EEG-based Emotion Classification Using Deep Learning Methods

Yerzhan Orazayev, Yerniyaz Tolegen

***Abstract*—In recent decade, accurate interpretation of the emotional state of a human has obtained a great significance in brain-computer interface (BCI). The advantage of BCI in this problem is that it has an access to brain activity which can provide significant insight into the person's emotional state. The objective of this project is to accurately classify human emotion by using electroencephalographic (EEG) signals and deep learning algorithms. We use open-source dataset DEAP (a Dataset for Emotion Analysis using EEG, Physiological, and video signals), which is benchmark dataset in emotion classification study. In this project, we propose a convolutional neural network (CNN) model using a spectrograms of EEG signals that can reach classification accuracy of 70% on test data. Also, we analyzed the predictive capacity of time points using various recurrent neural networks (RNN). In this regard, we have found that pure time points are not effective.**

I. Introduction

Emotions play a significant role in human life. Its significance not only limited with human-to-human communication, but also greatly contributes many other aspects of our everyday life such as rational and intelligent behavior [1]. There is an increased interest in automatic emotion recognition from brain-computer interaction (BCI) community. BCI systems enable to establish non-muscular communication channel between human brain and external devices [2]. The basic idea of BCI is to measure the neuronial activities of the user and to predict certain aspects of the user's cognitive state [3], [4]. Electroencephalography (EEG)-based emotion classification is an important research field in BCI.

Apart from EEG-based emotion recognition there are other methods to study human emotion such as voice [5], facial [6], [7] and gesture [7], [8] clues. However, EEG signals in this problem are considered to be more reliable because of its high accuracy and objective evaluation compared to above-mentioned human expressions [9]. In addition, different psychophysiology studies have shown the correlations between human emotions and EEG signals [10], [11]. Moreover, in recent years due to technological advancement in BCI such as proliferation of wireless EEG devices, advances in computational intelligence techniques, and in machine learning there are increased opportunities for automatic emotion recognition systems [12].

Accurate emotion classification based on EEG is capable of contributing to the development of different fields as there are various applications such as in entertainment [13], safe driving [14], [15], health care [16], social security [17] and etc. In [18] the authors proposed to implement emotion classification for establishing communication with a locked-in patient. In [19], the application of emotion recognition in psychiatry and neurology as an indicator of emotional disorder is mentioned.

However, many present-day BCI systems are unable to adequately interpret emotional cues of human and suffer from the lack of emotional intelligence. Namely, they are flawed to accurately identify emotional states of human [20]. Therefore, it is important to further improve the performance of existing models or propose better ones.

Often in EEG-based emotion classification traditional machine learning algorithms are used, such as support vector machine (SVM), K-nearest neighbor (kNN), linear discriminant analysis (LDA), random forest and Naïve Bayes (NB) [12]. However, SVM based on frequency domain features as power spectral density (PSD) is most commonly used machine learning algorithm in EEG-based emotion classification task [21], [22]. Nevertheless, the current state-of-the-art approaches in the field use deep learning methods. In [23] Zheng and Lu proposed deep belief network (DBN) model that outperformed classical SVM and K-NN methods. In addition, other different methods such as recurrent neural network (RNN) [6], [24], multilayer perceptron (MLP) and convolutional neural network (CNN) [25] are used.

In this project we propose a CNN model using spectograms of EEG signals to classify human emotion. Besides, spectogram data we analyzed classification using time points of EEG signals, which did not show accurate predictive performance. To validate our model we use DEAP dataset [20], which is an open source and benchmark dataset in emotion classification study.

The article is organized as follows: Sec. II presents the dataset used in the study. In Sec III and Sec IV the classification using time points and spectogram are given, respectively. Next, results of these two methods are presented in Sec V. Finally, Sec. VI concludes the paper.

II. Dataset

Publicly available DEAP dataset includes EEG signals obtained from 32 healthy subjects aged between 19 and 37. For collection of different emotions, subjects were exposed to 40 music videos, each corresponding to various emotional genre. One stimuli corresponded to one trial. Once the trial is finished, the subjects reflect their emotional state on a two-dimensional emotional space proposed by Russell [26]. In the aforementioned space, two dimensions are Arousal and Valence. The former ranges from relaxed to arousal, while the latter ranges from pleasant to unpleasant. In each dimension, ratings are given continuously from 1 to 9. We decided to

classify each trial into two classes according to the following rules: pleasant > 5, unpleasant ≤ 5; aroused > 5,relaxed ≤ 5 for Valence and Arousal, respectively. For each subject EEG and peripheral signals were recorded at a sampling rate of 512Hz. Next, these EEG data were downsampled to 128Hz and eye artifacts were removed. The same downsampling was performed to peripheral signals.

TABLE I
DATASET STRUCTURE

Name	Array Shape	Contents
data	40 x 40 x 8064	trial x channel x time points
labels	40 x 4	trial x label

As it can be seen from the Table I each subject has an data array of 40 videos $\times$ 40 (EEG+peripheral) channels $\times$ 8064 (63 seconds$\times$128Hz) readings. In this study we use only EEG channels and last 60s of readings since first 3s correspond to pre-trial baseline. Also, each subject has a labels array indicating four emotion labels of each trail.

In this project we deal with two different binary classification problems: low/high arousal and low/high valence.

III. Classification Using Time Points

A. Gated Recurrent Unit (GRU)

Gated Recurrent Unit was proposed in 2014 as a solution to overcome the limitations of a vanilla recurrent neural network, which fails to capture long short-term dependicies due to vanishing gradient problem [27].

GRU consists of several gates starting with an update gate. Mathematically one could describe the gate's behaviour as follows:

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \tag{1}$$

As it can be seen from the equation 1, based on the input the update gate decides how much information should be passed over from previous hidden state to the current one.

Another gate is called reset gate. The hidden state will be forced to 0 and reset with the current input only, everytime the output from the gate is close to 0. The gate is computed as follows:

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{2}$$

Eventually, the dependecies between the above-mentioned gates and the behaviour of the hidden unit is descrbied in the equations below:

$$\tilde{h_t} = tanh(W_{xh}x_t + W_h(h_{t-1} \odot r_t) + b_h) \tag{3}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h_t} \tag{4}$$

It can be concluded that GRU adaptively forgets and remembers its state based on the input signal and the gates, which in turn allows to capture the long short-term dependencies.

B. Long Short Term Memory (LSTM)

LSTMs are another variation of RNN and also proven to deal with long term dependencies [28]. The main idea behind this network is the presence of the memory cell state. By adjusting gates, LSTM can add and remove information from the cell state.

LSTM consists of a memory cell, an input gate, a forget gate and an output gate. While gates control the alterations made to memory content, the memory cell carries that memory content. First two gate, namely the input and the forget gate decide the amount of new content that needs to be memorized or forgotten. They are computed in the following way:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{6}$$

Mathematically, the dependencies between gates and the memory cell are decribed in the equations below:

$$c_t = i_t \odot \tilde{c_t} + f_t \odot c_{-1} \tag{7}$$

where, $\tilde{c_t}$ is given in equation 8

$$\tilde{c_t} = tanh(W_{xc}x_t + W_{hc}(h_{t-1} + b_c)) \tag{8}$$

Based on the equations above, the following summary of the memory cell behaviour can be proposed:

input gate	forget gate	memory cell behaviour
0	1	remember the previous value
1	1	add to the previous value
0	0	erase the value
1	0	overwrite the value

Eventually, the current state output is calculated by a sigmoid layer from the output gate (refer to eq.9) and the result of cell update:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{9}$$

$$h_t = o_t \odot tanh(c_t) \tag{10}$$

C. Construction of GRU and LSTM models

In this project, we implemented GRU as well as LSTM models to achieve accurate classification results. For both GRU and LSTM, the number of hidden layers and the segment duration was chosen as model parameters. Since the original data was 60 seconds long, we decided to segment the data by 64 and 128 time points, which corresponded to 0.5 and 1 seconds, respectively. Data segmentation was done for the sake of the increase in sample size. The number of hidden layers that were implemented in model selection was 3, 5 and 7.

Fig. 1 shows the sample LSTM model. This model consists of fully connected three (LSTM) layers, two dropout layers, and dense layer. The dropout layer is used to reduce the overfitting by preventing units from co-adapting too much. The LSTM and dropout layers are used to learn features from raw EEG signals and dense layer is used for classification.

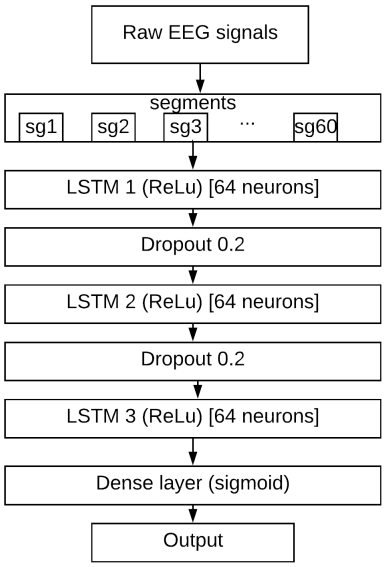

Fig. 1. Proposed neural network architecture

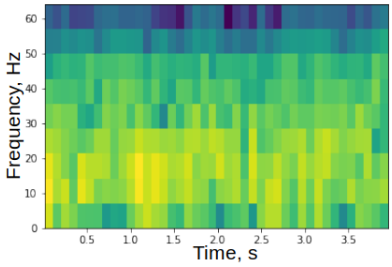

Fig. 2. Sample spectrogram

As a result of the model selection, it was shown that both GRU and LSTM presented low accuracy results, which in turn engendered the necessity to come up with another neural network model.

IV. Classification Using Spectrogram

A. The Generation of Spectrogram

A spectrogram is a visual representation of the spectrum of frequencies of the signal as they vary with time. In this project, we generate spectrograms for signal segments from each channel and treat them as an individual sample. The spectrum of each section is computed by Fast Fourier Transform (FFT) in python using *specgram* function from *matplotlib* package. The example of spectogram generated is shown in Fig. 2.

The segment duration is a model parameter to be selected. In this regard, we segment the time points from each channel into 2 second duration segment (256 time points) and 4 seconds duration segment (512 time points). As a result, for 2 second duration segmentation we obtain in total 38400 samples (30 segments $\times$ 32 channels $\times$ 40 trails) and for 4 second duration segmentation we obtain in total 19200 samples (15 segments $\times$ 32 channels $\times$ 40 trails). Next, these spectrograms are used in CNN.

B. Convolutional Neural Networks (CNN)

The CNN has a great success in computer vision, speech recognition, and natural language processing, etc. It allows to process data with grid-like structure and capable to extract multiple kinds of features automatically.

Usually a CNN model is composed of one or more stacked convolutional layers. Each of these layers typically consists of three processing stages: convolutional stage, detector stage and pooling stage [29]. In the first stage, convolutional stage, convolutional filters are applied to the given input of 2D shape and multiple feature maps are acquired from this input data. Convolutional stage is characterized by sparse connectivity and parameter sharing. The next stage, detector stage, is a non-linear transformation – such as a Sigmoid or ReLU activation function – of the obtained output from previous stage. The final stage, pooling stage, is an operation called pooling (e.g., Max Pooling and Average Pooling). This stage is used to make the representation to be invariant to translation to the next convolutional layer. Also, if next layer is fully connected layer, then its units can be reduced significantly by using pooling stage.

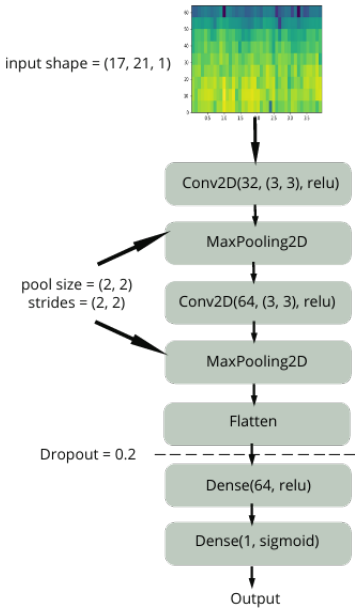

Fig. 3. CNN Architecture

C. Construction of CNN model

We propose a CNN model shown in Fig. 3 to classify human emotion using spectogram images of EEG singals. First convolutional layer takes 2D spectrogram image as input and the convolution operation uses 32 convolutional filters of and a convolutional kernel of size 3$\times$3. The first convolutional layer uses 'RElu' activation function. The next layer is a MaxPooling layer over 2$\times$2 blocks. It finds the maximum numerical value from previous layer's feature map and inserts it into the pooled feature map. In other words, it performs downsampling of the feature map. Also, it has stides of 2. It controls the distance between two successive windows in

max pooling. The second convolutional layer layer is set as 64 different filter with a size of 2×2 without overlap between strides. This setting helps to further fuse the information of a specific scale range from the prior features. Then, a max pooling stage is added as with first convolutional layer. Before connecting to fully connected layer, a flatten operation is adopted which is followed by dropout with 0.2 drop probability. Flatten is used to transform the final features into a one-dimensional feature vector, while dropout is used to avoid overfitting. In fully connected layer we have a dense layer with 64 units with 'RElu' activation function. Finally, one-unit dense layer with 'sigmoid' activation function is added for binary classification. Then this model is compiled with following parameters: optimizer is RMSprop, loss function is binary cross entropy and performance metric is accuracy.

D. Segment Size Selection for Spectrogram

As mentioned previously we select segment size of 2s or 4s duration for spectrogram generation. Using the above model we selected 4s long segment size since it gave the better performance during the validation as shown in Fig. 4 and 5. As can be seen, the validation accuracy for valence label of CNN model with spectrograms of 4s long segment size is better than 2s long segment size. Namely, the former one reaches validation accuracy of about 60%, whereas latter one goes to about 70%.

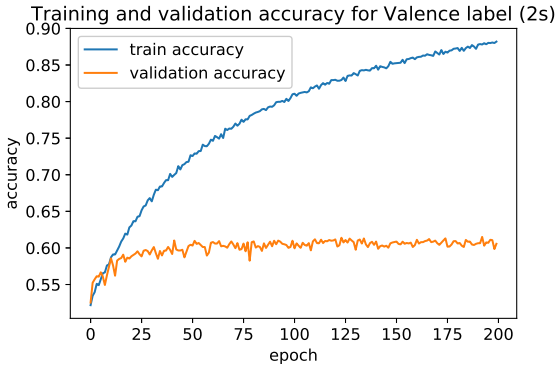

Fig. 4. Training and validation accuracy of CNN model for Valence label (2s).

V. Results

Models are trained and validated using 80% (from this 75% for training and 25% for validation) of the total data for 200 epochs. The remaining 20% of the data is used to test the model.

In Fig 7, the plot represents the results for LSTM model with 3 hidden layers and 64 time points for valence class. One can notice that the model is not training since there is no improvement in testing accuracy, while validation accuracy accounted for approximately 0.5 and remained stable with the increase of number of epochs.

Fig 8 presents the result for GRU model with 3 hidden layers and 64 time points for valence class. In this case, the same pattern as in fig.7 can be observed. Overall, the general

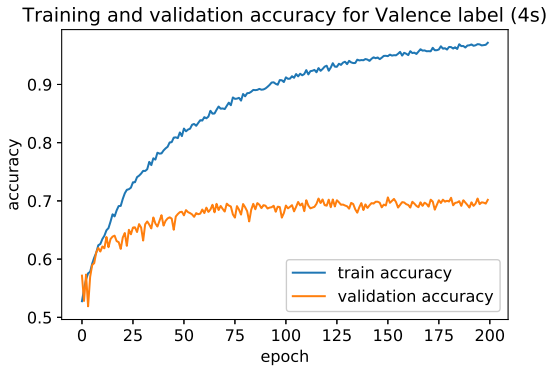

Fig. 5. Training and validation accuracy of CNN model for Valence label (4s).

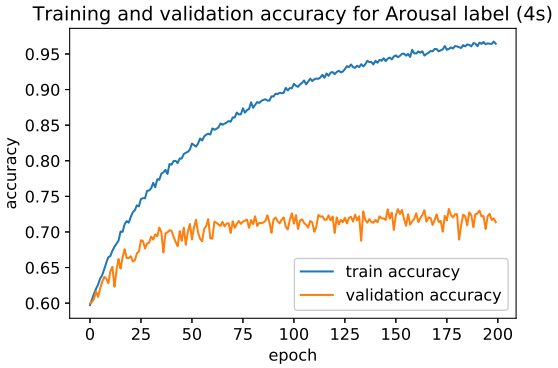

Fig. 6. Training and validation accuracy of CNN model for Arousal label (4s).

pattern holds true for both architectures with the various model parameters. To avoid repetition of superfluous figures and show the general pattern, the results for only 2 cases were shown.

Fig. 5 and 6 show the validation accuracy for both classes to be about 70%. From the presented models the best performance during the validation is achieved by CNN model.

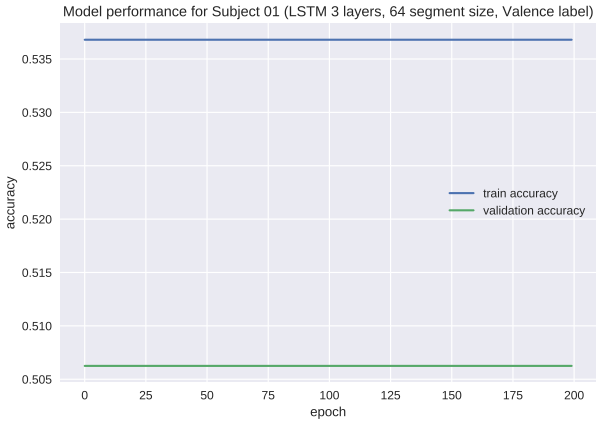

Fig. 7. LSTM results for Valence class

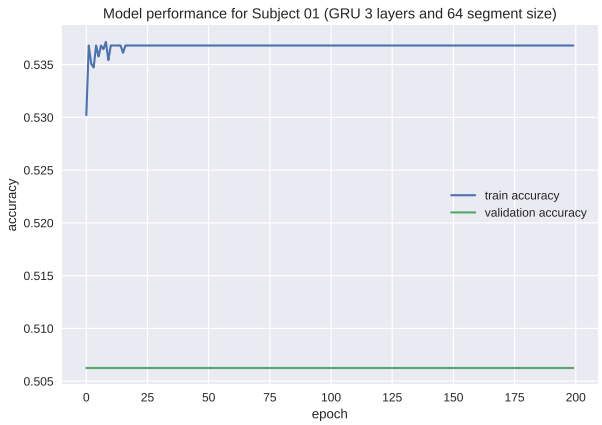

Fig. 8. GRU results for Valence class

Hence, our final model for this task is chosen to be a CNN model. Its performance on test data is shown in Table II.

TABLE II
PERFORMANCE OF CNN MODEL ON TEST DATA

	Valence Label	Arousal Label
Accuracy on Test Data	70.55%	70.83%

VI. CONCLUSION AND DISCUSSION

This study aimed to obtain accurate EEG-based emotion classification using deep learning methods. By testing on an open-source dataset DEAP, recurrent and convolutional neural networks were built and trained. After a thorough model selection procedure, the final model was chosen to be a CNN. The architecture consists of two convolutional layers followed by respective maxpooling layers, after which flatten and two dense layers are implemented. The results of recurrent neural networks failed to meet expected classification accuracy not reaching the value of more than 50 per cent, while CNN presented significant accuracy improvement and accounted for 70 per cent. From the RNN and CNN results, we can claim that time points have no predictive capacity, whereas "images" of spectral features can have a predictive capacity, in fact, prediction was done using a segment from a single channel. Thus, future work could be focused on the implementation of feature selection to find the best combination of channels.

REFERENCES

[1] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] T. Pun, T. I. Alecu, G. Chanel, J. Kronegg, and S. Voloshynovskiy, "Brain-computer interaction research at the computer vision and multimedia laboratory, university of geneva," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 210–213, 2006.

[4] E. T. Esfahani and V. Sundararajan, "Using brain–computer interfaces to detect human satisfaction in human–robot interaction," *International Journal of Humanoid Robotics*, vol. 8, no. 01, pp. 87–101, 2011.

[5] V. A. Petrushin, "Detecting emotions using voice signal analysis," May 22 2007, uS Patent 7,222,075.

[6] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of eeg signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, no. 1, pp. 17–28, 2016.

[7] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, 2007.

[8] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2007, pp. 71–82.

[9] G. L. Ahern and G. E. Schwartz, "Differential lateralization for positive and negative emotion in the human brain: Eeg spectral analysis," *Neuropsychologia*, vol. 23, no. 6, pp. 745–755, 1985.

[10] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007.

[11] G. G. Knyazev, J. Y. Slobodskoj-Plusnin, and A. V. Bocharov, "Gender differences in implicit and explicit processing of emotional facial expressions as revealed by event-related theta synchronization." *Emotion*, vol. 10, no. 5, p. 678, 2010.

[12] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on eeg brain-computer interface system research: A systematic review," *Applied Sciences*, vol. 7, no. 12, p. 1239, 2017.

[13] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behaviour & information technology*, vol. 25, no. 2, pp. 141–158, 2006.

[14] J. Healey, R. W. Picard *et al.*, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[15] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, no. 3, pp. 502–512, 2008.

[16] C. D. Katsis, N. S. Katertsidis, and D. I. Fotiadis, "An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders," *Biomedical Signal Processing and Control*, vol. 6, no. 3, pp. 261–268, 2011.

[17] B. Verschuere, G. Ben-Shakhar, and E. Meijer, *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press, 2011.

[18] C. Neuper, G. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller, "Clinical application of an eeg-based brain–computer interface: a case study in a patient with severe motor impairment," *Clinical neurophysiology*, vol. 114, no. 3, pp. 399–409, 2003.

[19] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel eeg data through convolutional recurrent neural network," in *Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on*. IEEE, 2016, pp. 352–359.

[20] 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.

[21] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "Eeg-based emotion recognition during watching movies," in *Neural engineering (NER), 2011 5th international IEEE/EMBS conference on*. IEEE, 2011, pp. 667–670.

[22] X.-W. Wang, D. Nie, and B.-L. Lu, "Eeg-based emotion recognition using frequency domain features and support vector machines," in *International conference on neural information processing*. Springer, 2011, pp. 734–743.

[23] 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.

[24] 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.

[25] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset." in *AAAI*, 2017, pp. 4746–4752.
[26] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
[27] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.