Objective Valence and Arousal Detection Using Extreme Learning Machines on EEG

Amir Eldesoky, Matthew DeSilva

University of New Brunswick, Department of Electrical and Computer Engineering

Abstract— **Advancements in machine learning are opening new doors in the field of Human-Computer Interaction, and specifically the area of Affective Computing. A significant cornerstone in Affective Computing is automated emotion detection to allow technology to adapt to the emotional state of the user. One of the most commonly used models for emotion detection is the valence arousal model to detect the user's emotional state based on their position on the valence arousal scale. Most of the research that has been done in this area has focused on subjective non-physiological signals like facial expressions and voice tones; However, there has been a recent push to use physiological signals as an objective metric for emotion detection. EEG emerged as one of the most popular physiological signals for emotions detection due to it being a noninvasive signal with good temporal and special resolutions. The majority of the classification approaches used in EEG classification can be divided into statistical supervised approaches and deep learning approaches. The statistical supervised approaches have fast training and inference times and low accuracies while the deep learning approaches have high accuracies with long training and inference times. As a result, a need was identified for a classifier that can obtain accuracies comparable to that of deep learning approaches while having a training and inference time similar to that of the statistical supervised methods. This research will explore using Extreme Learning Machines to satisfy that need.**

Key Terms— Affective Computing; Extreme Learning Machines; Support Vector Machines; Pattern Recognition; EEG; DEAP Dataset.

I. INTRODUCTION

With the recent advancements in machine learning, a lot of research has been conducted in the field of affective computing. This is a field in the area of human-computer interaction that enables technology to adapt and change based on the emotional state of the user to create a more natural, effective, and immersive user experience [1]. Most of the research done in affective computing focuses on analyzing non-physiological signals like facial expressions, body language, and voice tones to detect emotions. However, strides have been taken recently towards using physiological signals to further the capabilities of Brain-Computer Interface (BCI) technologies in emotion detection and classification [2]. The disadvantage to the non-physiological signals approach is its subjectivity because the way people express emotions varies from one person to the other. On the other hand, using physiological signals provides objective metrics that are intrinsic and independent of the individual's control, which makes it a more reliable approach to emotion detection. As seen in the work done by Li et al in [3], EEG emerged as one of the best physiological signals for emotion detection research because it is noninvasive with a good temporal resolution and an acceptable spatial resolution.

The research that has been conducted so far on EEG for emotion recognition has focused heavily on statistical approaches like Support Vector Machines similar to the work done by Wichakam et al in [4] and K-nearest Neighbors similar to the work done by Li et al in [3]. The other approach to emotion recognition from EEG is using deep and convolutional neural networks as seen in the work done by Tripathi et al in [5]. The deep learning classifiers surpassed the statistical supervised classifiers in terms of accuracy of emotion detection, but they require considerably long training and classification times due to the complex computations involved in their algorithms.

In order to be able to use these classification models in a real-world application, a new classification approach will need to be taken to obtain accuracies similar to that of the deep learning models in a time similar to that of the statistical approaches. This is a gap in the current existing solutions, and this research will aim to fill it. In this research, we propose an objective way of emotion detection using Extreme Learning Machine (ELM) on EEG. This paper will demonstrate how ELM can obtain accuracies similar to that of statistical methods with a significantly shorter training time, and with good features, it can provide a happy medium between deep learning and the statistical approaches.

II. DATASET

This research used the Dataset for Emotion Analysis using EEG, Physiological and Video Signals (DEAP) [6]. The DEAP dataset observed the physiological responses of 32 subjects as they watched one-minute extracts of 40 music videos. Following each video, the subject would rate the video based on their feelings of arousal, valence, dominance, and liking. For arousal, the subjects would rate the video on a scale of 1-9, with 1 being bored and 9 being excited. Regarding valence, the subjects would again rate the video on a scale of 1-9, with 1 being unhappy/sad and 9 being happy/joyful [4].

The dataset consists of raw physiological data (EEG, EOG, EMG, GSR, Respiration belt, Plethysmograph, and Temperature), preprocessed physiological data, participant ratings, participant questionnaire, and a video list [6]. During the experiment, the EEG signal was obtained from 32 different channels placed according to the international 10-20 system [7], the EOG signal was taken from 4 channels, the EMG was taken from 4 channels, and the GSR, Respiration, Plethysmograph, temperature were each taken from one channel. To produce the preprocessed EEG data, Koelstra et al applied a number of modifications to the raw data [8]. Such modifications included: downsampling the data from 512 Hz to 128 Hz, removing EOG artefacts, filtering the frequency from 4 to 45 Hz using a bandpass filter, the data was averaged to the common reference, the data was segmented into 60 second trials with a 3 second pre-trial baseline removed, and the trials were reordered [6]. Six frequency bands reside within the range of 4 to 45 Hz, these include the: delta, theta, alpha, slow alpha, beta, and gamma bands. The components of the dataset used in this experiment were the preprocessed EEG data, and participant ratings.

III. METHODS

This experiment will look at emotion detection by creating two classifiers using the DEAP dataset: one to detect valence and the other to detect arousal. The preprocessed EEG data from DEAP was used, so no further preprocessing was required. Afterwards, feature extraction and selection were conducted to obtain a subset of features that are representative of the way the EEG data correlates to the valence and arousal labels. Lastly, valence and arousal were classified using SVM, KNN, and ELM and their accuracies along with the training time were compared.

A. Feature Extraction

The Toolbox for Emotion Analysis using Physiological signals (TEAP) presented in the work of Soleymani et al in [9] was used to load the data and extract statistical features from it. In addition, the signals Power Spectral Density (PSD) for each frequency band, the energy, entropy, and fractal dimensions were obtained for these signals.

In order to be able to run the kernel trick for KELM on a local machine, the number of data points had to be reduced. Based on the work illustrated by Li et al in [3], only 18 channels out of the 32 were required to obtain high classification accuracies. The channels that showed high correlation with valence and arousal in Li's work were Fp1, Fp2, F3, F4, F7, F8, FC5, FC6, FC1, FC2, AF3, AF4, C3, C4, T7, T8, Fz, and Cz.

A summary of the features extracted for the 18 chosen channels and how they were obtained can be seen below.

i. Power Spectral Density (PSD)

The PSD for each of the six frequency bands was obtained using the TEAP toolbox, which obtains the PSD using Welch energy method with a window size of 4 seconds [10].

ii. Statistical Features

The TEAP toolbox was used to obtain the statistical features, and the extracted features were the mean, standard deviation, Kurtosis, and Skewness which were used in the statistical classifiers shown in [3, 4, 6, 11].

iii. Energy

The energy of the signal was obtained using the TEAP toolbox, which finds the energy by obtaining the squared sum of signal values over time.

iv. Entropy

The entropy of the signal calculates its degree of disorder, and it was extracted based on the work done by Li et al in [3].

v. Fractal Dimension

The Katz Fractal Dimension information was extracted based on the work shown by Liu and Sorina in [12].

B. Feature Selection

Following the feature extraction step, a total of 13 features were extracted. Two matrices of features scatters were plotted for valence and arousal as seen in Appendix I to explore the separability obtained from the combination of features. The features demonstrated very low interclass separability for both valence and arousal, which illustrates that the real challenge in this classification problem is feature extraction.

Afterwards, a subset of the features that correlate to the labels for valence and arousal were selected using a combination of Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS). The criterion for SFS and SBS was the ratio of error in the classification with a selected subset of features using Quadratic Discriminant Analysis (QDA). SFS algorithm starts with the feature that results in the smallest criterion value, and it stops when the criterion value increases. In contrast to SFS, SBS starts with calculating the criterion for all the features then it keeps dropping one feature at a time until the criterion value increases. Both algorithms were applied on the features, and the combinations of features that produced the lowest criterion values were selected. Based on the results of SFS, delta PSD, theta PSD, slow alpha PSD, alpha PSD, and gamma PSD were selected for Valence. For Arousal, delta PSD, slow alpha PSD, alpha PSD, Kurtosis, Skewness, and Entropy were selected using SBS.

C. Classification

The selected features were used to train KNN, SVM, ELM, and KELM classifiers. Two classifiers for each type were trained: one for arousal detection (high or low) and one for valence detection (high or low). All the classifiers were trained using K-fold cross-validation with a K value of 10. The average accuracy for each classifier and the training time were recorded and compared. The KNN classifiers were trained with varying values of K, and the K value that had the best performance was 1000 for both arousal and valence.

The SVM classifier used a kernel trick to add more dimensions to the problem to help with the interclass separability of the features. The SVM classifiers were tested with linear and radial basis function (RBF) kernels. Based on the classifiers results, RBF was selected because it resulted in the best classification performance.

ELM and KELM classifiers were trained for valence and arousal detection as well. The ELM classifier was optimized by trying different number of hidden neurons and testing the performance. The highest accuracy was reached with 200 hidden neurons, and no major changes were noticed in accuracy when more neurons were added after that point. Additionally, KELM classifiers were optimized by evaluating their performance with both linear and RBF kernels. Similar to SVM, the RBF kernel outperformed the linear kernel for KELM.

Lastly, the accuracies of the classifiers were recorded along with the training time for each classifier. The results were used to show that ELM has the ability to obtain the same results as KNN and SVM in a significantly shorter amount of time.

IV. Results

The results obtained from this experiment have been summarized in Figures 1, 2, and 3. As seen in Figure 1, each of the classifiers functions at a prediction accuracy between 50-60%. Since all 6 of the classifiers are heavily dependent on the extracted features, they all performed comparably.

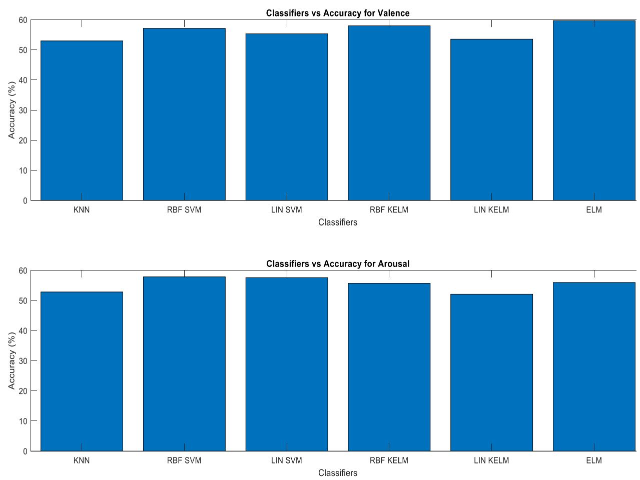

Figure 1: Classifiers vs Accuracy

Regarding Figures 2 and 3, the optimal classifier would appear in the top left corner of the scatter plot. A classifier present in this position would convey a high accuracy with a short training time. As seen in Figure 2, the ELM classifier can be found in the top left region of the scatter plot. Consequently, the ELM classifier demonstrated the highest classification accuracy with a training time of less than one second when classifying valence.

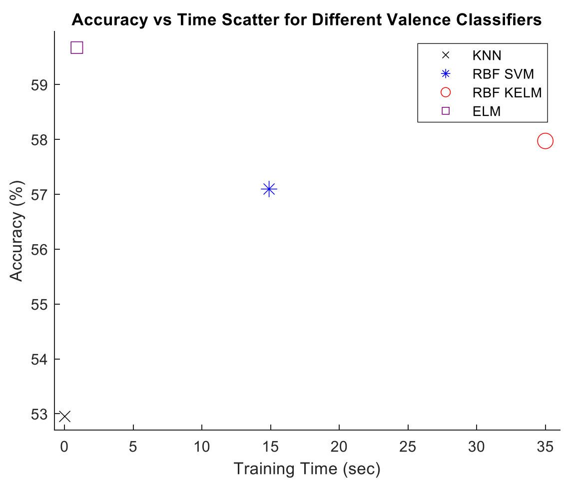

Figure 2: Accuracy vs Time for Valence Classifiers

On the other hand, as seen in Figure 3, the ELM classifier ranked lower than the RBF SVM classifier. Although the ELM classifier did not perform as accurately, the training time was nearly 20 times less than the RBF SVM classifier. Additionally, the classification accuracy for the ELM classifier was only 2% less than the RBF SVM classifier. This, for the purpose of this experiment, can be considered a comparable performance.

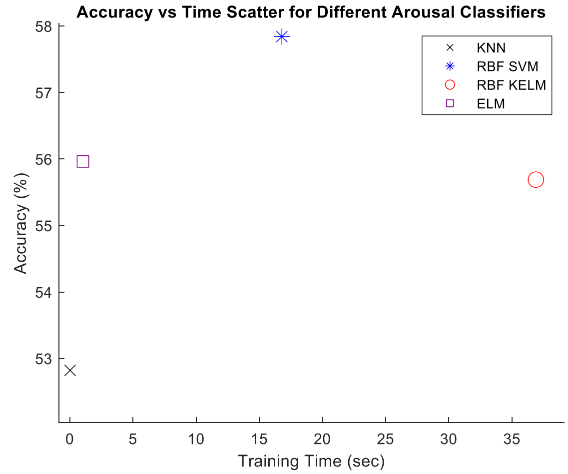

Figure 3: Accuracy vs Time for Arousal Classifiers

V. Discussion

We managed to train and compare multiple classifiers on the EEG signals from the DEAP dataset. Based on the initial results, ELM showed very promising results in this classification problem and it demonstrated its ability to obtain accurate results in a very short amount of time. However, many challenges were identified with this classification problem.

The most significant challenge with this problem is in the feature extraction phase. As seen in Appendix I, the features extracted for this problem displayed no signs of interclass separability for valence or arousal detection. These features were chosen based on the previous work done on the DEAP dataset. The accuracy obtained in the previous work using statistical supervised methods ranged from 60 to 70% [4, 6], with the exception to the 80% achieved in the work done by Li et al in [3]. However, something to be noted about the approach taken by Li et al in [3] is that they used a K value of 3 and they did not have a validation set, so their classifier was over tuned for the dataset which might have resulted in overfitting. All the classifiers demonstrated in this research rely on the extracted features, and in order to obtain better classification results, a breakthrough in feature extraction will need to be achieved.

The second challenge was using the preprocessed data instead of the raw data. The preprocessed MATLAB data has been modified and downsampled to a low sampling rate of 128Hz compared to the original sampling rate of 512Hz [6]. This significant drop in the sampling rate makes the resulting signal susceptible to aliasing and a huge loss in information, which can drastically affect the classification results. Additionally, the 4 to 45 Hz bandpass filter cut off half of the gamma band information and nearly the whole delta band information.

The third area of challenge is the way the data in DEAP was acquired. The data was acquired by showing 60-second music videos to subjects, and each video was given valence and arousal scores by the subjects. The EEG signal was recorded for each subject for the duration of the video, and the whole signal was given one label. The problem with this method of data acquisition is that the music videos used were not geared specifically for emotion detection and 60 seconds is a long time for the subject to have the same emotion. As a result, the whole EEG signal is given one score for valence and arousal, but the subject did not necessarily have those feelings for the whole duration of the video.

Lastly, a challenge that was identified is the shortcomings of the TEAP toolbox. It was observed that TEAP is specifically tailored for the DEAP dataset and most of the functions from TEAP were hardcoded for it. This did not affect the classification in this research but the lack of a toolbox that can generalize to different datasets can hinder the advancements in emotion classification based on physiological signals.

VI. CONCLUSION

ELM showed promising results in valence and arousal detection using EEG. It manages to achieve similar accuracies to the statistical supervised methods that have been previously applied on the DEAP dataset like KNN and SVM. The real gain from using ELM is the time saved in the training and inference while obtaining comparable results to the statistical supervised methods. However, in order for ELM to compete with the results observed using deep learning as seen in the work done by Tripathi et al in [5], a breakthrough in feature extraction will need be achieved since ELM's performance is directly affected by the extracted features.

APPENDIX I – MATRICES OF SCATTERPLOTS

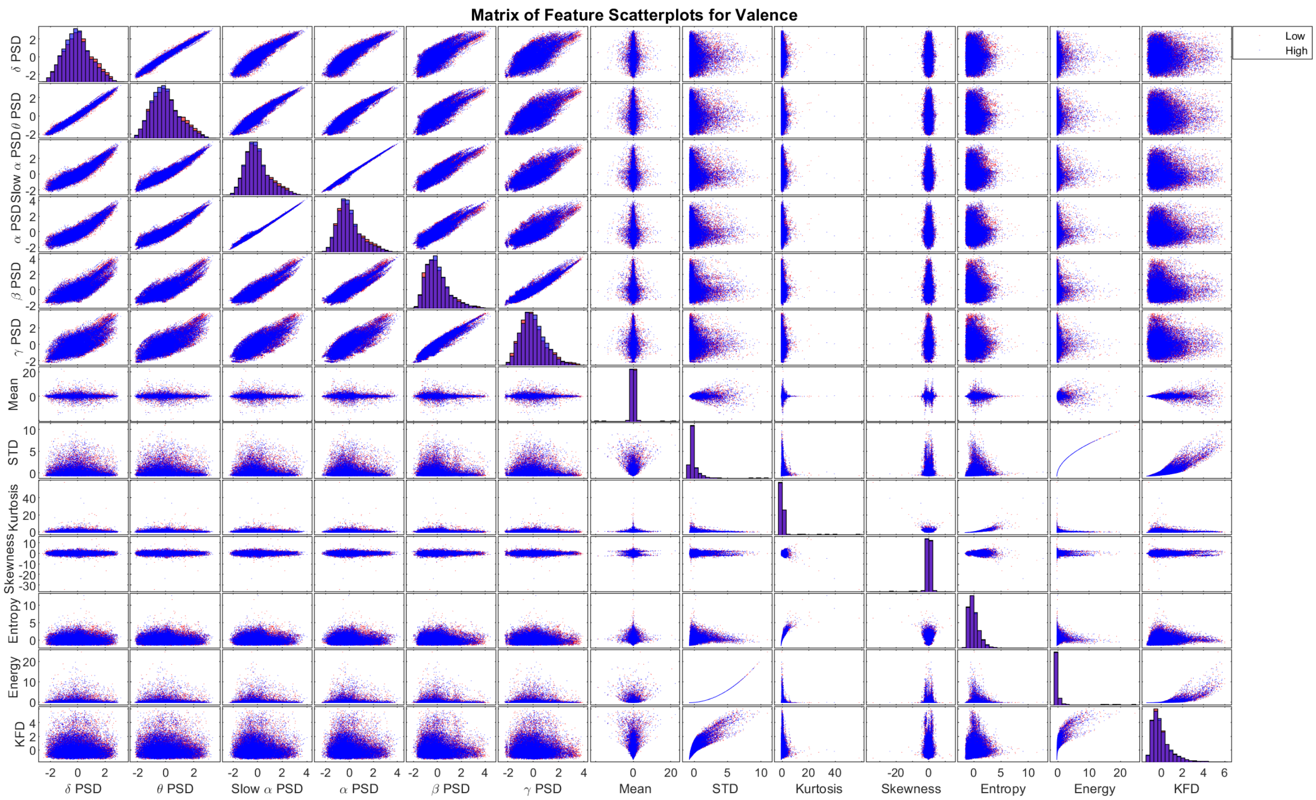

Figure 4: Matrix of Feature Scatterplots for Valence

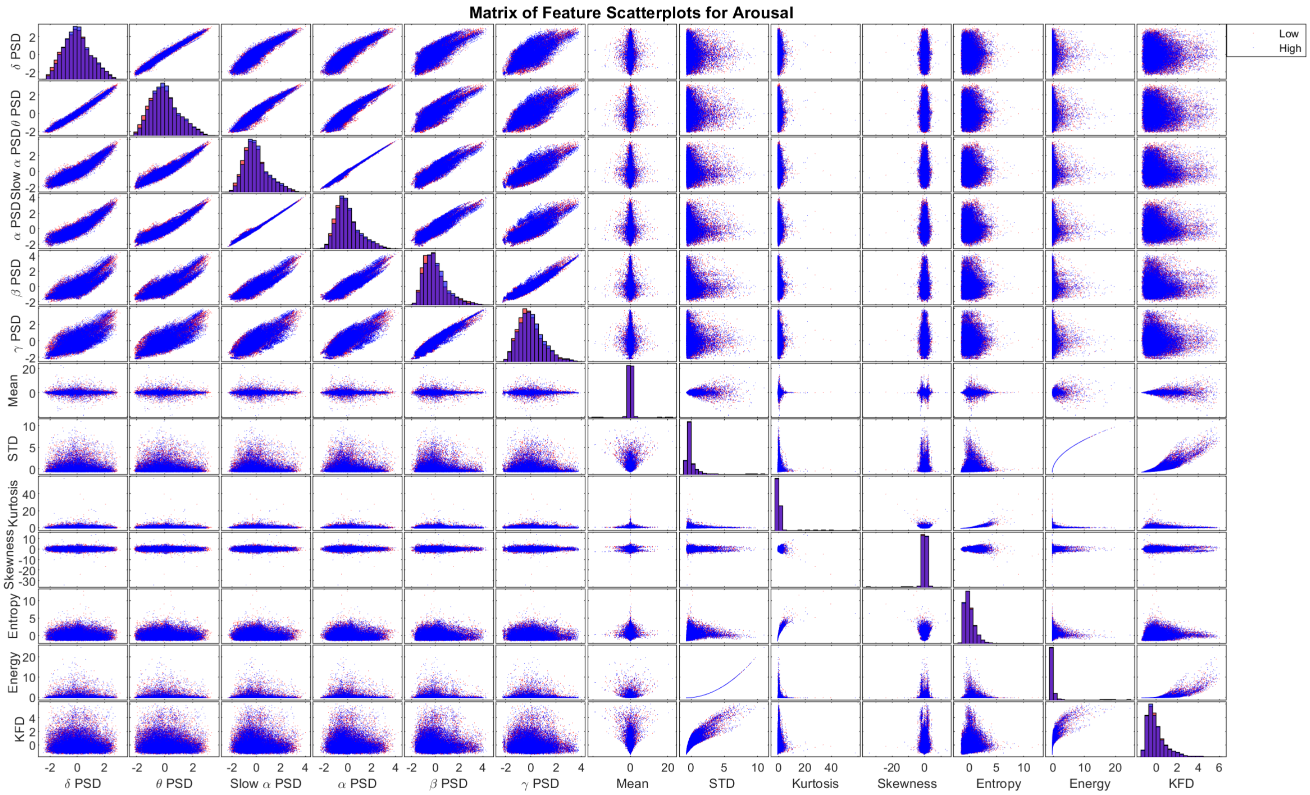

Figure 5: Matrix of Feature Selection for Arousal

ACKNOWLEDGEMENT

We would like to take the opportunity to thank Mr. Evan Campbell, Dr. Ankoon Phinyomark, and Dr. Erik Scheme for their continuous support and guidance through the duration of this experiment.

REFERENCES

[1] J. Marín-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E. P. Scilingo, M. Alcañiz and G. Valenza, "Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors," September 2018. [Online]. Available: https://www.nature.com/articles/s41598-018-32063-4. [Accessed December 2018].

[2] A. Al-Nafijan, M. Hosny, A. Al-Wabil and Y. Al-Ohali, "Classification of Human Emotions from Electroencephalogram (EEG) Signal using Deep Neural Network," *International Journal of Advanced Computer Science and Applications,* vol. 8, no. 9, pp. 419-425, 2017.

[3] M. Li, H. Xu, X. Liu and S. Lu, "Emotion recognition from multichannel EEG signals using K-nearest neighbor classification," 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/29758974. [Accessed December 2018].

[4] I. Wichakam and P. Vateekul, "An Evaluation of Feature Extraction in EEG-Based Emotion Prediction with Support Vector Machines," in *The 11th International Joint Conference on Computer Sciences and Software Engineering (JCSSE 2014)*, Pattaya, Thailand, 2014.

[5] 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 *Twenty-Ninth AAAI Conference on Innovative Application*, San Francisco, 2017.

[6] S. Koelstra, C. Mühl, M. Soleymani, A. Yazdani and J. s. Lee, "DEAP: A Dataset for Emotion Analysis using Physiological and Audiovisual Signals," 2011. [Online]. Available: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html. [Accessed 2018].

[7] J. N. Acharya, A. Hani, J. Cheek, P. Thirumala and T. N. Tsuchida, "American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature," *Journal of Clinical Neurophysiology,* vol. 33, no. 4, pp. 308-311, 2016.

[8] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt and I. Patras, "DEAP: A Database for Emotion Analysis using," *IEEE Transactions on Affective Computing,* vol. 3, no. 1, pp. 18-31, 2011.

[9] M. Soleymani, F. Villaro-Dixon, T. Pun and G. Chanel, "Toolbox for Emotional feAture extraction from Physiological signals (TEAP)," *Frontiers in ICT,* 2017.

[10] M. Soleymani, F. Villaro-Dixon and T. Pun, "TEAP (Toolbox for Emotion Analysis Using Physiological Signals) doc," 2017. [Online]. Available: http://www.teap.science/doxygen/files.html. [Accessed December 2018].

[11] A. Samara, M. L. R. Menezes and L. Galway, "Feature Extraction for Emotion Recognition and Modelling using Neurophysiological Data," in *2016 15th International Conference on Ubiquitous Computing and Communications*, Granada, Spain, 2016.

[12] Y. Liu and O. Sourina, "Real-Time Fractal-Based Valence Level Recognition from EEG," January 2013. [Online]. Available: https://www.researchgate.net/publication/288986001_Real-Time_Fractal-Based_Valence_Level_Recognition_from_EEG. [Accessed December 2018].