

Study on an effective cross-stimulus emotion recognition model using EEGs based on feature selection and support vector machine

Shuang Liu¹ · Jingjing Tong¹ · Jiayuan Meng¹ · Jiajia Yang¹ · Xin Zhao¹ · Feng He¹ · Hongzhi Qi¹ · Dong Ming¹

Received: 28 October 2015 / Accepted: 19 September 2016 / Published online: 27 September 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Electroencephalographic (EEG)-based emotion recognition has received increasing attention in the field of human–computer interaction (HCI) recently, there however remain a number of challenges in building a generalized emotion recognition model, one of which includes the difficulty of an EEG-based emotion classifier trained on a specific stimulus to handle other stimuli. Little attention has been paid to this issue. The current paper is to study this issue and determine the feasibility of coping with this challenge using feature selection. 12 healthy volunteers were emotionally elicited when watching the video clip. Power spectral density (PSD) and brain asymmetry (BAY) were extracted as EEG features. Support vector machine (SVM) classifier was then examined under within-stimulus conditions (samples extracted from one video were sent to both training set and testing set) and cross-stimulus conditions (samples extracted from one video were merely sent to one set, training set or testing set alternatively). The within-stimulus 5-class classification performed fairly well (accuracy: 93.31 % for PSD and 85.39 % for BAY). Cross-stimulus classification, however, deteriorated to low levels (46.22 % and 46.2 % accordingly). Trained and tested with the most robust feature subset selected by SVM-recursive feature elimination (RFE), the mean 5-class performance of cross-stimulus classifier was significantly improved to 68.89 and 64.44 % for PSD and BAY respectively. These

results suggest that cross-stimulus emotion recognition is reasonable and feasible with proper methods and brings EEG-based emotion recognition models closer to being able to discriminate emotion states in practical application.

Keywords Emotion recognition · Electroencephalographic (EEG) · A generalized classifier · Support vector machine (SVM)

1 Introduction

Emotion is a psycho-physiological process triggered by the conscious and/or unconscious perception of an object or a situation, which is often associated with mood, temperament, personality disposition and motivation [12]. Recently emotion recognition has received increasing attention in the field of human–computer interaction (HCI), there is evidence that if machines could understand a person’s emotional state when interacting with people, HCI may become more intuitive, smoother, and more efficient. Additionally, negative emotions, such as depression, anxiety, and chronic anger, have been shown to impede the work of the immune system, making people more vulnerable to viral infections, and slowing healing from surgery or disease [16], even affecting the people’s performance and efficiency severely.

To date, various physiological measures have been used to estimate emotional states, including electroencephalogram (EEG), electromyogram signal (EMG), respiratory volume, skin temperature, skin conductance and heart rate and so forth [10–12]. Among them, EEG-based emotion recognition has caught the most attention since EEGs could directly reflect emotional states with high temporal resolution, and EEGs tend to be less mediated by cognitive and

✉ Jiajia Yang
jiajia.yang@tju.edu.cn

✉ Dong Ming
richardming@tju.edu.cn

¹ Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China

social influences. More importantly, the signals are from the central nervous system which is where emotions originate. Recently, numerous studies were conducted to measure the brain emotional states by analyzing EEGs under the emotional stimuli that have occurred. Moreover, many different materials have been used to elicit emotions in the laboratory such as facial expressions [5], pictures [15], texts [1], music [11], and movies [19]. Specially, affective pictures, music and videos are the three most popular evoked stimuli, and got some acceptable recognition rates. Yohanes et al. [20] proposed discrete wavelet transform coefficients from EEGs in response to emotional pictures for emotion recognition, achieving a maximum accuracy of above 84.6 % for two emotional states. Koelstra et al. [13]. presented some results in classification of emotions induced by watching music videos, an average (maximum) classification rate of 55.7 % (67.0 %) for arousal and 58.8 % (76.0 %) for valence was obtained for EEG. For the emotional features, a number of these studies have focused on the question of asymmetrical activation of the cerebral hemisphere in the past few decades. And asymmetrical activation was repeatedly reported to be a good indicator in distinguish some specific emotions [3, 7, 17]. Hidalgo-Munoz et al. showed that the left temporal region has revealed to play an important role in the affective valence processing [8], and Baumgartner et al. [3] detected a pattern of greater EEG activity over the left hemisphere in the happy condition compared to negative emotional conditions. Additionally, spectral power in various frequency bands was also proven to be a distinguishable emotion indicator [2, 18].

Nearly all existing studies tend to provide evidence for the feasibility of using EEG measures in the monitoring of emotional states. However, recorded EEGs induced by the evoked stimulus contain not only the emotion-relative information, but also emotion-irrelative information, such as those related to the information processing delivered by the evoked stimulus. In the pattern recognition theory, parts of samples are sent to build the classifier, and the rest are used to test the classifier, called training set and testing set. If the samples extracted from one evoked stimulus (such as a video) are sent to both training set and testing set, this may bring about a potential problem, that is, shared emotion-irrelative information in the samples from one video would help the classifier recognize the samples more easily, resulting in an inflated classification accuracy. This was ignored in the previous studies. This paper was to cope with the problematic effects of stimulus-to-stimulus variations, and then tried to find out whether the performance of cross-stimulus emotion recognition can be improved with proper feature selection.

This paper is organized as follows. Section 2 addresses the methodology, including the experimental setup and

data analysis. Section 3 details the results analysis. The conclusion is stated in Sect. 4.

2 Materials and methods

2.1 Experimental setup

2.1.1 Subjects

A group of 12 healthy participants (6 female, 6 male, 20–26 years) were enrolled in this study. They are undergraduate students and postgraduate students in Tianjin University. All participants had normal or corrected-to-normal vision and normal hearing, and none of them had a history of severe medical treatment, psychological or neurological disorders. A signed consent was obtained from each subject before the experiment was carried out.

2.1.2 Emotional elicitation

In this study, we used the video induction method, that is, to record EEG signals while the subjects were watching different pieces of video clips to experience five emotional states of happy (H), neutral (N), tense (T), sad (S) and disgust (D) states. Eliciting emotional reactions from subjects is a difficult task and selecting the most effective stimulus materials is crucial. Before the experiment, 58 subjects who did not take part in the experiment participated in a questionnaire survey to verify the effectiveness of these elicitors. Finally, 15 of 72 movie clips were selected with 3 clips for each emotional state, respectively. The experimental procedure was depicted in Fig. 1. The experiment consisted of 15 sessions, and in each session, a movie clip was displayed for about 5 min, preceded by a 5 s red circle as the hint of start. At the end of each clip, the subjects were required to rate valence, arousal and the specific emotion they had experience during movie viewing. Each session was followed by a short break, and the recordings took place whenever the subject was ready to watch the next video.

For the experiment, a quiet listening room was prepared in order to ensure that the subjects would not be disturbed to experience the emotions evoked by the videos. Prior to the experiment, each participant was informed of the experiment protocol and the meaning of the different scales

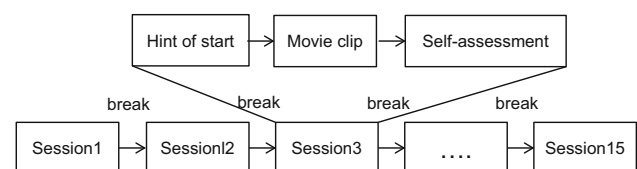


Fig. 1 The experimental procedure

used for self-assessment. They were allowed to be familiar with the task by an unrecorded preliminary experiment in which a short video was shown which would not appear in the real experiment. The subject was seated approximately 1 m from the screen. During the task, the StereoPhilips speakers were used and the video volume was set at a loud but comfortable level.

2.1.3 Data acquisition

During the experiment, 30-channels EEG signals were recorded continuously using a Neuroscan 4.5 amplifier system. The electrodes were placed on the scalp according to the extension of the international 10–20 electrode positioning system [14]. And Fig. 2 shows 30-channels EEG cap layout used in this study. All channels were referenced to right mastoid and grounded central region. The signals were digitized at 1000 Hz and stored in a PC for offline analysis.

2.2 Data processing

2.2.1 Preprocessing

Prior to calculating features, a preprocessing stage of the EEG signals is required. All channels were re-referenced to bilateral mastoid, and down-sampled to 128 Hz. EOG artifacts were removed using independent component analysis (ICA). Valid data were picked out according to the subjects' self-report about which period of time they felt the emotion strongly. EEG data were then split into 5-s, non-overlapping epochs in the following step of feature extraction.

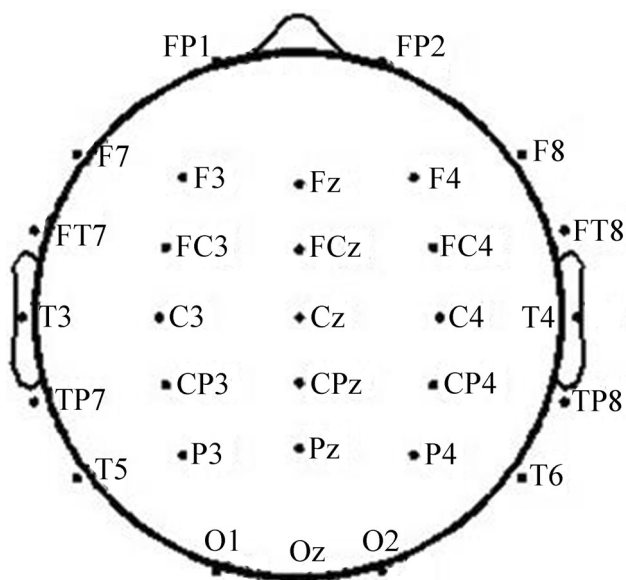


Fig. 2 EEG cap layout of 30 channels

2.2.2 Feature extraction

Power spectral density (PSD) was estimated using Burg's method (the order of an autoregressive model was 8). The sums of PSD in 6 frequency bands (θ : 4–8 Hz, α : 8–13 Hz, β_1 : 13–18 Hz, β_2 : 18–30 Hz, γ_1 : 30–36 Hz, γ_2 : 36–44 Hz) were extracted, which contributed to 6 features of each channel for all 30 channels, that is, 180 (6 per channel \times 30 channels) features were included in each feature set.

An asymmetry index (AI) representing relative right versus left sided activation was also used. The relative difference between the hemispheres may play an important role in the emotion recognition. AI was defined as follows:

$$AI(i) = \frac{P_L^i}{P_L^i + P_R^i}$$

P_L^i and P_R^i here represent the power in the left and right hemispheres for the i th pair of channels respectively. The value, larger than 0.5, indicates greater activities in the left than in the right hemisphere and vice versa. 12 pairs of channels (FP1–FP2, F7–F8, F3–F4, FT7–FT8, FC3–FC4, T3–T4, C3–C4, TP7–TP8, CP3–CP4, P3–P4, T5–T6, O1–O2) were used in the section of pattern recognition.

2.2.3 Pattern classification

SVM is a supervised learning algorithm which uses a discriminant hyperplane to identify classes. The goal of SVM is to find an optimal hyperplane with the maximal margin between two classes of data [14]. In the emotion recognition process, the features were mapped using Gaussian radial basis function into high-dimensional kernel space.

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

The penalty parameter C was set to 1 as the default value in the LibSVM toolkit developed by Chin-Jen Lin [4].

In this paper, we used two strategies to train and test SVMs for each participant. “WS” labelled the within-stimulus condition, in which the samples from one video were sent to both training and testing sets, while “CS” labelled the cross-stimulus condition, in which the samples from one video were merely sent to a set, training set or testing set alternatively.

2.2.4 Feature optimization

Since not all the features carry significant information, features selection is necessary for decreasing and discarding redundant features that can potentially deteriorate classification performance. SVM-RFE was

proposed by Guyon et al. [6]. and was based on the concept of margin maximization. In this case, however, the ranking criterion was modified as follows: In an N-dimension feature set, each feature was removed once and then got N performances with the remaining N-1 features. The feature, without which the feature set obtained the best accuracy, was considered as the one with the minimum contribution. At each iteration step we remove the feature with the minimum contribution from the feature set until only one feature remained. The features were removed one at a time, and there was a corresponding feature ranking. But it should be noted that the features that are top ranked are not necessarily the ones that are individually most relevant. Only taken together the features of a subset are optimal in some sense.

3 Results

3.1 Classification rates using different strategies

Figure 3 presented the 5-class classification rates for two strategies (WS, CS) based on different feature sets (PSD, BAY). The individual accuracy and the mean accuracy were shown. Obviously, mean within-stimulus (WS) performances were 93.31 and 85.39 % when using PSD and BAY features. But mean cross-stimulus (CS) performances were just 46.22 and 46.2 % accordingly, that is, the decrease of classification performance occurred when training sets and testing sets come from different emotional videos for emotion recognition. The results suggested within-stimulus emotion recognition may inflate the classification accuracies, this may because of shared emotion-irrelevant EEG features in samples from one video, such as content information processing delivered by the video.

3.2 Cross-stimulus classification results using SVM-RFE

Cross-stimulus classification performances may suffer from some effects, such as EEG patterns induced by specific stimulus, mismatched emotional intensity between training and testing stimulus and temporal effects. It was expected that SVM-RFE did at least pick out the most emotion-relevant features and improve the classification accuracies. Figure 4 shows the mean cross-stimulus classification rates using SVM-RFE. It can clearly be seen that SVM-RFE can significantly improve the mean accuracies and recognition rates jumped to average accuracy of 68.89 and 64.44 % based on the feature set of PSD and BAY, respectively.

Confusion matrices in Table 1 afforded a closer look at the sensitivity of five emotional states (happy, neutral, disgust, sad, tense). In these confusion matrices, the row represents the classified label and each column represents the true label. The tense state was better recognized, 82.6 and 75.9 % for PSD and BAY feature sets respectively. But disgust state was wrongly recognized with a relative higher proportion. The relatively lower accuracy of disgust state may be related to effect of cognitive avoidance: highly disgust-sensitive individuals avoided the formation of a mental representation of an aversive scene [9].

3.3 Finding the best emotion-relevant EEG features

Another point that should be addressed is finding the best emotion-relevant EEG frequency ranges. The features that were repeatedly selected by classifiers yielding good classification rates could be considered more robust than those that were rarely selected. The feature subset obtaining the best accuracy was regarded as the most salient feature set. The results of the automatic feature selection using RFE are presented in Fig. 5, which shows the contribution rate (CR) of each EEG frequency band to the most relevant

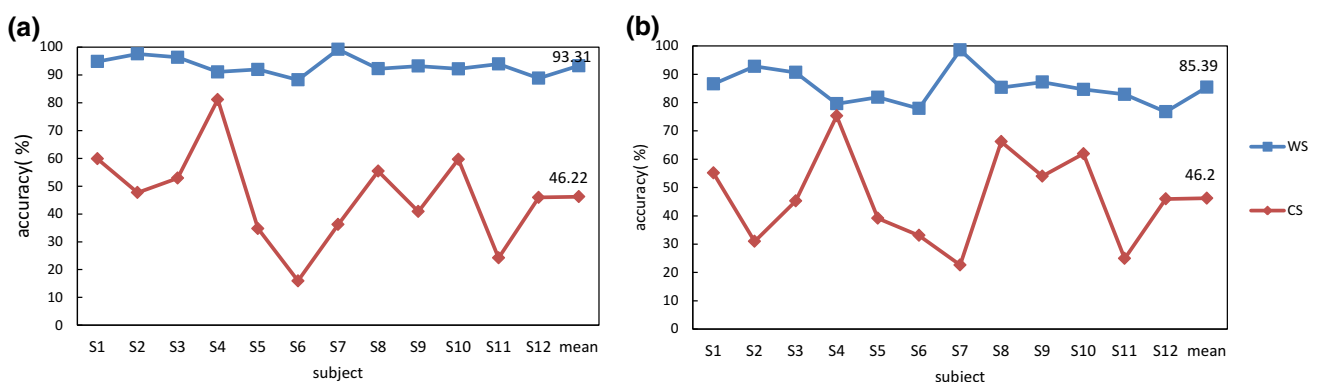


Fig. 3 5-class classification rates for two strategies (WS, CS) based on different feature sets of **a** PSD and **b** BAY respectively

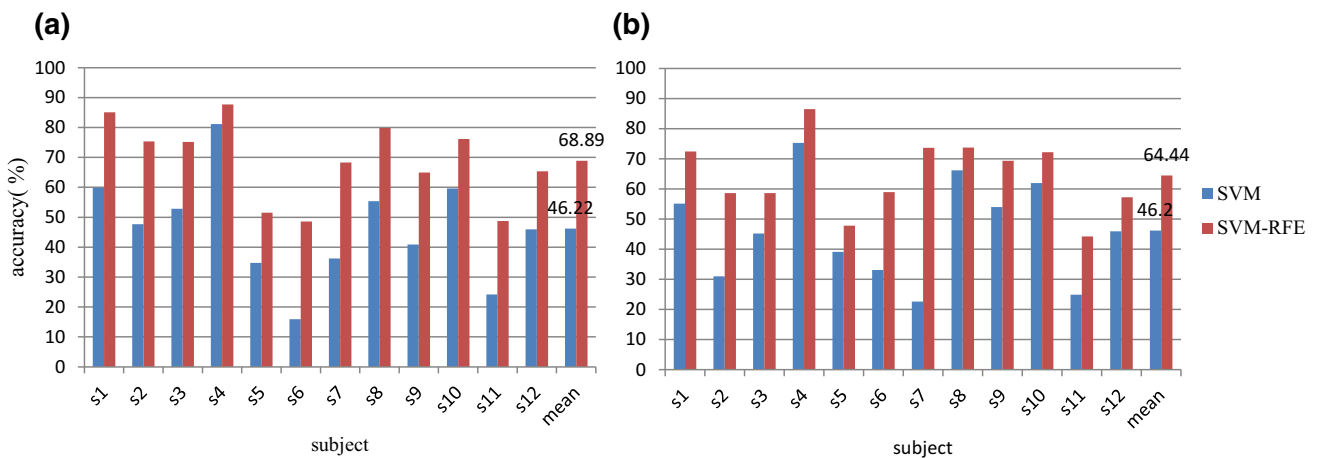


Fig. 4 Cross- stimulus classification using SVM-RFE based on the feature sets of **a** PSD and **b** BAY respectively

Table 1 Classification results achieved by using SVM-RFE

Acc (%)	PSD feature set					BAY feature set				
	Happy	Neutral	Disgust	Sad	Tense	Happy	Neutral	Disgust	Sad	Tense
Happy	0.617	0.078	0.064	0.086	0.153	0.555	0.136	0.107	0.079	0.122
Neutral	0.097	0.650	0.036	0.050	0.167	0.035	0.682	0.096	0.095	0.091
Disgust	0.178	0.148	0.488	0.111	0.075	0.191	0.141	0.413	0.193	0.061
Sad	0.036	0.065	0.020	0.745	0.134	0.012	0.118	0.048	0.640	0.181
Tense	0.065	0.019	0.068	0.022	0.826	0.051	0.05	0.039	0.101	0.759

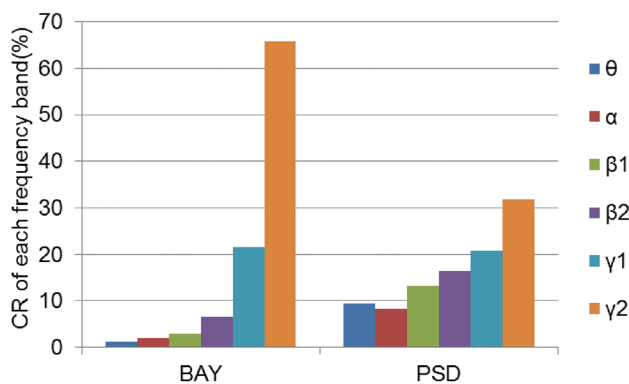


Fig. 5 Contribution rate (CR) of each frequency band to the features selected by RFE in cross-stimulus classification

features selected by RFE for both feature subsets in cross-stimulus classification. It could be found that the higher frequency band roughly occupied the larger contribution part. Notably, the contribution rate of γ_2 is as high as approximately 65 % when BAY was employed.

4 Discussion

It is encouraging to use the neurophysiological measurement in the emotion recognition system. The emotion recognition system should be able to accurately recognize the emotional

state independent of the type of inducing material the operator would encounter. This is because in a real-world application, stimuli are most likely to be different between the classifier construction and the emotion recognition.

Most previous studies used within-stimulus classification during emotion recognition. However, neglecting the shared non-emotional information of the induced material would falsely inflate the classification rates. This paper has tested the overinflated effect of within-stimulus recognition. To our knowledge, the recorded EEG responses contain not only the information related to the emotional state, but also others related to the basic sensory information processing and information processing delivered by the stimulus material which may play a significant role in the classification. Shared non-emotional information in the samples from the same inducing material would make the classifier easier to recognize the testing sample accurately. So the accuracy was significantly decreased once the cross-stimulus method employed.

For within-stimulus, the PSD features look better than BAY features for emotion recognition, while the two features corresponded to similar recognition accuracies for cross-stimulus, that is, more serious overinflated effect would be obtained when PSD features were employed. This may be partly related to the number of features in each feature set. 180 and 72 features were got for PSD and BAY feature sets respectively. Some channels were contained in

PSD feature set but not in BAY feature set, such as Cz, Oz and so on. The shared non-emotional information in these channels may contribute to more serious overinflated accuracies.

In addition to the importance in the emotion field, studies on overinflated accuracies may also be necessary in other fields related to the extremely complex cognitive processes such as mental workload, mental fatigue and vigilance. Though it is still a tremendous problem in EEG-based emotion recognition, these results provide a promising solution and take EEG-based model one step closer to being able to discriminate emotions in practical application.

5 Conclusion

The results of the current study indicate that within-stimulus emotion recognition would inflate the classification accuracies, and cross-stimulus performances can be improved by the optimized feature subset chosen by the feature selection method. Instead of a stimulus-specific classifier, the cross-stimulus classifier model was built to handle other stimulus. The current study demonstrated the feasibility of a model trained on one stimulus to handle another and took the first step towards a generalized model that can handle a new stimulus.

Acknowledgments This research was supported by National Natural Science Foundation of China (No. 81222021, 61172008), National Key Technology R&D Program of the Ministry of Science and Technology of China (No. 2012BAI34B02) and Program for New Century Excellent Talents in University of the Ministry of Education of China (No. NCET-10-0618). The authors sincerely thank all subjects for their voluntary participation.

References

- Alm CO, Roth D, Sproat R (2005) Emotions from text: machine learning for text-based emotion prediction. Paper presented at: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (Association for Computational Linguistics)
- Balconi M, Lucchiari C (2006) EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. *Neurosci Lett* 392:118–123
- Baumgartner T, Esslen M, Jäncke L (2006) From emotion perception to emotion experience: emotions evoked by pictures and classical music. *Int J Psychophysiol* 60:34–43
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *Acm Trans Intell Syst Technol* 2:389–396
- Edwards J, Jackson HJ, Pattison PE (2002) Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clin Psychol Rev* 22:789–832
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
- Henriques JB, Davidson RJ (1991) Left frontal hypoactivation in depression. *J Abnorm Psychol* 100:535
- Hidalgo-Muñoz A, López M, Pereira A, Santos I, Tomé A (2013) Spectral turbulence measuring as feature extraction method from EEG on affective computing. *Biomed Signal Process Control* 8:945–950
- Köchel A, Plichta MM, Schäfer A, Leutgeb V, Scharmüller W, Fallgatter AJ, Schienle A (2011) Affective perception and imagery: a NIRS study. *Int J Psychophysiol* 80:192–197
- Khalili Z, Moradi M (2008) Emotion detection using brain and peripheral signals. Paper presented at: Biomedical Engineering Conference, 2008 CIBEC 2008 Cairo International (IEEE)
- Kim J, André E (2008) Emotion recognition based on physiological changes in music listening. *Pattern Anal Mach Intell IEEE Trans* 30:2067–2083
- Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) Deap: a database for emotion analysis; using physiological signals. *Affect Comput IEEE Trans* 3:18–31
- Koelstra S, Yazdani A, Soleymani M, Mühl C, Lee J-S, Nijholt A, Pun T, Ebrahimi T, Patras I (2010) Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos. In: Yao Y, Sun R, Poggio T, Liu J, Zhong N, Huang J (eds) *Brain informatics*. Springer, Berlin, Heidelberg, pp 89–100
- Lotte F, Congedo M, Lécuyer A, Lamarche F (2007) A review of classification algorithms for EEG-based brain–computer interfaces. *J Neural Eng* 4:R1–R13
- Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Anal Mach Intell IEEE Trans* 27:1226–1238
- Picard RW, Vyzas E, Healey J (2001) Toward machine emotional intelligence: analysis of affective physiological state. *Pattern Anal Mach Intell IEEE Trans* 23:1175–1191
- Tomarken AJ, Davidson RJ, Wheeler RE, Kinney L (1992) Psychometric properties of resting anterior EEG asymmetry: temporal stability and internal consistency. *Psychophysiology* 29:576–592
- Verma GK, Tiwary US (2014) Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. *Neuro Image* 102:162–172
- Wang X-W, Nie D, Lu B-L (2014) Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129:94–106
- Yohanes RE, Ser W, Huang GB (2012) Discrete Wavelet Transform coefficients for emotion recognition from EEG signals. Paper presented at: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (IEEE)