that, compared to standard shallow ESNs, DeepESNs on the one hand are able to achieve higher accuracy under homogeneous environmental conditions (in which the model is trained and assessed on data coming from the same set of environments) and on the other hand they can better generalize to environmental conditions completely unseen at the training stage.

Averaged results obtained by DeepESNs and shallow ESNs on the Human Activity Recognition task are reported in Table 3 (averages and standard deviations are computed on the different reservoir guesses). As it can be seen, DeepESNs outperform shallow ESNs, being able to better fit the training data, and at the same time reaching a higher performance on the validation and test sets, on both of which it is obtained a 7-class accuracy of 77%.

The results obtained by DeepESNs on both the AAL tasks are relevant also in relation to those reported in literature on the same tasks. In particular, as regards the Indoor Movement Prediction task, the test performance of DeepESN reported in this paper is higher than those reported in [9] in the closest experimental setting⁶, even with respect to 5 times larger RC networks. Furthermore, the 7-class accuracy on the test set reported here for DeepESN on the Human Activity Recognition task is also comparable with the performance reported in [37] under the closest experimental setting⁷, although the latter has been achieved with possibly much larger RC networks and in correspondence

⁶ Non-local 4 setting of the heterogeneous task, as reported in [9].

⁷ RSS-based setting of the activity recognition system, as reported in [37].

of a less thorough scheme for performance assessment than the one considered here⁸.

Overall, the results of the experimental analysis provided in this section clearly point out the convenience of a hierarchical organization of the recurrent dynamical part of the neural network architecture in learning tasks from streams of sensorial data within the AAL domain. From a practical perspective, our results showed that given a recurrent network architecture with a limited size of 100 units (which realistically allows a direct embedding into a node of a WSN, as discussed also in [9]), and given the training efficiency characterization common of the RC framework, the layered DeepESN organization is preferable to a standard shallow one.

4 Conclusions

In this paper we have proposed an experimental investigation aimed at assessing the introduction of the DeepESN methodology for applications in the area of AAL from temporal data originated by a network of sensors. To this aim we have conducted experiments on two real-world benchmark datasets related to the identification of human indoor behavior from temporal streams of RSS information. On both the considered tasks DeepESNs were able to overcome the results achieved by standard ESNs under the same experimental settings and number of trainable parameters. Limiting our analysis to a total number of reservoir units that represent a suitable case for embedding into small devices, DeepESNs led to increase of classification accuracy, with respect to shallow ESNs, of 6% on the Indoor Movement Forecasting (binary classification) task, and of 3% (in terms of 7-class accuracy) on the Human Activity Recognition (multi-classification) task.

The analysis proposed in this paper pointed out the performance advantage brought about by a layered RNN architecture in dealing with AAL tasks: given the same amount of recurrent units it is indeed a good idea to stack them into a layered network when dealing with tasks in this domain. Moreover, the very good result achieved by DeepESNs in the analyzed cases also provide an interesting insight on the appropriateness of the structured feature representations developed by hierarchical reservoirs when excited by temporal data gathered from sensors. This would in turn indicate the relevance, at least partial, of a multiple time-scales nature of the temporal data involved in this application field. Overall, the experimental assessment presented in this paper indicates that the DeepESN approach, yet inheriting the desirable RC characterization of training efficiency, is able to further enhance the already good predictive ability of RC networks, and it is therefore put forward as an effective methodology for learning in temporal domains for future real-world AAL applications.

⁸ Results reported in [37] have been obtained with reservoirs up to 500 units and considering an hold-out cross-validation scheme, such that the test set performance therein reported refers to a smaller set of samples than the one considered in this paper.

References

1. Amato, G., Bacciu, D., Broxvall, M., Chessa, S., Coleman, S., Di Rocco, M., Dragone, M., Gallicchio, C., Gennaro, C., Lozano, H., McGinnity, T., Micheli, A., Ray, A., Renteira, A., Saffiotti, A., Swords, D., Vairo, C., Vance, P.: Robotic ubiquitous cognitive ecology for smart homes. *Journal of Intelligent & Robotic Systems* 80, 57 (2015)
2. Amato, G., Bacciu, D., Chessa, S., Dragone, M., Gallicchio, C., Gennaro, C., Lozano, H., Micheli, A., MP, G.O.G., Renteria, A., Vairo, C.: A benchmark dataset for human activity recognition and ambient assisted living. In: *Ambient Intelligence-Software and Applications-7th International Symposium on Ambient Intelligence (ISAmI 2016)*. pp. 1–9. Springer (2016)
3. Bacciu, D., Chessa, S., Ferro, E., Fortunati, L., Gallicchio, C., La Rosa, D., Llorente, M., Micheli, A., Palumbo, F., Parodi, O., Valenti, A., Vozzi, F.: Detecting socialization events in ageing people: The experience of the doremi project. In: *12th International Conference on Intelligent Environments (IE)*. pp. 132–135. IEEE (2016)
4. Bacciu, D., Chessa, S., Gallicchio, C., Micheli, A.: On the need of machine learning as a service for the internet of things. In: *Accepted for the International Conference on Internet of Things and Machine Learning (IML)* (2017)
5. Bacciu, D., Chessa, S., Gallicchio, C., Micheli, A., Barsocchi, P.: An experimental evaluation of reservoir computation for ambient assisted living. In: *Neural Nets and Surroundings, Smart Innovation, Systems and Technologies*, vol. 19, pp. 41–50. Springer (2013)
6. Bacciu, D., Chessa, S., Gallicchio, C., Micheli, A., Ferro, E., Fortunati, L., Palumbo, F., Parodi, O., Vozzi, F., Hanke, S., Kropf, J., Kreiner, K.: Smart environments and context-awareness for lifestyle management in a healthy active ageing framework. In: *Portuguese Conference on Artificial Intelligence (EPIA)*. pp. 54–66. Springer (2015)
7. Bacciu, D., Chessa, S., Gallicchio, C., Micheli, A., Pedrelli, L., Ferro, E., Fortunati, L., La Rosa, D., Palumbo, F., Vozzi, F., Parodi, O.: A learning system for automatic berg balance scale score estimation. *Engineering Applications of Artificial Intelligence* 66, 60 – 74 (2017)
8. Bacciu, D., Gallicchio, C., Micheli, A., Di Rocco, M., Saffiotti, A.: Learning context-aware mobile robot navigation in home environments. In: *The 5th International Conference on Information, Intelligence, Systems and Applications, IISA 2014*. pp. 57–62. IEEE (2014)
9. Bacciu, D., Barsocchi, P., Chessa, S., Gallicchio, C., Micheli, A.: An experimental characterization of reservoir computing in ambient assisted living applications. *Neural Computing and Applications* 24(6), 1451–1464 (2014)
10. Baronti, P., Pillai, P., Chook, V.W., Chessa, S., Gotta, A., Hu, Y.F.: Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and zigbee standards. *Computer communications* 30(7), 1655–1695 (2007)
11. Barsocchi, P., Chessa, S., Micheli, A., Gallicchio, C.: Forecast-driven enhancement of received signal strength (rss)-based localization systems. *ISPRS International Journal of Geo-Information* 2(4), 978–995 (2013)
12. Chessa, S., Gallicchio, C., Guzman, R., Micheli, A.: Robot localization by echo state networks using rss. In: *Recent Advances of Neural Network Models and Applications, Smart Innovation, Systems and Technologies*, vol. 26, pp. 147–154. Springer International Publishing (2014)

13. Dragone, M., Amato, G., Bacciu, D., Chessa, S., Coleman, S., Rocco, M.D., Gallicchio, C., Gennaro, C., Lozano, H., Maguire, L., McGinnity, M., Micheli, A., O'Hare, G., Renteria, A., Saffiotti, A., Vairo, C., Vance, P.: A cognitive robotic ecology approach to self-configuring and evolving AAL systems. *Engineering Applications of Artificial Intelligence* 45, 269–280 (2015)
14. Dragone, M., Gallicchio, C., Guzman, R., Micheli, A.: Deep reservoir computing: A critical analysis. In: *Proceedings of the 24th European Symposium on Artificial Neural Networks (ESANN)*. pp. 71–76. i6doc.com (2016)
15. Gallicchio, C.: Short-term memory of deep rnn. In: *Proceedings of the 26th European Symposium on Artificial Neural Networks (ESANN)* (2018)
16. Gallicchio, C., Martin-Guerrero, J., Micheli, A., Soria-Olivas, E.: Randomized machine learning approaches: Recent developments and challenges. In: *Proceedings of the 25th European Symposium on Artificial Neural Networks (ESANN)*. pp. 77–86. i6doc.com (2017)
17. Gallicchio, C., Micheli, A.: Architectural and markovian factors of echo state networks. *Neural Networks* 24(5), 440–456 (2011)
18. Gallicchio, C., Micheli, A.: Deep reservoir computing: A critical analysis. In: *Proceedings of the 24th European Symposium on Artificial Neural Networks (ESANN)*. pp. 497–502. i6doc.com (2016)
19. Gallicchio, C., Micheli, A.: Deep echo state network (DeepESN): A brief survey. *arXiv preprint arXiv:1712.04323* (2017)
20. Gallicchio, C., Micheli, A.: A reservoir computing approach for human gesture recognition from kinect data. In: *Proceedings of the Workshop Artificial Intelligence for Ambient Assisted Living (AI*AAL 2016)*, co-located with the 15th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2016). vol. 1803, pp. 33–42. *CEUR Workshop Proceedings* (2017)
21. Gallicchio, C., Micheli, A., Barsocchi, P., Chessa, S.: 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, pp. 151–168. Springer Berlin Heidelberg (2012)
22. Gallicchio, C., Micheli, A., L.Pedrelli: Deep echo state networks for diagnosis of parkinson's disease. In: *Proceedings of the 26th European Symposium on Artificial Neural Networks (ESANN)* (2018)
23. Gallicchio, C., Micheli, A., Pedrelli, L., Fortunati, L., Vozzi, F., Parodi, O.: A Reservoir Computing Approach for Balance Assessment, *Lecture Notes in Computer Science*, vol. 9785, pp. 65–77. Springer International Publishing (2016)
24. Gallicchio, C., Micheli, A., Silvestri, L.: Local lyapunov exponents of deep echo state networks. *Neurocomputing* p. (Accepted) (2017)
25. Gallicchio, C., Micheli, A., Silvestri, L.: Local Lyapunov Exponents of Deep RNN. In: *Proceedings of the 25th European Symposium on Artificial Neural Networks (ESANN)*. pp. 559–564. i6doc.com (2017)
26. Gallicchio, C., Micheli, A.: Echo state property of deep reservoir computing networks. *Cognitive Computation* 9, 337–350 (2017)
27. Gallicchio, C., Micheli, A., Pedrelli, L.: Deep reservoir computing: a critical experimental analysis. *Neurocomputing* 268, 87–99 (2017)
28. Gallicchio, C., Micheli, A., Pedrelli, L.: Hierarchical temporal representation in linear reservoir computing. In: *Proceedings of the 27th Italian Workshop on Neural Networks (WIRN)* (2017), *arXiv preprint arXiv:1705.05782*
29. Haykin, S.: *Neural networks and learning machines*, vol. 3. Pearson (2009)

30. Jaeger, H.: The "echo state" approach to analysing and training recurrent neural networks - with an erratum note. Tech. rep., GMD - German National Research Institute for Computer Science, Tech. Rep. (2001)
31. Jaeger, H., Haas, H.: Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* 304(5667), 78–80 (2004)
32. Jaeger, H., Lukoševičius, M., Popovici, D., Siewert, U.: Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Networks* 20(3), 335–352 (2007)
33. Kolen, J., Kremer, S.: A field guide to dynamical recurrent networks. John Wiley & Sons (2001)
34. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials* 15(3), 1192–1209 (2013)
35. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Computer Science Review* 3(3), 127–149 (2009)
36. Lukoševičius, M., Jaeger, H., Schrauwen, B.: Reservoir computing trends. *KI-Künstliche Intelligenz* 26(4), 365–371 (2012)
37. Palumbo, F., Gallicchio, C., Pucci, R., Micheli, A.: Human activity recognition using multisensor data fusion based on reservoir computing. *Journal of Ambient Intelligence and Smart Environments* 8(2), 87–107 (2016)
38. Palumbo, F., La Rosa, D., Ferro, E., Bacciu, D., Gallicchio, C., Micheli, A., Chessa, S., Vozzi, F., Parodi, O.: Reliability and human factors in ambient assisted living environments. *Journal of Reliable Intelligent Environments* pp. 1–19 (2017)
39. Vermesan, O., Bröring, A., Tragos, E., Serrano, M., Bacciu, D., Chessa, S., Gallicchio, C., Micheli, A., Dragone, M., Saffiotti, A., Simoens, P., Cavallo, F., Bahr, R.: Internet of robotic things: converging sensing/actuating, hypoconnectivity, artificial intelligence and iot platforms. In: *Cognitive hyperconnected digital transformation: internet of things intelligence evolution*, pp. 1–35 (2017)
40. Verstraeten, D., Schrauwen, B., d’Haene, M., Stroobandt, D.: An experimental unification of reservoir computing methods. *Neural networks* 20(3), 391–403 (2007)
41. Yildiz, I.B., Jaeger, H., Kiebel, S.J.: Re-visiting the echo state property. *Neural networks* 35, 1–9 (2012)