EEG feature learning with Intrinsic Plasticity based Deep Echo State Network

Rahma Fourati
REGIM-Lab.: REsearch Groups in Intelligent Machines,
University of Sfax,
National Engineering School of Sfax (ENIS),
BP 1173, Sfax, 3038, Tunisia
rahma.fourati@ieee.org

Boudour Ammar
REGIM-Lab.: REsearch Groups in Intelligent Machines,
University of Sfax,
National Engineering School of Sfax (ENIS),
BP 1173, Sfax, 3038, Tunisia
boudour.ammar@ieee.org

Yaochu Jin
Department of Computer Science
University of Surrey
Guildford, Surrey, GU2 7XH, UK
yaochu.jin@surrey.ac.uk

Adel M. Alimi
REGIM-Lab.: REsearch Groups in Intelligent Machines,
University of Sfax,
National Engineering School of Sfax (ENIS),
BP 1173, Sfax, 3038, Tunisia
adel.alimi@ieee.org

***Abstract*—In this paper, deep EEG feature learning method is proposed for emotion recognition. It is well known that EEG signals dramatically vary from person to person, thereby making subject-independent emotion recognition very challenging. To address the above challenge, this work presents a deep echo state network (DeepESN) to learn temporal representation from raw EEG data. DeepESN as an input-driven discrete time non-linear dynamical system allows to process the temporal information at each time step in a deep temporal fashion by means of a hierarchical composition of multiple levels of recurrent neurons. To make the DeepESN robust, we pre-train the reservoir connections with an unsupervised intrinsic plasticity rule to generate activities following a desired Gaussian distribution. Then, we propose a hybrid learning algorithm for training the output weights which benefits from both the ridge regression and the online delta rule. Our leaky DeepESN achieved encouraging results when tested on the well-known affective benchmarks DEAP and DREAMER.**

***Index Terms*—Deep echo state network, intrinsic plasticity, Gaussian distribution, electroencephalogram, emotion recognition, feature learning**

I. Introduction

The problem of classifying multi-channel Electroencephalogram (EEG) time series consists in assigning their representation to one of a fixed number of classes. This is a fundamental task in many healthcare applications, including anxiety detection ([1], [2] and [3]), epileptic seizures prediction [4] and also affective computing applications such as EEG-based emotion recognition ([5] and [2]). The problem has been tackled by a wealth of different approaches, spanning from the signal decomposition techniques of EEG signals to the feature extraction and feature selection algorithms as highlighted in the surveys [6], [7], [8], and [9].

Representation learning or feature learning [10] consists in automatically discovering the relevant representations for a classification or detection task directly from raw data such as the digits recognition ([11], [12], [13] and [14]). Consequently, the laborious handcrafted features are no longer needed since representation learning permits to both learn the features and use them to perform a specific task.

In this paper, we focus on EEG representation learning for emotion recognition using the reservoir computing approach in a subject-independent context. In comparison with feature-based work, few works were dedicated to EEG-based emotion recognition from raw data. Actually, Echo State Network (ESN) one of the proposed Recurrent Neural Network (RNN) within the reservoir computing paradigm allows to first process EEG time series through dynamic recurrent units (i. e. the so called ”reservoir”) and then the internal states activation generated over time is exploited to perform the classification. Neural networks have showed a great success in several applications ([15], [16], [17] [18], and [19]).

ESN is a non-gradient based RNN, that means there is no backpropagation algorithm involved in its learning. To tackle the problem of vanishing gradient and fast convergence, Jaeger ([20], [21], and [22]) proposed to randomly initialize input weights, recurrent weights and use linear regression to train only the output weights. It is worth noting that the initialized weights are not modified during the ESN training process which sometimes influence the global performance. To tackle the problem of initial randomness, several approaches based evolutionary computation were proposed to optimize the ESN hyper-parameters as well as the architecture in [23]. In the same context, other works tend to enhance the reservoir non-recurrent weights based on synaptic rule [24], [25], while in another work [26] the authors proposed to drive the activity of internal units to a desired distribution with the seek to maximize their entropy thanks to the Intrinsic Plasticity (IP) rule.

In [10], Bengio showed that deep representations handle more discriminative information which would increase the performance of the classifier. Along these lines, deep RNNs and especially deep reservoir computing received greater attention

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

in the neural networks community. The recently introduced Deep Echo State Network (DeepESN) [27] model paved the way to an extremely efficient approach handling deep temporal representations. The proposed works for DeepESN permits to shed light on the intrinsic properties of internal dynamics developed by hierarchical compositions of reservoir layers on the bias of depth in RNNs architectural design.

The current work deals with deep representation learning using DeepESN for EEG-based emotion recognition task. More specifically, the stacked reservoir layers are optimized by performing a bio-inspired algorithm, i. e. the IP rule. The latter tends to enhance the activity of each hidden neuron to respect a Gaussian distribution. The IP rule aims to maximize the entropy at each neuron of the reservoir layers, that means the neuron learns to maximize its information about the structure of the input sequence.

The rest of the paper is organized as follows. Section 2 gives an overview of existing works based on both feature and signal input for EEG-based emotion recognition. In Section 3, the DeepESN is detailed, followed by an explanation of the reservoir activity adaptation. The new hybrid learning algorithm is also presented. Section 4 presents the details of the experimental settings and then analyzes the achieved results. Section 5 recapitulates the presented work and outlines future affective research directions.

II. Related work on EEG-based Emotion Recognition

From a psychological point of view, Russell [28] developed the circumplex model of emotions which are distributed in a two-dimensional circular space, containing arousal and valence dimensions. Arousal represents the vertical axis and ranges from calm to excited. The valence represents the horizontal axis and ranges from unpleasant to pleasant. Mehrabian [29] showed that this circumplex is insufficient for the discrimination of anger and anxiety emotions. Hence, a third dimension called dominance is introduced which ranges from submissive to empowered. Low/High Valence (LVHV), Low/High Arousal (LAHA) and Low/High Dominance (LDHD) are examples of binary emotional classification.

To elicit emotions, various studies used picture or audio stimuli while others used video based emotion protocol. As consequence, affective benchmarks were proposed making the comparison of different emotion recognition models fair and possible. To the best of our knowledge, there are five recent affective benchmarks, to know, MAHNOB-HCI [30], DEAP dataset [31], SEED [32], DREAMER [33] and HR-EEG4EMO [34]. Among these datasets, DEAP is the most used. To add, DREAMER is an interesting dataset with low cost EEG device [35].

Power Spectral Density (PSD) features are extensively used in existing emotion recognition approaches ([36], [37] and [38]). The authors in [37] proposed a novel dynamical graph convolutional neural networks (DGCNN) to model the EEG features and then classify them with a softmax layer. The DGCNN model can dynamically learn the intrinsic relationship between different EEG channels, represented by an adjacency matrix, via training a neural network so as to benefit for more discriminative EEG features. Experiments on SEED dataset with differential entropy features achieved an average recognition accuracy of 90.4% for subject dependent scheme whereas 79.95% for subject independent cross-validation. Moreover, DGCNN model with PSD features in a subject-dependent experiment reached an average accuracy of 86.23%, 84.54% and 85.02% for the classification of LVHV, LAHA and LDHD on DREAMER dataset, respectively.

The work in [38] showed that the projection of PSD features on the scalp to take the spatial distribution in the form of a multiband feature matrix can considerably improve the performance. A capsule Network (CapsNet) is then used for the classification of valence, arousal and dominance levels on DEAP dataset. The subject-independent experiment achieves 66.73%, 68.28% and 67.25% for LVHV, LAHA and LDHD, respectively. Another point of view [39] consists in transforming the topographic PSD image through the pre-trained VGG 16 network of google team and reducing them with principal component analysis. Thereafter, a Long Short Term Memory (LSTM) model classifies the reduced PSD-VGG features. The proposed system achieves interesting results when tested on MAHNOB-HCI, DEAP and DREAMER datasets.

Most of the work done on EEG-based emotion recognition have suffered from finding informative features from EEG data. These findings have reshaped scientific understanding of EEG signals and inspired following works to analyze them directly instead of performing the feature extraction step. For example, feature learning was performed by feeding a one-second channel raw data from DEAP dataset to the Deep Belief Network (DBN) [36]. The new representation obtained from DBN is then fed to Support Vector Machine (SVM). The preliminary subject-dependent experiment results showed that the representation generated by DBN are comparable to the handcrafted PSD features for the classification of valence, arousal and dominance levels.

In a preliminary work [5], we proposed an ESN model for the classification of raw EEG data from DEAP dataset. The hidden layer allows to capture the temporal dynamic aspect of EEG channel signal and produce a new representation used by the output layer to classify emotions. One issue in ESN is that the random initialization of weights affects its performance. To alleviate this issue, the reservoir i. e. the hidden units of ESN is optimized with IP such that its activation follows a Gaussian distribution as suggested by [40]. Encouraging results are obtained, up to 71.03%, 68.28% for LVHV and LAHA, respectively. Even more, in [2] we investigated the impact of the plasticity form on the performance of the ESN model. Our findings showed that pre-training the reservoir with IP rule to result in an activation following a Gaussian distribution outperforms a reservoir weights pre-trained with the Anti-Oja rule [41] or the BCM rule [42] for the recognition of 3 emotions (Negative, Neutral and Positive) of SEED dataset. Another RNN-based work [43] proposed to decompose the EEG channel signal into 12 segments of 5 seconds duration and fed the sequence to LSTM model for emotion recognition on DEAP dataset. The model with two hidden layers considerably improve the performance in a subject-dependent experiment to attain 85.45% for LVHV and 85.65% for LAHA.

Inspired by the promising results achieved by our previous work [5], [2], we further analyze the ability of ESN enhanced with IP rule for deep EEG feature learning and emotion recognition. In our work, DeepESN is considered to ensure a robust representation from raw EEG data. More specifically, DeepESN with leaky integrator neurons is trained using ridge regression. A sensitivity analysis of the DeepESN hyper-parameters is performed to study the performance of the model. Our work can be considered as another proof of the success of deep reservoir computing approaches since the recognition of emotions from highly variable and personal EEG signals in a context of the subject-independent experiment is a challenging real-world application. Different from state-of-the-art works, the DeepESN model is trained in a hybrid fashion such that the output weights are first determined using ridge regression and in a second step they are trained with the online delta rule. Noticeably, the hybridization encompasses the problem of the random initialization of the output weights and it is not surprising that it achieves better results than the conventional learning method.

III. Classification and Representation learning with Intrinsic Plasticity based Deep Echo State Networks

In this section, the DeepESN architecture and its learning is presented. After that, the intrinsic adaptation of the reservoir activity is explained. The different ways for training DeepESN are detailed.

A. Deep Echo State Network model

Similar to the conventional shallow ESN model, the main component of DeepESN is the dynamical reservoir which embeds the input history into a rich state representation, and by a feed-forward readout layer that relies on the state encoded by the reservoir to compute the output. Crucially, DeepESN is organized into a hierarchy of stacked reservoirs where the output of each layer is fed to the next one, as depicted in Fig. 1. The topology of DeepESN is described by an input layer with size N_{in}, N reservoirs and an output layer with N_{out} neurons. The weights matrices are W^{in}, W_i^{res} where $i = \{1, 2, ...N\}$ and W^{out}, respectively. $W^{external}$ denotes the weights matrix of direct connections between successive reservoirs. At each time step t, after feeding the network with the input signal $u(t)$, the internal neurons from the first reservoir to the last one are activated according to the following equations:

$$\begin{aligned}
x_1(t+1) &= (1-\gamma)f^{res}(W_1^{res}x_1(t) + W^{in}u(t+1)) \\
&\quad +\gamma x_1(t), \\
x_2(t+1) &= \gamma x_1(t+1) + (1-\gamma)f^{res}(W_2^{res}x_2(t) \\
&\quad +W^{external}x_1(t+1)) + \gamma x_2(t), \\
&\dots \\
x_N(t+1) &= \gamma x_{N-1}(t+1) + (1-\gamma)f^{res}(W_N^{res}x_N(t) \\
&\quad +W^{external}x_{N-1}(t+1)) + \gamma x_N(t),
\end{aligned} \tag{1}$$

When $\gamma > 0$, the DeepESN is composed of leaky integrator neurons, otherwise if $\gamma = 0$ no leaky integrator is taken into consideration. The leaking rate γ of the internal hidden units can be assimilated to the speed of the reservoir update dynamics discretized in time. f^{res} is the activation function of the internal neurons, usually hyperbolic tangent or sigmoid. Initially, the weight matrices of all layers of the DeepESN are randomly initialized.

Subsequently, the output weights are computed following the ridge regression, also known as regression with $Tikhonov$ regularization:

$$W^{out} = Y^{target}X^T(XX^T + \beta I)^{-1}, \tag{2}$$

Where β is a regularization coefficient , and I is the identity matrix. Actually, large weights indicate that W^{out} exploits and amplifies tiny differences among the dimensions of $x(t)$, and can be very sensitive to deviations from the exact conditions in which the network has been trained. There are two objectives which are having a high training accuracy and obtaining small output weights. The regularization parameter β controls the compromise between these two objectives.

Ultimately, the output of DeepESN is expressed as follows:

$$y(t+1) = f^{out}(W^{out}[x(t+1); u(t+1)]). \tag{3}$$

where f^{out} is the activation function of the readout neurons, usually linear but it can be hyperbolic tangent or sigmoid.

B. Reservoir activity adaptation using Intrinsic Plasticity rule

Each reservoir layer of DeepESN is pre-trained with IP rule. Initially, Triesch [44] find out that the biological mechanism of a single neuron adapts its intrinsic excitability following the distribution of the given stimuli. While traditional learning algorithms update the weights of connections between neurons, the IP learning concerns the activation function of the neuron as highlighted in the first reservoir of DeepESN of Fig. 1.

More specifically, when the activation function is fermi, the target distribution of neuron activity is exponential as stated by Triesch [44]. Similarly, Schrauwen *et al.* [40] derived the IP rule for a Gaussian distribution and a hyperbolic tangent as activation function. IP rule is local, which means it is applied on each single neuron to maximize information about its input, thereby achieving an entropy maximization of the output. In addition, IP learning is unsupervised which implicitly tends to minimize the distance between the actual probability density of the neuron's output $\tilde{p}(x)$ and the desired probability density

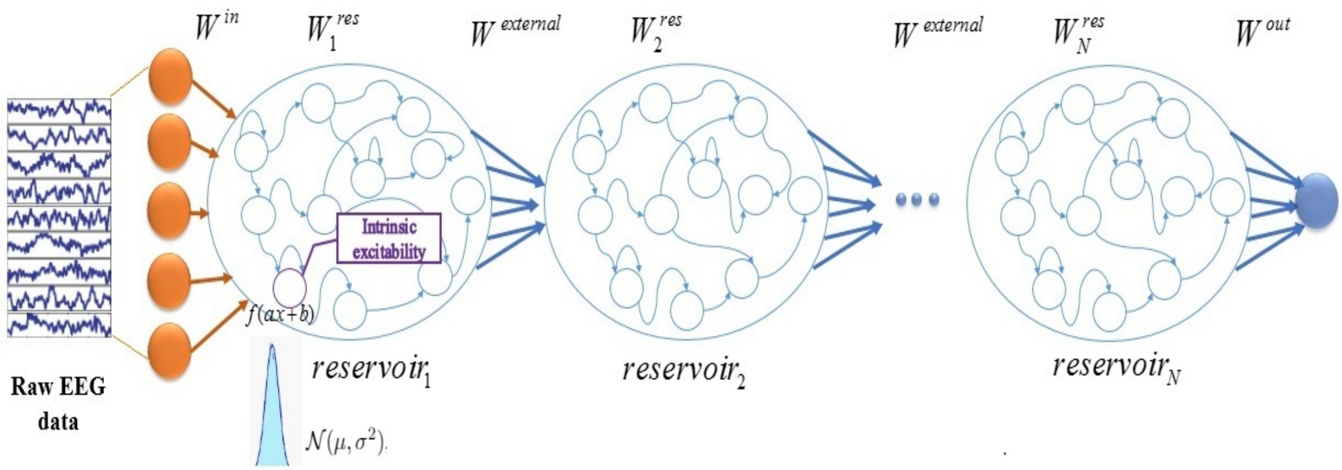

Fig. 1. The proposed system for EEG representation learning and emotion recognition.

$p_d(x)$ using the Kullback-Leiber divergence metric according to equation (4).

$$D_{KL}\left(\tilde{p}(x), p_d(x)\right) = \int p(x)\ log\left(\frac{p(x)}{p_d(x)}\right)\ dx. \quad (4)$$

For a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, the Kullback-Leiber divergence is quantified by equation (5).

$$\begin{aligned} D_{KL}\left(\tilde{p}(x) \parallel p_{\mathcal{N}}(x)\right) &= \int \tilde{p}(x)\ log\ \frac{\tilde{p}(x)}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}} dx \\ &= \int \tilde{p}(x)\ log\ (\tilde{p}(x))\ dx - \int \tilde{p}(x)\ log\ \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx \\ &= -H(x) + \frac{1}{2\sigma^2}\ E\left((x-\mu)^2\right) + log\ (\sqrt{2\pi}\sigma). \end{aligned} \quad (5)$$

Where $H(x)$ is the actual entropy and $E\left((x-\mu)^2\right)$ is the expectation value of the output distribution.

The stochastic gradient rule for adjusting the parameters a (gain) and b (bias) of the neuron's activation function at each time step t are expressed in equation (6) and (7).

$$\Delta a = \frac{\eta}{a} + \Delta b\left(W^{in}u(t) + W^{res}\ x(t-1)\right), \quad (6)$$

$$\Delta b = -\eta\left(-\frac{\mu}{\sigma^2} + \frac{x(t)}{\sigma^2}\left(2\sigma^2 + 1 - x(t)^2 + \mu\ x(t)\right)\right), \quad (7)$$

Hence, the activation of the reservoir neurons in the first layer is calculated by equation (8).

$$x(t) = f^{res}\left(diag(a)\left(W^{in}u(t) + W^{res}x(t-1)\right) + b\right). \quad (8)$$

The activation equation (1) is adapted according to the equation (8) for each reservoir R_i where $i = \{1, 2, ..., N\}$ internal neurons. (a_i, b_i) are the gain and the bias of the reservoir R_i characterizing each internal neuron.

C. Readout weight training

As mentioned above, the key ingredient of the success of the ESN and also the DeepESN models is the fact that only the output weights W^{out} have to be trained. In this context, we distinguish two modes: offline and online. The Moore-Penrose pseudoinverse is performed according to (9) and it is considered as an offline mode often known as linear regression:

$$W^{out} = Y^{Target} * X^T\left(XX^T\right)^{-1}, \quad (9)$$

When referring to equation (2) and setting $\beta = 0$ removes the regularization and makes the ridge regression a generalization of a regular linear regression. The training of the output weights aims to minimize the error between the desired output y^{Target} and the real output y when the training samples are presented in a sequential order. The delta rule is an online learning type performed by a stochastic gradient descent method originally proposed for the weights update of a single-layer perceptron model [45]. The weights update according to the delta rule is as follows:

$$\Delta W = \eta\left(y^{Target}(t) - y(t)\right)x(t), \quad (10)$$

$$W^{out} = W^{out} + \Delta W. \quad (11)$$

where η is the learning rate and t is the time step of the learning iterations, $t = \{1, 2, ..., T\}$. $x(t)$ is the vector of neuron firing activation states of x at time step t. This mechanism computes the incremental adaptation of readout weights. In our previous work [2], we proposed a new hybrid mode which first trains the output weights using linear regression and then trains them using the delta rule. In such a manner, the hybrid mode starts the training with the linear regression and proceeds with the delta rule, rather than randomly initialized weights. In the current work, the proposed system for the EEG-based recognition task relies on a leaky DeepESN pre-trained with IP rule and trained with the hybrid mode which combines the ridge regression and the delta rule.

TABLE I
DESCRIPTION OF DEAP AND DREAMER DATASETS

Experiment	DEAP dataset	DREAMER dataset
Stimuli	40	18
Subjects	32	23
EEG cap	Biosemi	Emotiv Epoc
Channels	32	14
Trial duration	one-minute	
SAM scales	1-9	[1,2,3,4,5]
Dimenions	Valence, Arousal and Dominance	

TABLE II
DETAILS OF DEEPESN HYPER-PARAMETERS

Hyper-parameter	Value
Input layer size	EEG time series size
W^{in} range	$[-0.1, 0.1]$
Reservoir size (N_R)	100, 150, 200, 300, 500
$W^{res} range$	$[-0.1, 0.1]$ $[-0.5, 0.5]$ $[-1, 1]$
Connectivity density α	0.1, 0.25, 0.5
Spectral radius ρ	0.5, 0.8, 0.9 0.95
Output layer size	Number of emotions to recognize
Leaking rate γ	0.1, 0.3, 0.5, 0.7, 1.5, 2
Regularization rate β	0.001, 0.01, 0.1
IP mean μ	0, 0.2
IP variance σ	0.1, 0.3, 1
IP learning rate η	0.0005, 0.005, 0.01
IP nb_iterations	5, 10

TABLE III
COMPARISON WITH STATE-OF-THE ART METHODS ON DEAP DATASET BASED-ON ACCURACY(%)

Study	Input	Classifier	LVHV	LAHA	LDHD
[38]	Multiband PSD matrix	CapsNet	66.73	68.28	67.25
[39]	PSD-VGG	LSTM	71.09	72.58	—
[5]	raw data	ESN-IP with offline mode	71.03	68.28	—
[2]	raw data	ESN-IP with hybrid mode	71.25	69.23	—
Ours	raw data	Leaky DeepESN-IP	**81.95**	**83.02**	**85.32**

TABLE IV
COMPARISON WITH STATE-OF-THE ART METHODS ON DREAMER DATASET BASED-ON ACCURACY(%)

Study	Input	Classifier	LVHV	LAHA	LDHD
[33]	PSD	SVM	62.49	62.17	61.84
[39]	PSD-VGG	LSTM	78.99	79.23	—
Ours	raw data	Leaky DeepESN-IP	**82.11**	**83.58**	**84.98**

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this work, we propose a new methodology to automatically discover relevant representation for an emotion recognition task. To validate the effectiveness of our methodology, we tested it on two public EEG benchmarks, namely, DEAP [31] and DREAMER [33] datasets. First, details about used datasets and the fixed parameters for DeepESN are provided. Next, we present the results for three classification problems, i.e. LVHV, LAHA, and LDHD. Note that the experiments here are subject-independent.

A. Experimental settings

The authors of DEAP dataset [31] elicited emotions through 40 videos of 32 participants. Biosemi EEG cap with 32 channels is used for the recording of EEG signals. The participants reported their feelings of the valence, arousal and dominance dimensions by rating a SAM questionnaire with a continuous scaling from 1 to 9. Recently, [33] proved the effectiveness of the Emotiv Epoc with 14 channels [35] for the collection of EEG signals in an emotion recognition protocol. The experiment involved 23 subjects stimulated with 18 videos. The participants rated their feelings through SAM questionnaire with discrete scaling from 1 to 5 for valence, arousal and dominance dimensions. Both trials in the two datasets are of one-minute length. The description of the two protocols details is depicted in Table I. The EEG trials were pre-processed following the process given in [31] and [33]. Each trial is divided into segments of one-second duration. The input of leaky DeepESN is one segment.

As mentioned previously, our methodology is inspired from biologically plausible reservoir computing paradigm (([20], [21], [22]), and [40]). In Table II, we highlight the tuning of hyper-parameters for DeepESN model with two reservoirs. The input size is determined by the length of the EEG channel. The number of emotions to discriminate is equivalent to the output size. In our work, the hidden units are responsible for the extraction of relevant features from raw EEG data. All the reservoir size are less than the input size. Actually, the reservoir computing approach performance is conditioned by the selection of a number of hyper-parameters which are the range of the input weights W^{in} and reservoir weights W_i^{res}, the number of neurons connected in the reservoir α and the spectral radius ρ. The authors in [20], [21], and [22] suggested to set $\rho(W^{res}) < 1$ in order to satisfy the echo state property. The reservoir layer is pre-trained using the IP rule. Numerous values were tested for IP parameters and here we report all the tested combinations.

B. Affective recognition results and interpretation

As mentioned in the related work section, most studies focus on the feature extraction step and few works attempt to use deep learning models to learn discriminative representation from raw EEG data. Table III depicts the comparison of recent works on DEAP dataset using features as input. Compared to the methodology based on PSD-VGG features reduced by PCA and classified by LSTM, our leaky DeepESN-IP achieves an improvement with almost 10.86%, 10.44% for LVHV and

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

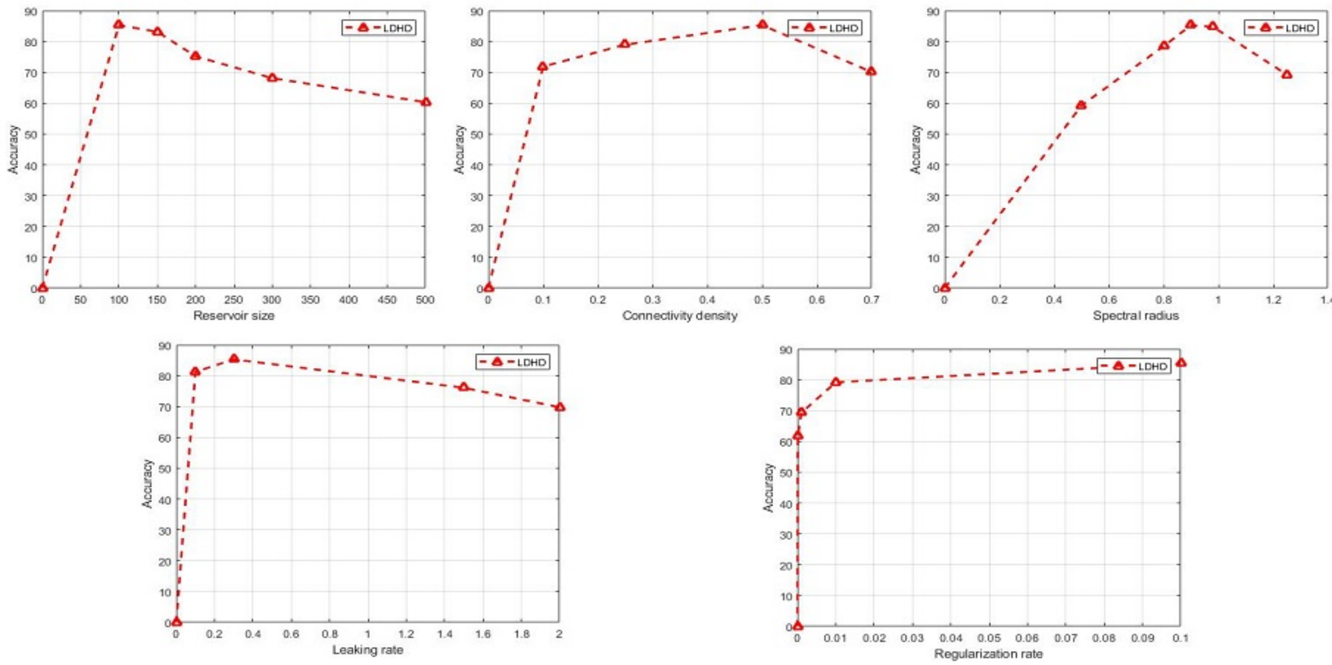

Fig. 2. The impact of DeepESN hyper-parameters on the performance using DEAP dataset.

LAHA, respectively. For the classification of LDHD, the proposed system outperforms the work based on the classification of the multiband PSD matrix with the CapsNet to reach 85.32%. It can be seen from Table III, the deep representations obtained with the leaky DeepESN-IP trained with the hybrid mode increased the recognition rate with 10.70% and 13.79% for LVHV and LAHA, respectively.

Since DREAMER dataset [33] is recently released, few works are conducted in the literature, especially for the subject-independent experiment. Table IV reports the basic work of the DREAMER authors using the popular PSD features with SVM classifier. The PSD-VGG features in [39] again showed better results than [33] for the classification of valence and arousal two levels. Our proposed DeepESN reached the higher accuracy on DREAMER dataset to obtain 82.11%, 83.58% for LVHV and LAHA, respectively.

According to Tables III and IV, it is remarkable that the highest accuracy is attained when classifying the two dominance levels. The classification task strongly depends on the labels of sample data. We think that the dominance level in the SAM questionnaire was clear for the participants such that their ratings were not overlapping.

The sensitivity analysis is very important for the reservoir computing approach. It actually gives us detailed insights into how changing the main hyper-parameters of the DeepESN changes the overall behaviour of the network. Fig. 2 illustrates the impact of changing the reservoir size, the connectivity density the spectral radius, the leaking rate and the regularization rate. Note that, these hyper-parameters are with same values for the two reservoir in our leaky DeepESN-IP. Since DEAP contains more trials than DREAMER, we report the evaluation on DEAP dataset. For instance, increasing the number of the hidden units in each reservoir leads to overfitting problem. The best achieved results in our work are handled with two dense reservoirs ($\alpha = 5\%$). For a greater value of α, the accuracy decreases to 70.10%. The spectral radius determines the forgetting factor of DeepESN. When it is near 0.9, the proposed system attains its maximum accuracy. Furthermore, a big value for the leaking rate γ influences negatively the accuracy. Finally, the highest the value of the regularization rate β the better the accuracy rate is.

V. Conclusion

In this work, we have proposed a new method based on deep ESNs to automatically extract suitable representations from raw EEG time series towards real-time application context. DeepESNs allow us to transfer the striking advantages of the ESN methodology to the case of deep recurrent architectures, resulting in an efficient approach to conceiving deep neural networks for temporal data. It is interesting to note that intrinsic plasticity can enhance the DeepESN's performance on emotion recognition validated on two well known affective benchmarks. The experiments are challenging since the classification is subject-independent, i.e., is subject to high variability of EEG signals between subjects.

Our findings suggested that DeepESN based on IP rule can be further enhanced by optimizing the synaptic weights inside each reservoir using the anti-Oja rule [41] or the BCM rule [42]. Synergies between different plasticity rules as shown in [25] will be beneficial for generating richer dynamics of the reservoir, thereby being for efficient in learning spatio-temporal representations raw EEG streams.

Acknowledgment

The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

References

[1] A. Baghdadi, Y. Aribi, and A. M. Alimi, "Efficient human stress detection system based on frontal alpha asymmetry," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 858–867.

[2] R. Fourati, B. Ammar, J. Sanchez-Medina, and A. M. Alimi, "Unsupervised learning in reservoir computing for eeg-based emotion recognition," *IEEE Transactions on Affective Computing, to be published*, 2020, doi:10.1109/TAFFC.2020.2982143.

[3] A. Baghdadi and Y. Aribi, "Effectiveness of dominance for anxiety vs anger detection," in *2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME)*. IEEE, 2019, pp. 1–4.

[4] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals," *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 2018.

[5] R. Fourati, B. Ammar, C. Aouiti, J. Sanchez-Medina, and A. M. Alimi, "Optimized echo state network with intrinsic plasticity for eeg-based emotion recognition," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 718–727.

[6] A. Baghdadi, Y. Aribi, and A. M. Alimi, "A survey of methods and performances for eeg-based emotion recognition," in *International Conference on Hybrid Intelligent Systems*. Springer, 2016, pp. 164–174.

[7] F. Movahedi, J. L. Coyle, and E. Sejdić, "Deep belief networks for electroencephalography: A review of recent contributions and future outlooks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 642–652, 2017.

[8] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, 2018.

[9] B. García-Martínez, A. Martinez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero, "A review on nonlinear methods using electroencephalographic recordings for emotion recognition," *IEEE Transactions on Affective Computing, to be published*, 2019, doi:10.1109/TAFFC.2018.2890636.

[10] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[11] D. Mellouli, T. M. Hamdani, and A. M. Alimi, "Deep neural network with rbf and sparse auto-encoders for numeral recognition," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*. IEEE, 2015, pp. 468–472.

[12] D. Mellouli, T. M. Hamdani, M. B. Ayed, and A. M. Alimi, "Morph-cnn: a morphological convolutional neural network for image classification," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 110–117.

[13] C. B. Chaabane, D. Mellouli, T. M. Hamdani, A. M. Alimi, and A. Abraham, "Wavelet convolutional neural networks for handwritten digits recognition," in *International Conference on Health Information Science*. Springer, 2017, pp. 305–310.

[14] D. Mellouli, T. M. Hamdani, J. J. Sanchez-Medina, M. B. Ayed, and A. M. Alimi, "Morphological convolutional neural network architecture for digit recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2876–2885, 2019.

[15] S. Bousnina, B. Ammar, N. Baklouti, and A. M. Alimi, "Learning system for mobile robot detection and tracking," in *2012 International Conference on Communications and Information Technology (ICCIT)*. IEEE, 2012, pp. 384–389.

[16] H. Brahmi, B. Ammar, F. Cherif, and A. M. Alimi, "On the dynamics of the high-order type of neural networks with time varying coefficients and mixed delay," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 2063–2070.

[17] R. Fourati, C. Aouiti, and A. M. Alimi, "Improved recurrent neural network architecture for svm learning," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*. IEEE, 2015, pp. 178–182.

[18] H. Brahmi, B. Ammar, F. Chérif, and A. M. Alimi, "Stability and exponential synchronization of high-order hopfield neural networks with mixed delays," *Cybernetics and Systems*, vol. 48, no. 1, pp. 49–69, 2017.

[19] B. Ammar, H. Brahmi, and F. Chérif, "On the weighted pseudo-almost periodic solution for bam networks with delays," *Neural Processing Letters*, vol. 48, no. 2, pp. 849–862, 2018.

[20] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note," *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, vol. 148, no. 34, p. 13, 2001.

[21] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, "Optimization and applications of echo state networks with leaky-integrator neurons," *Neural Networks*, vol. 20, no. 3, pp. 335–352, 2007.

[22] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009.

[23] N. Chouikhi, B. Ammar, A. Hussain, and A. M. Alimi, "Bi-level multi-objective evolution of a multi-layered echo-state network autoencoder for data representations," *Neurocomputing*, vol. 341, pp. 195–211, 2019.

[24] M.-H. Yusoff, J. Chrol-Cannon, and Y. Jin, "Modeling neural plasticity in echo state networks for classification and regression," *Information Sciences*, vol. 364, pp. 184–196, 2016.

[25] X. Wang, Y. Jin, and K. Hao, "Evolving local plasticity rules for synergistic learning in echo state networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2019.

[26] ——, "Echo state networks regulated by local intrinsic plasticity rules for regression," *Neurocomputing*, vol. 351, pp. 111–122, 2019.

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

[28] J. A. Russell, "A circumplex model of affect." *Journal of Personality and Social Psychology*, vol. 39, no. 6, p. 1161, 1980.

[29] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[30] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[31] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[32] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from eeg," *IEEE Transactions on Affective Computing, to be published*, 2017, doi:10.1109/TAFFC.2017.2712143.

[33] S. Katsigiannis and N. Ramzan, "Dreamer: a database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2018.

[34] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, and L. Albera, "Emotion recognition based on high-resolution eeg recordings and reconstructed brain sources," *IEEE Transactions on Affective Computing, to be published*, 2017, doi:10.1109/TAFFC.2017.2768030.

[35] H. Ekanayake, "P300 and emotiv epoc: Does emotiv epoc capture real eeg?" *Web publication http://neurofeedback. visaduma. info/emotivresearch. htm*, 2010.

[36] X. Li, P. Zhang, D. Song, G. Yu, Y. Hou, and B. Hu, "Eeg based emotion identification using unsupervised deep feature learning," in *SIGIR2015 Workshop on Neuro-Physiological Methods in IR Research*, 2015.

[37] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing, to be published*, 2018, doi:10.1109/TAFFC.2018.2817622.

[38] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion recognition from multiband eeg signals using capsnet," *Sensors*, vol. 19, no. 9, p. 2212, 2019.

[39] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Transactions on Affective Computing, to be published*, 2019, doi:10.1109/TAFFC.2019.2916015.

[40] B. Schrauwen, M. Wardermann, D. Verstraeten, J. J. Steil, and D. Stroobandt, "Improving reservoirs using intrinsic plasticity," *Neurocomputing*, vol. 71, no. 7-9, pp. 1159–1171, 2008.

[41] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, 1982.
[42] G. Castellani, N. Intrator, H. Shouval, and L. Cooper, "Solutions of the bcm learning rule in a network of lateral interacting nonlinear neurons," *Network: Computation in Neural Systems*, vol. 10, no. 2, pp. 111–121, 1999.
[43] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *Emotion*, vol. 8, no. 10, 2017.
[44] J. Triesch, "A gradient rule for the plasticity of a neuron's intrinsic excitability," in *International Conference on Artificial Neural Networks*. Springer, 2005, pp. 65–70.
[45] J. C. Pemberton and J. J. Vidal, "When is the generalized delta rule a learning rule? a physical analogy," in *IEEE International Conference on Neural Network*, vol. 1, 1988, pp. 309–315.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE