Deep Echo State Networks for Functional Ambulation Categories Estimation

L. Pedrelli¹, E.Bergamini², M.Tramontano^{2,3}, G. Vannozzi², A.Mannini^{1,4*}

 1-The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy.
2-Fondazione Santa Lucia IRCCS, Rome, Italy.
3-Dept. of Movement, Human and Health Sciences, University of Rome Foro Italico, Rome, Italy.
4-IRCCS Fondazione Don Carlo Gnocchi, Firenze, Italy.

In this work, we introduce a novel application for the au-Abstract. tomatic estimation of Functional Ambulatory Category (FAC) based on deep Echo State Networks (ESNs). FAC is a clinical scale for assessing the gait ability used in post-stroke rehabilitation and, in general, for disease monitoring. In this application, the estimation is performed automatically by analyzing signals gathered from wearable sensors (located on both tibiae, pelvis, trunk and head) during the execution of a walking test. This is performed by analysing the whole time-series through the DeepESN model without preprocessing. The experimental results show that the use of a deep recurrent neural network allows the model to exploit the richness contained in the whole raw temporal signal improving the performance w.r.t. the shallow recurrent model. Overall, our approach obtained 0.37 of mean absolute error with a maximum error of 0.78 resulting very accurate in the classification of the gait ability through the estimation of the FAC value. Considering the experimental results obtained, the proposed approach represents a good baseline for medical applications based on the automatic estimation of the FAC scale.

1 Introduction

Functional Ambulatory Category (FAC) is a clinical scale considered a valid, reliable and responsive tool for assessing the gait ability [1]. The FAC scale is used especially in post-stroke patients and in general for assessing the severity of diseases and for rehabilitation monitoring. However, like most clinical questionnaires/scales, the FAC requires dedicated and expert medical staff to be administered and its reliability and validity strongly depends on the operator's experience. In this context, machine learning models are good candidates in supporting gait analysis for medical purposes. For instance, the analysis of data gathered from wearable sensors during a walking test through machine learning approaches has been successfully tested for pathology classification, gait pattern

^{*}This research was partly supported by the Italian Ministry of University under the TRAINED project (mulTifeature analysis of heaRt rate variability and galt features in cliNical Evaluation of Depression, PRIN 2017 grant number 2017L2RLZ2). The work of authors at Scuola SantâAnna was also supported by the Dept. of Excellence in Robotics & A.I.

analysis and rehabilitation monitoring [2, 3]. Up to our knowledge, even if similar approaches based on the automatic estimation of Berg Balance Scale were previously pursued [4, 5], no earlier studies targeted an automatic estimation of the FAC scale values from wearable sensors data.

To fill this gap, we propose here a new machine learning approach for the automatic estimation of the FAC scale from raw temporal signals based on Recurrent Neural Networks (RNNs). In particular, we consider the Reservoir Computing (RC) [6] paradigm which showed good abilities to address raw and noisy temporal signals achieving good results in clinical scale estimation [5] through the use of Echo State Networks (ESNs). Recently, within the RC paradigm, the introduction of Deep Echo State Networks (DeepESN) [7] showed a good ability to represent complex multiple time-scales dynamics, outperforming the standard ESN in Parkinson's Disease diagnosis [8].

Based on such considerations, in this application we analyzed the whole temporal data gathered from body-mounted wearable inertial sensors during the 10 minute walking test through a DeepESN model. We evaluate the FAC estimation approach on a dataset composed of raw signals gathered during the walking test of 42 stroke patients. Moreover, we assess the relevance of layered recurrent architectures in this application by comparing the DeepESN (the multi-layered version) model with the standard ESN (the 1-layered version). Finally, we show how we can improve the performance of the estimation through the ensembling of DeepESN instances.

2 Deep Echo State Networks

The DeepESN [7] model is an extension of Leaky Integrator ESNs (LI-ESNs) [9]. It is a deep recurrent architecture (see Fig. 1 for an example) composed of a hierarchy of N_L recurrent layers (called reservoirs). The input at time step t is denoted by $\mathbf{u}(t) \in \mathbb{R}^{N_U}$, while, the state of the l-th reservoir layer at time step t is denoted by $\mathbf{x}^{(l)}(t) \in \mathbb{R}^{N_R}$. In the following formulas we avoid the bias term for the ease of notation. The state of the network of the first layer is computed as:

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

while for layers l > 1, the state of the network is computed as:

$$\mathbf{x}^{(l)}(t) = (1 - a^{(l)})\mathbf{x}^{(l)}(t - 1) + a^{(l)} \mathbf{tanh}(\mathbf{W}^{(l)}\mathbf{x}^{(l-1)}(t) + \hat{\mathbf{W}}^{(l)}\mathbf{x}^{(l)}(t - 1)), \quad (2)$$



Fig. 1: An example of DeepESN architecture.

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.

where $\mathbf{W}_{in} \in \mathbb{R}^{N_R \times N_U}$ is the input weights matrix, $\hat{\mathbf{W}}^{(l)} \in \mathbb{R}^{N_R \times N_R}$ is the recurrent weights matrix of layer l, $\mathbf{W}^{(l)} \in \mathbb{R}^{N_R \times N_R}$ is the inter-layer weights matrix that connects the layer l - 1 to the layer l, $a^{(l)}$ is the leaky parameter of layer l and **tanh** is the hyperbolic tangent which represents the activation function. The global state of the DeepESN is computed by concatenating all the layer states $\mathbf{x}(t) = (\mathbf{x}^{(1)}(t), \dots, \mathbf{x}^{(N_L)}(t)) \in \mathbb{R}^{N_L N_R}$.

The values of matrices \mathbf{W}_{in} and $\{\mathbf{W}^{(l)}\}_{l=2}^{N_L}$ are randomly initialized from a uniform distribution and re-scaled such that $\|\mathbf{W}_{in}\|_2 = \sigma$ and $\|\mathbf{W}^{(l)}\|_2 = \sigma$ where σ is the scaling factor of the input weights and the inter-layer weights. The values of matrices $\{\hat{\mathbf{W}}^{(l)}\}_{l=1}^{N_L}$ are initialized according to the Echo State Property of DeepESNs [10].

In the task considered in this work, the output of the DeepESN is computed for each input sequence. First, the mean state mapping is computed as: $\chi(\mathbf{s}) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{x}(t)$, where $\mathbf{s} = [\mathbf{u}(1), ..., \mathbf{u}(n)]$ is an input sequence of length *n*. Second, the output is computed as follows: $\mathbf{y}(\mathbf{s}) = \mathbf{W}_{\text{out}}\chi(\mathbf{s})$, where $\mathbf{W}_{\text{out}} \in \mathbb{R}^{N_Y \times N_L N_R}$ is the matrix of the readout layer weights. The readout weights are trained by means of pseudo-inversion or ridge regression [9, 6] as in standard RC.

In the following the LI-ESN (equivalent to a 1-layered version of DeepESN) is referred as *ESN*.

3 DeepESNs for Functional Ambulation Estimation

Here, we experimentally assess the proposed DeepESN approach for Functional Ambulation Category (FAC) estimation. Moreover, we compare the DeepESN (the multi-layered model) with the standard ESN (the 1-layered model) in order to show the effectiveness of layering in RNN models in this kind of applications. The dataset was acquired at Fondazione Santa Lucia [11]. It is composed of 185 sequences characterized by raw signals gathered during the walking test of 42 stroke patients. Each temporal sequence is associated to a FAC value measured by the clinical staff during the execution of the walking test. In this work, we considered three of the possible FAC values (i.e. 3, 4 and 5). The value represents the walking ability (the more the better) of the patient. Our approach consists in estimating the FAC value through the DeepESN model starting from a raw signal gathered during the walking test performed by the patient. The FAC estimation is performed by ridge regression on the FAC values similarly to other approaches based on the estimation of clinical scales [5]. Each time step of a sequence is a vector of 65 values gathered from 5 wearable sensors (located on both tibiae, pelvis, trunk and head). For each sensor, we consider the accelerometer (3 components), the gyroscope (3 components), the magnetometer (3 components) and the orientation quaternions (4 components). In our approach, the DeepESN is fed by consider the whole raw signal without preprocessing. Fig. 2 a) and 2 b) show an excerpt of the signals gathered from the trunk accelerometer (x, y and z components) at each time step during the execution of the walking test by a patient with FAC 3 (Fig. 2 a)) and a patient with FAC 5 (Fig. 2 b)).



Note that this kind of data is rather noisy, then the estimation of FAC without preprocessing needs an approach suitable for noisy time-series.

Fig. 2: An excerpt of the signals gathered from the trunk accelerometer (x, y and z components) at each time step during the execution of the walking test by a patient with FAC 3 (Fig. a)) and a patient with FAC 5 (Fig. b)).

We evaluated the model through a double cross validation approach with 2 inner folds (validation sets) and 3 outer folds (test sets). Each validation and test fold is composed of 18 time series gathered from 6 patients with balanced labels. Following the approach proposed by [5], the evaluation of the model is performed on patients which are not present in the training set. In this way, the generalization ability of the model does not depend on the patient. For the model selection, we optimized the hyperparameters through a random search by considering the spectral radius $\rho^{(l)}$ from a uniform distribution in [0, 1], the leaky integrator $a^{(l)}$ from a uniform distribution in [0, 1], the input norm $\sigma^{(l)}$ from a logarithmic distribution in $[10^{-4}, 1]$ and the number of recurrent units N_R from a uniform distribution in [5, 100]. We considered a DeepESN with 10 N_L layers and the same number of N_R units per layer. For the standard ESN (i.e., $N_L = 1$), we considered the N_R range [50, 1000] in the model selection in order to have a fair comparison by considering the same number of total recurrent units of the DeepESN model. In the following we use the mean absolute error (MAE) as metrics to evaluate the performance of the models. In the following experiments, for each hyper-parametrization we generated 5 random instances that we call *guesses*. Then, the results are averaged on such guesses.

In order to improve the performance from a point of view of the FAC application, we exploited the variability provided by different guesses of the selected model by means of an ensemble approach. The ensemble model is obtained by averaging the output of the guesses of the DeepESN model. Moreover, we evaluated the classification ability of the model considering the 3 values of the FAC scale (3, 4 and 5) as 3 classes. Then, the output of the model is converted to the class 3 if $\mathbf{y}(\mathbf{s}) < -\tau$, to the class 5 if $\mathbf{y}(\mathbf{s}) > \tau$ and to the class 4 otherwise. The τ value is optimized in the validation set by performing a search grid in the range [0.05, 0.1, 0.15, ..., 1].

4 Experimental Results

Here, we present the results obtained by the considered models on the FAC task. Table 1 shows the comparison between DeepESN and ESN. The DeepESN outperforms ESN obtaining a test MAE of 0.39 while ESN achieved a test MAE of 0.59. These results highlight that the ability of layering in RC architectures

Table 1: Test MAE (mean (standard deviation)) obtained by DeepESN and ESN on FAC task.

Architecture	Test MAE
DeepESN	0.39(0.033)
ESN	0.59(0.026)

in providing a hierarchical temporal representation of the input sequences can be exploited to improve the model performance in this kind of applications.

Here we show the experimental results obtained by the ensemble approach by using the selected DeepESN model. Table 2 shows the test MAE, test classification accuracy (Test Acc), rate of the class 3 (3PR), 4 (4PR) and 5 (5PR) obtained by DeepESN performing the ensemble approach.

Table 2: Test MAE, classification accuracy (Acc), positive rate of the class 3 (3PR), 4 (4PR) and 5 (5PR) obtained by DeepESN performing the ensemble approach.

Architecture	Test MAE	Test Acc	3PR	4PR	5PR
DeepESN	0.37	77.93%	83.33%	66.67%	83.33%

By using the ensemble approach the estimation improves from 0.39 to 0.37 test MAE. The maximum test MAE obtained by the model is 0.78. This means that the model estimations are closer to the correct FAC value for all cases in the considered experimental setting. From a classification point of view the ensemble model is able to obtain 77.93% of accuracy. In particular, by considering a test set which contains only the class 3 and 5 with a threshold $\tau = 0$ the model achieved the 100% of accuracy (in the considered experimental setting). Therefore, the proposed approach is remarkably able to discriminate between mild and severe stroke patients.

5 Conclusions

In this work, we proposed a novel approach for the automatic estimation of functional ambulation based on DeepESN model. We considered the whole time

series produced by the wearable sensor during the walking test without the use of preprocessing. The experimental assessment is performed on a dataset composed of 185 time-series of raw signals from 42 stroke patients. The DeepESN outperforms the standard ESN (the 1-layered) on FAC estimation task. This highlights the ability of deep recurrent models to exploit the richness of information presents in the raw signal. Moreover, the ensemble approach achieved a remarkable classification ability in the discrimination between mild and severe stroke patients. To the best of our knowledge, this work represents the first attempt to develop an approach based on machine learning for the automatic estimation of the FAC scale using gait measures with wearables. The proposed solution represents a preliminary study, toward intelligent systems based on wearable devices able to provide early detection of gait alterations. Such capability let us envision mass screening potentials of the joint use of machine learning and wearables in the field of gait analysis. In future works, it could be interesting to investigate this novel problem with other models and more data in order to improve the results interpretation and the estimation ability.

References

- Jan Mehrholz, Katja Wagner, Katja Rutte, Daniel Meiβner, and Marcus Pohl. Predictive validity and responsiveness of the functional ambulation category in hemiparetic patients after stroke. Archives of physical medicine and rehabilitation, 88(10):1314–1319, 2007.
- [2] Andrea Mannini, Diana Trojaniello, Andrea Cereatti, and Angelo M. Sabatini. A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntingtonâs disease patients. *Sensors*, 16(1), 2016.
- [3] Hong yin Lau, Kai yu Tong, and Hailong Zhu. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Human Movement Science*, 28(4):504–514, 2009.
- [4] H. Simila, J. Mantyjarvi, J. Merilahti, M. Lindholm, and M. Ermes. Accelerometry-based berg balance scale score estimation. *Biomedical and Health Informatics, IEEE Journal* of, 18(4):1114–1121, 2014.
- [5] D. Bacciu, S. Chessa, C. Gallicchio, A. Micheli, L. Pedrelli, E. Ferro, L. Fortunati, D. La Rosa, F. Palumbo, F. Vozzi, and O. Parodi. A learning system for automatic berg balance scale score estimation. *Engineering Applications of Artificial Intelligence*, 66:60-74, 2017.
- [6] D. Verstraeten, B. Schrauwen, M. d'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural networks*, 20(3):391–403, 2007.
- [7] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep reservoir computing: a critical experimental analysis. *Neurocomputing*, 268:87–99, 2017.
- [8] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep echo state networks for diagnosis of parkinson's disease. In *Proceedings of the 26th European Symposium on Artificial Neural Networks (ESANN)*, pages 397–402. i6doc.com, 2018.
- [9] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Networks*, 20(3):335–352, 2007.
- [10] C. Gallicchio and A. Micheli. Echo state property of deep reservoir computing networks. Cognitive Computation, 9(3):337–350, 2017.
- [11] Elena Bergamini, Marco Iosa, Valeria Belluscio, Giovanni Morone, Marco Tramontano, and Giuseppe Vannozzi. Multi-sensor assessment of dynamic balance during gait in patients with subacute stroke. *Journal of Biomechanics*, 61:208–215, 2017.