Deep Echo State Networks for Diagnosis of Parkinson's Disease

Claudio Gallicchio and Alessio Micheli and Luca Pedrelli

Department of Computer Science, University of Pisa
Largo Bruno Pontecorvo 3 - 56127 Pisa, Italy

Abstract. In this paper, we introduce a novel approach for diagnosis of Parkinson's Disease (PD) based on deep Echo State Networks (ESNs). The identification of PD is performed by analyzing the whole time-series collected from a tablet device during the sketching of spiral tests, without the need for feature extraction and data preprocessing. We evaluated the proposed approach on a public dataset of spiral tests. The results of experimental analysis show that DeepESNs perform significantly better than shallow ESN model. Overall, the proposed approach obtains state-of-the-art results in the identification of PD on this kind of temporal data.

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disease that mainly affects the extrapiramidal motor system causing tremor, bradykinesia, rigidity and loss of postural reflexes [1]. The analysis of motor capacities such as handwriting and sketching abilities of patients is performed to assess and diagnose PD [2]. While handwriting abilities are influenced by language capacities, sketching abilities involved in the execution of a spiral test (i.e., drawing a spiral with a pen) are considered as measures independent of education [2]. An example of application of deep learning models for PD classification, applied directly to images of spiral tests, is introduced in [3]. In [4] it is introduced a study on PD diagnosis through statistical methods based on the analysis of velocity and the pen pressure data collected from a tablet device during the execution of the spiral test.

In this paper, we propose a novel approach for diagnosis of PD based on Recurrent Neural Networks (RNNs) that are able to robustly exploit the whole time course of noisy and heterogeneous time-series data collected from a tablet device during the execution of spiral tests. In particular, we consider the Reservoir Computing (RC) [5, 6] framework that obtained state-of-the-art results in clinical assessment [7] related to the study of neurological diseases. In the application presented in [7], signals characterized by noisy and heterogeneous time-series are analyzed for postural stability assessment by using Echo State Networks (ESNs). Recently, the introduction of DeepESN model as the extension of the Reservoir Computing (RC) [5, 6] paradigm to the deep learning framework allowed to study the intrinsic properties of stacked RNNs [8] and to build efficiently trained models able to develop a hierarchical temporal representation of the input signals, resulting in a suitable method to deal with tasks featured by multiple time-scales dynamics [9, 10]. Based on such considerations,

we think that the DeepESN model can be particularly suitable for the study of PD through the analysis of tablet signals.

We assess our approach for diagnosis of PD on a publicly available dataset of spiral tests introduced in [11]. Moreover, in order to investigate the practical relevance of using layered recurrent architectures in this domain, the experimental analysis is also conducted in comparison to the standard ESN. Finally, we evaluate the results achieved by ensembling the realizations of the selected model, in order to improve the classification performance of the proposed approach.

2 Deep Echo State Networks

DeepESNs [9] extend Leaky Integrator ESNs (LI-ESNs) [5] to the deep learning framework. A DeepESN architecture (see Fig. 1 for an example) is composed by a hierarchy of N_L reservoirs. Here we denote by $\mathbf{u}(t) \in \mathbb{R}^{N_U}$ and $\mathbf{x}^{(l)}(t) \in \mathbb{R}^{N_R}$, respectively, the input and state of the l-th reservoir layer at step t. Avoiding the bias term for the ease of notation, the state transition function of the first layer is computed by the following equation:

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

while for higher layers $l > 1$, the state transition function is described as follows:

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

In the above eq. 1 and 2, $\mathbf{W}_{in} \in \mathbb{R}^{N_R \times N_U}$ is the matrix of input weights, $\hat{\mathbf{W}}^{(l)} \in \mathbb{R}^{N_R \times N_R}$ is the matrix of the recurrent weights of the layer l, $\mathbf{W}^{(l)} \in \mathbb{R}^{N_R \times N_R}$ is the matrix of weights relative to the inter-layer connections from layer $l-1$ to layer l, $a^{(l)}$ is the leaky parameter at layer l and $\mathbf{tanh}$ is the activation function of recurrent units implemented by a hyperbolic tangent. Finally, the state of the DeepESN is given by the concatenation of all the states encoded in the recurrent layers of the architecture $\mathbf{x}(t) = (\mathbf{x}^{(1)}(t), \ldots, \mathbf{x}^{(N_L)}(t)) \in \mathbb{R}^{N_L N_R}$.

Let σ and $\hat{\sigma}$ be the scaling of the external input and the inter-layer scaling, then the weights in 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 $\|\mathbf{W}^{(l)}\|_2 = \hat{\sigma}$ respectively. According to the necessary condition of Echo State Property for DeepESNs [8], values of matrices $\{\hat{\mathbf{W}}^{(l)}\}_{l=1}^{N_L}$ are randomly initialized from
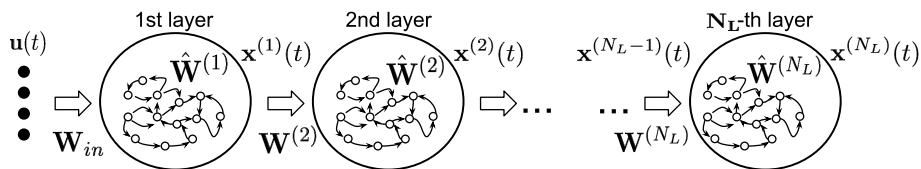

Fig. 1: The hierarchy of reservoirs in the architecture of a DeepESN.

uniform distribution and rescaled such that $\max\limits_{1 \leq l \leq N_L} \rho\left((1 - a^{(l)})\mathbf{I} + a^{(l)}\hat{\mathbf{W}}^{(l)}\right) < 1$, where ρ denote the spectral radius of the matrix (i.e., the largest absolute value of its eigenvalues). Note that considering a number of layers $N_L = 1$ the recurrent architecture described above is equivalent to a LI-ESN, that is referred to as *shallow ESN* in the following.

For the classification task considered in this work, the DeepESN computes an output in correspondence of an entire input sequence. Accordingly, we compute a mean state mapping as follows: $\chi(\mathbf{s}) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}(t)$, where $\mathbf{s} = [\mathbf{u}(1), ..., \mathbf{u}(n)]$ is an input sequence of length n. Then, the output of the network is computed as $\mathbf{y}(\mathbf{s}) = \mathbf{W}_{out}\chi(\mathbf{s})$, where $\mathbf{W}_{out} \in \mathbb{R}^{N_Y \times N_L N_R}$ is the weights matrix relative to the readout layer. The linear combination between the states encoded in the recurrent layers and the weights of the readout allows to differently weight the contribution of the multiple dynamics developed in the network state. As in the standard RC framework, only the readout layer is trained, typically by means of direct method such as pseudo-inversion or ridge regression [5, 6].

3   DeepESNs for Diagnosis of PD through Spiral Tests

We evaluated the proposed DeepESN model on the spiral dataset described in [11]. This dataset is publicly available on the UC Irvine Machine Learning Repository[1]. The dataset is composed by spiral tests executed on a tablet device by 61 PD patients and 15 control patients without PD. For each time-step, the time-series gathered from the tablet contain pen position (x and y components), pen pressure and grip angle. In our method, the models were fed directly with such signals without feature extraction or data preprocessing. Note that, the whole temporal signal generated by the tablet potentially contains a richer information than static preprocessed features, thereby a RNN model can be more effective on the analysis of this kind of data. Fig. 2 a), 2 b), 2 c) and 2 d) show (x,y) pen position, pen pressure and grip angle gathered at each time-step from tablet during the execution of the test by a control (Fig. 2 a), 2 c)) and a PD (Fig. 2 b), 2 d)) patient. As we can note, by visual inspection the sketches showed in Fig. 2 a) and 2 b) are quite similar and consequently the classification task is not trivial in such cases. Moreover, from Fig. 2 c) and 2 d) it is possible to see that the signals relative to pen pressure and grip angle are rather noisy, therefore, the analysis of such kind of data without feature extraction can be challenging since we need a noise-robust approach to perform a correct classification.

We evaluated the generalized performance of the proposed approach through a 3-fold cross validation, inserting in each fold 20 (or 21) PD and 5 control patients. We considered a 10-layered DeepESN (i.e. $N_L = 10$) with the same number of N_R units per layer and a leaky integrator of $a^{(l)} = 0.1$ for each layer l. The rest of the DeepESN hyper-parametrization were selected from the ranges

---

[1]<https://archive.ics.uci.edu/ml/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet>
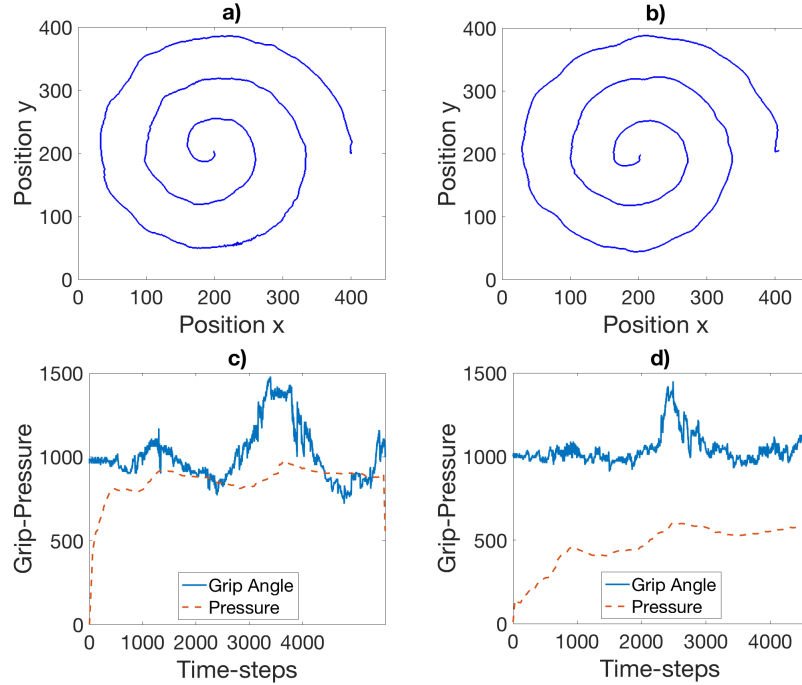
Fig. 2: Pen positions (1st row), pressure and grip angle (2nd row) gathered for a control (fig. a), c)) and a PD (fig. b), d)) patient.

reported in Tab. 1, on a validation set by an extra level of 5-fold validation on each fold. For each hyper-parametrization, we generated 10 reservoir guesses, averaging the results on such guesses.

Hyper-parameter	
recurrent units per layer N_R	$10, 20, 30, 40, 50$
input scaling σ	0.1, 0.5, 1, 2
inter-layer scaling $\hat{\sigma}$	0.1, 0.5, 1, 2
spectral radius ρ	0.7, 0.8, 0.9, 1.0
readout regularization λ_r	$0, 10^{-10}, 10^{-9}, ..., 10^0$

Table 1: Range of DeepESN hyper-parameters values for model selection.

Tab. 2 shows the accuracies and the standard deviations on reservoir guesses (in parenthesis) obtained by DeepESN and shallow ESN in PD classification task. In order to perform a fair comparison, we selected the shallow ESN (i.e., 1-layered DeepESN) considering the same hyper-parameters ranges as for Deep-ESN and the same range of total number of recurrent units (i.e. 100-500), where the specific choice depends on the fold of the 3-fold cross validation. Interest-

Model	TR	VL	TS
DeepESN	**94.27% (1.18%)**	**86.57% (2.62%)**	**87.20% (2.78%)**
shallow ESN	91.60% (0.56%)	83.62% (3.39%)	84.13% (2.63%)

Table 2: Accuracies and standard deviations on reservoir guesses (in parenthesis) obtained by DeepESN and shallow ESN in validation, training and test.

ingly, DeepESN model outperformed shallow ESN in training, validation and test set of 2.67%, 2.95% and 3.07% of accuracy respectively. This comparison result suggests that the ability of DeepESN in providing a hierarchical temporal representation of input signals with complex dynamics allows us to improve the performance on this kind of tasks. For completeness, we statistically compared the accuracies obtained on test set by DeepESN and shallow ESN models through McNemar's test (a χ^2 variant). Classification accuracies obtained by DeepESN and shallow ESN resulted significantly different with a p-value of 0.0032.

To exploit the variability provided by different reservoir guesses, we evaluated the selected model considering an ensemble approach. Accordingly, the classification was performed averaging the output of the different guesses of the selected model. Results are reported in Tab. 3, and show that the ensemble approach allowed DeepESN to achieve improved performance in training, validation and test set of 0.40%, 2.95% and 2.13% of accuracy respectively. Overall, our proposed automatic system obtained a test accuracy of 89.33% with a sensitivity (percentage of PDs correctly classified) and a specificity (percentage of controls correctly classified) in test set of 90.00% and 80.00% respectively.

Model	TR ACC	VL ACC	TS ACC	TS SEN	TS SPEC
DeepESN	94.67%	89.52%	**89.33%**	**90.00%**	**80.00%**

Table 3: Accuracy (ACC), sensitivity (SEN) and specificity (SPEC) obtained by ensemble of DeepESN in training (TR), validation (VL) and test (TS) set.

For what concerns the comparison with the state-of-the-art, results of DeepESN in this paper compare well with literature approaches recently devised on the same type of input data. In particular, in [4] it is introduced a method for diagnosis of PD based on the analysis of spiral tests gathered from a similar tablet device, achieving a classification accuracy in PD identification of 79.1%. Interestingly, our approaches outperform such method obtaining a test accuracy of 84.13%, 87.20% and 89.33% achieved by shallow ESN, DeepESN and ensemble of DeepESNs respectively. This comparison further indicates that, contrary to what might appear at first glance from the example in Fig. 2, simple statistics on the input signals, e.g. pressure data, are not rich enough to accurately discriminate PD[2], while our approach can capture relevant information from the whole temporal signal, allowing to effectively improve such results.

---

[2]Correlation between averaged pressure and PD in the dataset used in our paper is weak (-0.31).

4 Conclusions

In this paper, we proposed a novel approach for diagnosis of PD based on DeepESN model. The deep recurrent model is fed by the whole time-series gathered from a tablet during the sketching of spiral tests. We performed a comparative assessment of our approach on a public dataset containing spiral tests executed by PD and control patients. Results showed that the predictive accuracy obtained by DeepESN significantly outperforms the result of shallow ESNs, highlighting the potential benefits of deep recurrent architectures in the treatment of temporal signals for PD diagnosis. Moreover, the use of ensemble method on the selected DeepESN model led to a further performance enhancement. Overall, the proposed approach compared well also with respect to state-of-the-art results, further stressing the potentiality of exploiting the whole richness of temporal signals for PD diagnosis.

At the best of our knowledge, this work represents the first attempt to develop an approach for diagnosis of PD by using recurrent models, such as DeepESN, able to develop hierarchical temporal representations from tablet signals without the need of feature extraction and data preprocessing.

References

[1] J. Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, 2008.

[2] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, and S. L. Pullman. Validity of spiral analysis in early parkinson's disease. *Movement disorders*, 23(4):531–537, 2008.

[3] C. R. Pereira, D. R. Pereira, J. P. Papa, G. H. Rosa, and X. Yang. Convolutional neural networks applied for parkinson's disease identification. In *Machine Learning for Health Informatics*, pages 377–390. Springer, 2016.

[4] P. Zham, D. K. Kumar, P. Dabnichki, S. Poosapadi Arjunan, and S. Raghav. Distinguishing different stages of parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. *Frontiers in Neurology*, 8:435, 2017.

[5] 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.

[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] 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.

[8] C. Gallicchio and A. Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9(3):337–350, 2017.

[9] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep reservoir computing: a critical experimental analysis. *Neurocomputing*, 268:87–99, 2017.

[10] C. Gallicchio, A. Micheli, and L. Pedrelli. Hierarchical temporal representation in linear reservoir computing. In *27th Italian Workshop on Neural Networks (WIRN)*, 2017. arXiv preprint arXiv:1705.05782.

[11] M. Isenkul, B. Sakar, and O. Kursun. Improved spiral test using digitized graphics tablet for monitoring parkinson's disease. In *ICEHTM*, pages 171–175, 2014.