



Emotion Recognition from Brain Signals While Subjected to Music Videos

Puneeth Yashasvi Kashyap Apparasu and S. R. Sreeja^(✉)

Department of Computer Science and Engineering, Indian Institute of Information Technology Sri City, Chittoor, Andhra Pradesh, India
{puneethyashasvi.a18,sreeja.sr}@iiits.in

Abstract. Emotions are simple, yet complex windows to the brain. Music and emotions are associated closely together. There are few things that stimulate the brain the way music does. It can be used as a powerful tool to regulate one's emotions. In recent years, emotion detection using brain waves has become an active topic of research. Various researchers are implementing different feature extraction techniques and machine learning models to classify the emotion by predicting the measurement of the electroencephalography signals. Many researchers are working on improving the accuracy of this problem and employing different techniques. In our study, we looked into achieving good scores by trying to predict the actual 4 emotional quadrants of the 2 dimensional Valence-Arousal plane. We evaluated and looked into various feature extraction approaches, modeling approaches and tried to combine best practices in our approach. We used the publicly available DEAP dataset for this study. Features from multiple domains were extracted from the EEG data and various statistical metrics and measures were extracted per channel. In our proposed approach, a one-dimensional convolutional neural network and a two-dimensional convolution neural network model were combined and fed through a neural network to classify the four quadrants of emotions. We did extensive and systematic experiments on the proposed approach over the benchmark dataset. The research findings that may be of significant interest to the user adaptation and personalization are presented in this study.

Keywords: Electroencephalography (EEG) · Emotion recognition · Music · Feature extraction · Classification

1 Introduction

Emotions are the purest forms of communication. They transcend all languages. What exactly is emotion? According to Merriam-Webster dictionary [1], emotion is a conscious mental reaction (such as anger or fear) subjectively experienced as a strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body. It is a state of mind and a response to a particular stimulus. There are six primary emotions namely

happy, sad, fear, anger, surprise, and disgust. There are many secondary and tertiary emotions. To understand them we can use the colour analogy where we can consider the primary emotions as the primary colours which when mixed forms secondary colours and tertiary colours. An emotion wheel is a good representation of all the different emotions.

Music has a high impact on the brain. Music can change one's ability to perceive time, improve mood, evoke memories, assist in repairing brain damage and much more. Music has the ability to evoke and change emotions. Basic human emotion can be expressed through facial expressions [18], eyes or body movement but it is difficult for an untrained eye to decide what the emotions are and how strong they are. Primarily, emotions and behaviors are nothing but communication between neurons in our brains. Apart from analysing brain waves, emotions can also be detected using pupillometry, facial expressions, heart rate. Discrete emotions like happiness and anger can also be detected by speech patterns and handwriting. In this work, our main focus is to identify the emotions using Electroencephalogram (EEG) brain signals.

There are billions of cells in our brain. Half of them are neurons, while the rest help and ease the activity of neurons. Synapses are gaps that separate two neurons. Neurons communicate by releasing chemicals that travel across these gaps. Activity in the synapses generates a faint electrical impulse known as the post-synaptic potential. When a large number of neurons activate at once, an electric field that is strong enough to spread through tissue and skull is generated [5]. Eventually, it can be measured on the scalp. This electrical impulse is then captured by the BCI devices. Apart from EEG, there are various BCI noninvasive devices like positron emission tomography or PET, functional near infrared or fNIR, Magnetoencephalography or MEG, functional resonance imaging or fMRI have been accepted and employed to measure the brain signals in both medical and non-medical contexts. Although there are many ways of identifying emotions, methods like EEG have shown potential in recent times [5]. EEG is an effective tool to record the electrical impulses our brain cells produce to communicate with each other. It has small metal disks or electrodes that are attached to the subjects head where each electrode records the electrical impulse of the corresponding part of the brain they are located at.

Generally, Emotions can be described in a two-dimensional space of valence and arousal. Valence is a positive or negative effect whereas arousal indicates how exciting or calming the stimulation is. Arousal arises from the reptilian part of our brain. It stimulates a fight or flight response that helps in our survival [7]. The objective of this study is to detect human emotions by inspecting and analyzing the brainwaves, which are generated by synchronised electrical pulses from a bunch of neurons communicating with each other. The data points collected from analyzing the brainwaves would then be used to classify the emotions using various techniques in our arsenal.

2 Literature Survey

Emotions have a key role in determining the stress level and they can interpret the positive emotional effect on an individual as per the study on the emotional effect of music therapy on palliative care cancer patients [15]. Various researchers are implementing different feature extraction techniques and machine learning models to classify the emotion by predicting the measurement of the psychological metrics. Many research works have been conducted recently on emotion recognition using EEG signals. Many studies tried constructing models by using different forms and aspects of the electrical signals represented by EEG and observed the results and reported the efficiency of their experiments.

From the literature, it is observed that some works used the time domain features of the EEG signal. Some works used other features from the frequency domain, wavelet domain, etc. There are also approaches that used features from multiple domains and fused them. In another work [18], the author used power spectral density (PSD) extracted from theta, alpha, beta and gamma bands as features and used them to train LSTM. In [22], the author used Gabor filter to extract the spatial and frequency domain features to train the SVM classifier. In [6], human emotion related features were extracted using multi-wavelet transformation. Normalized Renyi entropy, ratio of norms based measure and Shannon entropy were also used to measure the multi-wavelet decomposition of EEG signals. However, it is still a research problem, whether the extracted features are optimal features or not.

There are a good number of research papers available where human emotions were classified using different machine learning and deep learning algorithms [2, 3, 11, 12, 19–21]. In [3], to classify human emotions from EEG signals, the author used deep neural network and achieved an accuracy of 66.6% for valence, 66.4% for arousal classification. A recent work on EEG-based emotion detection system [2], used a Neucube-based SNN as a classifier, with an EEG dataset of fewer than 100 samples. They achieved an accuracy of 66.67% for valence classification and 69.23% for arousal classification. Even though, many researchers have made significant contributions and improvements in this field of research, most of the reported works used the classifiers for binary emotion classification problem.

3 Data and Method

This section will describe the EEG-based emotion dataset used in this research and then the proposed model for emotion classification.

3.1 Dataset Description

DEAP Dataset [13] is a multimodal dataset for the analysis of human affective states. The EEG and peripheral physiological signals of 32 participants were

recorded as each watched 40 one-minute long excerpts of music videos. Participants rated each video in terms of the levels of arousal, valence, like/dislike, dominance and familiarity. For 22 of the 32 participants, frontal face video was also recorded. There are two parts in the DEAP dataset: the ratings rated by 14–16 volunteers based on arousal, valence and dominance when subjected to 120 one-minute music video extracts in an online self-assessment and the physiological recordings and face video of an experiment where 32 volunteers watched a subset of 40 music videos. The actual duration of each EEG signal is of 63 s, which includes the preparation time of 3 s. After pre-processing, the sampling frequency is down-sampled to 128 Hz (Table 1). The 2D valence-arousal emotional space is given in Fig. 1.

Table 1. Characterization of the dataset.

No. of stimuli	40
Stimuli duration	60 s each
Rating scale	Valence, arousal, liking, familiarity
Number of participants	32
Recorded signal properties	32 channel 512 Hz EEG
Total data points	1280 data points

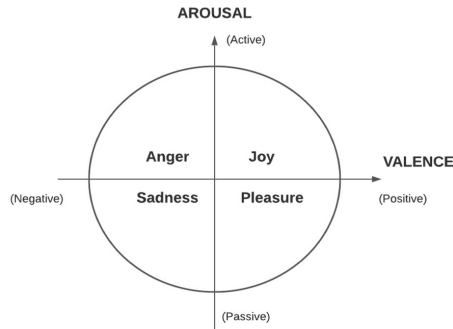


Fig. 1. The valence-arousal emotional plane.

3.2 Methodology

In this study, only the channels indicating EEG data were considered. A deep learning approach was used for emotion classification. The features from time domain, frequency domain, wavelet domain and statistical measures are extracted from EEG channels. For the target, the valence and arousal were separated into four emotions, namely *anger*, *joy*, *sadness* and *pleasure* as shown in Fig. 1. Both the multi-dimensional features and the statistical features are

passed through two networks, a one-dimensional convolutional neural network and a two-dimensional convolutional neural network which are later merged together with the help of a merge layer and output results were observed. The methodology followed is outlined in Fig. 2.

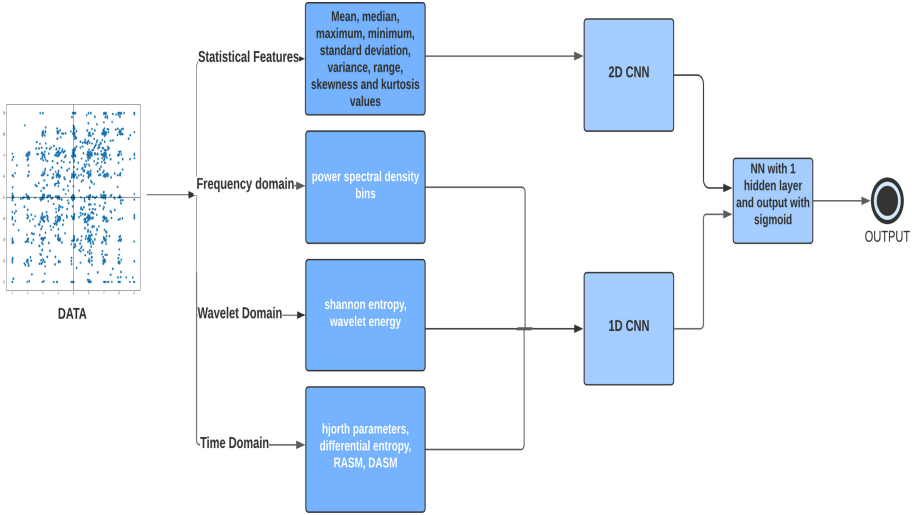


Fig. 2. The flowchart of the proposed methodology.

Channel Selection: The EEG dataset consists of 32 channels. Out of 32, the mandatory 16-channels of EEG are selected for further processing. The selected channels includes the *seven* EEG channels from the frontal lobe (Fp1, Fp2, F3, F4, F7, F8, Fz) that are responsible for the high level of cognitive, emotional, and mental functions; *seven* EEG channels from the parietal and occipital lobes (CP2, P3, P7, P8, PO3, Pz, O1) responsible for the auditory and visual information processing; and *two* EEG channels (C3, C4) from the central region were selected [8,12].

Feature Extraction

Frequency Domain: As EEG data is a time-series data, which proves to be difficult for an in-depth analysis. Frequency domain is helpful in doing much deeper analysis as it captures the changes in the EEG data. Hence, Short-Time Fourier Transform (STFT) was used as it returns a time-frequency distribution that specifies complex amplitude versus time and frequency for any signal [16]. STFT is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time [17]. In STFT, the signal is divided into shorter segments of equal lengths and fourier transform is performed on each segment. In our experiment, an STFT with a

hanning window and a segment length of 64 with 50% overlap was performed per channel of the EEG data and the power spectrum features were extracted. The power spectrum is divided into 5 bands: Delta (0–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–30 Hz), Gamma (30–45 Hz).

Time Domain: According to some researches, there is evidence that the asymmetry ratios between the left and right hemisphere of the brain affects the emotions [9]. Hence the differential asymmetry (DASM) and rational asymmetry (RASM) feature was calculated from the electrodes Fp1–Fp2 [12]. Hjorth parameters of activity, complexity and mobility and differential entropy of the signal were also calculated in the time domain.

Wavelet Domain: A wavelet is an oscillation, similar to a wave, which has an amplitude that starts at zero, rises, and comes back to zero. It is like a short oscillation similar to the recordings of a heart monitor or a seismograph. In general, wavelets are designed to have particular properties that make them beneficial for signal processing. As a mathematical tool, they can aid in extraction of information from a variety of data. Sets of wavelets are needed to inspect a data completely. A set of harmonious wavelets will decompose data in a stretch, making the decomposition process mathematically reversible [17]. Daubechies (*db1* and *db6*) wavelets which are usually used to decompose EEG signals, were used to transform the EEG data in our experiment. Daubechies wavelet with order 1 gave a better result in this use case. After performing discrete wavelet transform, the wavelet energy and Shannon entropy were extracted from the transformed data.

Statistical: Analogous to the methods in [20], *nine* statistical features, such as mean, median, maximum, minimum, standard deviation, variance, range, skewness and kurtosis were calculated for each channel data. Each EEG channel of 63 seconds has $63 * 128 = 8064$ data points, which was divided into 10 batches and then compressed to $(10 * 9) 90$ data points.

Therefore, for 32 users and 40 stimuli ($32 * 40 = 1280$), *three* Hjorth parameters, *sixteen* differential entropies, *one* wavelet energy, *one* wavelet entropy, RASM, DASM and *five* bands of power spectrum features and statistical features (1280, 90) were calculated.

Classification

As the dimensions of the statistical features and the multi-domain features are different, they cannot be passed through the same model without embedding or reshaping one of them. We did an experiment by reducing the dimensions using auto-encoder compression techniques and concatenating both the features, but the results were slightly on the lower end.

Hence, both the multi-domain features and the statistical features were taken as input to a one-dimensional convolutional neural network and a two-dimensional convolutional neural network and the outputs of these CNNs were

flattened and merged together by passing it through a layer that concatenates both the outputs. The steps followed by the classification model is shown in Fig. 3. The combined output is then passed through a two layer neural network that is analogous with logistic regression (LR), *i.e.*, one hidden layer and an output layer with sigmoid activation function which predicts if it is the said emotion or not. We did experiment with creating a single neural network that classified all the four emotions with a help of a softmax activation function, but the accuracy achieved was very less.

In order to improve the performance of the model, a stochastic gradient descent (SGD) with learning rate of 0.00001 with decay of $1e-6$ and momentum of 0.9 was used to reduce the binary cross entropy loss of the model. The hyper-parameters of the models were tuned with help of keras-tuner [14], where a search space was created for each hyper-parameter, like number of filters in a convolutional layer, kernel shape, number of dense units and different values. The obtained optimal values were substituted to the model and the best model was taken by comparing the validation accuracies.

4 Experimental Results

From the selected 16 channels, multi-domain features and statistical features have been extracted and it is fed as input to two-dimensional CNN and one-dimensional CNN, respectively. From the extracted data, 33% of data was used for validation and the rest for training. With an early stopping call back, the models were trained with the data corresponding to the emotion it has to predict. The hyper-parameters for the classification model are tuned with keras-tuner [14]. For each emotion, the obtained optimized hyper-parameter are listed in Table 2.

The best classification model was selected based on the validation accuracies. The best accuracy obtained are noted and it is listed in Table 3. Except Joy (high arousal-positive valence), other emotions achieved an accuracy above 75% and we got the maximum accuracy for the emotion - pleasure (low arousal-positive valence), which achieved an accuracy of $\approx 81\%$.

Most of the existing works on emotion prediction concentrate on different signal transformation methods or different machine learning techniques. As most of the work on this DEAP standard dataset is on binary classification problem. Hence, in the following, we have listed some of the works where they achieved good accuracy on classifying multi-class emotions. The work by T. B. Alakus *et al.* [4] gave accuracies of 75.00% for High Arousal Positive Valence; 96.00% for High Arousal Negative Valence; 71.00% for Low Arousal Positive Valence; 79.00% for Low Arousal Negative Valence, respectively. Another work by Y. Huang *et al.* [10] which gave 74.84% Anger; 71.91% Boredom; 66.32% Fear; 63.38% Joy; 69.98% Neutral; 72.68% Sadness. Our proposed approach was also able to produce better results in most of the emotions.

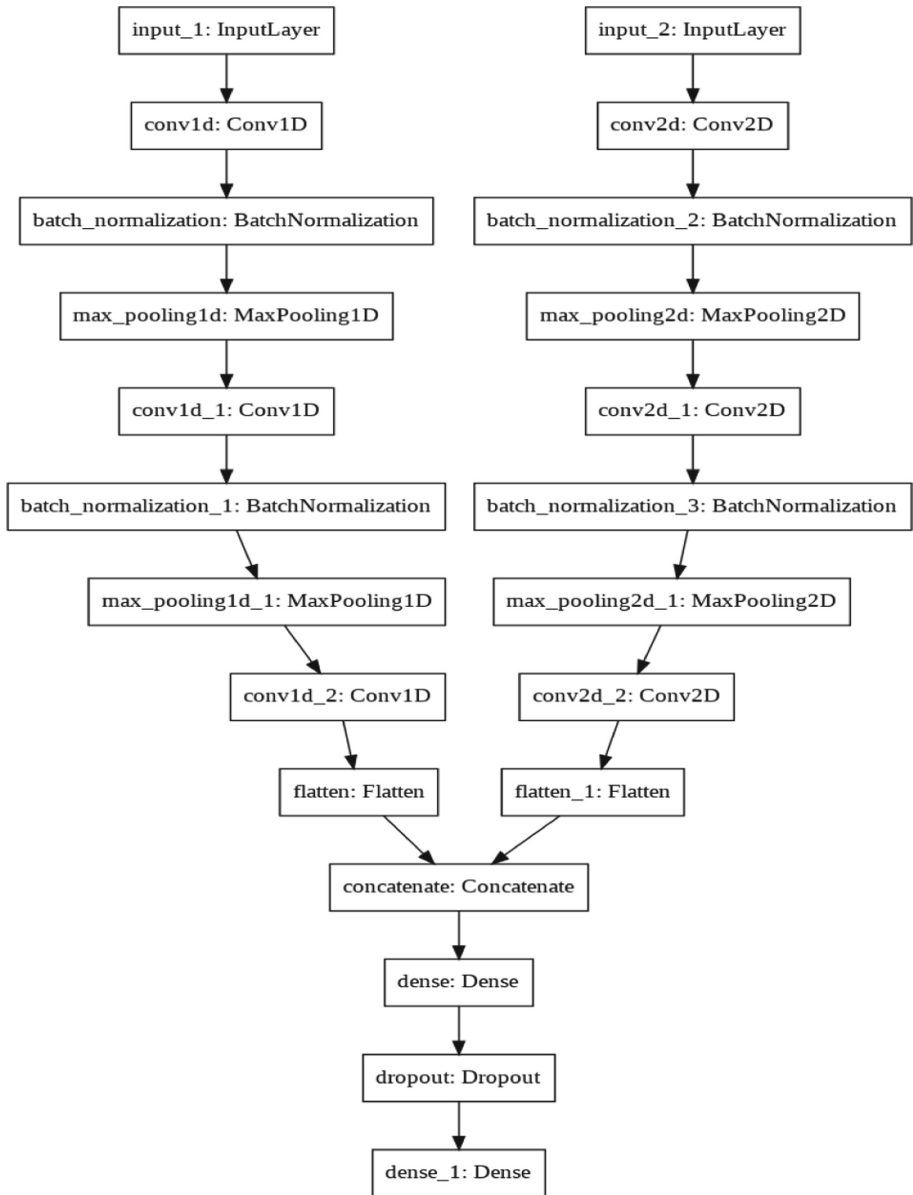


Fig. 3. The steps followed by the classification model.

Table 2. One dimensional CNN hyper-parameters after hyper-parameter tuning for the emotion detection models of Joy, Anger, Sadness and Pleasure (J, A, S, P)

Hyper-parameter	Value (J, A, S, P)
conv 1 filter	40, 224, 200, 64
conv 1 kernel	3, 3, 5, 3
conv 1 activation	ReLU, tanh, ReLU, ReLU
conv 2 filter	280, 424, 312, 416
conv 2 kernel	3, 3, 5, 3
conv 2 activation	tanh, ReLU, ReLU, ReLU
conv 3 filter	168, 488, 384, 400
conv 3 kernel	5, 3, 5, 3
conv 3 activation	ReLU, ReLU, ReLU, ReLU
conv_stats 1 filter	440, 384, 152, 192
conv_stats 1 kernel	3, 5, 3, 5
conv_stats 1 activation	tanh, tanh, ReLU, tanh
conv_stats 2 filter	240, 480, 312, 344
conv_stats 2 kernel	5, 3, 3, 5
conv_stats 2 activation	tanh, ReLU, tanh, ReLU
conv_stats 3 filter	264, 488, 176, 304
conv_stats 3 kernel	5, 3, 5, 4
conv_stats 3 activation	tanh, tanh, ReLU, ReLU
dense units	32, 128, 256, 32
dense activation	tanh, tanh, ReLU, ReLU

Table 3. Emotions and its corresponding Accuracy

Emotion	Accuracy
JOY (high arousal-positive valence)	68.79%
ANGER (high arousal-negative valence)	76.12%
SADNESS (low arousal-negative valence)	80.14%
PLEASURE (low arousal-positive valence)	81.32%

5 Conclusion

Unlike many approaches, in this work statistical features and the features extracted from time, frequency and wavelet domain were combined and used as inputs to a set one-dimensional convolution and two-dimensional convolution models which were combined in a later layer to give an output. From the output, we can see that the model was able to give better accuracy for pleasure compared to other emotions. Future work includes experiment to broaden the

emotion spectrum and improve the accuracies of the current emotions along with creating a *Music Recommendation system* that recommends music according to the emotions of a user.

References

1. Merriam-webster (n.d.). Emotion, in merriam-webster.com dictionary (2021). <https://www.merriam-webster.com/dictionary/emotion>
2. Al-Nafjan, A., Alharthi, K., Kurdi, H.: Lightweight building of an electroencephalogram-based emotion detection system. *Brain Sci.* **10**(11), 781 (2020)
3. Al-Nafjan, A., Hosny, M., Al-Wabil, A., Al-Ohali, Y.: Classification of human emotions from electroencephalogram (EEG) signal using deep neural network. *Int. J. Adv. Comput. Sci. Appl.* **8**(9), 419–425 (2017)
4. Alakus, T.B., Gonen, M., Turkoglu, I.: Database for an emotion recognition system based on EEG signals and various computer games-GAMEEMO. *Biomed. Signal Process. Control* **60**, 101951 (2020)
5. Arvaneh, M., Tanaka, T.: Brain-computer interfaces and electroencephalogram: basics and practical issues. *Signal Process. Mach. Learn. Brain-Mach. Interfaces* (2018)
6. Bajaj, V., Pachori, R.B.: Detection of human emotions using features based on the multiwavelet transform of EEG signals. In: Hassanien, A.E., Azar, A.T. (eds.) *Brain-Computer Interfaces*. ISRL, vol. 74, pp. 215–240. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-10978-7_8
7. Birkett, A.: Valence, arousal, and how to kindle an emotional fire (2020). <https://cxl.com/blog/valence-arousal-and-how-to-kindle-an-emotional-fire/>
8. Choi, E.J., Kim, D.K.: Arousal and valence classification model based on long short-term memory and DEAP data for mental healthcare management. *Healthc. Inform. Res.* **24**(4), 309–316 (2018)
9. Davidson, R.J.: Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn.* **20**(1), 125–151 (1992)
10. Huang, Y., Tian, K., Wu, A., Zhang, G.: Feature fusion methods research based on deep belief networks for speech emotion recognition under noise condition. *J. Ambient. Intell. Humaniz. Comput.* **10**(5), 1787–1798 (2017). <https://doi.org/10.1007/s12652-017-0644-8>
11. Islam, M.R., et al.: Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques. *IEEE Access* **9**, 94601–94624 (2021)
12. Khateeb, M., Anwar, S.M., Alnowami, M.: Multi-domain feature fusion for emotion classification using DEAP dataset. *IEEE Access* **9**, 12134–12142 (2021)
13. Koelstra, S., et al.: DEAP: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **3**(1), 18–31 (2011)
14. O'Malley, T., et al.: Kerastuner (2019). <https://github.com/keras-team/keras-tuner>
15. Ramirez, R., Planas, J., Escude, N., Mercade, J., Farriols, C.: EEG-based analysis of the emotional effect of music therapy on palliative care cancer patients. *Front. Psychol.* **9**, 254 (2018)
16. DSP Related: The short-time fourier transform. https://www.dsprelated.com/freebooks/sasp/Short_Time_Fourier_Transform.html

17. Sejdić, E., Djurović, I., Jiang, J.: Time-frequency feature representation using energy concentration: an overview of recent advances. *Digit. Signal Process.* **19**(1), 153–183 (2009)
18. Soleymani, M., Asghari-Esfeden, S., Fu, Y., Pantic, M.: Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* **7**(1), 17–28 (2015)
19. Teo, J., Hou, C.L., Mountstephens, J.: Deep learning for EEG-based preference classification. In: *AIP Conference Proceedings*, vol. 1891, p. 020141. AIP Publishing LLC (2017)
20. 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: *Twenty-Ninth IAAI Conference* (2017)
21. Yin, Y., Zheng, X., Hu, B., Zhang, Y., Cui, X.: EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Appl. Soft Comput.* **100**, 106954 (2021)
22. Zamanian, H., Farsi, H.: A new feature extraction method to improve emotion detection using EEG signals. *ELCVIA: Electron. Lett. Comput. Vis. Image Anal.* **17**(1), 29–44 (2018)