A Comparative Study of Subject-Dependent and Subject-Independent Strategies for EEG-Based Emotion Recognition using LSTM Network

Conference Paper · March 2020

DOI: 10.1145/3388142.3388167

6 authors, including:

Debarshi Nath

Anubhav Anubhav
Delhi Technological University

Divyashikha Sethia
Delhi Technological University

Diksha Kalra
Delhi Technological University

A Comparative Study of Subject-Dependent and Subject-Independent Strategies for EEG-Based Emotion Recognition using LSTM Network

Debarshi Nath
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
debarshinath94@gmail.com

Anubhav
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
anubhav2901@gmail.com

Mrigank Singh
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
mriganksingh1@gmail.com

Divyashikha Sethia
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
divyashikha@dtu.ac.in

Diksha Kalra
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
kalradiksha11@gmail.com

S. Indu
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
s.indu@dce.ac.in

ABSTRACT
This paper addresses the problem of EEG-based emotion recognition and classification and investigates the performance of classifiers for subject-independent and subject-dependent models separately. The results are compared with other classifiers and also with existing work in the concerned domain as well. We perform the experiments on the publicly available DEAP dataset with band power as the feature and classification accuracies are found pertaining to the widely accepted Valence-Arousal Model. The best results were reported by the LSTM model in case of the subject-dependent model with accuracies of 94.69% and 93.13% on valence and arousal scales respectively. SVM performed the best for the subject-independent model with accuracies of 72.19% on valence scale and 71.25% on arousal scale.

CCS Concepts
• **Human-centered computing →Human computer interaction (HCI);** • **Computing methodologies → Machine learning; Artificial intelligence; Feature selection; Classification and regression trees; Machine learning algorithms; Machine learning approaches; Support vector machines; Neural networks.**

Keywords
Electroencephalography (EEG); Brain Computer Interface (BCI); Emotion Recognition; Valence-Arousal Model; Long Short-Term Memory network;

ICCDA 2020, March 9–12, 2020, Silicon Valley, CA, USA

ACM ISBN 978-1-4503-7644-0/20/03…$15.00
https://doi.org/10.1145/3388142.3388167

1. INTRODUCTION
Assessment of emotion is an area that has been dealt with extensively in recent literature in the domain of Brain-Computer Interface. Recent researches have targeted mainly for people suffering from disorders that hinder or alter their ability to express emotions effectively [10]. Different modalities have facilitated the study of emotions like facial expressions, human speech and tone, and physiological signals. One of such modalities is the electroencephalogram (EEG). The human brain comprises of billions of cells, of which a large number are neurons, and others which aid and facilitate the activity of neurons. Whenever any activity occurs, it generates an electrical impulse in the brain due to which thousands of neurons fire in sync. The activations of neurons within our cranium determine emotions and responses. Brain waves are generated by electrical pulses fired in sync from billions of neurons communicating with each other in the nervous system. Brain waves can be detected using electrodes placed on the scalp using invasive, non-invasive or semi-invasive procedures. We study these brain waves in the form of Electroencephalography (EEG). EEG in recent years has become an essential tool with its applications and utility in neuroscience, Brain-Computer Interfaces (BCI's), and commercial applications as well. Analytical tools prevalent in EEG studies frequently use machine learning techniques to uncover relevant information for neural classification and neuroimaging.

Emotion, as easy as it is to feel, is equally challenging to study and measure objectively. Researchers for its objective analysis have pro-posed several models. Most prominent of those is Valence-Arousal Model by Russell [11] that represents emotions on a 2-D circular space where arousal represents the vertical axis, and valence represents the horizontal axis. Various emotions circumscribe the centre of the valence-arousal plane. This centre represents the neutral valence and a medium value for arousal. This model is widely accepted and used extensively in emotion recognition studies. Figure 1 illustrates thevarious ratings of valence and arousal for different emotions.

We also explore other models available for emotion representation. Bradley et al. [2] proposed another model known as Approach and Withdrawal Model or the Vector model. This model is also a 2-D

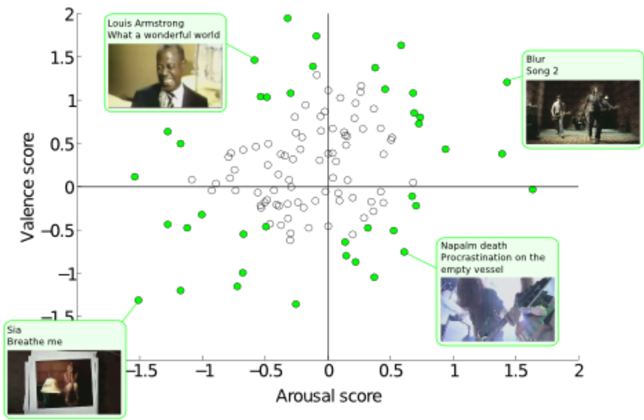

Figure 1. Valence and Arousal ratings corresponding to different emotions as reported in DEAP dataset. [6]

Table 1: Different frequency bands and their significance

Bands	Frequency Range (Hz)	Significance
Delta	0.5 – 4	Deep Sleep, Deepest level of Relaxation
Theta	4 – 8	REM Sleep, Deep and Raw Emotions, Cognitive processing
Alpha	8 – 13	Drowsy state, Relaxation, Calmness
Beta	13 – 30	Conscious state, Thought process
Gamma	> 30	Two different senses at same time

model where the valence determines the direction of emotion. Here, a positive value of valence turns the emotion in the top vector, and on the contrary, the negative value of valence turns the emotion in the down vector. Watson et al. [4] describes a Positive-Negative Model. In this model, the horizontal axis represents low to high negative affect, and the vertical axis represents low to high positive affect.

A typical EEG device measures the electrical activity with the help of electrodes which are in contact with the head of a subject, according to the 10-20 International System. The raw EEG signal is composed of many frequency bands and noise. Hence we need to filter the raw EEG signal then decompose it into constituent frequency bands like Delta (0-4 Hz), Theta (4-8 Hz), Alpha (8-13Hz), Beta (13-30 Hz), and Gamma (>30 Hz) for emotion recognition studies. These bands contain useful information of concern for brain activities. Table 1 describes the significance of these frequency bands. Therefore, EEG signals have broad applications ranging from emotion recognition to diseases and disorders prediction like Sleep Apnea, Epilepsy, and Alzheimer's disease.

2. RELATED WORKS

This section comprises of a detailed review of some prominent researches related to emotion recognition from EEG Signals. For efficient classification of emotions, various factors plays a vital role like pre-processing of brain signals to remove artefacts, feature extraction, feature reduction techniques, and lastly the classifiers used for classification of emotions. Feature extraction plays a vital in emotion recognition as these features represent the information stored in EEG signals. Broadly these features are of three types:

- Time Domain Features
- Frequency Domain Features
- Time-Frequency Domain Features

Jenke et al. [5] have reviewed various feature extraction methods like ReliefF and Min-Redundancy-Max-Relevance (mRMR) for recognition of emotions using EEG signals. Various time-domain features like Event-Related Potentials (ERPs), Statistics of signal such as mean, standard deviation, power, and Hjorth features, Fractal Dimension (FD), Higher-Order Crossings (HOC) are studied. Also, frequency domain features like band power, and Higher-Order Spectra (HOS) as well as time-frequency domain features like Hilbert Huang Spectrum (HHS), Discrete Wavelet Transform (DWT) are described in detail. From our review of the existing literature, we infer that there are two different types of approaches towards emotion recognition algorithm:

- Subject-Dependent
- Subject-Independent

In both the algorithms feature extraction methodology remains the same. However, in the case of subject-dependent on the model algorithm, the classifier is trained for each subject individually, whereas, in the subject-independent model, the classifier is trained for several subjects. We discuss some notable researches related to subject-dependent and subject-independent emotion recognition strategies. Pandey and Seeja [9] used the DEAP Dataset for creating a subject-independent model using wavelet transform as a feature and deep neural network (DNN) as the classifier, achieving a maximum classification accuracy of 62.5% for valance, and 64.25% for arousal. Liu and Sourina [8] proposed a real-time subject dependent emotion recognition technique from EEG Signals using fractal dimension (FD) in combination with statistical and higher Order Crossings (HOC) as feature and SVM as classifier. 14-channel Emotiv device was used for data collection. Sixteen subjects were shown visual stimuli and eight different emotions were recognized. Maximum mean accuracy of 77.81% was achieved for classifying two different emotions. The proposed method is validated on the DEAP Dataset, achieving a mean accuracy of 85.38% for classifying two emotions. Various deep learning applications in Natural Language Processing (NLP) demonstrates the memorizing capability of the LSTM model. Hence, LSTM model can be instrumental in emotion recognition. We also describe other researches employing LSTM model for the task of emotion recognition. Alhagry et al. [1] employed the DEAP dataset for emotion recognition. LSTM model was used for the classification of emotions of 32 subjects individually and achieved 85.65%, 85.45%, and 87.99% accuracy corresponding to arousal, valence, and liking. Li et al. [7] classified human emotions from EEG signals. RASM is used as a feature which is extracted from the DEAP dataset. LSTM network is used as the classifier, achieving a mean accuracy of 76.67% for valence.

Yang et al. [14] employs a parallel combination of Convolutional Neural Network (CNN) and LSTM Network to derive features from EEG and physiological signals in DEAP dataset. Then, trains the softmax classifier for emotion classification using the subject-dependent strategy, which achieves a mean accuracy of 90.80% and 91.03% for valence and arousal. In this paper, we explore two different training strategies, namely subject-dependent and subject-independent. For both environments, we use band power feature extracted from the raw EEG signals and test different classification techniques like KNN, SVM, Decision Tree, Random Forest, and LSTM model. The next section describes the methodology adopted in this research.

3. METHODOLOGY

3.1 Dataset

The DEAP dataset [6] is a publicly available multimodal dataset designed specifically for emotion analysis. This dataset describes the recordings of EEG and other physiological signals for 32 participants while they watched 40 one-minute-long excerpts of selected videos. A standard EEG headset with 32 channels and sensors like EOG, EMG, and temperature record signals with a sampling frequency of 128 Hz. Then the recorded signals are passed from a bandpass filter to remove noise and artefacts like eye blinks. For assigning an emotion to the signals in terms of valence, arousal, dominance, and liking the participants rated each video with values ranging from 1-9.

3.2 Preprocessing

We observe the raw EEG signals of the DEAP dataset and try to find any similarity between EEG signals of each subject, and also any similarity between different trials for the same subject. We infer that the EEG signal of each individual is unique. There exists no similarity in the EEG signals of different individuals, even if the signals are from the same trial. This is intuitive as each individual possesses different emotional limits and affinities.

However, it is also worth noting that there exist similarities in the EEG signals of the same subject across various trials. From these two observations, we conclude that there is a scope for developing two separate models for studying the emotional states for the subjects, which is also validated by the availability of separate researches into the two models. A subject-dependent model would facilitate the recognition of the emotional state of a particular individual the model is trained on the same individual in prior. A subject-independent model will help determine the emotional state of a person whose previous EEG signals have never been recorded. These are two possible use cases of the two models.

The signals reported in the DEAP data set have a time duration of 63s (3s prior and 60s for the video). Although the 3s prior signal is not affected by the stimulus video, the signals displayed during the stimulus and the final value of emotion may have some correlation with the prior signals. So in this study, we do not remove prior signals before feature extraction. As previous studies suggest, we explore the frequency domain features where we extract the band power of different bands of the EEG signals. We use the Welch method with the Hanning window of 1s to determine Power Spectral Density as displayed in Figure 2. To obtain minute samples of information, we employ a stride of 0.25s for the entire signal. Thus, we obtain 249 band power values for different time instances. We use only the EEG signals for feature extraction and modelling to determine regions of brains responsible for a specific emotion.

Even though the anatomy of brains is similar for every human, but the consciousness of everyone is unique. This fact explains the diversity in the activations of brain, which is validated by the variance in EEG signals. So to train generalized models for the emotion recognition task, we require an extensive database. As the DEAP dataset reports EEG and physiological signals of 32 participants only, it is not currently possible to account for such diversity.

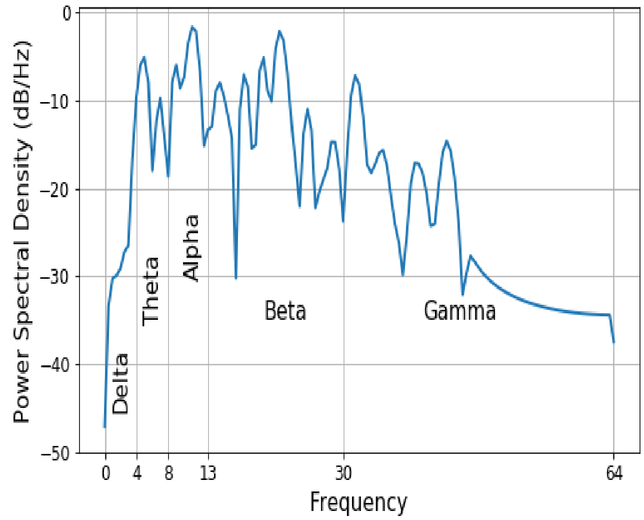

Figure 2. Power Spectral Density of EEG signal

This lack of database motivates to contrast the subject-dependent and subject-independent strategies to train the classifiers for emotion recognition task using EEG signals.

3.3 LSTM Model

In this paper, we test the LSTM network as an efficient tool for the prediction of an individual's emotion. In Natural Language Processing (NLP), the LSTM networks play an influential role in remembering long-term as well as short-term dependencies. The LSTM cell utilizes the following equations for employing the memory retention property.

$$\hat{c}^{<t>} = tanh(\, W_c\, [a^{<t-1>} : x^{<t>}] + b_c\,) \tag{1}$$

$$\Gamma_u = \sigma(W_u\, [a^{<t-1>} : x^{<t>}] + b_u\,) \tag{2}$$

$$\Gamma_f = \sigma(W_f\, [a^{<t-1>} : x^{<t>}] + b_f\,) \tag{3}$$

$$\Gamma_o = \sigma(W_o\, [a^{<t-1>} : x^{<t>}] + b_o\,) \tag{4}$$

$$c^{<t>} = \Gamma_u * \hat{c}^{<t>} + \Gamma_f * c^{<t-1>} \tag{5}$$

$$a^{<t>} = \Gamma_o * tanh(c^{<t>}) \tag{6}$$

Here, the input and output values of the LSTM cell for time t is $x^{<t>}$, and $a^{<t>}$. The symbols W and b represent Weights and Biases of different gates and candidate value for memory. Γ_u,Γ_f, Γ_o represent the Update, Forget, and Output gate for regulating the state of a LSTM cell. $c^{<t>}$ represents the memory of the cell while the $\hat{c}^{<t>}$ represents the candidate value for memory update. The following equations demonstrate the σ(x) and tanh(x) non-linear activation functions:

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{7}$$

$$tanh(x) = \frac{e^{x}+e^{-x}}{e^{x}-e^{-x}} \tag{8}$$

Since the previous activations of brain and events critically affect the subsequent activations of the brain, thus remembering past events may enable the LSTM model to gain insights from the patterns of EEG signals. Hence, we test the LSTM model for the emotion recognition task using the EEG signals. Figure 3 displays the best LSTM model configuration for subject-dependent training strategy. To ensure GPU support, we implement the LSTM

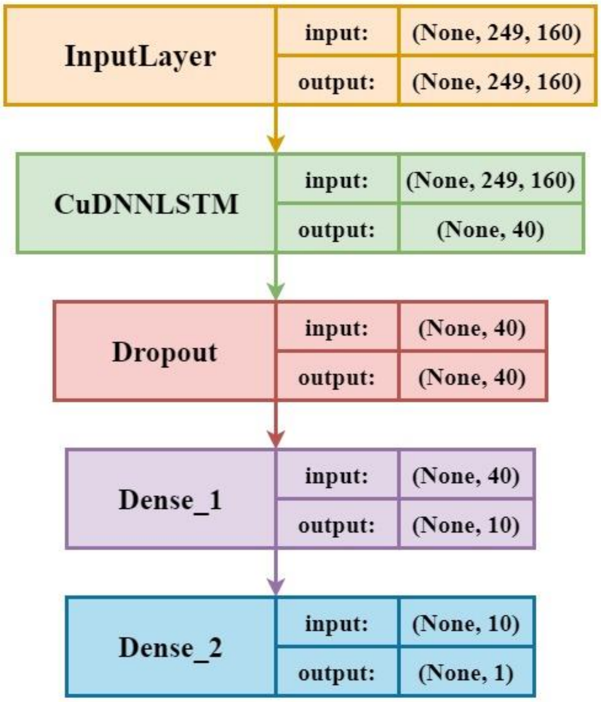

Figure 3. Proposed LSTM Model

model in Python 3 using Keras library with Tensorflow backend on the Google Colab platform. We experiment with several configurations of following LSTM model parameters:

- Number of LSTM Nodes: 32, 40
- Number of Dense 1 Nodes: 10, 20, 24, 32, 40
- Activation for Dense 1: 'relu', 'sigmoid', 'tanh'
- Learning Rate: 1, 1e-1, 1e-2, 1e-3, 1e-4
- Learning Rate Decay: 1e-5, 1e-6

We get the best testing accuracies for the model having 40 nodes in LSTM layer, 10 nodes in Dense 1 layer with 'tanh' activation function, and a single node with 'sigmoid' activation function in Dense 2 layer. To avoid over-fitting, we use 25% dropout between the LSTM Layer and Dense 1 Layer. This model minimizes the log-loss function using the Stochastic Gradient Descent (SGD) optimization algorithm. Following equation displays log-loss function:

$$L(y,\ \hat{y}) = -y\ log(\hat{y}) - (1-y)\ log(1-\hat{y}) \qquad (9)$$

Best results are obtained with the configuration of SGD with learning rate=$1 * e^{-2}$, learning rate decay constant=$1 * e^{-5}$, and momentum constant= 0.9.

4. RESULTS

To contrast the subject-dependent and subject-independent strategies, we examine the performances of different classifiers such as KNN, SVM, Decision Tree, Random Forest, and LSTM model. We train the LSTM model, and other classifiers on the similarly pre-processed dataset having band power features extracted from the EEG signals. The DEAP dataset reports the values of Valence and Arousal on a scale of 1-9; we use 5 as the threshold value for categorizing the low and high values. For determining the accuracy in prediction, we convert the predicted values of different classifiers into labels 0 (low) and 1 (high). Thus,

Table 2. Testing accuracies for Subject-Dependent and Subject-Independent models

Models	Subject-Dependent		Subject-Independent	
	Valence	Arousal	Valence	Arousal
KNN	86.03	79.64	70.86	68.36
SVM	76.56	72.66	**72.19**	**71.25**
Decision Tree	71.10	67.97	58.13	55.63
Random Forest	81.25	81.19	61.95	61.25
LSTM	**94.69**	**93.13**	70.31	69.53

we calculate the prediction accuracy using the equation:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (10)$$

Here, we determine *True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP)* values from the confusion matrix. We follow both the subject-dependent and subject-independent strategies by separately training and testing the classifiers. We test these classifiers several times to ensure that the results observed are significant. Table 2 highlights the average testing accuracies for both the subject-dependent and subject-independent training strategies.

On analyzing the results, we observe that for the subject-dependent strategy, the LSTM model outperforms the other classifiers by a large margin. Whereas in the subject-independent strategy, the SVM model performs better. In subject-dependent conditions, we observe a significant increment of about 16% and 18% for valence and arousal scales, respectively when comparing the performance of LSTM model with other classifiers. We obtain a maximum improvement in average testing accuracy of 18% for valence scale and 20% for arousal scale when comparing the SVM classifier with the LSTM model.

We also contrast our results with the results obtained by Yang et al. [14] following a similar experimental procedure with a parallel combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) model. Here, we observe a gain of 4% for valence and 2% for arousal in average testing accuracies for all the subjects. Proposed LSTM model for subject-dependent approach observes a significant increment of 9% and 7.5% in valence and arousal when compared with Alhagry et al. [1] and 14% and 19% in valence and arousal when compared with Xing et al. [13]. Figure 4 describes the average testing accuracies with positive and negative deviations for valence and arousal.

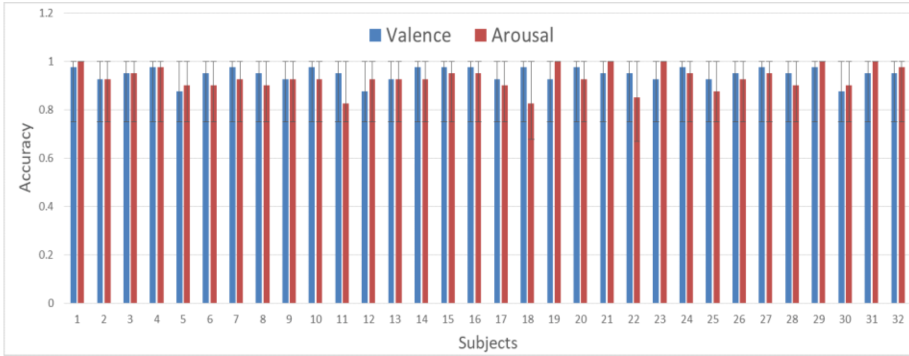

Figure 4. Testing accuracies for 32 subjects using LSTM model for subject-dependent model

For subject-independent conditions, we observe a drop in prediction accuracy of all classifiers. Nevertheless, the performance of the LSTM model drops most significantly, i.e. almost 24% for valence and arousal. This drop in performance indicates that the LSTM model requires an extensive data set for achieving generalization. For the subject-independent model, the best results were achieved by SVM, showing better performance by approximately 6% on the valence scale and 7% on the arousal scale over other classifiers.

5. CONCLUSION

In this work, we evaluate power spectral density over the 32 channels of EEG from the DEAP dataset. We decompose them into five bands of frequencies, Delta, Theta, Alpha, Beta, and Gamma to derive the band power of each band of frequency. We use this band power over each trial as a feature for classifying valence and arousal rating of the subject. We evaluate the performance of classification using KNN, SVM, Decision Tree, Random Forest, and LSTM as our classifiers and compare our results. We find an average increment of 16% for valence and 17% for arousal in the average testing accuracies for the subject dependent model by LSTM model over other classifiers. The maximum classification accuracy of 94.69% for valence and 93.13% for arousal was achieved using the LSTM classifier, which outperforms the state-of-the-art classifiers. For the subject-independent model, the best results were achieved by SVM, showing an average increment of approximately 6% on the valence scale and 7% on the arousal scale over other classifiers. Thus, to create a generalized Deep Learning models for excellent performance on emotion recognition, we require an extensive database.

For further work, we can extend the proposed experimental setup for real-time applications. For improvements further, we can add more features from EEG channels along with physiological features and test for their performance. A subset of the channels for feature generation, rather than using all the channels of EEG, may perform better in terms of accuracy as demonstrated by the works of Wichakam and Vateekul [12]. Other emotion models beyond the Valence-Arousal Model can also be addressed using the strategy used in this paper, like the 3-D emotion model used in the work of Dabas et al. [3].

6. REFERENCES

[1] Salma Alhagry, Aly Aly Fahmy, and Reda A El-Khoribi. 2017. Emotion Recognition based on EEG using LSTM Recurrent Neural Network.Emotion8, 10 (2017), 355–358.

[2] Margaret M Bradley, Mark K Greenwald, Margaret C Petry, and Peter J Lang. 1992.Remembering pictures: Pleasure and arousal in memory. Journal of experimental psychology: Learning, Memory, and Cognition18, 2 (1992), 379–390.

[3] Harsh Dabas, Chaitanya Sethi, Chirag Dua, Mohit Dalawat, and Divyashikha Sethia. 2018. Emotion Classification Using EEG Signals. In Proc. ACM Int. Conf. Computer Science and Artificial Intelligence. ACM, 380–384.

[4] Watson David and Tellegen Auke. 1985. Toward a consensual structure of mood. Psychological bulletin 98, 2 (1985), 219–235.

[5] Robert Jenke, Angelika Peer, and Martin Buss. 2014. Feature Extraction and Selection for Emotion Recognition from EEG.IEEE Trans. on Affective Computing 5, 3 (2014), 327–339.

[6] Sander Koelstra et al.2011. Deap: A Database for Emotion Analysis; Using Physiological Signals. IEEE Trans. Affective Computing 3, 1 (2011), 18–31.

[7] Zhenqi Li, Xiang Tian, Lin Shu, Xiangmin Xu, and Bin Hu. 2017. Emotion Recognition from EEG using RASM and LSTM. In International Conference on Internet Multimedia Computing and Service. Springer, 310–318.

[8] Yisi Liu and Olga Sourina. 2014. Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm. Springer Berlin Heidelberg, Berlin, Heidelberg, 199–223.

[9] Pallavi Pandey and K.R. Seeja. 2019. Subject independent emotion recognition from EEG using VMD and deep learning. Journal of King Saud University -Computer and Information Sciences (2019).

[10] Robert Plutchik. 1991. The Emotions. University Press of America.

[11] James A Russell. 1980. A circumplex model of affect. Journal of personality and social psychology39, 6 (1980), 1161–1178.

[12] Itsara Wichakam and Peerapon Vateekul. 2014. An evaluation of feature extraction in EEG-based emotion prediction with support vector machines. In Proc. IEEE Int. Conf. Joint conference on computer science and software engineering. IEEE, 106–110.

[13] Xiaofen Xing, Zhenqi Li, Tianyuan Xu, Lin Shu, Bin Hu, and Xiangmin Xu. 2019. SAE+LSTM: A New Framework for Emotion Recognition From Multi-Channel EEG. Frontiers in Neurorobotics13 (2019), 37.

[14] Yilong Yang, Qingfeng Wu, Ming Qiu, Yingdong Wang, and Xiaowei Chen. 2018. Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent

Neural Network. Proc. IEEE Int. Conf. Joint Conference on Neural Networks(2018)