Unsupervised Learning in Reservoir Computing for EEG-based Emotion Recognition

Rahma Fourati, *Student Member, IEEE,* Boudour Ammar, *Senior Member, IEEE,* Javier Sanchez-Medina, *Senior Member, IEEE* and Adel M. Alimi, *Senior Member, IEEE*

Abstract—In real-world applications such as emotion recognition from recorded brain activity, data are captured from electrodes over time. These signals constitute a multidimensional time series. In this paper, Echo State Network (ESN), a recurrent neural network with a great success in time series prediction and classification, is optimized with different neural plasticity rules for classification of emotions based on electroencephalogram (EEG) time series. Actually, the neural plasticity rules are a kind of unsupervised learning adapted for the reservoir, i.e. the hidden layer of ESN. More specifically, an investigation of Oja's rule, BCM rule and gaussian intrinsic plasticity rule was carried out in the context of EEG-based emotion recognition. The study, also, includes a comparison of the offline and online training of the ESN. When testing on the well-known affective benchmark "DEAP dataset" which contains EEG signals from 32 subjects, we find that pretraining ESN with gaussian intrinsic plasticity enhanced the classification accuracy and outperformed the results achieved with an ESN pretrained with synaptic plasticity. Four classification problems were conducted in which the system complexity is increased and the discrimination is more challenging, i.e. inter-subject emotion discrimination. Our proposed method achieves higher performance over the state of the art methods.

Index Terms—Emotion recognition, Electroencephalogram, Echo state network, synaptic plasticity, intrinsic plasticity.

✦

1 INTRODUCTION

AFFECTIVE Computing [1] is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science [2]. In 1990s, Picard pointed out in her book "Affective Computing" [3], that imbuing machines with the ability of detecting, recognizing, and processing the human emotion is necessary to further enhance human machine interaction. Toward that more reliable interaction, Picard defined three major applications of affective computing: (i) systems that detect and recognize emotions, (ii) systems that express emotions (e.g., avatars, agents), and (iii) systems that feel emotions. Specifically, Emotion Recognition (ER) drew the most attention of scientists.

In the early stage of affective computing, the proposed works focused on the recognition and synthesis of facial expression, and the synthesis of voice inflection. After that, a variety of physiological measurements are available which would yield clues to ones hidden affective state. Affective wearables offer possibilities of new health and medical research opportunities and applications. Medical studies could move from measuring controlled situations in laboratory, to measuring more realistic situations in life.

Electroencephalogram (EEG) is a direct measurement of brain activity. EEG gains interest thanks to its excellent temporal resolution, i.e. it is recorded on a millisecond. Note that, human cannot control his brain activity, it is unconscious. In addition, emotion recognition from other modalities which are external manifestations can lead to inaccurate interferences particularly where the included subjects in the experiment control their emotions and suppress them or alter them to exhibit false emotions. Nevertheless, the evoked issue is not available for the case of EEG modality. Emotion recognition based on EEG classifies the inner human emotions. The issue for EEG-based emotion recognition comes down to the variability from an individual to another of the recorded EEG signals in response to a stimulus, which remains the inter-subject emotion discrimination a challenging task.

Diverse writers from psychology defined affect or emotion differently. Ekman [4] distinguished six basic emotions which are happiness, sadness, surprise, disgust, anger, and fear. Whereas, Russel [5] defined two dimensions of emotion. Although the precise names vary, the two most common categories for the dimensions are arousal (calm/excited), and valence (negative/positive). Mehrabian [6], [7] showed that a third dimension is required to discriminate between anger and anxiety. The third dimension tends to be called dominance. It ranges from submissive (or without control) to dominant (or in control/empowered). Another interesting representation of emotion belongs to Sun *et al..* [8]. They improved the valence-arousal space in order to handle emotions by computers. The adaptation was improved by the introduction of the typical fuzzy subspace.

One way to design emotion recognition system is the use of neural networks. There are two types of neural networks: feedforward and recurrent. The FeedForward Neural Networks (FFNNs) are characterized by their activation, they fed forward from input to output through "hidden

- *R. Fourati, B. Ammar and A. M. Alimi are with the Research Groups in Intelligent Machines, Department of Electrical and Computer Engineering, National Engineering School of Sfax, University of Sfax, Sfax 3038, Tunisia*
  E-mail: {rahma.fourati, boudour.ammar, adel.alimi}@ieee.org
- *J. J. Sanchez-Medina is with the Innovation Center for the Information Society, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain.*
  E-mail:javier.sanchez.medina@ieee.org

layers" like "Multi-Layer Perceptrons" (MLP), "Radial Basis Function Network" (RBFN) [9], etc. However, natural neural networks are usually connected in more complex structures, like recurrent synaptic connections. Recurrent Neural Networks (RNNs) try to emulate that recurrence, which is more suited for dynamical systems modeling.

ESNs are a class of RNN. They are characterized by being three-layered NNs, with a high level of recurrence. The hidden layer is called reservoir. ESNs are well fitted for a number of applications, like time series prediction [10], [11], [12], [13] or robot control [14], usually showing high performance. As the EEG is temporal signal, ESNs are a natural match for emotion recognition. In our proposed methodology, we used the reservoir to encode the spatio-temporal information of the EEG signals. It is well known that the major drawback of the ESN is its random initialization which affects both the convergence state and the performance rate. In the current work, we circumvent this problem by adding an unsupervised learning of the random reservoir before the learning of the output layer. Here, we investigate the reservoir adaptation with two different techniques which are the synaptic plasticity and the intrinsic plasticity (IP). While the formal adapts the weights of reservoir synapses, the latter adapts the intrinsic excitability of each neuron in the reservoir to the emotion recognition task.

In this paper, we proposed an enhanced ESN for EEG-based emotion recognition adopting the two-dimensional model of emotions. This work is distinguished by neglecting the feature extraction step and feeding ESN directly with the raw EEG as input. To value this choice, we have also extracted the power band features in the time-frequency domain in order to compare the influence of input type on the classification performance. The main contributions of this work are (1) an empirical study of reservoir behaviour improved with unsupervised learning step in an EEG-based emotion recognition task and (2) an extensive validation of the proposed methodology where four classification problems were conducted raising different level of complexity, i.e. Low/High Valence discrimination, Low/High Arousal discrimination, Stress/Calm discrimination and 8 emotional states discrimination.

The outline of this paper is composed of 4 sections. Section 2 first describes the existing affective benchmarks and next overviews state of the art methods for EEG-based emotion recognition. Section 3 begins with the ESN model, and next introduces the different rules of plasticity and details the proposed approach. Section 4 illustrates the experimental results and discussion. Section 5 summarizes the paper and outlines our future work.

2 LITERATURE REVIEW OF EEG-BASED EMOTION RECOGNITION

Several works for EEG-based emotion recognition were proposed including signal processing and machine learning methodologies. In this section, we present the DEAP dataset and the recent works validated using it.

TABLE 1
Emotional states projection in the VAD space

VAD levels	Emotional State
HVLALD	Protected
HVLAHD	Satisfied
HVHALD	Surprised
HVHAHD	Happy
LVLALD	Sad
LVLAHD	Unconcerned
LVHALD	Frightened
LVHAHD	Angry

2.1 Existing Affective benchmarks

At first blush, there are several works on EEG-based for emotion recognition. Unfortunately, each proposed approach was validated on a specific EEG experiment, i.e. a private EEG dataset. To the best of our knowledge, there are five publicly accessible affective benchmarks: MAHNOB-HCI [15], DEAP [16], SEED [17], DREAMER [18] and HR-EEG4EMO [19]. The DEAP dataset is the most used one.

The elicitation protocol of DEAP dataset used 40 emotional videos. It was conducted on 32 participants (16 male and 16 female) wearing Biosemi Active 2 acquisition system with 32 EEG channels and 8 peripheral sensors. For each trial, participant ratings were expressed in the three-dimensional space such that arousal, valence and dominance. Consequently, varied classification problems were yielded such as Low/High valence (LVHV), Low/High arousal (LAHA) and Low/High Dominance (LDHD). The authors of [20] defined two criteria for discrimination of anxiety levels. Stress label is obtained if valence<=3 and Arousal>=5. Calm label is obtained if 4<=Valence<=6 and Arousal<4. Decomposing the valence arousal space into 4 quadrants leads to 4 emotions which are LALV, LAHV, HALV and HAHV [21]. Combination of the three dimensions leads to 8 emotional states [22] as shown in Table 1.

2.2 Related work on DEAP dataset

In the literature, most of the existing methods begins with the signal decomposition and the feature extraction steps. As EEG signals contain both time and frequency information, feature extraction methods also differ and belongs to three kinds: time-domain, frequency-domain and time-frequency domains. Several works were proposed for classification of 2 valence or arousal levels as in [24], [25] and [26]. There is a lack of proposed works for classification of more than 2 emotions and that's mainly due to the poor achieved results.

A very important aspect in the classifier is whether the task is user dependent or independent. Portioning training data and test data form the same subject and adapting the classifier on a specific subject makes the task subject dependent which has improved results. For instance, in [24] classification of statistical features extracted from DEAP dataset achieves 82.76% and 82.77% for 2 valence and arousal levels, respectively with k-Nearest Neighbor (k-NN). But, here the lack of generalization is the cost to pay. Hence, proposing

TABLE 2
Previous works on DEAP dataset

Study	Year	Subject-Independent	Input	#Channels	Classifier	Affective states	Performance (%)
[23]	2019	Yes	DWT coefficients	1	MLP	Happy/Sad	58.50
[24]	2018	No	Statistical features	32	k-NN	LAHA	82.77
						LVHV	82.76
[25]	2018	Yes	Entropy and energy features of 4s	10	k-NN	LAHA	89.81±0.46
				14			92.24±0.33
				18			93.69±0.30
				32			95.69±0.21
				10		LVHV	89.54±0.81
				14			92.28±0.62
				18			93.72±0.48
				32			95.70±0.62
[20]	2017	Yes	Entropy features	32	SVM	Stress and Calm	81.31
[26]	2017	No	IMF features of 5s	8	SVM	LAHA	71.99
				32			72.10
				8		LVHV	69.10
				32			70.41
[27]	2017	Yes	Raw EEG	32	ESN	LAHA	68.28
						LVHV	71.03
						8 emotions	68.79
[21]	2017	Yes	Differential entropy of 1s	32	GELM	LALV, HALV, LAHV, and HAHV	69.67
[28]	2017	No	Raw EEG segment	32	LSTM	LAHA	85.65
						LVHV	85.45
[29]	2015	No	PSD of 1s	32	SVM	LAHA	64.30
						LVHV	58.20
			DBN output from 1s raw EEG			LAHA	64.20
						LVHV	58.40
[30]	2014	Yes	Raw EEG	32	HMM	LAHA	55.00±4.5
						LVHV	58.75±3.8
[31]	2014	Yes	Bandpower features	10	SVM	LAHA	64.90
				32			62.90
			PSD features	10			63.00
				32			63.40
			Bandpower features	10		LVHV	64.90
				32			62.30
			PSD features	10			56.40
				32			60.00
[22]	2013	Yes	Fractal dimension features	16	SVM	8 emotions	65.63
				32			69.53

a model trained and tested with data from independent users will lead to an easy application to new subject, i. e. no requirement to design a new model for the new subject.

To achieve this merit, several works focused on the feature extraction step aiming to find the most relevant features for EEG-based ER. Power spectral features (PSD) using Short-Time Fourier Transform as in [29] and [31] are considered the baseline features and the most used. Fractal Dimension (FD) [22], entropy and energy features [25] have shown acceptable results as depicted in Table 2. Other works tested features of decomposition techniques such as Discrete Wavelet Transform (DWT) [23] or Intrinsic Mode Functions (IMF) [26]. Most of the work done on EEG-based ER 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, [30] classified the EEG preprocessed signals using the Hidden Markov Model (HMM). Likewise, feature learning was performed by feeding the raw channel data to the Deep Belief Network (DBN) [29]. The new representation obtained from DBN is then fed to Support

Vector Machine (SVM). A comparison with PSD features showed that raw EEG data achieved better results. The task here is subject-dependent and the length of the trial is 1 second. Note that, EEG signals are acquired from a number of channels.

Investigation of the impact of specific channels on the ER performance is very important. [23] finds that one channel F4 is sufficient for an EEG-ER task. Also, the authors in [31] showed that bandpower features from 10 channels handle better results than using 32 channels for classifying 2 levels of valence and arousal, respectively. In the current work, ESN has as input the preprocessed EEG raw and the classification is also performed by the readout layer.

The use of ESN for processing EEG data is not new at all, but the works that are based on ESNs are really few. In fact, there are only two works [32] and [33]. We should notice that the input is event-related potential features. Bozkhov *et al.* [32] have pretrained the reservoir with IP in order to have optimal values of IP parameters, i.e. the gain and the bias. The next step is to compute the steady states of the reservoir. Bozkhov *et al.* [32] proved that the discrimination of the steady states is more efficient due to the adequate representation of the input data with the IP. Therefore, the best performance 76.9% found with using Linear Discriminant Analysis (LDA).

Bozkhov *et al.* have also integrated a feature selection step by using projection of 2D, 3D and 4D of the steady states. The input of the ESN is a feature vector with 252 attributes. The experiment was conducted on 26 females seeing positive and negatives emotional pictures from IAPS and wearing an EEG cap with 21 channels. The best recognition rate up to 98.1% was by using 4D projections with SVM. In a follow up work of Bozkhov *et al.* [33], it was demonstrated that the data representation with reservoir computing outperforms the autoencoder. In fact, the highest classification accuracy achieved is 81% with 2 layers autoencoders.

Recently, [27] proposed ESN pretrained with IP and fed it with preprocessed EEG data. The aim of the work is to show the effectiveness of IP rule on the reservoir layer for performing feature extraction step. The classification of arousal levels, valence levels and 8 emotional states resulted in 68.28%, 71.03% and 68.79% respectively.

Thus, the use of ESN and its promising results in ER encourage us to further use the ESN for feature extraction and classification of emotions from EEG signals. We are different from state-of-the-art methods [32] and [33]. We use ESN as an architecture for both representation of input data and its classification. The projection and the classification ranking in [32] and [33] are more complex and hence the computational time is high.

While our work focus only on the EEG modality, we emphasize a recently published work using ESN pretrained with IP having as input both EEG and physiological signals [34]. After calculating the asymmetry index from the frontal lobe channels, and based on its values some signals were discarded in the recognition process. Next, Wavelet Packet Transform (WPT) is applied to extract 4 sub-bands. After that, k-means is used to cluster the generated WPT coefficients of each window in each channel. This work used ESN for feature selection step that means to reduce the size of
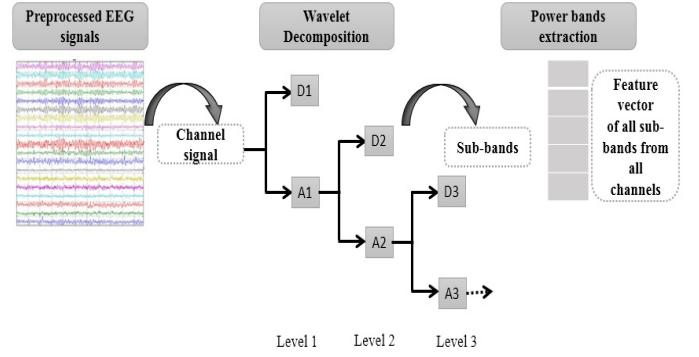

Fig. 1. Power bands feature extraction process.

the first feature vector. The output of the reservoir states is then fed to SVM with RBF kernel. The authors in [34] tested the proposed methodology on DEAP dataset labeled with the 4 quadrants LALV, HALV, LAHV and HAHV to yield 78.2% as recognition rate. Compared to our work, our feature extraction step is more simple and benefit from all information of the raw EEG data.

Our objective goes more than one form of plasticity, we further investigate a second form and two techniques of ESN training which are the offline and the online modes. Even if in [35] the authors studied the effect of synaptic plasticity with these two modes, it looks that the study was evaluated on time series of small size and the IP form was ignored. Briefly, we distinguish our work by illuminating the important roles of both adapting the reservoir and training mode to a more challenging task, i. e. EEG-ER.

3 METHODOLOGY

In our methodology, we consider the preprocessed EEG signals. To justify such choice, we also perform a feature extraction step and fed power bands features to ESN to recognize emotions. Note that, our methodology consists in three steps which are feature extraction, reservoir pretraining with one kind of plasticity rule and readout layer training for the classification.

3.1 Feature Extraction

EEG signals are not stationary. To handle this hypothesis, it was shown that wavelet decomposition is more convenient than Fourier transform [36]. In our study, we applied DWT to decompose EEG channel signal into different bands. The DWT coefficients are defined as follows:

$$C_{x(t)}(l,n) = \int_{-\infty}^{\infty} x(t)\Psi_{l,n}(t)dt \qquad (1)$$

$$\Psi_{l,n}(t) = 2^{-(l+1)}\Psi(2^{-(l+1)}(t - 2^{-1}n)) \qquad (2)$$

The EEG signal, $x(t)$, is correlated with a wavelet function $\Psi_{l,n}(t)$. The variable l and n are the scale and translation variables of the wavelet function, respectively. A dyadic scale is performed as in (2) to chose the scale and translation variables. Note that, the original signal is reconstructed if orthogonality is ensured [37]. The scale variable provides analysis in frequency domain: compressed version of the
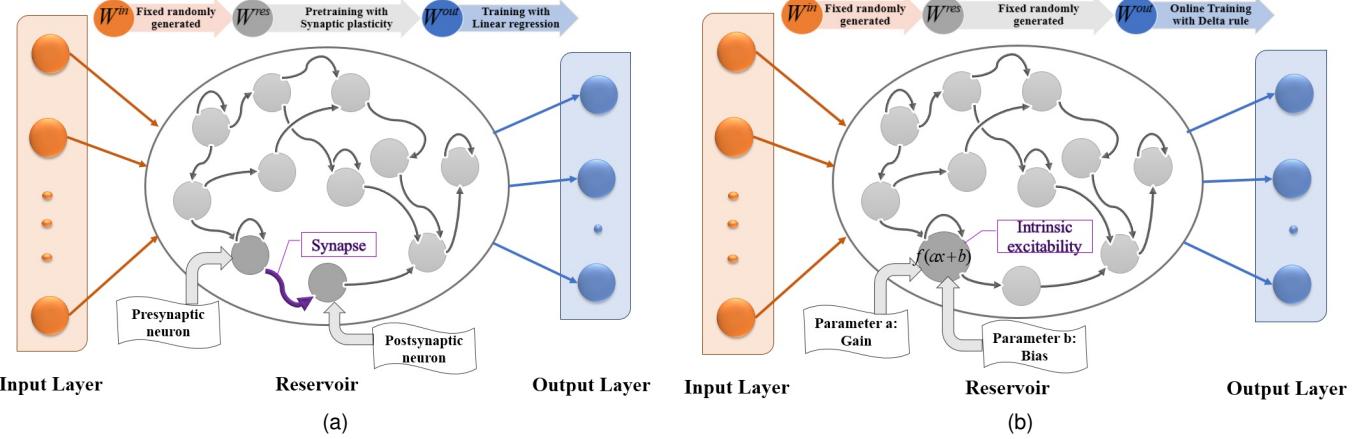
Fig. 2. Unsupervised learning of Reservoir layer using (a) Synaptic plasticity rule and offline mode and (b) Intrinsic plasticity rule and online mode.

wavelet function corresponds to the high frequency components of the original signal while the stretched version corresponds to the low frequency components; and the translation variable provides analysis in the time domain. The decomposition of DWT leads to coefficients of high frequency which are "detail" and coefficients of coarse approximation of original signal in time domain which are "approximation". After decomposition, power features from all bands are extracted. Note that, power bands features are the most popular features in the context of EEG-based emotion recognition. The Definition of EEG frequency bands differs slightly between studies. Finally, a feature vector is formed through the concatenation of all power features from all channels of the same EEG trial. Thus, the size of the feature vector is equal to the number of the bands multiplied by the number of channels. Fig. 1 illustrates the feature extraction process adapted in our methodology.

3.2 ESN Model

ESNs are a type of recurrent neural networks proposed by Jaeger [38]. It is composed of input layer, a reservoir and an output layer often called readout layer as depicted in Fig. 2a. The recurrence in ESNs is handled through the recurrent connections between hidden units and the possible feedback connection from the output layer to the reservoir. Direct connection from input to output layer can be added.

The basic structure proposed for solving computational intelligence problems was the FFNN. However, for specific problems which have dynamic and temporal nature, FFNNs are not able to handle complex temporal machine learning problems. As a solution, recurrent connections are added to the structure, giving birth to RNNs for handling several problems such as SVM leaning process [39]. Training RNNs was done by extending the canonical backpropagation algorithm to be Back-Propagation-Through-Time (BPTT) [40]. One of the major limitations of BBTT is its high computational cost and its slow convergence. Hence, Jaeger *et al.* [38] proposed the reservoir computing approach to enhance the learning process.

The reservoir is composed of sparsely connected neurons with cycled connections. The weights from input to

reservoir and the inner weights of reservoir are randomly initialized and remain unchanged during training phase. The simplicity of ESN turns out in the need of training only the weights of the readout layer which is a linear regression. Consider a topology of I input neurons, R internal neurons and O output neurons. The first step in training ESN is to collect the matrix of activation states of each neuron in the reservoir. The activation equation at each time step is expressed in (3):

$$x(t) = f_{res}\left(W^{in}u(t) + W^{res}x(t-1)\right) \qquad (3)$$

where W^{in} and W^{res} are the weights of the input and the reservoir layers respectively. f_{res} is the non-linear activation function, usually, a sigmoid. The second step consists in calculating the output weights. Here, we distinguish two modes: offline and online. Linear regression is calculated as follows in (4) and it is considered as an offline mode:

$$W^{out} = Y_{target} * \left(X^T X\right)^{-1} \qquad (4)$$

The training in the online mode is ensured through the minimization between the target output and the produced output. The presentation of the training samples is in the sequential form.

Originally proposed for MLP model, the delta rule is a stochastic gradient descent method used for the update of one layer of MLP neural network [41]. It is expressed as follows:

$$\Delta W = \eta \left(y^{desired}(t) - y(t)\right) x(t) \qquad (5)$$

$$W^{out} = W^{out} + \Delta W \qquad (6)$$

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.

Then, the output of the ESN can be generated according to (7) using linear regression if f_{out} is a linear function:

$$y(t) = f^{out}\left(W^{out} * x(t)\right) \qquad (7)$$

For a specific input sample, the neuron having the highest activation score wins, while the other neurons in the same

layer are inhibited. This mechanism is called winner-take-all. Assuming that a trial is composed of N channels, the ESN is fed sequentially with these N signals one by one. As a matter of fact, the ESN output is a channel label. Each successive N channels labels are combined through majority voting and the class label receiving the most number of votes is regarded as the trial label.

3.3 Plasticity Rules for Unsupervised Learning of Reservoir Layer

The generation of fixed random reservoir makes the RNN training fast. Meanwhile, other studies in neuroscience have reported that the modification of the connection strength endows neural networks with a powerful learning ability. This mechanism is called Synaptic plasticity illustrated in Fig. 2a. Here we detail the most used existing ones, to know, the Oja's rule and the BCM rule. Recent findings showed that the neuron is able to change its intrinsic excitability to fit the distribution of the input as shown in Fig. 2b. We also explain the gaussian intrinsic plasticity mechanism.

3.3.1 Oja's rule

Oja learning rule, proposed by Erkki Oja [42], is a model of how neurons in the brain or in artificial neural networks alter the strength of connections, or learn, over time. Oja's rule is an extension of hebbian learning rule [43] proposed in Hebb book "The Organization of Behaviour". In classical hebbian learning rule the update scheme for weights may result in very large weights when the number of iterations is large.

Oja's rule is based on normalized weights, the weights are normally normalized to unit length. This simple change results in a totally different but more general and stable weight update scheme compared to classical hebbian learning scheme. It can also be extended to non-linear neurons as well. It has been mathematically proven that when the input data is centered at the origin and when Oja's update rule converges it results in the neuron learning the first principal component of the training data, hence Oja's rule is important in feature based recognition systems [42].

The Oja learning rule can be described as follows in (8):

$$\Delta W_{kj}(t) = \xi y_k(t)\left[x_j(t) - y_k(t)W_{kj}(t)\right] \tag{8}$$

where W_{kj} is the change of the synaptic weight between the postsynaptic neuron, y_k, and the presynaptic neuron, x_j, at time, t. ξ is the learning rate. Note that Eq. (8) is a modified version of the anti-Hebbian rule by adding a forgetting factor to limit the growth of the synaptic weight to avoid the saturation of W_{kj} .

3.3.2 BCM rule

The BCM rule [44], named for Bienenstock, Cooper and Munro, follows the Hebbian learning principle, with a sliding threshold as a stabilizer function to control the synaptic alteration.

The main idea was that the sign of weight modification should be based on whether the postsynaptic response is above or below a threshold. Responses above the threshold should lead to strengthening of the active synapses, responses below the threshold lead to weakening of the active

synapses. The threshold varies as a nonlinear function of the average output of the postsynaptic neuron, which is the main concept of the present BCM model providing stability properties.

The BCM rule has several variants and here the one suggested in [44] is adopted as given by (9) and (10):

$$\Delta W_{kj}(t) = y_k(y_k - \theta_M)x_j/\theta_M \tag{9}$$

$$\theta_M = E\left[y_k^2\right] = \sum p_k y_k^2 \tag{10}$$

where θ_M is the modification threshold of the postsynaptic neuron, y_k, and p_k is the probability of choosing vector of from the dataset, $E[\,]$ is the temporal average, W_{kj} is the adjustment of the synaptic weight between postsynaptic neuron, y_k and presynaptic neuron, x_j, at time t.

3.3.3 Intrinsic plasticity rule

From a biological point of view, Triesch [45] stated that the biological neuron does not adapt its synapses, rather it adapts its intrinsic excitability. While traditional learning algorithms update the weights of connections between neurons, the IP rule algorithm update the activation function of the neuron. Particularly, Triesch derived the IP rule for fermi activation function and for an exponential desired distribution. In the same manner, Schrauwen et al. [46] extended IP considering also hyperbolic tangent as activation function with Gaussian desired distribution.

IP rule is local, that means it is applied on each single neuron to maximize information about its input. From information theory, entropy measure allows us to realize the maximization of information. Equation (11) measures the distance between the actual probability density of the neurons output and the targeted probability density using the Kullback-Leiber divergence metric.

$$D_{KL}\left(p(x), p_d(x)\right)) = \int p(x) log\left(\frac{p(x))}{p_d(x)}\right) \tag{11}$$

The Kullback-Leiber divergence can be developed into (12) for a Gaussian distribution with a mean μ and a standard deviation σ.

$$D_{KL}\left(p(x), p_d(x)\right)) = -H(x) + \frac{1}{2\sigma^2}E\left((x-\mu)^2\right) + log\frac{1}{\sigma\sqrt{2\pi}} \tag{12}$$

A balance is achieved between the maximization of the actual entropy H and the minimization of the expected entropy E. The update of the gain a and bias b is handled using (13) and (14).

$$\Delta a = \frac{\eta}{a} + \Delta b\left(W^{in}u + W^{res}x\right) \tag{13}$$

$$\Delta b = -\eta\left(-\frac{\mu}{\sigma^2} + \frac{x}{\sigma^2}\left(2\sigma^2 + 1 - x^2 + \mu x\right)\right) \tag{14}$$

Each neuron in the reservoir layer will be activated using (15).

$$x(t) = f_{res}\left(diag(a)\left(W^{in}u(t) + W^{res}x(t-1)\right) + b\right) \tag{15}$$

Schrauwen et al. [46] showed that updating IP parameters in a gaussian distribution with hyperbolic tangent

TABLE 3
Wavelet Decomposition OF EEG channel signal

Bandwith (Hz)	Frequency Band	Decomposition level
64-128 Hz	Noise	D1
32-64 Hz	Gamma	D2
16-32 Hz	Beta	D3
8-16 Hz	Alpha	D4
4-8 Hz	Theta	D5
1-4 Hz	Delta	A5

activation function is similar to the update in an exponential distribution with fermi activation function. Hence, there is no dependence between the chosen non-linearity function and the targeted distribution.

A very interesting unsupervised rule, the IP, is able to make reservoir computing more robust in a fashion that its internal dynamics can autonomously tune themselves independently of the randomly generated weights or the scaling of input to the optimal dynamic regime for emotion recognition. For reservoir enhancing purpose, we follow previous works in which they used the IP rule [45], [46] and [47], [48].

4 EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present and discuss our results using different configurations. To validate our method, we tested it on DEAP benchmark in order to be able to compare our results to the current state-of-the-art methods. Details of experimental settings are provided. An interpretation of emotion recognition results is done. Finally, a sensitivity analysis of ESN to the hyperparameters is detailed.

4.1 Experimental Settings

For the feature extraction step, we chose daubechies function db5 with 5 levels for wavelet decomposition. We choose 5 levels because the sampling rate of recorded EEG signals in DEAP dataset is 128Hz. Besides, a filter pass band [4-45 Hz] was done when preprocessing DEAP dataset, therefore the delta band was not considered in our work. Power features are generated from 4 bands using the 32 channels resulting in a feature vector with 128 as size. Table 3 shows details about frequency bands and corresponding wavelet levels. To train and test our model, we used 80% as a training partition and the remaining 20% as a test partition.

4.2 Emotion Recognition Results

In this subsection, we present the results of 4 classification problems using DEAP dataset.

For arousal classification, using ESN with offline training which is the linear regression yields better results than the online training using the delta rule. An exception is made for the case of the reservoir pretrained with the Oja's rule using the online training of output weights as depicted in Table 4. The comparison is handled with the work in [30] since it belongs to the same context as the current work.

We highlight that using feature as input achieves higher results while using BCM rule. But, the best accuracy

TABLE 4
Arousal Discrimination Results

System	Input type	Offline	Online	Hybrid
ESN-Oja rule	Feature	59.77%	49.22%	60.01%
	Signal	54.30%	58.98%	**60.29%**
ESN-BCM rule	Feature	61.72%	54.14%	**62.17%**
	Signal	56.25%	50.00%	59.34%
ESN-IP rule	Feature	61.21%	59.11%	62.39%
	Signal	68.28% [27]	62.98%	**69.23%**
SVM with PSD features [31]	Feature	63.40%		
HMM [30]	Signal	55.00±4.5%		

TABLE 5
Valence Discrimination Results

System	Input type	Offline	Online	Hybrid
ESN-Oja rule	Feature	59.77%	52.73%	60.81%
	Signal	61.26%	54.92%	**62.13%**
ESN-BCM rule	Feature	57.42%	46.88%	58.26%
	Signal	56.25%	41.67%	**59.31%**
ESN-IP rule	Feature	53.52%	55.86%	57.94%
	Signal	71.03% [27]	66.23%	**71.25%**
SVM with bandpower features [31]	Feature	62.30%		
HMM [30]	Signal	58.75±3.8%		

TABLE 6
Emotional states Discrimination Results

System	Input type	Offline	Online	Hybrid
ESN-Oja rule	Feature	35.49%	42.23%	48.29%
	Signal	54.29%	58.12	**59.29%**
ESN-BCM rule	Feature	32.42%	33.98%	41.58%
	Signal	56.69%	59.81%	**60.23%**
ESN-IP rule	Feature	38.22%	44.65%	49.58%
	Signal	68.79% [27]	69.25%	**69.95%**
SVM with FD features [22]	Feature	69.53%		

TABLE 7
Stress/Calm Discrimination Results

System	Input type	Offline	Online	Hybrid
ESN-Oja rule	Feature	65.45%	47.27%	60.36%
	Signal	61.82%	65.45%	**67.27%**
ESN-BCM rule	Feature	**65.45%**	50.91%	64.63%
	Signal	49.09%	54.55%	58.29%
ESN-IP rule	Feature	69.06%	65.45%	**76.15%**
	Signal	41.82%	49.09%	61.27%
SVM with entropy features [20]	Feature	81.31%		

achieved is 69.23% with ESN pretrained with IP using the hybrid mode inputted directly with EEG channel signal.

Note that, training the output weights in a first step using linear regression and using the delta rule as a second step represents the hybrid mode. The idea here is that instead using the online mode with random initial weights, we trained them first with linear regression. As a result, the performance increased up to 14% over the state of the art method [30].

For valence classification, the reservoir pretrained with IP using the hybrid training reaches the highest result up to 71.25%. Our system, again, outperforms the existing work using HMM with signal as input [30] up to 13% as shown in Table 5. Furthermore, Oja's rule achieves higher result with signal as input 62.13% than the feature 60.81%. While, the BCM rule achieves higher performance 59.31% with signal than the feature input 58.26%.

The discrimination of 8 emotional states is considered as the most challenging task in our work since the complexity of the system is increased. The online mode is more efficient here than the offline mode for both signal and feature input. The hybridization allows us to achieve the best result which is 69.95%. In the literature, most of the existing works classify arousal and valence levels. There is only one work which classifies 8 emotional states [22]. Our ESN trained with linear regression followed with the delta rule outperforms the SVM classifier with FD features.

According to Table 6, the ESN is more robust when using signal instead of feature. When discriminating stress and calm, we remark that ESN with synaptic plasticity achieves the same accuracy with signal as well feature input which is 65.45%. Combining linear regression and delta rule training enhanced the performance from 61.82% to 67.27%. While using ESN with IP, we can rise the accuracy to 76.15% with bandpower features. When inputting signal to ESN with IP, we obtained an accuracy of 41.82% and 49.09% with the offline mode and online mode, respectively. The hybrid mode shows its effectiveness in this case to reach 61.27% as recognition rate.

We believe that existing work combining SVM classifier with entropy features achieves the best result up to 81.31%, but we highlight that we proved that feeding ESN with EEG signal directly can achieve better result than bandpower features which is the aim of our work as depicted in Table 7.

Overall, the online mode is better than the offline mode for the classification of the emotional states. But, for the other problems the offline outperforms the online mode. We can conclude that the delta rule is more efficient when the complexity of the architecture is high than the linear regression. Pretraining the reservoir with Synaptic plasticity rules achieves higher results using the feature vector as input than the raw EEG signal. The only case the synaptic plasticity reaches good accuracy rate with signal is the classification of stress / calm states using Oja's rule.

In all cases, pretraining ESN using intrinsic plasticity rule has achieved the best results either with using feature as input or raw EEG signal. As a consequence, we recommend the use of IP rule and the hybrid mode for classification of EEG signals.
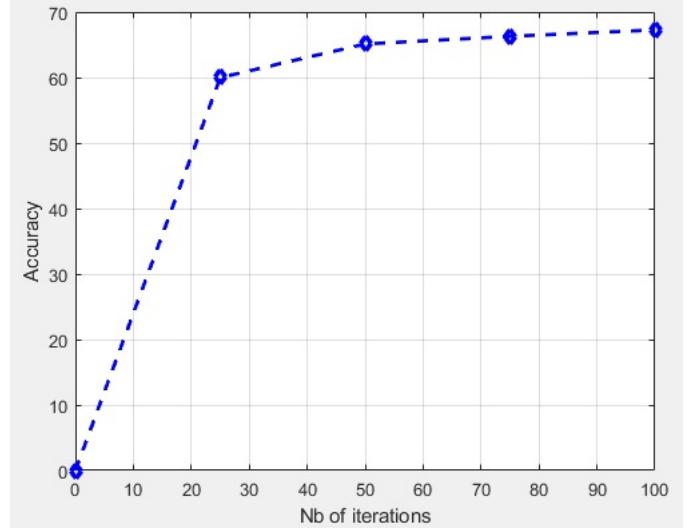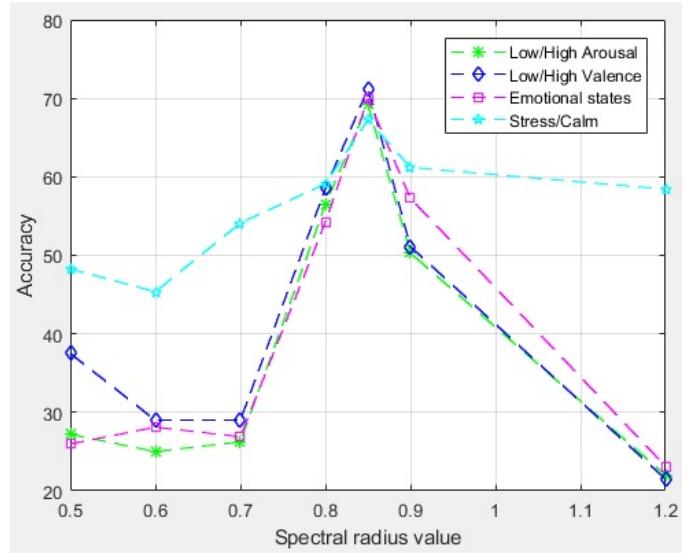

Fig. 3. Impact of Oja's parameter on the performance.

Fig. 4. Impact of spectral radius on the performance.

4.3 Impact of Plasticity Parameters

Through the aforementioned results, we deduced that in three cases the IP rule achieves better results that SP rules, i. e. arousal, valence and 8 emotional states discrimination. But, for the discrimination of stress and calm states the Oja's rule achieves the highest result.

In our study, we delve into the plasticity parameter which is the number of iterations. Our results showed that the previous findings are the same for our case for the IP rule. After 10 iterations, the IP is stable and adding more iterations does not improve the result. Nevertheless, performing more iterations up to 100 of Oja's rule increases the recognition rate as shown in Fig.3.

4.4 Impact of spectral radius and reservoir size

One of the most important parameters in the ESN model is the spectral radius. We recall that Jaeger [38] recommended
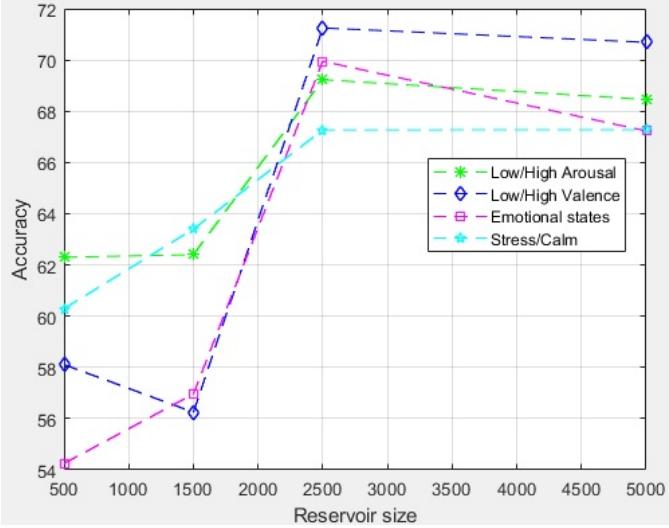
Fig. 5. Impact of reservoir size on the performance.

synaptic plasticity for EEG signals classification. We suggest that synergestic learning of the reservoir could yield ESN model with benefits from each paradigm. We are expecting that these suggestions will result in a more robust reservoir computing approach for EEG-based emotion recognition.

that the reservoir should satisfy the ESP. It relates asymptotic properties of the excited reservoir dynamics to the driving signal.

Intuitively, the ESP states that the reservoir will asymptotically wash out any information from initial conditions. The ESP is granted for any input if this spectral radius is smaller than unity. In Fig. 4, it is shown that the best value of spectral radius is 0.85 for our classification problems. In order to evaluate the sensitivity of our proposed system, we also varied the reservoir parameter from 500 to 5000 neurons as depicted in 5. The choice was not arbitrary, rather we tried to take into consideration the channel data length which is 8064. The role of the reservoir here is to compress the input data and to provide a new representation for classification purpose.

The smaller size 500 does not result in a good representation of channel data leading to low accuracy results. For a medium size 1500, the performance is acceptable for both emotional states and stress/calm discrimination. But, for arousal and valence discrimination the performance is very low. Using 2500 as a reservoir size enhances the performance for classification problems. That means the reservoir is able to make a new representation of the raw EEG signal making the new feature vector more discriminative. When increasing the reservoir size to 5000 neurons, the performance is slightly lower. This is mainly due to the increased complexity of the system. It can also include that the reservoir is stable with 2500 neurons.

5 CONCLUSION

This paper presented the prominent computational models of plasticity and their current applicability to empirical improvements as neural network adaptation mechanisms. The presented forms of plasticity are applied to randomly connected recurrent reservoirs to learn the structural information in EEG signals, achieve sparsity in neural connectivity and enhance learning performance of emotion recognition. We advocate the use of intrinsic plasticity for reservoir pretraining since it achieves best results in comparison with

REFERENCES

[1] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, 2017.
[2] J. Tao and T. Tan, "Affective computing: A review," in *International Conference on Affective computing and intelligent interaction*. Springer, 2005, pp. 981–995.
[3] R. Picard, *Affective Computing*, ser. Inteligencia artificial. The MIT Press, 1997. [Online]. Available: https://books.google.es/books?id=N1qqQgAACAAJ
[4] P. Ekman, "Basic emotions in handbook of cognition and emotions (t. dalgleish and m. power, eds.)," *John Wiley& Sons Ltd*, 1999.
[5] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
[6] A. Mehrabian, "Framework for a comprehensive description and measurement of emotional states." *Genetic, social, and general psychology monographs*, 1995.
[7] ——, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.
[8] K. Sun, J. Yu, Y. Huang, and X. Hu, "An improved valence-arousal emotion space for video affective content representation and recognition," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*. IEEE, 2009, pp. 566–569.
[9] D. Mellouli, T. M. Hamdani, and A. M. Alimi, "Deep neural network with rbf and sparse auto-encoders for numeral recognition," in *Intelligent Systems Design and Applications (ISDA), 2015 15th International Conference on*. IEEE, 2015, pp. 468–472.
[10] N. Chouikhi, B. Ammar, N. Rokbani, A. M. Alimi, and A. Abraham, "A hybrid approach based on particle swarm optimization for echo state network initialization," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2896–2901.
[11] N. Chouikhi, R. Fdhila, B. Ammar, N. Rokbani, and A. M. Alimi, "Single-and multi-objective particle swarm optimization of reservoir structure in echo state network," in *Neural Networks (IJCNN), 2016 International Joint Conference on*. IEEE, 2016, pp. 440–447.
[12] N. Chouikhi, B. Ammar, N. Rokbani, and A. M. Alimi, "Pso-based analysis of echo state network parameters for time series forecasting," *Applied Soft Computing*, vol. 55, pp. 211–225, 2017.
[13] N. Slama, W. Elloumi, and A. M. Alimi, "Distributed recurrent neural network learning via metropolis-weights consensus," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 108–119.
[14] B. Ammar, N. Chouikhi, A. M. Alimi, F. Chérif, N. Rezzoug, and P. Gorce, "Learning to walk using a recurrent neural network with time delay," in *International Conference on Artificial Neural Networks*. Springer, 2013, pp. 511–518.
[15] 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.
[16] 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.
[17] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[18] 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.

[19] 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*, no. 1, pp. 1–1, 2017.

[20] B. García-Martínez, A. Martínez-Rodrigo, R. Zangróniz, J. M. Pastor, and R. Alcaraz, "Symbolic analysis of brain dynamics detects negative stress," *Entropy*, vol. 19, no. 5, p. 196, 2017.

[21] 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*, 2017.

[22] Y. Liu and O. Sourina, "Eeg databases for emotion recognition," in *Cyberworlds (CW), 2013 International Conference on*. IEEE, 2013, pp. 302–309.

[23] P. Pandey and K. Seeja, "Emotional state recognition with eeg signals using subject independent approach," in *Data Science and Big Data Analytics*. Springer, 2019, pp. 117–124.

[24] L. Piho and T. Tjahjadi, "A mutual information based adaptive windowing of informative eeg for emotion recognition," *IEEE Transactions on Affective Computing*, 2018.

[25] M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel eeg signals using k-nearest neighbor classification," *Technology and Health Care*, no. Preprint, pp. 1–11, 2018.

[26] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, "Emotion recognition from eeg signals using multidimensional information in emd domain," *BioMed research international*, vol. 2017, 2017.

[27] 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.

[28] 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.

[29] X. Li, P. Zhang, D. Song, G. Yu, Y. Hou, and B. Hu, "Eeg based emotion identification using unsupervised deep feature learning," 2015.

[30] C. A. Torres-Valencia, H. F. Garcia-Arias, M. A. A. Lopez, and A. A. Orozco-Gutiérrez, "Comparative analysis of physiological signals and electroencephalogram (eeg) for multimodal emotion recognition using generative models," in *Image, Signal Processing and Artificial Vision (STSIVA), 2014 XIX Symposium on*. IEEE, 2014, pp. 1–5.

[31] I. Wichakam and P. Vateekul, "An evaluation of feature extraction in eeg-based emotion prediction with support vector machines," in *Computer science and software engineering (JCSSE), 2014 11th international joint conference on*. IEEE, 2014, pp. 106–110.

[32] L. Bozhkov, P. Koprinkova-Hristova, and P. Georgieva, "Learning to decode human emotions with echo state networks," *Neural Networks*, vol. 78, pp. 112–119, 2016.

[33] ——, "Reservoir computing for emotion valence discrimination from eeg signals," *Neurocomputing*, vol. 231, pp. 28–40, 2017.

[34] F. Ren, Y. Dong, and W. Wang, "Emotion recognition based on physiological signals using brain asymmetry index and echo state network," *Neural Computing and Applications*, pp. 1–11, 2018.

[35] 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.

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

[37] P. S. Addison, *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press, 2017.

[38] H. Jaeger, *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach*. GMD-Forschungszentrum Informationstechnik Bonn, 2002, vol. 5.

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

[40] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[41] 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.

[42] E. Oja, "Simplified neuron model as a principal component analyzer," *Journal of mathematical biology*, vol. 15, no. 3, pp. 267–273, 1982.

[43] D. Hebb, J. Martinez, and S. Glickman, "The organization of behavior-a neuropsychological theory-hebb, do," 1994.

[44] 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.

[45] J. Triesch, "A gradient rule for the plasticity of a neurons intrinsic excitability," in *International Conference on Artificial Neural Networks*. Springer, 2005, pp. 65–70.

[46] 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.

[47] J. J. Steil, "Online reservoir adaptation by intrinsic plasticity for backpropagation–decorrelation and echo state learning," *Neural Networks*, vol. 20, no. 3, pp. 353–364, 2007.

[48] M. Wardermann and J. J. Steil, "Intrinsic plasticity for reservoir learning algorithms." in *ESANN*, 2007, pp. 513–518.

Rahma Fourati (IEEE Graduated Student Member 15). She was born in Sfax, she is a PhD student in Computer Systems Engineering at the National Engineering School of Sfax (ENIS), since October 2015. She received the M.S. degree in 2011 from the Faculty of Economic Sciences and Management of Sfax (FSEGS). She is currently a member of the REsearch Group in Intelligent Machines (REGIM).Her research interests include Recurrent Neural Networks, Affective computing, EEG signals analysis, Support Vector Machines, Simulink Modeling.

Boudour Ammar (IEEE Student Member08, Member13, Senior Member'17) was born in Sfax, Tunisia. She graduated in computer science in 2005. She received the Master degree in automatic and industrial computing from the National School of Engineers of Sfax, University of Sfax in 2006. She obtained a PhD degree in recurrent neural network learning model for a biped walking simulator with the Research Group on Intelligent Machines (REGIM), University of Sfax, since February 2014. She is currently an assistant professor with the Department of Computer Engineering and Applied Mathematics at National Engineering School of Sfax (ENIS). Her research interests include iBrain (Artificial neural networks, Machine learning, Recurrent neural Network) and i-health (Autonomous Robots, Intelligent Control, Embedded Systems, medical applications, EEG, ECG).

Javier Sanchez-Medina (M13) received the M.E. degree from Telecommunications Faculty in 2002 and the Ph.D. degree from the Computer Science Department, Universidad de Las Palmas de Gran Canaria, in 2008. He has authored over 30 international conference and 15 international journal papers. He is interested in the application of evolutionary and parallel computing techniques for intelligent transportation systems. He is also very active as a Volunteer of the IEEE ITS Society, where he has been serving in a number of different positions. He is the Editor-in-Chief of the ITS Podcast, the ITS Newsletter Vice President of the IEEE ITSSs Spanish chapter, and the General Chair for the IEEE ITSC2015.

Adel M. Alimi (IEEE Student Member91, Member96, Senior Member00).He was born in Sfax (Tunisia) in 1966. He graduated in Electrical Engineering 1990, obtained a Ph.D. and then an HDR both in Electrical and Computer Engineering in 1995 and 2000 respectively. He is now a Professor in Electrical and Computer Engineering at the University of Sfax. His research interest includes applications of intelligent methods (neural networks, fuzzy logic, evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. He focuses his research on intelligent pattern recognition, learning, analysis and intelligent control of large scale complex systems. He is an Associate Editor and Member of the editorial board of many international scientific journals (e.g. Pattern Recognition Letters, Neurocomputing, Neural Processing Letters, International Journal of Image and Graphics, Neural Computing and Applications, International Journal of Robotics and Automation, International Journal of Systems Science, etc.). He was a Guest Editor of several special issues of international journals (e.g. Fuzzy Sets and Systems, Soft Computing, Journal of Decision Systems, Integrated Computer Aided Engineering, Systems Analysis Modeling and Simulations). He is an IEEE senior member and member of IAPR, INNS and PRS