Biologically Inspired Cognitive Architectures 24 (2018) 98–106

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

EEG based emotion classification using "Correlation Based Subset Selection"

Debashis Das Chakladar[a,*], Sanjay Chakraborty[b]

[a] Computer Science and Engineering Department, Institute of Engineering and Management, Kolkata, India
[b] Department of information Technology, Techno India, Kolkata, India

ARTICLE INFO

Keywords:
Electroencephalography (EEG)
Emotion classification
Human machine interaction
Machine learning
Brain–computer interface (BCI)
Higher Order Statistics (HOS)

ABSTRACT

Emotion detection is one of the popular research topics in "Brain–Computer Interfacing" where researchers are trying to find the various emotional states of people. EEG signal is widely used for detecting different categories of emotions. The EEG signal is captured through multiple electrode channels, very few of them are useful for emotion detection. In our paper, a "Correlation-based subset selection" technique is introduced for dimension reduction. Then we proceed with classification process using "Higher Order Statistics" features of the reduced set of channels. However, we have classified four classes of emotions (positive, negative, angry and harmony) in our paper. The execution time of our proposed algorithm is O(n^2 + 2n). The classification accuracy of this model with the reduced set of channels is 82%. Finally, we compare our proposed model with some popular emotion classification models and the result shows that our model substantially outperforms all the previous models. However, the proposed model helps physically disabled people to express their feelings with minimum time and cost-effectively.

1. Introduction

In real-life, emotion plays an important role in the human–human interaction. Considering the evolving nature of the machine, we look forward human–machine interaction through BCI nowadays. Brain-Computer Interface (BCI) is an interface between brain and outside world without the direct intervention of any muscular activity. BCI technology has mainly consisted of four sections, such as Signal acquisition, Signal processing, Feature extraction and classifications, Application Interface. BCI has several applications like P300 speller, wheelchair control, robotic arm movement, and cursor movement. In our previous paper (Chakladar & Chakraborty, 2017), we had proposed an efficient algorithm of cursor movement to reach the desired target in minimum time. For building an effective human–machine interaction system like Brain-computer interfacing, the most prerequisite is to develop an efficient emotion recognition system (Wang, Nie, & Lu, 2014). According to the two-dimensional model of emotion described by Davidson, Schwartz, Saron, Bennett, and Goleman, 1979, emotion is represented in two-dimensional space (arousal and valence) as shown in Fig. 1. Valence represents the quality of emotions from negative to positive due to higher frontal activity in alpha power, whereas arousal represents quantitative excitation from passive to the active emotional state of higher beta power in the parietal lobe. The recognizability of emotions depends on how well the EEG features are mapped to the arousal-valence 2D model (Bos, 2006). Emotion can be explained in many ways (1) visual (images/pictures), (2) audio-visual (clips/video clips) and (3) audio (songs/sounds) (Murugappan, Ramachandran, & Sazali, 2010) etc. Various studies show that peripheral physiological signals like Electrocardiogram (ECG), Skin Conductive Resistance (SCR), and Blood Volume Pressure (BVP) can also change the emotions (Picard, 2000). Davidson and Fox, 1982 suggested that frontal brain activity is related to positive and negative emotions. A probabilistic classifier and "perceptron convergence" algorithm are used for emotion classification (Yoon & Chung, 2013). Previously, several works have been done on dimension reduction of the input features, but this paper focuses on a new dimension reduction technique (Correlation-based subset selection) of input channels and classifies those input based on four different types of emotions (positive, negative, depressed and harmony). For the sake of simplicity, here we choose four important emotions from the arousal-valence model. We select the positive, negative, angry and harmony emotion from high arousal-high valence (HAHV), low arousal-low valence (LALV), high arousal-low valence (HALV) and low arousal-high valence (LAHV) region of the arousal-valence model (Fig. 1) respectively. After performing dimension reduction process, we select emotion-specific channels for each class which gives a true prediction rate over the test data that leads to a good classification accuracy. The rest of the paper is organized as follows. In Section 2, we have described the overall flowchart of our proposed work and the algorithm of dimension reduction technique. In Section 3, we analyze the proposed algorithm and perform emotion classification

* Corresponding author.
E-mail addresses: ddaschakladar@gmail.com (D.D. Chakladar), schakraborty770@gmail.com (S. Chakraborty).

https://doi.org/10.1016/j.bica.2018.04.012
Received 18 March 2018; Received in revised form 24 April 2018; Accepted 26 April 2018
2212-683X/

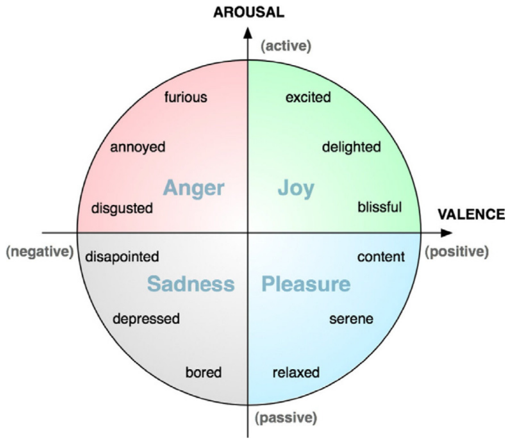

Fig. 1. Two dimensional emotional model (Davidson et al., 1979).

of thirty-two healthy subjects. Section 4 illustrates the discussion of the entire work and list out all the observations of our work. In Section 5, we have compared our proposed work with some previous studies related to different parameters of classification. Finally, Section 6 gives the conclusion of this paper.

2. Methods

2.1. Experimental protocol

Thirty-two healthy subjects (aged between 19 and 37), participate in the experiment. Before the actual experiment takes place, participants have been informed the prior instructions of the experiment (each step in the trial). Initially, the EEG signal is recorded using 13 electrode channels (FP1, FP2, F3, F4, Fz, C3, T7, T8, PO3, P3, P4, Pz, O1) through which the raw signal is stored during the experiment. The sampling rate of the used EEG signal is 512 Hz. The detail description of the experimental protocol is given in Table 1.

2.2. Experimental setup

In this paper, we have classified different emotions of the subjects using the famous "DEAP" dataset (Placidi, Di Giamberardino, Petracca, Spezialetti, & Iacoviello, 2016). In "DEAP" dataset, the EEG signal is recorded using 32 electrodes but here, we consider 13 electrode channels for sake of simplicity. The main goal of the experiment is to identify the emotional states of human based on visual emotional stimuli. In each trial, the subject is performing the following steps.

(1) Subject watch a music video for a 1-min duration. (2) The subject is asked to assign arousal, valence ratings based on visualization of the video. The duration of rating the video is 40 sec. (3) A black image of 1 sec is appeared for refreshing the brain. (4) The same process (1–3) is continued for 5 videos. (5) Subject takes rest for 4 sec and the experiment has ended. The same process has been continued for the rest of the subjects. The timing diagram of each trial in the experiment has been shown in Fig. 2. The experiment consists of total 40 trials and in each trial, one video is used for test set and remaining are used for the training set. We have performed leave-one-out cross-validation (LOOCV) scheme for identifying the number of training and testing sets.

Table 1
Description of experimental protocol.

Number of Subjects	Male	Female	Number of input channels	Number of useful channels	Sampling frequency (Hz)
32	16	16	13	4	512

During this experiment, we have captured the different emotional responses of subjects based on arousal, valence, dominance rating. The arousal, valence, dominance rating are obtained from self-assessment manikin (SAM) (Bradley & Lang, 1994). We have plotted the SAM in Fig. 3. The detail description of the experimental setup is given in Table 2.

- *Analysis of EEG data through signal processing*
  The EEG signal is preprocessed to check the presence of artifacts and that artifacts are removed using adaptive filtering. After filtering, we have applied our proposed dimension reduction technique for identifying the reduced set of channels and proceed for emotion classification phase.
- *Framework of the Proposed work*
  In the experiment, EEG signal is captured from the multi-electrode array and recordings of all the subject have been stored in the dataset for future use. Usually "n" number of recordings and "m" number of channels (with amplitude (μv) value) of different cortex in the brain are stored in the dataset. But all the channels are not useful for emotion detection. So we proposed a new dimension reduction technique (Correlation-based subset selection) of input channels. After dimension reduction technique, we retrieve a reduced set of emotion-specific channels from a large number of input channels. Then we extract the important features related to emotion classification from the reduced set of channels. Here, we consider the features related to Higher Order Statistics (HOS), like mean, standard deviation, skewness, and kurtosis. In this paper, we have used Linear Discriminant Analysis(LDA) classifier because the variance of each output emotional class is same and the output variable (value of each emotion class) is categorical. We have described flowchart of the entire proposed work in Fig. 4.

2.2.1. Dimension reduction using "correlation-based subset selection (CSS)" and emotion classification

In our proposed work, we have captured the brain signal using 13 channels electrode array and stored in a dataset. Out of 13 channels, only some are useful for emotion classification task and those are mapped into four different emotional classes. In the dataset, we have stored amplitude (μv) value of all the input channels for a specific time interval. After the data preprocessing task, the filtered data has been transferred into our proposed dimension reduction model for extracting the useful channels. For selecting emotion-specific channels, we have implemented a new fitness function based on the number of channels, feature class correlation of each channel (using entropy function in the machine learning) and an inter-channel correlation between two channels. Entropy defines the expected information to classify a data point in a class so maximum value of the entropy function refers to the maximum purity of the channels for a class. Initially, the search space contains all the 13 input channels and output subset contains null. We have computed the fitness function for all the channels and choose the channel with a highest fitness value and insert that channel in the output subset. The process of choosing the channel with highest fitness value and putting it into the output subset is continued (in a loop) until all the desired channels of emotion classification are inserted in the subset and no further changes will happen in the subset. Once the output subset (reduced set of channels) is formed, we can extract HOS features from those reduced set of channels and proceed for emotion classification task. The inter-channel correlation between channels is computed by "Pearson correlation" coefficient (Eq. 1). The fitness function (F_c) is described in the Eq. (2).

$$\rho_{v1,v2} = \frac{cov(v1,v2)}{\sigma_{v1}\sigma_{v2}} \quad (1)$$

where cov (v1,v2) is the covariance between every pair of channels, σ_{v1} and σ_{v2} represents the standard deviation of each channel (v1 and v2) where v1, v2 $\subset$ V.

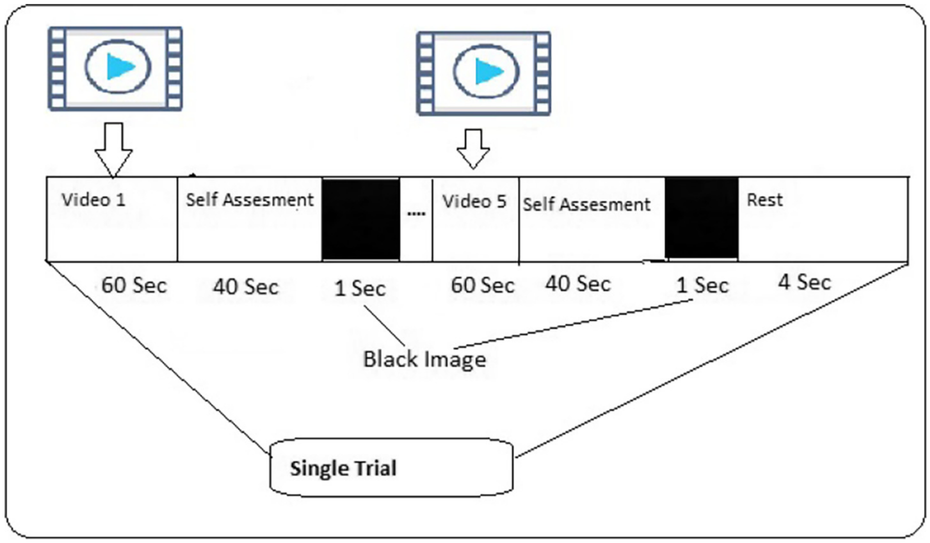

Fig. 2. The timing diagram of a trial in the experiment.

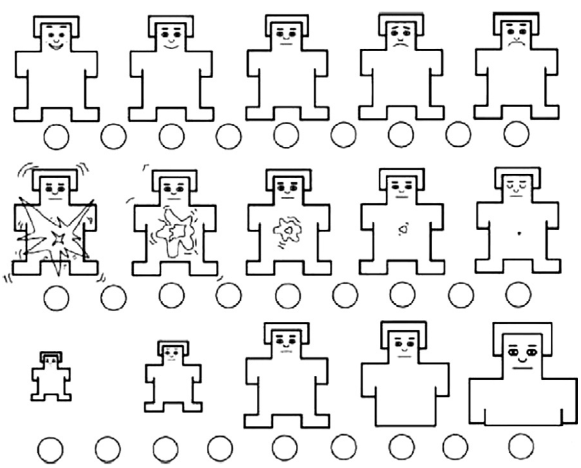

Fig. 3. Self-assessment manikin. From the top: valence SAM, arousal SAM, dominance SAM.

Table 2
Description of experimental setup.

Total number of trials	Videos/trial	Number of training set videos	Number of testing set videos	Duration/trial (sec)
40	5	160	40	509

$$F_c = \frac{n \cdot C_{fc}}{\sqrt{\rho_{v1,v2}}} \tag{2}$$

where n = number of channels, C_{fc} is the feature class correlation and $\rho_{v1,v2}$ is the inter-channel correlation.

From Eq. (2), we understand that the fitness function (F_c) is directly proportional to the number of channels, the value of the entropy function of a channel respectively and inversely proportional to the inter-channel correlation between two channels. The maximum entropy value of a channel refers maximum purity of class level partition which leads to the higher value of the fitness function. The inter-channel correlation between two channels is discussed in the similarity matrix. From that matrix, we can identify that correlation between two channels in the same lobe is higher than the different lobe of the brain. So if we choose one channel (FP1) from a frontopolar lobe then we don't need to select another channel (FP2) from the same lobe due to high correlation and low fitness value. As the fitness function is inversely proportional to the inter-channel correlation, the highly correlated redundant channels in the same lobe of the cortex are removed from the reduced subset. As the fitness function is directly proportional to the feature class correlation (symmetrical uncertainty/entropy), so the emotion-specific channels with high entropy value are selected as the most relevant channel in the subset. The proposed algorithm of dimension reduction of input channels is discussed in Algorithm 1.

2.2.2. Similarity matrix: represent inter-channel correlation

Similarity matrix: represent inter-channel correlation

$$\mathrm{Mat}_{Sim} = \begin{array}{c} \\ P3 \\ FP2 \\ F3 \\ F4 \\ T7 \\ T8 \\ C3 \\ Fz \\ FP1 \\ Pz \\ PO3 \\ O1 \\ P4 \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} P3 & FP2 & F3 & F4 & T7 & T8 & C3 & Fz & FP1 & Pz & PO3 & O1 & P4 \end{array} \\ \begin{bmatrix} 1 & .77 & .90 & .11 & .20 & .01 & .26 & .89 & .80 & .95 & .36 & .09 & .94 \\ & 1 & .9 & .14 & .16 & .019 & .30 & .86 & .89 & .70 & .51 & .21 & .85 \\ & & 1 & .93 & .40 & .21 & .45 & .95 & .91 & .78 & .27 & .049 & .90 \\ & & & 1 & .79 & .81 & .87 & .93 & .40 & .32 & .66 & .38 & .29 \\ & & & & 1 & .89 & .81 & .46 & .41 & .43 & .38 & .63 & .42 \\ & & & & & 1 & .77 & .27 & .23 & .35 & .46 & .57 & .24 \\ & & & & & & 1 & .44 & .53 & .43 & .64 & .33 & .45 \\ & & & & & & & 1 & .85 & .82 & .86 & .23 & .92 \\ & & & & & & & & 1 & .80 & .36 & .008 & .93 \\ & & & & & & & & & 1 & .084 & .20 & .87 \\ & & & & & & & & & & 1 & .84 & .18 \\ & & & & & & & & & & & 1 & .15 \\ & & & & & & & & & & & & 1 \end{bmatrix} \end{array}$$

The similarity matrix is symmetric in nature (correlation (FP1,FP2) = correlation(FP2,FP1)). so the similarity matrix looks like a upper triangular matrix.

After the formation of similarity matrix of inter-channel correlation, we compute the fitness function and create the subset of reduced channels. The formation of the reduced subset (X) has been shown in Fig. 5. Initially, the subset (X) contains NULL and we have applied forward selection method to search the best channel in each run. The optimal channel selection is done by the fitness function. The most optimal channel in each run has been marked with green color and after the 4th run, we have checked that all the required channels of emotion detection (FP1: negative, F3: angry, T7: positive and PO3 for harmony emotion) has been added in the subset (X) so we stop our process. In Fig. 6, we have marked the specific channels (reduced subset refers to Fig. 5) in the 10–20 electrode placement system for better observation.

Algorithm 1. Dimension reduction of input channels using CSS.

Input : Number of input channels(n), amplitude(μv) of input channels.
Output: Reduced set(X) of channels.
1 Search space(S): Contains all the input channels(n) with amplitude value for a specific time instance.
2 |S|=n and |X|=NULL.
3 Fitness function(F_c)=$\frac{n.C_{fc}}{\sqrt{\rho_{v1,v2}}}$
4 root_channel(r) = NULL.
5 **while** *((channels in the X has changed) AND (all emotion-specific channels are not present in X))* **do**
6   **foreach** *channel i in S* **do**
7   Calculate feature class correlation/entropy function(C_{fc}) of i^{th}channel.
8   **if** *(r == NULL)* **then**
9             $F_c(i)$ = n.$C_{fc}(i)$
10   **else**
11   Compute the inter-channel correlation between r and other (n-1) channels.
12   Calculate the Fitness function $F_c(i, M)$, where M $\in$ (i+1,i+2,...n-2,n-1)
13   **end**
14   **end**
15   Choose the channel(i) having Maximum Fitness value.
16   r $\leftarrow$ i.
17   X= X $\cup$ {i}.
18 **end**
19 return X;

We have described our proposed algorithm of dimension reduction in Algorithm 1. Initially, we take all the input channels (n) in the search space (S) and based on fitness function we have added the most optimal channel in the reduced subset (X). We have considered root_channel (r) for the most effective channel in each run. In the first run, when X = NULL then the fitness function depends only on input channel and feature class correlation (C_{fc}) as inter-channel correlation in the same lobe of the brain is always 1 and we choose the channel with the highest fitness function based on feature class correlation. From the 2nd run onwards, the value of the fitness function (F_c) computed using line number 3 of the proposed algorithm. Finally when the subset (X) not changes anymore and all the required channels of emotion classification present in the X then the algorithm is stopped and the reduced set of channels (X) has been returned. After the dimension of input

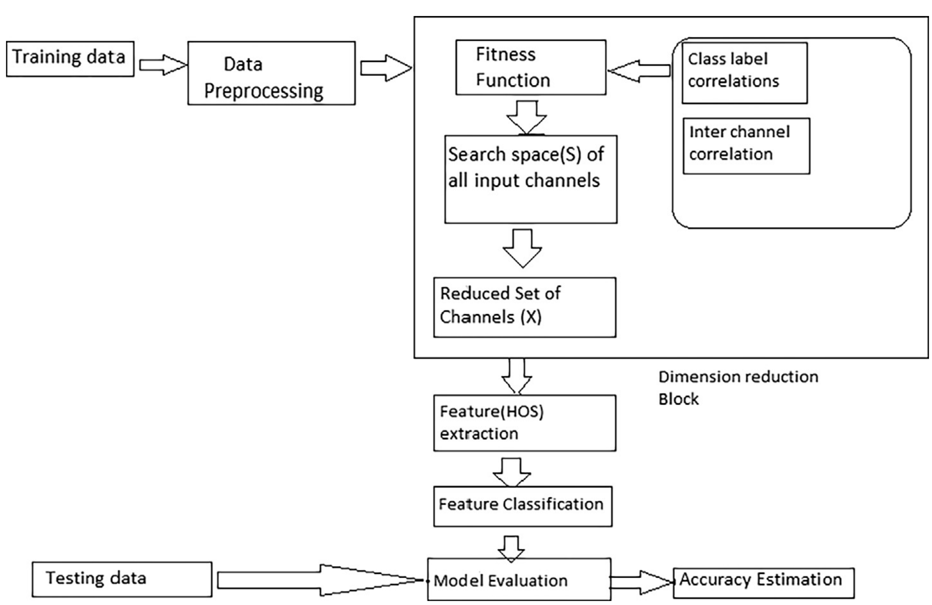

Fig. 4. Flowchart of the proposed work.

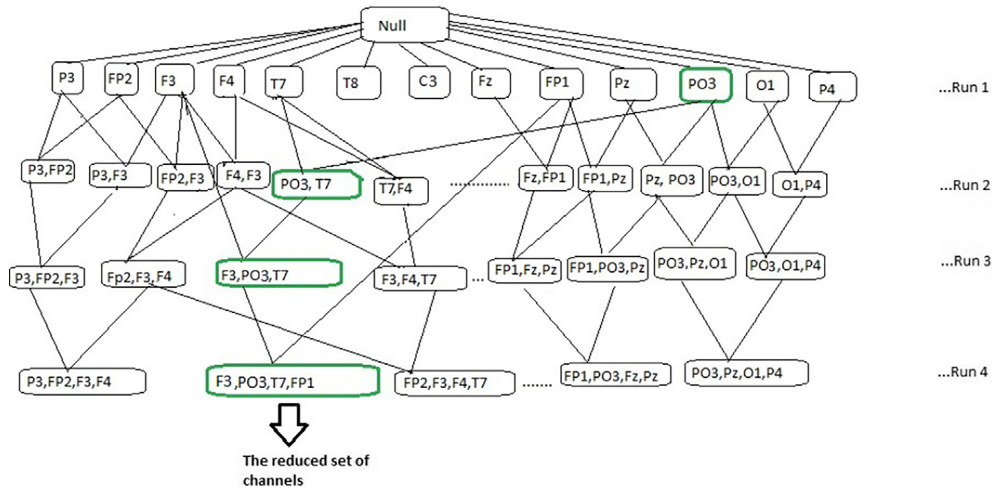

Fig. 5. Reduced set of channels after dimension reduction process.

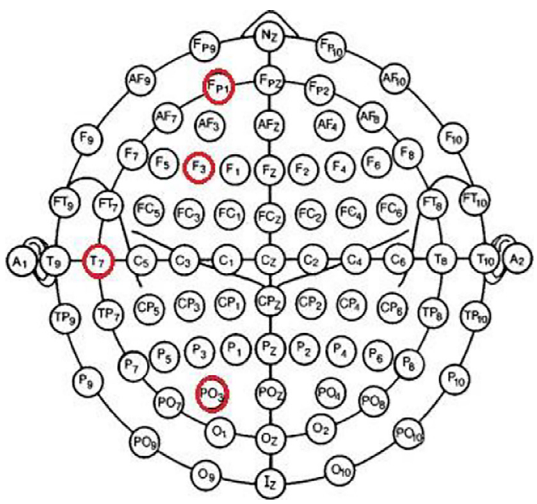

Fig. 6. Electrode placement after dimension reduction method as per 10–20 system (Seeck et al., 2017).

channels are reduced by our proposed algorithm (CSS) then we extract the useful HOS features from those reduced set of channels and proceed for emotion classification. In this paper, we have used LDA classification based on statistical features of the channels and computed the performance analysis of the classification.

3. Result analysis

The proposed dimension reduction technique and LDA classification are performed on the following computing platform. *Application Environment*: R (version-3.3.2), SignalPlant (version-1.2.3.4) (Plesinger, Jurco, Halamek, & Jurak, 2016) *Hardware Environment*: Operating system-Windows7 (64 bit), Processor- Intel Core(TM) i3, RAM-4 GB, Clock speed-2.26 GHz.

3.1. Preprocessing

After completion of the experiment, we need to remove the artifacts/noise of the captured EEG signal (eye blinks, eye movement, hair movement, etc.). For preprocessing (artifacts removal, filtering, etc.) of the raw EEG signal, we have used "Signalplant" software (Plesinger et al., 2016). We have applied "Bandpass filtering" on the raw EEG signal to remove the noise from that signal. The raw and filtered EEG signals are plotted in Figs. 7 and 8 respectively. After the filtering

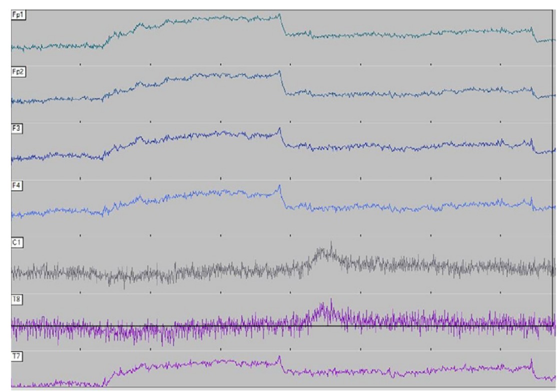

Fig. 7. Raw EEG signal.

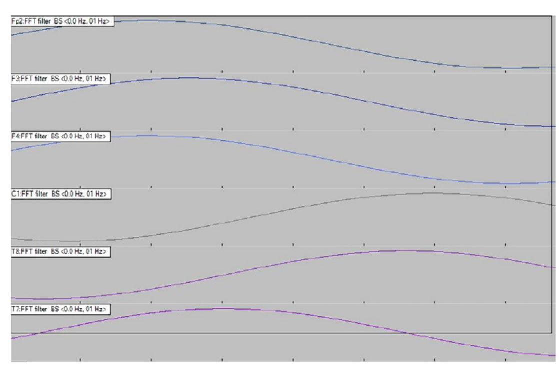

Fig. 8. EEG signal after BandPass filtering.

process of the EEG signal is completed, we proceed for dimension reduction method of input channels. Then we extract the useful HOS features from the reduced set of channels and perform classification.

3.2. Analysis of dimension reduction method

In our proposed algorithm (correlation-based subset selection), we reduce the number of input channels based on the fitness function. In a particular lobe of the brain, fitness function chooses the channel with high-class label correlation and less inter-channel correlation. High-class label correlation defines more accuracy and selects the most relevant channel for the emotion-specific class and low inter-channel correlation removes the redundant channel from the same class. From Fig. 9, we verify that we receive maximum classification accuracy using

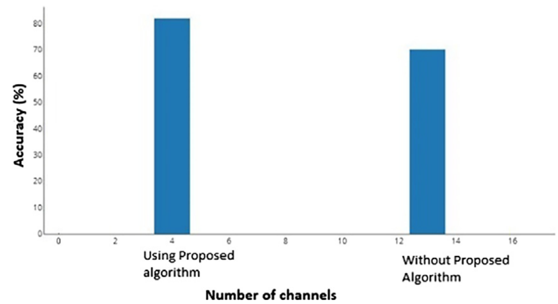

Fig. 9. Performance analysis of classification using the proposed algorithm.

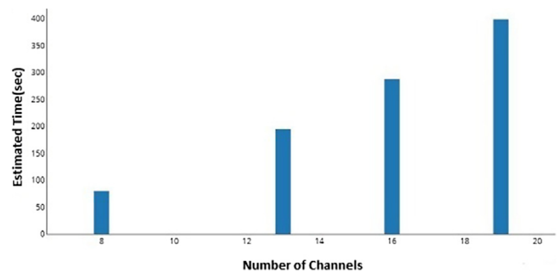

Fig. 10. Estimated time of the proposed algorithm based on number of channels.

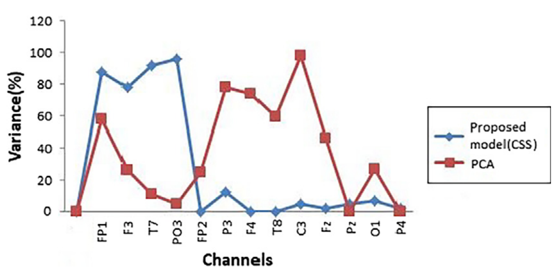

Fig. 11. Comparative analysis of proposed dimension reduction method (CSS) with PCA based on variance of different channels.

our proposed algorithm with least number of channels. We receive higher accuracy because our proposed algorithm removes the redundant channels and selects the most relevant channels for the emotion classification. In the Fig. 10, we plot the estimated time of our proposed algorithm against the number of input channels. From Fig. 10, we find that the estimated time of our proposed algorithm is increased with the number of input channels.

- *Comparative analysis of proposed dimension reduction model(CSS) with PCA*

  In Figs. 11 and 12, we have compared our proposed dimension reduction model (CSS) with well-known dimension reduction model principal component analysis (PCA). In dimension reduction process, the important low dimensional set of features are extracted from the high dimensional set of features. The important features are extracted from the input feature set such that maximum information can be obtained. Here variance of each channel has been computed and channel with larger variance refers to the more important channel for the input feature set. In Fig. 11, we have computed the variance of every input channels using our proposed dimension reduction method (CSS) and PCA. From Fig. 11, we can conclude that our proposed method has a larger variance than PCA for all emotion-specific channels (FP1, F3, T7, PO3) so information gain will be much better for CSS than PCA. In Fig. 12, we have computed the computational time (CPU time) with the number of data points for CSS and PCA dimension reduction method. From Fig. 12, we can conclude that PCA consumes higher computational time than CSS for a large number of data points due to the cubic computational complexity of PCA. Large computational time leads to low response time and low CPU utilization of the system so our proposed dimension reduction method (CSS) gives the better throughput of the system than PCA.

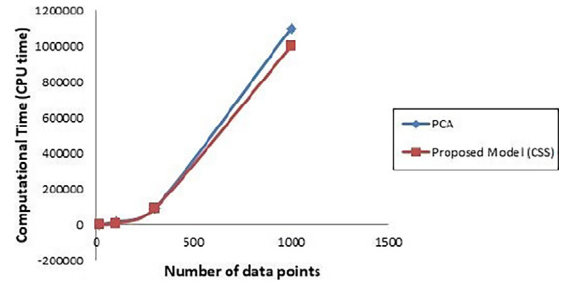

Fig. 12. Comparative analysis of proposed dimension reduction method (CSS) with PCA based on computational time and data points.

- *Computational complexity of the Proposed algorithm (For "n" number of channels)*

  Time Complexity
  Finding inter-channel correlations between "n" channels: $O(n^2)$
  Compute the feature class correlation,fitness function: $O(n)$
  Time to select channel with highest fitness function: $O(n)$
  Total Time complexity: $O(n^2 + 2n)$
  Space Complexity Space for storing "n" number of channels: O(n)
  Total Space complexity: O(n)

The time complexity of our proposed algorithm is quadratic in nature, so the estimated time will be increased for a large number of channels. The space complexity is linear in nature as the maximum "n" number of channels can store in memory in each run.

3.3. Feature extraction and selection

After completing the dimension reduction process, we extract all the statistical features from those reduced set of channels. However, among all the statistical features (minimum, maximum, mean, mode, median, standard deviation, skewness and kurtosis) only some are useful for emotion detection task. So we select the important features (mean, standard deviation, skewness and kurtosis) from those reduced set of channels and proceed for the classification process.

- *Statistical analysis of features*

The mean computes the average amplitude value of the collected EEG data samples within the selected area. The mean is computed using Eq. (3).

$$mean = \frac{1}{n_e - n_s} \sum_{i=n_s}^{n_e-n_s} X_i \tag{3}$$

where n_s, n_e is the starting point and ending point of the sample data. A total number of samples is $n_e - n_s$ and X_i represents the value of the curve at the vertical axis.

The standard deviation computes the deviation of the EEG signal from the mean value within selected endpoints. Standard deviation is computed using Eq. (4).

$$Stddev = \sqrt{\frac{1}{(n_e - n_s) - 1} \sum\nolimits_{i=n_s}^{n_e-n_s} (X_i - mean)^2} \tag{4}$$

Skewness is a measure of lack of symmetry. It shows the degree of asymmetry in a data distribution. A distribution is symmetric if it looks same from both ends (left and right) with respect to the mean point. The

Table 3
Classifications based on statistical features.

Emotional state	Statistical feature's value (μv)			
	Mean	Standard deviation	Skewness	Kurtosis
Positive	303.249	32.869	.149	.166
Negative	−728.133	48.549	−.250	−.350
Angry	−869.379	37.496	.368	−.332
Harmony	184.758	27.860	.122	.120

skewness is measured using Eq. (5).

$$Skewness = \frac{\sum_{i=n_s}^{n_e} (X_i - mean)^3 \big/ N}{(Stddev)^3} \quad (5)$$

where "N" is the total number of sample points. Kurtosis defines whether the data is heavily-tailed or lightly-tailed relative to the normal distribution. Channel value with high kurtosis means the presence of noise in the data when the low kurtosis value refers to lack of outliers. The kurtosis is measured using Eq. (6).

$$Kurtosis = \frac{\sum_{i=n_s}^{n_e} (X_i - mean)^4 \big/ N}{(Stddev)^4} \quad (6)$$

3.4. Classification

Once the important statistical features are identified, then we perform the classification based on the feature's value of useful channels (FP1, F3, T7, PO3). The detail description of all the useful features along with related emotional states are mentioned in Table 3. The LDA classification has been performed based on "HOS" features of selected channels while the subject is watching the video during the experiment. For evaluation of our model, we perform leave-one-out cross-validation scheme.

Here we measure the value of statistical features using "SignalPlant" software. The statistical features (mean, median, skewness, kurtosis) of each emotional state are measured using the specific channel (FP1, F3, T7, PO3, etc.) values taken from different samples.

- *Performance Analysis of LDA classification*

In this paper, we analyze the performance of the classification process using confusion matrix. The confusion matrix is a table describing the performance of a classification model on a set of test data for which true values are already known. For each class, confusion matrix gives the number of correct and incorrect predictions in a summarized form. With the help of confusion matrix, we can simply observe how much our classification model gets "confused" during predictions. Here, we have evaluated the classification accuracy of our proposed model using the confusion matrix (Mat_{con}).

Confusion Matrix:

$$\mathrm{Mat}_{con} = \begin{array}{r} \\ PositiveEmotion \\ NegativeEmotion \\ Angry \\ Harmony \end{array} \begin{array}{c} \begin{array}{cccc} PositiveEmotion & NegativeEmotion & Angry & Harmony \end{array} \\ \begin{bmatrix} 32 & 42 & 4 & 4 \\ 14 & 611 & 3 & 5 \\ 17 & 31 & 18 & 10 \\ 18 & 30 & 7 & 163 \end{bmatrix} \end{array}$$

Sensitivity/recall refers the true prediction of each class (emotional state) when it is actually true. Specificity refers false prediction of each class (emotional state) when it is actually false. Precision means how many of the truly classified samples are relevant in each class. Prevalence is the ratio between the summation of actual true samples in each class and the total number of samples. F-measure computes the accuracy of the multi-class problem, it is calculated using precision and recall value. From the Table 4 and Fig. 14, we can conclude that emotional state "Harmony" (class 4) has the maximum sensitivity so the classifier shows the maximum true prediction rate for that class, whereas emotional state "Positive emotion" (class 1) has the maximum specificity or false prediction rate. Emotional state "Negative emotion" (class 2) refers the maximum occurrence of actual true samples out of total samples due to the highest prevalence value. "Negative emotion" (class 2) also has the highest precision value indicating the maximum number of correctly classified samples which belong to the relevant class. Class 2 refers to maximum accuracy based on highest F-measure. In the Fig. 13, we have verified the performance of multi-class classification based on three accuracy parameters. Balanced accuracy is useful for the balanced dataset (number of training samples is equal to test samples in each class) whereas Kappa value is a metric which compares an observed accuracy and expected accuracy. From Fig. 13, we have identified that our model has good accuracy for the balanced dataset (average accuracy: 82% and balanced accuracy: 67.16%) but

Table 4
Performance analysis of each class using statistical parameters.

Parameters (%)	Positive emotion	Negative emotion	Angry	Harmony
Sensitivity	39	85.57	56.25	89.56
Precision	39	96.52	23.68	74.77
Specificity	94.61	92.54	94.06	93.34
Prevalence	8.028	70.76	3.171	18.04
F-measure	28.18	43.27	5.15	5.24

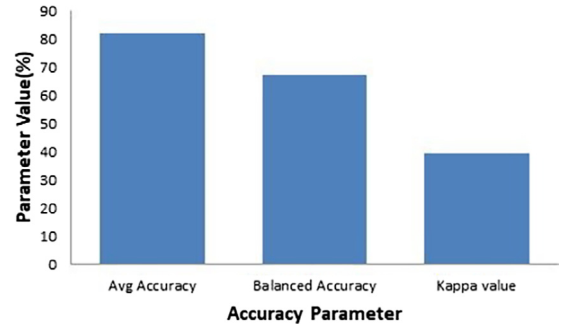

Fig. 13. Performance analysis of classification based on accuracy parameters.

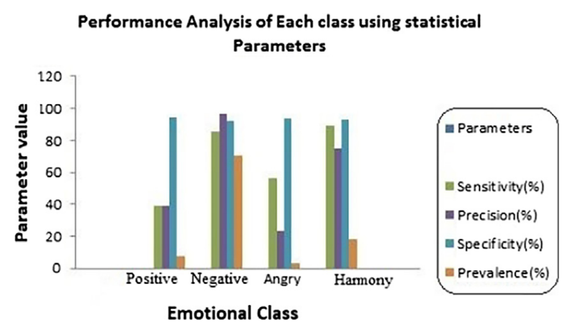

Fig. 14. Performance analysis of classification based on statistical parameters.

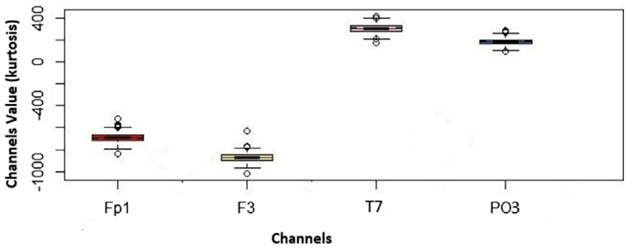

Fig. 15. Boxplot of useful channels based on kurtosis value.

less compatible with the imbalanced dataset (40%).

From Fig. 15, we can conclude that larger value of kurtosis in channels T7, PO3 indicates positive and harmony state due to the large dispersion in temporal & parieto-occipital area of the brain and smaller value in channels F3, FP1 indicates angry and negative state due to decreased power of kurtosis value in left frontal and frontopolar section of the brain.

4. Discussion

Nowadays emotion detection from the different psychological signal is a very popular research topic. Here we have used higher order statistical features (mean, skewness, kurtosis, etc.) for classification. We implement the signal processing and statistical analysis using "SignalPlant" software (Plesinger et al., 2016). We can predict the performance of our classification model based on new dimension reduction process (correlation based subset approach). We have evaluated different classification parameters (sensitivity, specificity, precision, prevalence) from the confusion matrix. Based on those parameters, we have analyzed the performance of our predicted model. The classification accuracy of our model reaches to 82% which is quite larger than the other emotional detection model (Anh, Van, Ha, & Quyet, 2012; Chanel, Rebetez, Betrancourt, & Pun, 2011; Jirayucharoensak, Pan-Ngum, & Israsena, 2014). The accuracy of the classification process is dependent on various factors such as the type of signals, number of subjects, the attention of subjects, placement of electrodes and so on (Selvaraj, Murugappan, Wan, & Yaacob, 2013). Though we have implemented the emotion classification over the able-bodied subjects, this processes can also be implemented over the paralyzed people in the near future. So the paralyzed people can express their emotions during a fixed time interval which will improve the effectiveness of BCI system.

5. Comparison with previous studies

- *Comparison based on Classification accuracy*

In paper Bhardwaj, Gupta, Jain, Rani, and Yadav, 2015, the author has described six different emotional states (excluding neutral state) and he uses LDA and SVM classifiers for detecting different emotional states. In the paper Zong and Chetouani, 2009, the author has mentioned that non-linear approach based classification reaches the maximum accuracy 76%. The maximum mean classification rate for four classes has been achieved 70% (Kim & Andre, 2008). In paper Chai, Ling, Hunter, and Nguyen, 2012, the author uses the artificial neural network (ANN) as classification method and achieves the average accuracy 73%. Other recent well-known emotion classification model (Anh et al., 2012; Jatupaiboon, Pan-ngum, & Israsena, 2013; Jirayucharoensak et al., 2014; Wijeratne & Perera, 2012; Yoon & Chung, 2013) has lower classification accuracy than our proposed methodology which gives the average accuracy of 82%.

- *Comparison based on Statistical features*

In paper Islam, Ahmed, Mostafa, Yusuf, and Ahmad, 2013, the author has used the statistical parameters to classify different emotional states, but the paper does not cover the specific channels from which the statistical features need to be extracted. The paper also does not men-

Table 5
Comparison of proposed model with other popular models.

Study	Property of study					
	Number of electrodes	Sampling frequency (Hz)	Classifier	Number of emotions	Stimuli	Accuracy (%)
Proposed	4	512	LDA	4	Video	82
Wang et al., 2014	128	200	SVM	3	Music, Video	90
Jirayucharoensak et al., 2014	32	128	SVM	3	Music,Video	49.52
Jatupaiboon et al., 2013	64	128	SVM	3	Image,Video	65.12
Bastos-Filho et al., 2012	32	256	kNN	2	Video	69.5
Wijeratne and Perera, 2012	32	256	ANN	3	Video	75
Anh et al., 2012	16	256	SVM	5	Image	70.5
Chanel et al., 2011	19	256	LDA	3	Video games	63

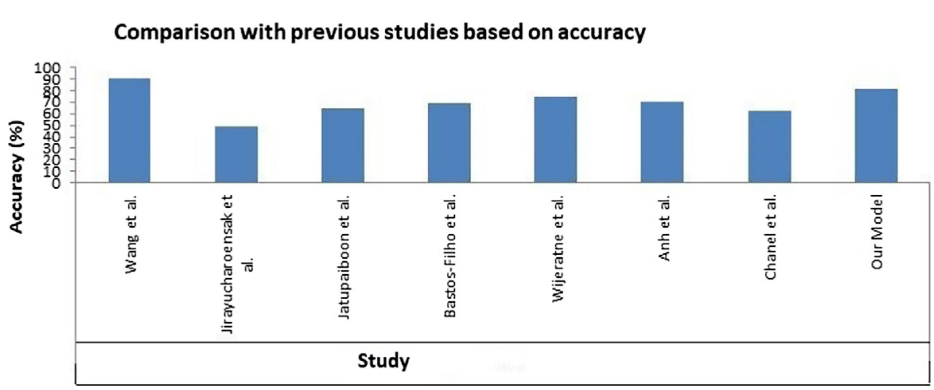

Fig. 16. Comparative analysis with previous studies based on classifier's accuracy.

tion the classification accuracy of the emotional state detection process. We have mentioned all the statistical parameters value of useful channels in the specific cortex of the brain. The comparison of our classification process with other processes is described in Table 5.

In Fig. 16, we have compared our proposed model with other models(mentioned in Table 5) based on the accuracy and showed that our model beats all the previous models except the model refer by Wang et al., 2014 but the number of emotions recognized by our model is greater than Wang et al. So considering all the parameters, we can conclude that our proposed classification model outperforms all the other models of emotion detection.

6. Conclusion and Future work

In this paper, we have investigated different emotional states of human based on higher-order statistical parameters. We have introduced a new dimension reduction technique (correlation based subset selection) of input channels and select the most relevant emotion-specific channels for classification. Then we extract useful HOS features from those relevant channels and perform LDA classification. The execution time of our proposed algorithm of dimension reduction is O(n^2 + 2n). We have performed the performance analysis of classification process using the proposed algorithm (dimension reduction process). The classification accuracy has been computed using the confusion matrix and different statistical parameters (sensitivity, specificity, prevalence, precision). After comparing with other well-known models, we can conclude that our proposed model outperforms others. In the future, we try to develop the classification process for a large number of non-linear data using different SVM kernels and figure out the efficient kernel for emotion classification.

Acknowledgment

No research funding has been received for this work.

References

Anh, V. H., Van, M. N., Ha, B. B., & Quyet, T. H. (2012). A real-time model based support vector machine for emotion recognition through EEG. In *2012 International conference on control, automation and information sciences (ICCAIS), November* (pp. 191–196). IEEE.

Bastos-Filho, T. F., Ferreira, A., Atencio, A. C., Arjunan, S., & Kumar, D. (2012). Evaluation of feature extraction techniques in emotional state recognition. In *2012 4th International conference on intelligent human computer interaction (IHCI), December* (pp. 1–6). IEEE.

Bhardwaj, A., Gupta, A., Jain, P., Rani, A., & Yadav, J. (2015). Classification of human emotions from EEG signals using SVM and LDA classifiers. In *2015 2nd International conference on signal processing and integrated networks (SPIN), February* (pp. 180–185). IEEE.

Bos, D. O. (2006). EEG-based emotion recognition. *The Influence of Visual and Auditory Stimuli, 56*(3), 1–17.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.

Chai, R., Ling, S. H., Hunter, G. P., & Nguyen, H. T. (2012). Mental task classifications using prefrontal cortex electroencephalograph signals. In *2012 Annual international conference of the IEEE engineering in medicine and biology society (EMBC), August* (pp. 1831–1834). IEEE.

Chakladar, D. D., & Chakraborty, S. (2017). Study and analysis of a fast moving cursor control in a multithreaded way in brain computer interface. In *International conference on computational intelligence, communications, and business analytics, March* (pp. 44–56). Singapore: Springer.

Chanel, G., Rebetez, C., Betrancourt, M., & Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 41*(6), 1052–1063.

Davidson, R. J., Schwartz, G. E., Saron, C., Bennett, J., & Goleman, D. J. (1979). Frontal versus parietal EEG asymmetry during positive and negative affect. In *Psychophysiology, January* (Vol. 16(2), pp. 202–203). 40 West 20th Street, New York, NY 10011-4211: Cambridge Univ Press.

Davidson, R. J., & Fox, N. A. (1982). Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science,* 1235–1237.

Islam, M., Ahmed, T., Mostafa, S. S., Yusuf, M. S. U., & Ahmad, M. (2013). Human emotion recognition using frequency & statistical measures of EEG signal. In *2013 International conference on informatics, electronics & vision (ICIEV), May* (pp. 1–6). IEEE.

Jatupaiboon, N., Pan-ngum, S., & Israsena, P. (2013). Real-time EEG-based happiness detection system. *The Scientific World Journal. Hindawi*. http://dx.doi.org/10.1155/2013/618649 12 pages, Article ID 618649.

Jirayucharoensak, S., Pan-Ngum, S., & Israsena, P. (2014). EEG-based emotion recognition using deep learning network with principal component-based covariate shift adaptation. *The Scientific World Journal. Hindawi*. http://dx.doi.org/10.1155/2014/627892 10 pages, Article ID 627892.

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(12), 2067–2083.

Murugappan, M., Ramachandran, N., & Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering, 3*(04), 390.

Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM Systems Journal, 39*(3.4), 705–719.

Placidi, G., Di Giamberardino, P., Petracca, A., Spezialetti, M., & Iacoviello, D. (2016). Classification of emotional signals from the DEAP dataset. In *International congress on neurotechnology, electronics and informatics, November* (Vol. 2, pp. 15–21). SCITEPRESS < http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html > .

Plesinger, F., Jurco, J., Halamek, J., & Jurak, P. (2016). SignalPlant: An open signal processing software platform. *Physiological Measurements, 37*(7), N38–N48 < https://signalplant.codeplex.com/ > .

Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., & Beniczky, S. (2017). The standardized EEG electrode array of the IFCN. *Clinical Neurophysiology, 128*(10), 2070–2077.

Selvaraj, J., Murugappan, M., Wan, K., & Yaacob, S. (2013). Classification of emotional states from electrocardiogram signals: A non-linear approach based on hurst. *Biomedical Engineering Online, 12*(1), 44.

Wang, X. W., Nie, D., & Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing, 129*, 94–106.

Wijeratne, U., & Perera, U. (2012). Intelligent emotion recognition system using electroencephalography and active shape models. In *2012 IEEE EMBS conference on biomedical engineering and sciences (IECBES), December* (pp. 636–641). IEEE.

Yoon, H. J., & Chung, S. Y. (2013). EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Computers in Biology and Medicine, 43*(12), 2230–2237.

Zong, C., & Chetouani, M. (2009). Hilbert-Huang transform based physiological signals analysis for emotion recognition. In *2009 IEEE international symposium on signal processing and information technology (ISSPIT), December* (pp. 334–339). IEEE.