Journal of Universal Computer Science, vol. 29, no. 10 (2023), 1116-1138
submitted: 13/12/2022, accepted: 11/5/2023, appeared: 28/10/2023

Enhancing EEG-based emotion recognition using PSD-Grouped Deep Echo State Network

Samar Bouazizi
(Research Groups in Intelligent Machines Lab, BP 3038, Sfax, Tunisia
https://orcid.org/0000-0002-8793-1128, samar.bouazizi@enis.tn)

Emna Benmohamed
(Department of Cyber Security, College of Engineering and Information Technology, Onaizah Colleges, P.O. Box 5371, Onaizah, Kingdom of Saudi Arabia
https://orcid.org/0000-0002-3934-3962, emna.benmohamed@enis.tn)

Hela Ltifi
(Research Groups in Intelligent Machines Lab, BP 3038, Sfax, Tunisia
https://orcid.org/0000-0003-3953-1135, hela.ltifi@ieee.org)

Abstract: Emotions are a crucial aspect of daily life and play a vital role in shaping human interactions. The purpose of this paper is to introduce a novel approach to recognize human emotions through the use of electroencephalogram (EEG) signals. To recognize these signals for emotion prediction, we employ a paradigm of Reservoir Computing (RC), called Echo State Network (ESN). In our analysis, we focus on two specific classes of emotion recognition: H/L Arousal and H/L Valence. We suggest using the Deep ESN model in conjunction with the Welch Power Spectral Density (Wlech PSD) method for emotion classification and feature extraction. Furthermore, we feed the selected features to a grouped ESN for recognizing emotions. Our approach is validated on the well-known DEAP benchmark, which includes the EEG data from 32 participants. The proposed model achieved 89.32% accuracy for H/L Arousal and 91.21% accuracy for H/L Valence on the DEAP dataset. The obtained results demonstrate the effectiveness of our approach, which yields good performance compared to existing models of emotion analysis based on EEG.

Keywords: EEG signals, Recognition, Echo State Networks (ESN), Grouped Deep ESN, Power Spectral Density, Welch method
Categories: I.2., I.5, E.1
DOI: 10.3897/jucs.98789

1 Introduction

Recognizing emotions using EEG signals is a challenging task due to the complexity and richness of the data, and the difficulty of detecting internal emotions from outward manifestations [Bouazizi et al. 2021] [Bouazizi et al. 2023]. In fact, the development of systems that can identify, comprehend, analyse, and replicate human emotions is known as Affective Computing [Poria et al. 2017], which encompasses computer science, psychology, and cognitive science [Khemakhem et al. 2020] . As Picard noted in his book ”Affective Computing” [Picard 1997] in the 1990s, improving human-machine connection requires giving machines the ability to detect, recognize, and understand human emotions [Ellouzi et al. 2015]. Researchers have mostly focused on Emotion Recognition (ER) in their work [Fourati et al. 2022].

The electroencephalogram (EEG) is a tool for directly assessing brain activity. EEG is intriguing because it has exceptional temporal resolution, capturing data every millisecond. Recognizing emotions from outward manifestations can result in inaccurate interferences, particularly when participants have control over their emotions. However, the EEG modality overcomes this problem. The goal of EEG- based emotion recognition is to classify internal human emotions. The idea of applying machine learning techniques is driven by the complexity and richness of EEG data. Machine learning involves using computer system models and logic to explore a particular goal. The primary families of models used in EEG-based applications are nearest neighbour classifiers, neural networks, nonlinear Bayesian classifiers, and linear classifiers.

The motivation behind this study is rooted in the challenge of accurately recognizing emotions using EEG signals. To address this challenge, the usage of Neural Networks (NN) has presented promising results [Fourati et al. 2022]. It allows for the development of sophisticated models that can effectively analyse the complex patterns in EEG data and classify emotions with high accuracy. Therefore, the utilization of EEG and NNs can significantly improve our understanding of emotions and their underlying neural processes. A powerful type of NN for time-series analysis and forecasting is Recurrent Neural Networks (RNN), which includes Echo State Networks (ESNs). They are distinguished by their high amount of recurrence and being three-layered NNs, with the hidden reservoir layer being particularly important. ESNs are ideally suited for several applications, such as time series prediction [Chouikhi et al. 2017], [Slama et al. 2017] or robot control [Ammar et al. 2013], and typically perform well. ESNs are a perfect fit for emotion identification since EEG is a temporal signal.

The machine learning community has become increasingly interested in the potential of deep learning models to cope with the difficulties of modelling complex tasks by representing them in a hierarchical fashion. Deep learning models have the capacity to learn data representations at various levels of abstraction. For this reason, we propose the use of the Deep ESN for EEG-based emotion recognition. However, several research works applied Deep ESN for temporal prediction. Most of the existing models have been improved to achieve efficient emotion classification. The suggested approach employs multiple deep models to attain higher accuracy than the existing methods. We propose increasing the depth of the reservoir network in Deep ESN by using the Grouped Deep ESN architecture to further improve its performance.

Furthermore, it is commonly acknowledged that the process of EEG signal recognition involves several steps, one of which is feature extraction. This indispensable step is designed to reduce the loss of significant signal-embedded information. In this context, we propose integrating the Power Spectral Density (PSD) based on the welch method for feature extraction with the Grouped Deep ESN algorithm for EEG signal recognition. The aim of combining PSD, Welch method, and Grouped Deep ESN is to overcome two challenges in EEG signal processing: (1) extracting meaningful information from the signals through feature extraction, which involves transforming the raw data into a set of features that capture relevant characteristics. To achieve this, we integrate the PSD technique, which measures the power distribution across different frequency bands in the signal, providing information about the relative power of different frequency components and identifying patterns or abnormalities. The Welch method is used to estimate the PSD by dividing the signal into overlapping segments, computing the PSD of each segment using the Fourier transform, and averaging the estimates to obtain a smoother PSD estimate. (2) the second challenge is classifying the EEG signals, which is achieved using the Grouped Deep Echo State Network (ESN). It is commonly used for time-series prediction and classification tasks, capable of handling high-dimensional input data, i.e., EEG signals, and learning complex temporal patterns in that data. This algorithm aims to improve the performance of the ESN by dividing it into groups of neurons and training each group separately, enabling it to learn more diverse and complex representations of the input data.

To validate our proposal, named PSD-Grouped Deep ESN for EEG-based emotion recognition, we tested it on the frequently used “Database for Emotion Analysis using the Physiological Signals (DEAP)” benchmark, which comprises EEG signals of 32 participants. The results indicate better performance than recent existing related works. The outline of this article includes four sections. Section 2 presents our research context concerning the EEG-based emotion recognition. Section 3 provides an explanation of our PSD-Grouped Deep ESN model building. Section 4 details the experimental results and the evaluation. Finally, the article is summarized, and our perspectives are described in Sections 5 and 6.

2 Research context

The theoretical foundations and related research for EEG-based emotion recognition are presented in this section.

2.1 EEG-based emotion recognition: theoretical background

Research is increasingly focusing on emotion recognition. In fact, the main scientific question in affective computing is how to enable computer systems to accurately analyse, understand, and interpret emotional information expressed by humans [Picart 1976]. It is a key research area in both ambient intelligence and artificial intelligence [Dunne et al. 2021]. Discoveries in emotion recognition can further advance the development of many fields, including computer science, engineering, and medical science. Emotion is a complicated psychological state that manifests in bodily movements and physiological changes [Adolphs et Anderson 2018]. The recent years have seen significant progress in the emotion recognition domain, which relies on affective data gathered from a variety of physical actions and physiological processes, including vocalizations from microphones, signals from devices that measure neurophysiological activity, videos from cameras, and texts from websites, etc.

Humans can sometimes control their facial expressions, but it is not easy to control their natural bodily responses, such as variations in body temperature, heart rate, or physiological signals [Rattanyu et al. 2010]. Biosignals, imaging techniques or emotional videos are some of the ways to recognize emotions. Because a biosignal is a physiological signal that is independent of human desire, it can be used to determine human emotions objectively [Wioleta 2013]. Therefore, our work focuses on EEG- based emotion recognition. EEG signals are a promising method for detecting emotional states. They are physiological signals derived from brain biosignals [Tripathi 2011] and are recorded using electrodes implanted in the scalp. These electrodes capture electrical energy potentials indicative of brain activity [Minguillon et al. 2017].

Numerous studies on EEG-based emotion recognition have been introduced in the literature, utilizing well-known open EEG datasets and applying machine learning and signal processing techniques.

2.2 EEG datasets

There are numerous studies on EEG-based emotion detection in the literature. Each study was tested and evaluated using a particular private EEG dataset. As far as we know,

there are five open EEG benchmarks: MAHNOB-HCI [Soleymani et al. 2012], DEAP [Koelstra et al. 2012], SEED [Zheng et Lu 2015], DREAMER [Katsigiannis et Ramzan 2018], and HR-EEG4EMO [Becker et al. 2017]. The most popular dataset is DEAP [Fourati et al. 2022]. Therefore, we used the DEAP affective benchmark to validate our proposed deep ESN for EEG-based emotion recognition.
Forty emotional videos were utilized in the DEAP dataset's elicitation methodology. Thirty-two participants, 16 men and 16 women, were involved in the study. They were outfitted with Biosemi Active 2 acquisition systems, which have 32 EEG channels and 8 sensors. Participants' judgments for each trial were stated based on the three emotion components: (1) arousal for emotion intensity incited by a stimulus, (2) valence for the pleasantness of a stimulus, and (3) dominance for the control level exerted by a stimulus.

Several research works have contributed to the EEG-based emotion recognition on the DEAP dataset. Therefore, we briefly discuss these related works in the next section to provide insight into available knowledge.

2.3 EEG-based emotion recognition related works

Research in the field of EEG-based emotion recognition is rich and varied in terms of classifier algorithms and EEG datasets. In our work, we are interested in related works that deal with the DEAP dataset.

Our literature review focused on the most recent contributions of emotion recognition between 2017 and 2022. For the classification of two valence or arousal levels, various works have been proposed (such as [Zhuang et al. 2017], [Piho et Tjahjadi 2018], and [Li et al. 2018]). Better outcomes were achieved by making the task subject-dependent and customizing the classifier for that subject by combining training and test data from the same subject. For instance, [Piho et Tjahjadi 2018] used the K-Nearest Neighbor (KNN) approach to classify statistical features from the DEAP dataset and achieved 82.76% and 82.77% accuracy rates for the two valence and arousal levels, respectively. [Piho et Tjahjadi 2018] underlined the importance of specific channels in affecting the emotion recognition performance. They asserted that just one F4 channel is enough for an EEG-based emotion recognition task. In 2019, [Chen et al. 2019] applied the hierarchical bidirectional GRU (Gated Recurrent Unit) based on the attention mechanism and obtained accuracy rates of 67.9% for valence and 66.5% for arousal. Although using ESN to process EEG data is not novel (such as [Bozhkov et al. 2016] [Bozhkov et al. 2017] and [Ren et al. 2018]), there are not many related works that use it on the DEAP data set. As far as we know, there are just few works: [Fourati et al.2017] and [Fourati et al. 2022]. [Fourati et al. 2017] suggested training ESN with Intrinsic Plasticity before applying it to pre-processed DEAP EEG data. Arousal levels, valence levels, and the eight emotional states were classified with accuracy rates of 68.28%, 71.03%, and 68.79%, respectively. The findings obtained with an ESN pretrained with synaptic plasticity, where the greatest accuracy result achieved was 76.15%, were then outperformed by those obtained with an ESN pretrained with intrinsic Gaussian plasticity, as discovered by [Fourati et al. 2022].

Recently, deep learning has attracted more attention as an intelligent technology to achieve better recognition performance. Deep learning for the EEG-based emotion recognition has presented interesting results in the literature. [Tripathi et al. 2017] and [Al-Nafjan et al. 2017] conducted CNN tests on the DEAP dataset with successful performance outcomes (average of 77,47% and 82% respectively). In 2020, [Kim et Choi 2020] applied LSTM (Long Short-Term Memory) technique combined with an attention mechanism to attribute weights to the affective states occurring at a specific instant. They achieved a best accuracy rate of 90.1%.

Cheng et al. (2020) [Cheng et al. 2020] proposed a deep forest-based approach for emotion recognition from multi-channel electroencephalogram (EEG) signals, reporting an accuracy of 92.89% for arousal recognition and 93.06% for valence recognition. In 2022, [Joshi et Ghongade 2022] applied the Bidirectional LSTM (BiLSTM) with Modified Differential Entropy. The best accuracy they obtained was 75%. [Qu et Zheng 2022] used physiological functions of EEG regions and the active scenario of band signals to select emotion-sensitive signals. These signals were then combined using LSTM and CNN to extract both temporal and spatial features. The merged features were classified to identify the emotion. The researchers evaluated the method on the DEAP dataset, and the average accuracy for recognizing valence and arousal dimensions was found to be 92.87% and 93.23%, respectively. [Hu et al. 2022] combined a Convolutional Neural Network (CNN), a Bidirectional Long and Short-Term Memory network (BiLSTM), and Multi-Head Self-Attention (MHSA). This model enables deep learning of time-series and spatial information of EEG emotion signals. CNN is used to smooth EEG signals and extract deep features, BiLSTM to learn emotion information from past and future time series, and MHSA to enhance recognition accuracy by reassigning weights to emotional features. The authors achieved an accuracy of 85.57%.

Recently, a study in 2023 [Lyer et al. 2023] introduces a hybrid model that combines. CNN and LSTM models for accurate detection. The approach allows extracting features using Differential Entropy (DE). These selected features are fed to CNN, LSTM, and CNN-LSTM models for emotion recognition. At the end, the predictions of the three models are combined by an ensemble model. The researchers validate the proposed approach on two datasets, SEED and DEAP, for EEG-based emotion analysis. The proposed method achieves a high 97.16% accuracy for emotion classification on the SEED dataset, outperforming compared methods for EEG-based emotion analysis.

However, for the DEAP dataset, the proposal achieves 65% accuracy. Upon reviewing the literature, we have found that deep learning has enabled the achievement of accuracies beyond 90%. As we have not come across any studies, to the best of our knowledge, utilizing deep ESN for EEG-based emotion recognition from the DEAP dataset, we propose to apply this method and enhance its performance through the use of the grouped deep ESN architecture.

3 PSD-based Grouped Deep ESN model building

The EEG-based emotion recognition process includes three principal steps: data preprocessing, feature extraction using Welch-PSD, and classification using the grouped deep ESN as illustrated in the following figure:

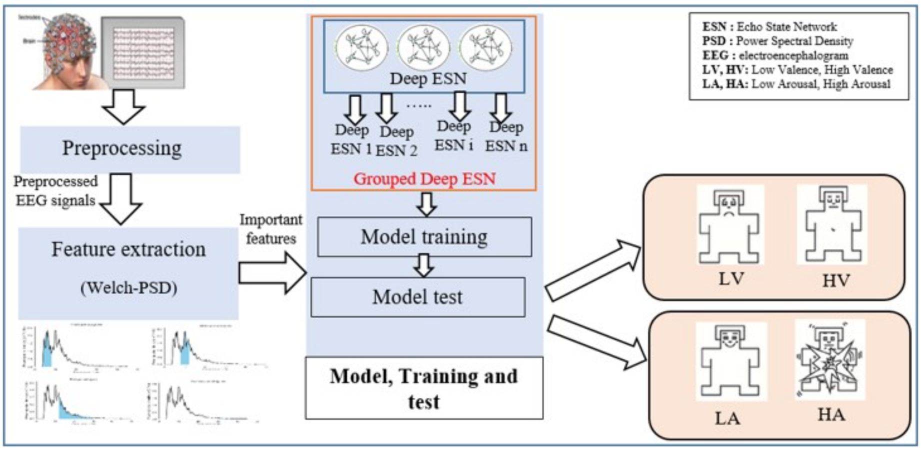

Figure 1: Flowchart of the proposed approach

3.1 Data pre-processing

As presented previously (cf. Section 2.2), we are interested to the DEAP affective dataset, which contains three dimensions, namely Valence, Arousal and Dominance. Consequently, numerous categorization problems were created, including Low/High Valence, Low/High Arousal, and Dominance.

Figure 2 presents the first 40 data rows for one participant in the DEAP dataset.

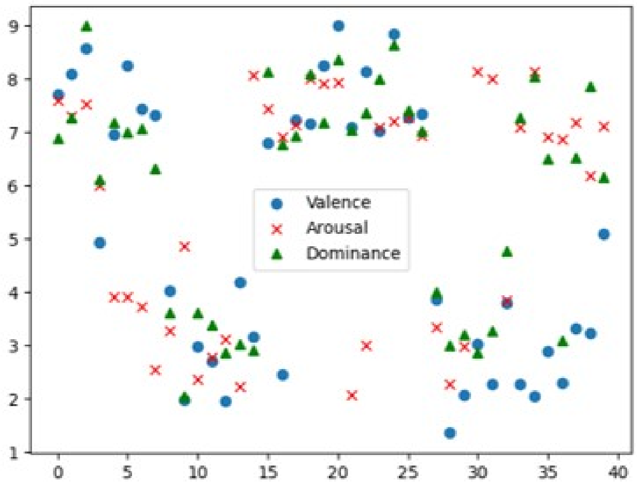

Figure 2: First 40 data rows for one participant

We have separated the EEG and non-EEG data. The DEAP dataset includes 32 EEG channels and 8 peripheral physiological channels. The peripheral signals consist of electrooculogram (EOG), electromyograms (EMG) of Zygomaticus and Trapezius muscles, GSR, respiration amplitude, blood volume by plethysmograph, and skin temperature. Emotional states can be derived from the combinations of valence and arousal, including high arousal and high valence (excited and happy), low arousal and high valence (calm and relaxed), high arousal and low valence (angry and nervous), and low arousal and low valence (Sad, Bored). Table 1 presents the number of trials for each group.

High Valence	571
Low Valence	509
High Arousal	540
Low Arousal	540

Table 1: Trials per group

Table 1 contains the count of trials that have been conducted for each group. Then, we generated the number of trials for the combinations between the four groups as shown in Table 2.

High Arousal High Valence	242
Low Arousal High Valence	215
High Arousal Low Valence	200
Low Arousal Low Valence	223

Table 2: Trials per group

3.2 Welch based PSD for Feature extraction

In this work, the adopted feature extraction method was Power Spectral Density (PSD). PSD is an effective stationary signal processing method that works well with narrowband signals. It is a typical signal processing approach that displays the energy strength as a function of frequency and distributes the signal power over frequency [Ong et Ibrahim 2018]. In this research, PSD was applied along with Welch's technique. To estimate power spectra, Welch's method (also known as the periodogram method) [Welch 1967] divides the time signal into sequential blocks, creates a periodogram for each block, and then averages the results. Equations (1) and (2) mathematically describe the Welch-based PSD. The first equation defines the power spectra density, and the second one calculates the Welch Power Spectrum, i.e., the average of the periodogram for each interval.

$$P(f) = \frac{1}{MU} | \sum_{n=0}^{M-1} x_i(n) e^{-j2\pi f} |^2 \tag{1}$$

where:

- P(f) is the power spectral density at frequency f

- MU is a normalization factor, typically the length of the signal
- xi(n) is the ith sample of the signal at time n
- W(n) is a window function applied to the signal to reduce spectral leakage

$$P_{welch}(f) = \frac{1}{L}\sum_{i=1}^{L-1} P(f) \tag{2}$$

Where L is the length of the signal or window used in the Welch method.

The accuracy of the traditional periodogram is increased by the Welch's approach (cf. Figure 3). The simple logic behind this is that EEG data is constantly time-varying. Therefore, it is highly improbable for the signal to appear as a perfect sum of pure sines when examining 30 seconds of EEG data. Instead, the neural activity occurring underneath the scalp gradually modifies the EEG's spectral composition over time. The issue with the conventional periodogram is that it requires the signal's spectral content to remain stationary (i.e., time-unvarying) during the time period taken into account to provide a genuine spectral estimate. In this research, we used the welch's method to extract theta, alpha, beta and gamma spectral power for each electrode (cf. respectively Fig. 4(a), (b), (c), and (d)). The frequency bands used are theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-64 Hz).

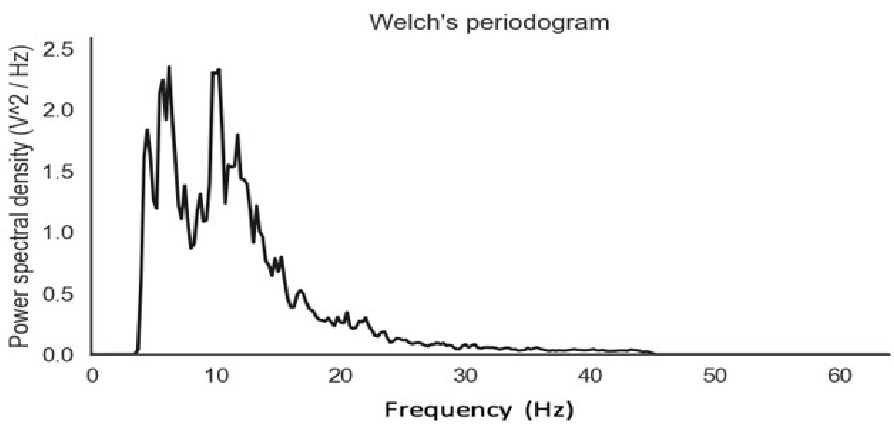

Figure 3: Welch's periodogram

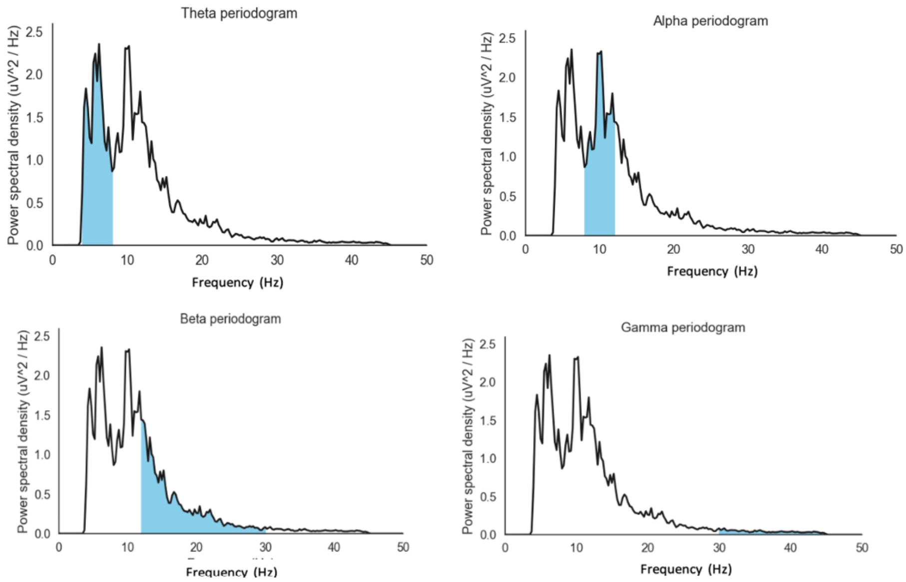

Figure 4: (a) Theta periodogram, (b) Alpha periodogram, (c) Beta periodogram, and (d) Gamma periodogram

The frequency vector is plotted on the x-axis (frequency bins), while the PSD (Power Spectral Density) vector is plotted on the y-axis. Typically, the units of the power spectral density for EEG data are micro-Volts-squared per Hz (uV2/Hz). It is essential to identify the frequency bins that cross the delta, theta, alpha, and beta frequency ranges. The welch periodogram generated is shown in Figures 3 and 4, with the filled area representing theta, alpha, beta, and gamma band power. The blue region in these figures denotes the band power. Since there isn't a closed-form formula to integrate this region, it must be approximated. The composite Simpson's rule is frequently used for this purpose [Weisstein 2003]. This region is divided into multiple parabolas, and the areas of each parabola are added to determine the total area.

Theta	5.434119660168186
Alpha	5.369595513295193
Beta	6.286556266834863
Gamma	0.9879159580139809

Table 3: Value of bands power

Table 3 presents the value of each band power (i.e., Theta, Alpha, Beta and Gamma).

Fig. 5 illustrates the power spectral density of each channel related to the participant's scalp. Figures 6, 7, 8 and 9 show the topographical map of the brain for the alpha, beta, theta, and gamma bands, respectively. The distinct colour lines show how much energy each electrode received from the frequency band wave. The colour bar used to show the topological brain map indicates the highest power intensity in red and the lowest power intensity in blue.

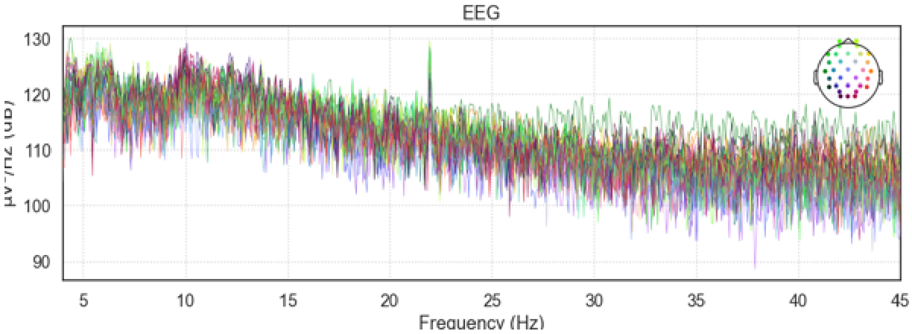

Figure 5: The power spectral density across channels

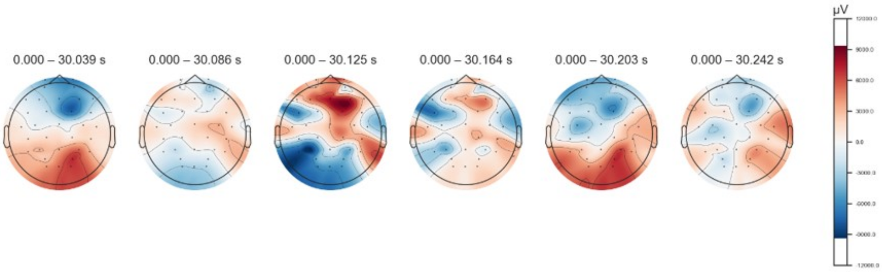

Figure 6: Theta band of frequency topo map

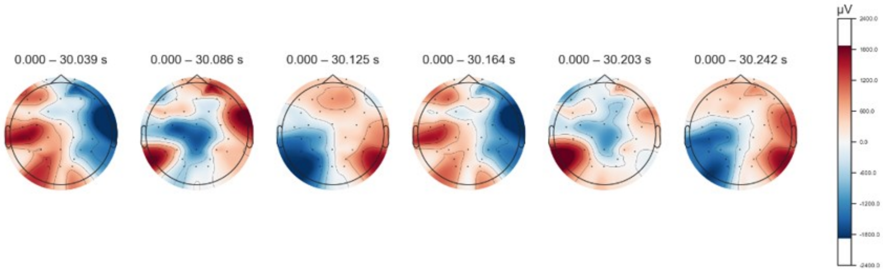

Figure 7: Alpha band of frequency topo map

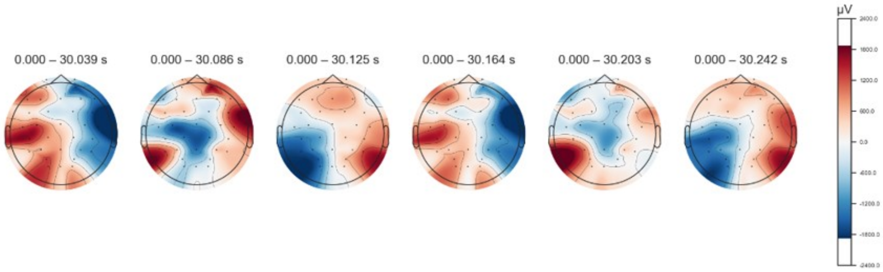

Figure 8: Beta band of frequency topo map

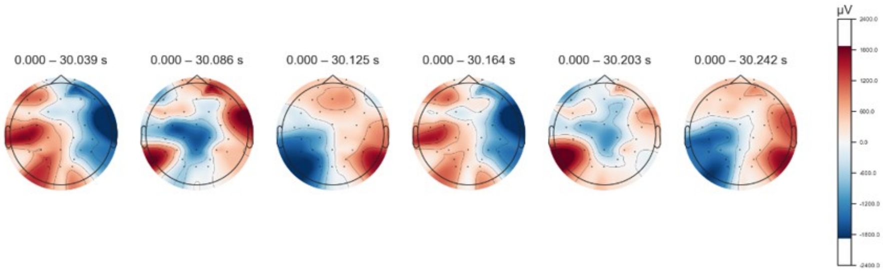

Figure 9: Gamma band of frequency topo map

Jaeger [Jaeger 2001] introduced Echo State Networks (ESNs), a specific class of Recurrent Neural Networks (RNN) that is part of the Reservoir Computing (RC) framework. ESNs consist of an input, a reservoir (hidden) layer, and an output layer. The structure of an ESN is illustrated in Figure 10.

Unlike traditional RNN, the connection weights between the input layer and the reservoir (i.e., Win) and the Reservoir weights (i.e., W) are both not trainable and randomly initialized. However, the connection weights between the reservoir and the output layer (Wout) must be trained throughout the learning process. The output layer and reservoir are connected by a Wback connection as well. The ESN is capable of multi-step predictions when Wback is present. If not, it can only forecast one step forward.

The Win and Ws are initialized arbitrarily during the initial network establishment and are left untouched during the training phase. Once the reservoir's condition and the ESN's output mode have been established, Wout can be calculated based on the desired output to minimize the difference between the predicted model output and the actual one. Learning ESN is a relatively quick linear regression problem [Liu et al., 2020]. There are numerous ways to calculate this straightforward linear regression [Liu et al., 2019].

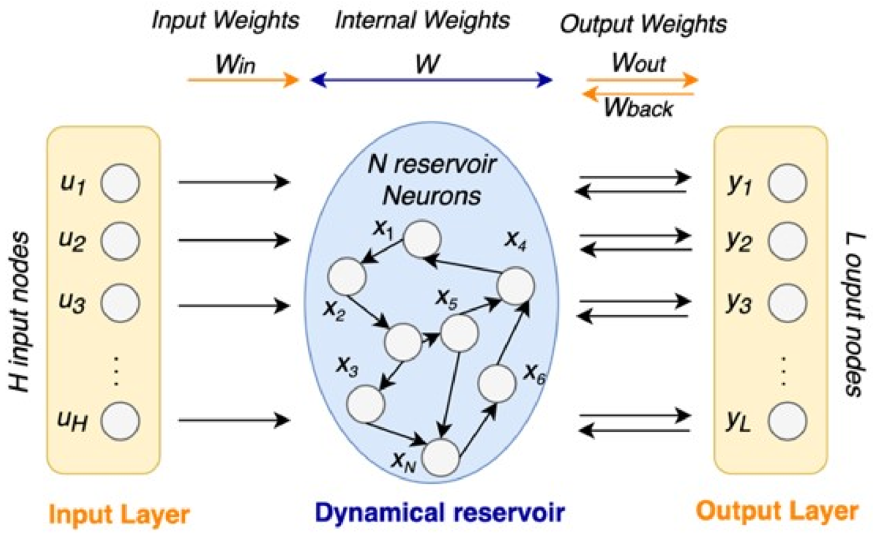

Figure 10: ESN Architecture inspired from [Li et al. 2015]

The equation (3) and (4) shows the ESN sate update:

$$x(i+1) = f(W_in * u(i+1) + W * x(i) + W_back * y(i)) \quad (3)$$

$$y(i+1) = g(W_out * x(i+1) \quad (4)$$

*With **f**: activations function for reservoir neurons*
*and **g**: activations function for output layer neurons. It can be linear or nonlinear.*
Recently, a few research works have examined deep ESNs to improve the capabilities of the ESN. The primary idea behind adding depth to an ESN is to support hierarchical representations of the temporal input while simultaneously capturing the multi-scale dynamics of the input characteristics (cf. Figure 11).

ESNs have been utilized in the literature to perform various temporal tasks [Soures et al. 2017]. To effectively utilize each reservoir's temporal kernel characteristic, [Ma et al. 2017] suggested the Deep-ESN architecture, which includes stacked reservoir layers (i.e., N Sub-Reservoirs for N layers) and unsupervised encoders. In memory capacity experiments, [Gallicchio et al. 2017] found that the Deep ESN network performed best both with and without using intrinsic plasticity.

Unlike a standard deep neural network, the output of all intermediate layers is concatenated for to produce the final result (cf. Figure 11).

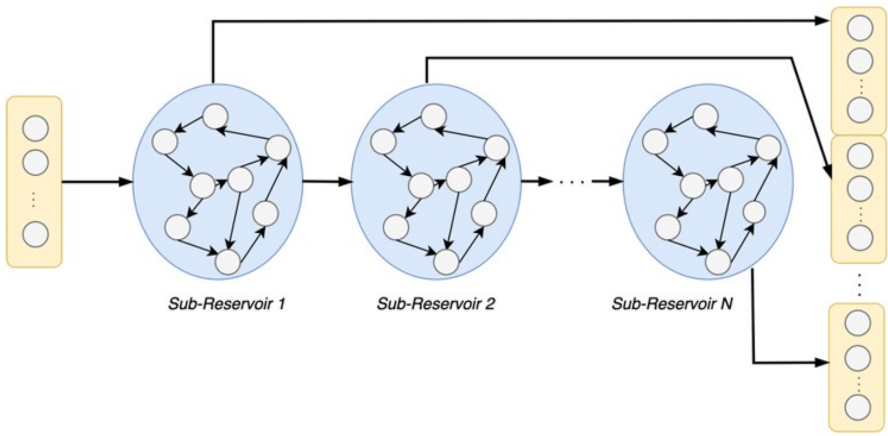

Figure 11: Deep ESN

To extract features with a greater order of complexity and capture multi-scale dynamics of time series data, a deeper reservoir network is necessary [Gallicchio et al. 2017]. In this context, we suggest the Grouped Deep ESN as an efficient solution to recognize EEG emotions (cf. Figure 12). The multi-layered topology of the Grouped deep ESN is shown in Figure 12. The Deep ESN sub-reservoirs' outputs are concatenated to generate the final grouped deep ESN result.

The grouped deep ESN uses a two-dimensional organization of sub-reservoirs for its architecture. Equations (5), (6) and (7) are used to characterize the grouped deep ESN state transitions:

$$x^{(i,j)}(n) = (1 - \alpha^{(i,j)} x^{(i,j)}(n-1) + \alpha(i,j)\bar{x}(i,j)(n) \tag{5}$$

$$\bar{x}(i,j)(n) = f(\mathbf{W}_{in}^{(i,j)}[1; \mathbf{v}^{(i,j}(n)] + \mathbf{W}_{x}^{(i,j)}(n-1)), \tag{6}$$

$$v^{(i,j)}(n) = \begin{cases} u(n)\ if j = 1 \\ x^{i,j-1)}(n)\ if j > 1, \end{cases} \tag{7}$$

Where N_g is the number of groups (1 < i <= N_g), N_1 the number of layers (1 < i <= N_1) and (1 < i <= α (i,j) is the leaking rate for i^{th} group,j^{th} layer.

$$\mathbf{y}(n) = \mathbf{W}_{out}[1; \mathbf{x}^{(1,1)}(n); ...; \mathbf{x}^{(1,N_1)}(n); ...; \mathbf{x}^{(N_g,1)}(n); ...; \mathbf{x}^{(N_g,N_1)}(n)] \tag{8}$$

Where $x^{(i,j)}$(n) are the node activations for all sub-reservoirs. They are concatenated the generate the Wout: $W_{out} \in R^(Ny \times (NxNgNl + 1))$ With Ny output units and Nx input units.

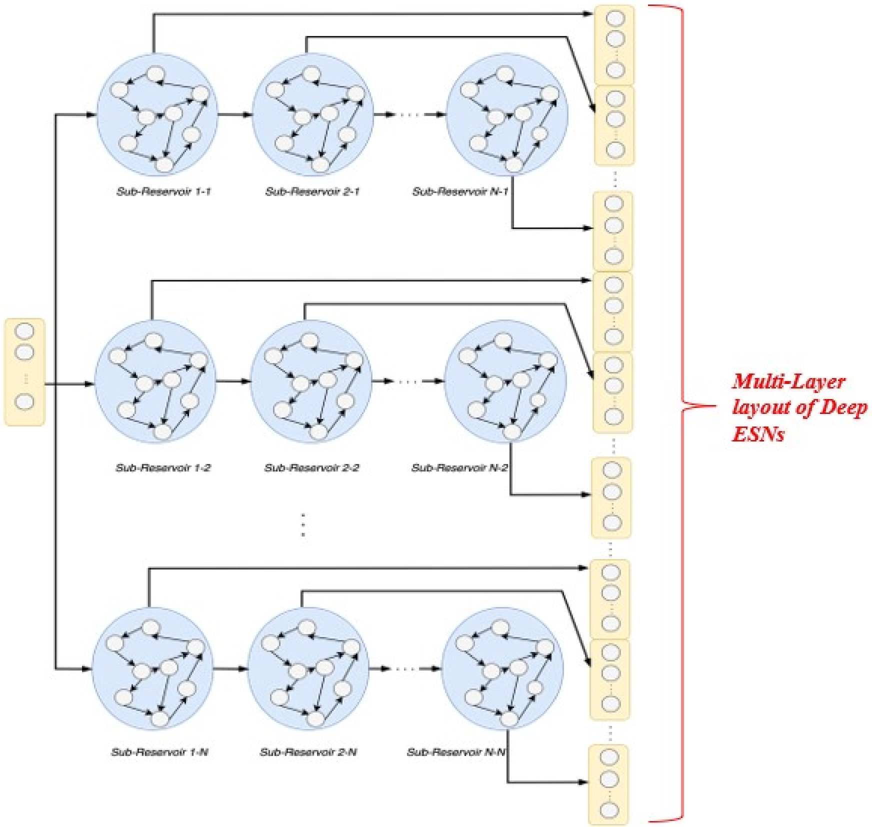

Figure 12: Grouped deep ESN

4 Experimental analysis and comparison

We already described the workflow for PSD-Grouped Deep ESN (cf. section 3). We highlight its effectiveness in this part by describing the experimental outcomes of its application to the DEAP dataset for EEG-based emotion recognition.

4.1 Experimental Setup

In our study, the experimental setup involves splitting the dataset into two parts - 80% for training the ESN, deep ESN and grouped deep ESN algorithms and 20% for testing their performance.

Additionally, a cross-validation for each algorithm was generated. Cross-validation is a technique used to validate the performance of a model. In our context, we generated 10-fold cross-validation. This involves splitting the dataset into 10 subsets, training the model on 9 subsets and testing it on the remaining subset. This process is repeated 10 times, with each subset used as the testing set exactly once. The results are then averaged to obtain an estimate of the model’s performance. Figure 13 presents the generated cross-validation.

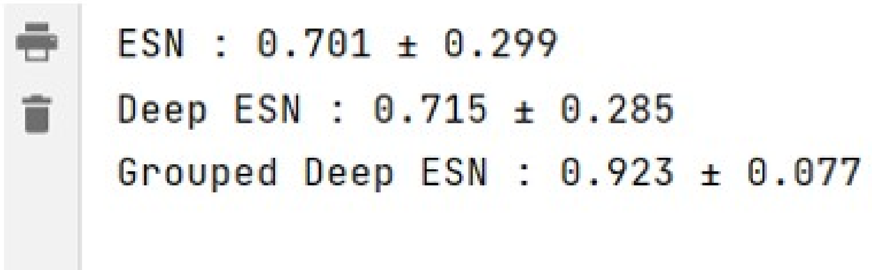

Figure 13: Generated 10-fold cross-validation results

The results provided in Figure 13 show the mean and standard deviation of the performance of the different models on the DEAP dataset using 10-fold cross-validation. The performance is measured using accuracy, which is commonly used in classification tasks. The mean accuracy of the ESN on the dataset is 0.701, and the standard deviation is 0.299. This means that the performance of the ESN varies widely across the 10 folds, with some folds achieving high accuracy and others achieving low accuracy. The mean accuracy of the Deep ESN is slightly higher than that of the ESN at 0.715, and the standard deviation is also similar at 0.285. This suggests that the Deep ESN is performing slightly better than the standard ESN, but the performance is still inconsistent across the folds. As for the Grouped Deep, the mean accuracy on the Grouped Deep ESN on the dataset is significantly higher at 0.923, and the standard deviation is much lower at 0.077. This indicates that the Grouped Deep ESN is performing much better than the other two models and is more consistent in its performance across the 10 folds.

4.2 ESN, deep ESN and grouped deep ESN comparison

A comparative analysis of the results of PSD-ESN, PSD-Deep ESN, PSD-Grouped deep ESN allows us to conclude the validity of the choice of this latter model. Mean Square Error (MSE), Root Mean Square Error (RMSE), and Normalized Root Mean Square Error (NRMSE) are important metrics for assessing the quality of a predictive model. The formulas following for calculating these metrics are given in Equations (9), (10) and (11):

$$MSE = \frac{1}{T}\sum_{i=1}^{T}(Y_i - P_i)^2 \tag{9}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y_i - P_i)^2} \tag{10}$$

$$NRMSE = \frac{1}{T}\sum_{i=1}^{T}\frac{(Y_i - P_i)^2}{T\sigma^2} \tag{11}$$

The target and predicted outputs are denoted by Yi and Pi, respectively. T represent the number of data samples, and σ^2 represents the variance of target values. Three models were used for performance comparisons: PSD-ESN, PSD-deep ESN and PSO-Grouped-deep ESN (cf. Table 4).

Model	**MSE**	**RMSE**	**NRMSE**
PSD-ESN	0.13960 3909616	0.37363606 614265527	0.74727213 202853105
PSD-Deep-ESN	1.997851453 7141146e-08	0.00060571 88466422778	0.00267438 0209248765
PSD-Grouped Deep ESN	1.40976765 53335032e-08	8.8770565 88940362e-05	0.0003919 445570371235

Table 4: MSE, RMSE and NRMSE comparison

By averaging the squared errors from data, the MSE is computed. The MSE measures how inaccurate statistical models are. It evaluates the average squared discrepancy between model predictions and observations. When a model is error-free, the MSE is equal to 0. Its value rises when model errors increase. As visible in Table 4, the smaller MSE (1.4097676553335032e-08) is of the PSD-Grouped deep ESN which proves its superiority in terms of estimation. The Root Mean Squared Error (RMSE) of the PSD-Grouped Deep ESN model is also the lowest among the three models, with a value of 8.877056588940362e-05, which suggests that this model has the smallest difference between predicted and actual values.

The Normalized Root Mean Squared Error (NRMSE) is a metric that measures the error between predicted and actual values, normalized by the range of the target variable. Among the three models, the PSD-Grouped Deep ESN model is the most accurate, with the smallest NRMSE value of only 0.039% of the range of the target variable.
Based on the results presented in Table 4, it can be concluded that the combination of PSD and Grouped Deep ESN is the most effective model for accurately predicting emotions in the DEAP dataset. Another metric frequently used to evaluate classification models is accuracy, which is interpreted as the proportion of correctly predicted values. It summarizes the performance of the model with a single value. Therefore, in Table 5, we present the accuracy of PSD-ESN, PSD-Deep ESN and PSD-Grouped deep ESN.

We note that the accuracy values of PSD-ESN and PSD-Deep ESN are close, indicating that there is no point in migrating to PSD-Deep ESN. However, the accuracy of the PSD-Grouped Deep ESN is significantly higher, indicating its effectiveness. We observed a notable increase in accuracy from 71.58% (for the deep version) to 92.39%, which is an encouraging result. Table 5. Accuracy model comparison

PSD-ESN	70.14%
PSD-Deep-ESN	71.58%
PSD-Grouped Deep ESN	92.39%

Table 5: Accuracy model comparison

The results presented in Tables 4 and 5 highlight the superiority of the grouped deep ESN model which proves its validity for EEG-based emotion recognition.

4.3 Performance measurement

The performance of our proposed PSD-Grouped Deep ESN method has been further verified using the performance measurements shown in Table 6. These metrics are expressed in terms of the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) parameters of the confusion matrix (see Figure 13).

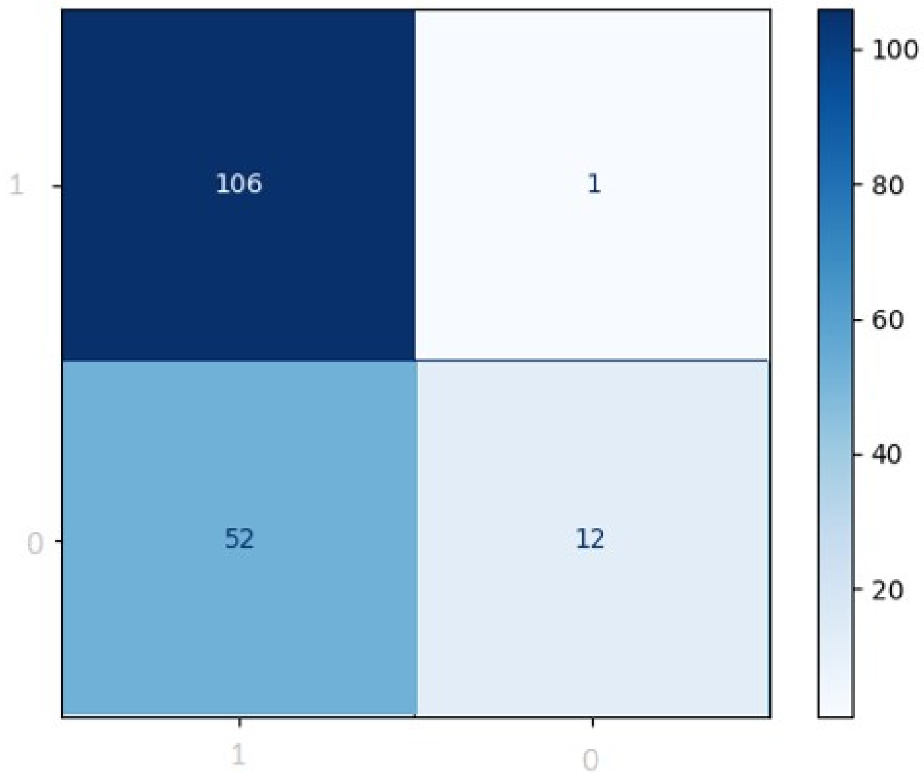

Figure 14: PSD-Grouped deep ESN confusion matrix

Metric	Description	Equation	Value
Accuracy (Acc)	Measures the model performance	$Acc = \frac{(TP + TN)}{(TP + TN + FN + FP)} * 100$	92.39%
Positive predictive value (PPV)	Measures the percent of predicted positives that are actually positive	$PPV = \frac{(TP)}{(TP + FP)} * 100$	99.06%
Negative predictive value (NPV)	Measures the percent of negative positives that are actually negative	$NPV = \frac{(TN)}{(TN + FN)} * 100$	81.25%
Sensitivity (Sen)	Measures the probability of a positive test, conditioned on truly being positive.	$Sen = \frac{(TP)}{(TP + FN)} * 100$	89.83%
Specificity (Spe)	Measures the probability of a negative test, conditioned on truly being negative	$Spe = \frac{(TN)}{(TN + FP)} * 100$	92.39%
F1-score	Measures the harmonic mean	$F1 - Score = 2 * \frac{(Sen \times PPV)}{(Sen + PPV)}$	94.21%
Matthews's correlation coefficient (MCC)	Measure the difference between the predicted values and actual values	$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	92.39%

Table 6: Performance metrics

The calculated performance results of our PSD-Grouped Deep ESN are interesting, with an accuracy of 92.39%, a PPV of 99.06% and NPV 81.25%. These values prove the significant prediction ability of our model. The sensitivity, specificity and F1-score rates (89.93%, 98.11% and 94.21%, respectively) inform about the preformance of the classification ability of the model. They can be considered encouraging results. In contrast, the MCC is a more trustworthy statistical measure that generates a high score only if the prediction did well in each of the four confusion matrix classes (TP, FN, TN, and FP) according to the amount of the dataset's positive and negative elements. In our case, the MCC rate is 84%, which is considered a high score, proving again the performance of the PSD-Grouped Deep ESN.

4.4 Comparison with related works

To judge the quality and validity of our results in the literature, we compare them to existing scientific conclusions (cf. Table 7), i.e., to related works that have already been presented in the background of this paper (cf. section 2.3).

Related work	Feature extraction	Classifier	Emotion	Accuracy
[Fourati et al. 2017]	Raw EEG	ESN	H/L Arousal H/L Valence 8 emotions	68.28% 71.03% 68.79%
[Zhuang et al. 2017]	Intrinsic Mode Functions	SVM	H/L Arousal H/L Valence	72.10% 70.41%
[Tripathi et al. 2017]	CNN	CNN	H/L Arousal H/L Valence	81.78% 73.1%
[Al-Nafjan et al. 2017]	DNN	DNN	H/L Arousal H/L Valence	82.00% (average)
[Piho et Tjahjadi 2018]	Statistical features	KNN	H/L Arousal H/L Valence	82.77% 82.76%
[Chen et al. 2019]	Attention mechanism	Hierarchical bidirectional GRU	H/L Arousal H/L Valence	66.5% 67.9%
[Kim et Choi 2020]	Attention mechanism	LSTM	H/L Arousal H/L Valence	87.9% 90.1%
[Chen et al 2020]	-	Deep forest	H/L Arousal H/L Valence	92.89% 93.06%
[Joshi et Ghongade 2022]	Modified Differential Entropy	BiLSTM	H/L Arousal H/L Valence	75% 73.5%
[Fourati et al. 2022]	Discrete Wavelet Transform	ESN-Intrinsic Plasticity	H/L Arousal H/L Valence 8 emotions	69.23% 71.25% 69.95%
[Qu et Zheng 2022]	LSTM CNN to extract temporal and spatial features	FC (SoftMax)	H/L Arousal H/L Valence	93.23% 92.87%
[Hu et al. 2022]	CNN BiLSTM MHSA	CNN BiLSTM MHSA	H/L Arousal H/L Valence	85.57%
Lyer et al. [2023]	Differential Entropy (DE)	Hybrid CNN-LSTM	H/L Arousal H/L Valence	65%
Our proposal	**Welch-PSD**	**K2 based ESN**	H/L Arousal H/L Valence	**93.59**%

Table 7: Precision, recall and F1-score.

Our proposed method for EEG-based emotion recognition on the DEAP dataset combines the Welch Power Spectral Density (PSD) method with Grouped deep echo state networks (ESNs), resulting in superior feature extraction and classification capabilities.

To validate its effectiveness, we compared its accuracy with that of previous works, all of which reported accuracy for the Arousal (High/Low) and Valence (High/Low) dimensions. Our method achieved an accuracy of 89.32% for H/L Arousal and 91.21% for H/L Valence, which is notably better than that of most previous studies. Although two works ([Chen et al. 2020] and [Qu et Zheng 2022]) achieved higher accuracy, our proposed approach still demonstrates significant improvement over related works.

The Welch PSD method improves our feature extraction capabilities by extracting accurate frequency information from EEG signals through noise reduction and enhanced spectral resolution. This is a significant improvement over previous studies. Additionally, our Grouped deep ESN technique leverages the hierarchical architecture of reservoirs to capture more complex temporal patterns, resulting in enhanced classification performance. Together, these techniques enable our proposed method to achieve state-of-the-art accuracy in emotion recognition on the DEAP dataset, surpassing the performance of most previous works. Therefore, our study demonstrates the effectiveness of combining Welch PSD and Grouped deep ESN techniques for EEG-based emotion recognition.

In addition, based on the provided results in terms of running time, it seems that the proposed model, PSD-Grouped deep ESN, has the lowest running time among the listed related works with a time of 5.35 seconds. This is can be attributed to the fact that ESNs are known to be fast and efficient in terms of training and testing time. This is because ESNs have a simple and fixed structure, with the reservoir weights randomly initialized and kept fixed during training. As a result, the computational cost of training an ESN is much lower than that of more complex architectures such as deep neural networks (DNNs) or convolutional neural networks (CNNs). Furthermore, our proposal, PSD- Grouped deep ESN, groups the input features based on their power spectral density (PSD) and uses separate ESNs for each group. This allows for a more efficient utilization of the available computational resources as each ESN is trained on a subset of the input features, reducing the overall complexity of the model.

For all these reasons we can say that our work improves the EEG-based emotion recognition compared to recent literature which encourages us to move forward in our research work in this direction.

5 Conclusion

Recognizing human emotions is proposed an improved deep method for emotion recognition in EEG signals. To achieve our objective, we used the Grouped Deep ESN model in conjunction with the Welch Power Spectral Density (Wlech PSD) method. This approach allows for the selection of important features to improve the recognition accuracy. In addition, we conducted tests on the well-known DEAP benchmark by employing the selected features to grouped Deep ESN model. The produced results, using 32 channels, show that the performance of our proposed model is better than many existing models, achieving 89.32% and 91.21% accuracy for H/L Arousal and H/L Valence classes, respectively. Furthermore, the introduced model has a significantly lower running time compared to the related works, taking only 5.35 seconds to run, while the other models have longer running times, ranging from 89.98 seconds to 693.4861 seconds. This indicates that our proposed model is not only effective in terms of accuracy but also efficient in terms of computational resources. Therefore, our approach is promising for real-time EEG- based emotion recognition applications, where speed and accuracy are both critical factors.

6 Future Work

In this work, we have proposed an improved deep learning method for EEG signals in the context of emotion recognition. In near future, we plan to apply the proposed method on other EEG-emotion datasets to conduct a comparative study. This will enable to analyse the performance of the model in different scenarios and provide a better understanding of its capabilities. Additionally, we aim to explore better ESN Reservoir Computing architectures that can efficiently process dynamic temporal graphs. This will allow for more efficient processing and analysis of data, leading to more accurate and reliable results.

References

[Adolphs et Anderson 2018] Adolphs, R. and Anderson, D.J.: The neuroscience of emotion: A new synthesis. Princeton University Press, 2018.

[Al-Nafjan et al. 2017] Al-Nafjan, A.; Hosny, M.; Al-Wabil, A.; Al-Ohali, Y.: Classification of human emotions from electroencephalogram (EEG) signal using deep neural network. International Journal of Advanced Computer Science and Applications, 8, 2017, 419–425.

[Ammar et al. 2013] Ammar B., Chouikhi N., Alimi A.M., Cherif F., Rezzoug N., and Gorce P.: Learning to walk using a recurrent neural network with time delay, in International Conference on Artificial Neural Networks. Springer, 2013, 511–518.

[Becker et al. 2017] Becker, H., Fleureau, J., Guillotel, P., Wendling, F., Merlet, I., and Albera, L.: Emotion recognition based on high-resolution eeg recordings and reconstructed brain sources. IEEE Transactions on Affective Computing, 8(1), 2017, 1–1.

[Bouazizi et al. 2021] Bouazizi, S., Ltifi, H. (2021). Improved Visual Analytic Process Under Cognitive Aspects. In: Barolli, L., Woungang, I., Enokido, T. (eds) Advanced Information Networking and Applications. AINA 2021. Lecture Notes in Networks and Systems, vol 225. Springer, Cham.

[Bouazizi et al. 2023] Bouazizi, S., Benmohamed, E., and Ltifi, H. (2023). Decision-making Based on an Improved Visual Analytics Approach for Emotion Prediction. Intelligent Decision Technologies, 2023, 1–20, DOI: 10.3233/IDT-220263

[Bozhkov et al. 2016] Bozhkov, L., Koprinkova-Hristova, P., and Georgieva, P.: Learning to decode human emotions with echo state networks. Neural Networks, 78, 2016, 112–119.

[Bozhkov et al. 2017] Bozhkov, L., Koprinkova-Hristova, P., and Georgieva, P.: Reservoir computing for emotion valence discrimination from eeg signals. Neurocomputing, 231, 2017, 28–40.

[Chen et al. 2019] Chen, J., Jiang, D., Zhang, Y.: A Hierarchical bidirectional GRU model with attention for EEG-based emotion classification, 2019, 118530–118540.

[Cheng et al. 2020] Cheng, J., Chen, M., Li, C., Liu, Y., Chen, X. Emotion recognition from multi-channel EEG via deep forest. IEEE Journal of Biomedical and Health Informatics, 2020, 24(10), 2881-2889.

[Chouikhi et al. 2017] Chouikhi, N., Ammar, B., Rokbani, N., and Alimi A.M.: Pso- based analysis of echo state network parameters for time series forecasting, Applied Soft Computing, 55, 2017, 211–225.

[Dunne et al. 2021] Dunne, R., Morris, T., and Harper, S.: A survey of ambient intelligence. ACM Computing Surveys (CSUR), 54(4), 2021, 1–27.

[Ellouzi et al. 2015] Ellouzi, H., Ltifi, H., Ayed, M. B. (2015). New multi-agent architecture of visual intelligent decision support systems application in the medical field. In 2015 IEEE/ACS 12th international conference of computer systems and applications (AICCSA), 1-8. IEEE.

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

[Fourati et al. 2022] Fourati, R., Ammar, B., Sanchez-Medina, J., Alimi, A-M.: Unsupervised Learning in Reservoir Computing for EEG-Based Emotion Recognition. IEEE Transactions on Affective, 13, 2022, 972-984.

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

[Hu et al. 2022] Hu, Z., Chen, L., Luo, Y., Zhou, J.: EEG-Based Emotion Recognition Using Convolutional Recurrent Neural Network with Multi-Head Self-Attention. Applied Science, 12, 2022, 11255.

[Iyer et al. 2023] Iyer, A., Das, S. S., Teotia, R., Maheshwari, S., Sharma, R. R.: CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings. Multimedia Tools and Applications, 82(4), 2023, 4883-4896.

[Jaeger 2001] Jaeger, H.: The 'echo state' approach to analysing and training recurrent neural networks, Technology GMD Technical Report 148, Germany National Research Center for Information Technology, 2001.

[Joshi et Ghongade 2022] Joshi, V.M., and Ghongade, R.B.: IDEA: Intellect database for emotion analysis using EEG signal. Journal of King Saud University – Computer and Information Sciences, 34, 2022, 4433–4447

[Katsigiannis et Ramzan 2018] Katsigiannis S. and Ramzan N.: 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, 2018, 22(1), 98–107.

[Khemakhem et al. 2020] Khemakhem, F., Ellouzi, H., Ltifi, H., Ayed, M. B. Agent-based intelligent decision support systems: a systematic review. IEEE Transactions on Cognitive and Developmental Systems, 2020, 14(1), 20-34.

[Kim et Choi] Kim, Y., Choi, A.: EEG-based emotion classification using long short-term memory network with attention mechanism. Sensors, 20, 2020, 6727

[Koelstra et al. 2012] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I.: Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing, 3(1), 2012, 18–31.

[Li et al. 2015] Li, G., Li, B.-J., Yu, X.-G. and Cheng, C.T.: Echo State Network with Bayesian Regularization for Forecasting Short-Term Power Production of Small Hydropower Plants. Energies, 8, 2015, 12228-12241.

[Li et al. 2018] Li, M., Xu, H., Liu, X., and Lu, S.: Emotion recognition from multichannel eeg signals using k-nearest neighbor classification: Technology and Health Care. Preprint, 2018, 1–11.

[Liu et al. 2019] Liu J., Huang, X., Huang, Y., Luo, Y., Yang, S.: Multi-objective spiking neural network hardware mapping based on immune genetic algorithm. In: Proceedings of the Artificial Neural Networks and Machine Learning: Theoretical Neural Computation – ICANN, 2019, 745–757.

[Liu et al. 2020] Liu, J., Sun, T., and Luo, Y., et al.: Echo state network optimization using binary grey wolf algorithm. Neurocomputing, 385, 2020, 310–318

[Ma et al. 2017] Ma, Q., Shen, L., Cottrell, G.W.: Deep-esn: A multiple projection-encoding hierarchical reservoir computing framework. arXiv preprint arXiv:1711.05255, 2017.

[Minguillon et al. 2017] Minguillon, J., Lopez-gordo, M.A., Pelayo F.: Biomedical Signal Processing and Control Trends in EEG-BCI for daily-life: Requirements for artifact removal. Biomedical Signal Processing and Control, 2017, 31, 97-101.

[Ong et Ibrahim 2018] Ong, Z.Y, and Ibrahim, Z.: Power Spectral Density Analysis for Human EEG-based Biometric Identification. In 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA), 2018, pp. 1–6

[Picard 1997] Picard, R.: Affective Computing, ser. Inteligencia artificial. The MIT Press, 1997. [Online]. Available: https://books.google.es/books? id=N1qqQgAACAAJ.

[Piho et Tjahjadi 2018] Piho, L., and Tjahjadi, T.: A mutual information based adaptive windowing of informative eeg for emotion recognition. IEEE Transactions on Affective Computing, 11 (4), 2018, 722-735.

[Poria et al. 2017] Poria, S., Cambria, E., Bajpai, R., and Hussain, A.: A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 2017, 98– 125.

[Qu et Zheng 2022] Qu, Z., Zheng, X.: EEG Emotion Recognition Based on Temporal and Spatial Features of Sensitive Signals. Computational Intelligence and Neuroscience, 2022, 5130184.

[Rattanyu et al. 2010] Rattanyu, K, Ohkura, M, Mizukawa, M.: Emotion Monitoring from Physiological Signals for Service Robots in the Living Space, 2010.

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

[Slama et al. 2017] Slama, N., Elloumi, W., and Alimi, A.M.: Distributed recurrent neural network learning via metropolis-weights consensus. International Conference on Neural Information Processing. Springer, 2017, 108–119.

[Soleymani et al. 2012] Soleymani, M., Lichtenauer, J., Pun, T., and Pantic M.: A multi- modal database for affect recognition and implicit tagging, IEEE Transactions on Affective Computing, 2012, 3(1), 42–55.

[Soures et al. 2017] Soures, N., Hays, L., Kudithipudi, D.: Robustness of a memristor based liquid state machine. In 2017 international joint conference on neural networks, 2017, May, pp. 2414-2420.

[Tripathi 2011] Tripathi, K.: Important Physiological Signals in the Body. Biomedical Engineering, 2011.

[Tripathi et al. 2017] Tripathi, S., Acharya, S., Sharma, R.D., Mittal, S, Bhattacharya, S.: Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In Proceedings of the Twenty-ninth IAAI conference, San Francisco, CA, USA, 4–9 February 2017.

[Weisstein 2003] Weisstein, E.W.: Simpson's rule, 2003. Retrieved from https://mathworld.wolfram.com/SimpsonsRule.html

[Welch 1967] Welch, P.: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics, 15(2), 1967, 70–73.

[Wioleta 2013] Wioleta, S.: Using Physiological Signals for Emotion Recognition. In 2013 6th International Conference on Human System Interactions (HSI), 2013, 556-561.

[Zheng et Lu 2015] Zheng, W.L., and Lu, B.L.: Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. IEEE Transactions on Autonomous Mental Development, 7(3), 2015, 162–175.

[Zhuang et al. 2017] Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H., and Yan, B.: Emotion recognition from EEG signals using multidimensional information in EMD domain. BioMed research international, 2017.