Biomedical Signal Processing and Control 70 (2021) 102979

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Feature and channel selection for designing a regression-based continuous-variable emotion recognition system with two EEG channels

Mahshad Javidan [a], Mohammadreza Yazdchi [a], Zahra Baharlouei [b], Amin Mahnam [a,*]

[a] Department of Biomedical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran
[b] Ragheb Isfahani Higher Education Institute, Isfahan, Iran

A R T I C L E I N F O

Keywords:
Electroencephalography (EEG)
Emotion recognition
Regression
RReliefF

A B S T R A C T

Objective: With deepened interactions between human and computer, the need for a reliable and practical system for emotion recognition has become significant. The aim of this study is to propose a practical system for estimation of a continuous measure of valence based on a few number of EEG channels.
Methods: A vast spectrum of time, frequency and coherence features were implemented with linear Regression (LR), Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) models and then ranked for the performance on DEAP database using a regression-based Relief filter. Regression outcomes were also classified to compare the performance of the proposed method with the literature. Finally, a video-based emotion recognition experiment was designed and conducted on 12 subjects using F7, F8, FC2 and T7 electrodes.
Results: Magnitude Squared Coherence Estimate(MSCE) on F7–F8 with SVR model provided the highest performance on DEAP dataset. Classification of the output led to an average accuracy of 67.5%. For the gathered data, combination of MSCE and Hilbert–Huang Spectrum provided the best performance with 0.22 root mean square error and 0.67 correlation with self-reported valence in the scale of 1–9.
Conclusion: MSCE could provide a good accuracy in estimation of Valence using 2 EEG channels on Deep dataset, and with addition of Hilbert–Huang Spectrum, it also demonstrated good accuracy and correlation with self-reported valence, in a completely different experiment.
Significance: Continuous-value estimation of the valence can be achieved with only 2 EEG channels for practical applications out of the lab.

1. Introduction

Emotions have an important role in human to human communications as well as in human daily activities. Emotion recognition has several applications including assistance to patients who are able to understand emotions but fail to show them in their face, psychological and psychiatric research and diagnosis, human–computer communication improvement and development of robots capable of understanding human emotions and reacting to them [1]. In this regard, recognizing the emotions has been of interest in the past decades. Emotions can be recognized from speech, motion, facial expression or body gestures. There are other methods that recognize emotions by measuring physiological data such as Electrocardiography (ECG), Heart Rate Variability (HRV), Electroencephalography (EEG), Facial Recognition (FR), Forehead Bio-Signal (FBS), Speech Recognition (SR), Skin Temperature (SKT), Blood Volume Pulse (BVP), and Respiration (RSP) [2]. Among these methods, EEG signals are the most accurate, reliable and informative [3,4], as brain is the source of any reaction to external stimuli. Using EEG signals, new researches on affective Brain–Computer Interface (aBCI) are conducted in which BCI is enhanced with the ability to detect, process and respond to humans' affective states [5]. During the last decade, a lot of studies has been done to propose features and machine learning techniques for EEG-based emotion recognition. Since there is no standard model for experimentation, the procedures, defined emotions, EEG set-ups and the stimuli used in these studies differ significantly and this makes a proper comparison of proposed techniques difficult. Here some of these studies are reviewed.

Most of studies consider discrete emotions and therefore need to use classifiers in their models to recognize the emotions. The most widely used classification methods in such studies are Support Vector Machine (SVM) [6–9] and K-Nearest Neighbor (KNN) [8–10]. However in some studies, a very small number of participants have been used, e.g. 6 participants in [11] and 5 in [12,13] which may not be enough for a strong statistical analysis.

* Corresponding author.
E-mail address: mahnam@eng.ui.ac.ir (A. Mahnam).

https://doi.org/10.1016/j.bspc.2021.102979
Received 9 December 2020; Received in revised form 20 June 2021; Accepted 7 July 2021
Available online 27 July 2021

Different types of stimuli with different duration times are applied in the experiments. For example, in [13,14] images are used as stimuli while in [15,16] audio, and in [17,18] video clips are used. The results of some researches show that music videos have more significant effects on the human emotions than image and audio do [19]. While most works use stimuli from international databases, it seems that the effectiveness of a stimulus is related to the culture and the behavioral habits of the society and thus should be selected from national cases.

The other parameters that differ in researches are the type and number of features extracted from EEG signals for emotion recognition. For example, in [20], Fractal Dimension (FD) and Hjorth has been used; [14] has used HOC features; [17] focuses on non-linear features; [11] compares three sets of features including power spectrum, wavelet and nonlinear dynamical features; and some papers use only frequency band [21,22] or time–frequency [23] features. The EEG signals are non-linear in nature resulting in unexpected changes in the features. Therefore, it seems that to validate an approach, different types of features should be extracted in order to compare their performances in emotion recognition.

In the literature, different emotional states have been defined. Many papers have referred to Russel model of emotions which describes emotions as two dimensional space of valence and arousal [24]. Although in this model, valence and arousal have been considered continuous variables, in emotion recognition literature, emotions have been considered as discrete states. For example in [20], three emotional states are considered to be recognized and those with low arousal level are not considered. Six levels of emotions in [14] and two levels are studied in [11]. [16] focuses only on like and dislike emotions.

Standard recorded EEG signals for emotion recognition such as DEAP [25] and SEED [26] databases are implemented in some papers [23,27] while in some studies, e.g. [14], new experiments have been conducted; however, the results of such models cannot be representative else the standard databases are used as well.

The number of EEG channels that usually lies between 2 and 257 is also another different point of the researches. A large number of EEG channels has been used in many researches to improve the accuracy of emotion recognition. However, more EEG channels would not necessarily improve the performance. Furthermore, in practical use of an emotion recognition system minimum number of electrodes is desired, as reliable electrode placement is difficult and time consuming. Thus, one of the most challenging issues in the EEG-based models is optimizing the number of channels [28]. In some papers, channel reduction has been studied. The significant electrodes were investigated in [22,23], but the results were not applied in the classification of the emotions. In [22], most top selected channels were at frontal and parietal lobes. Authors in [23] found that the lateral temporal region has higher activity for positive emotion, while negative emotion correlates with significantly higher delta responses at parietal and occipital sites and higher gamma responses at pre-frontal sites. Using previous studies, authors of [29] selected $AF3$, $F4$ and $FC6$ as the most significant electrodes, but the validity was not investigated in the experiment. Studies on emotion recognition from EEG are listed in Table 1 with their novelty, as well as the advantages and limitations of their work.

This study aims to develop an algorithm for the estimation of valence as a continuous variable, as described by the Russell model, using a minimum number of features and channels. This is the first step in developing a practical system that is able to recognize emotions and can be easily implemented for realistic applications out of the lab. According to Table 1, our work is the first attempt, to the best of our knowledge, to perform regression-based estimation of emotion where emotion is considered as a continuous variable. As we could not find any previous study with this approach, a classification stage was also performed on the regression results to compare our feature/channel selection results with other studies. Extracting a limited number and types of features from EEG signals, as implemented in some studies, may have led to the results that are not close to the optimum. In the current work, various features from different categories have been evaluated using a standard dataset and features and their corresponding channels with significant contribution in recognition of valence were selected. Some researchers have used standard databases to implement their method (e.g., [23,27]), while others only used their own recorded datasets (e.g., [14]). Here, we first use a standard dataset with different regression methods to find features and channels that work the best for a minimal number of electrodes, and then validate that by conducting an experiment with 12 participants. Therefore, in this work, the results are more representative since we use both data from standard databases as well as the data we recorded ourselves, and demonstrate how the algorithm derived from one dataset, works on another dataset recorded for the same purpose but with different setting and different culture.

In summary, the contribution of this paper is three fold. First, we developed a systematic approach for estimation of valence as a continuous variable which is novel to the extent of authors knowledge. Second, in this work a comprehensive evaluation of various features from different categories for different channels is presented to find the best features for emotion recognition with minimal number of channels. Third, feature and channel selection is performed on a standard dataset, but then the outcome is validated in another completely different dataset recorded as part of this study, to demonstrate the performance of selected features and reduced channels for implementing in practical applications.

2. Methods

This study consisted of two parts. In the first part, a comprehensive feature and channel selection is implemented on a publicly available dataset, to achieve the best 2 channels and features for recognition of valence. In the second part, an experiment was conducted with completely different audiovisual stimuli to validate the developed method in a separate set-up. For this purpose, first various time, frequency, time–frequency and mutual features were extracted and ranked for estimation of valence using DEAP dataset [25]. The regression-based feature selection method, RReliefF [31], was used for selecting the high value features and also the best EEG channels, as channels were considered as another dimension for the feature set. The elected features and channels were then fed to a SVR model for estimating the valence.

In the second part, experiments were conducted by recording the selected EEG channels in response to a set of video clips. Some previous experiences in this lab, have shown that the impact of the stimuli is affected by the cultural differences. For instance, emotions perceived by our participants from Deap videos were far from the labels associated to them. Therefore, a new set of video clips were used in this study that were culturally relevant. To select these pieces 20 individuals, non-participant in the main experiment, were asked to rate the valence with numbers between 1 (negative emotions) to 9 (positive emotions). Then out of sixty, 30 video clips that had the strongest volunteer ratings and at the same time a small variation were selected for the final experiment. Next, a set of high performance features along with several regression models were used for the estimation of valence value for each trial. The described steps applied in this study are shown in Fig. 1. In the rest of the section, we explain each step more precisely.

2.1. DEAP database

In order to explore the optimal features and subsequently optimal associated electrodes, we first evaluate performance of a vast range of implemented features using the publicly available emotion dataset, DEAP. In the DEAP dataset, there are 32 EEG channels; and as stimuli, 40 one-minute music videos are used to record eight physiological signals of 32 subjects aged between 19 and 37. The emotional states are presented in the valence-arousal emotion model [24] in which each emotional state is placed on a two-dimensional scale. The first dimension represents valence, ranged from negative to positive, and the second one is arousal, ranged from calm to exciting. In the DEAP, each video clip is rated from 1 to 9 by each subject after viewing, in the sense of arousal and valence separately [25].

Table 1
Comparison of the various studies on emotion recognition using EEG.

Study	Novelty	Advantages	Disadvantages \Limitations			
			Small number of participants in test	Only a specific group of features were compared	Using large number of channels	Discrete values were considered for the valence or arousal axis
[11]	Tracking the trajectory of emotion changes with manifold learning.	Visualizing trajectory of emotion changes by reducing subject-independent features.	✓			✓
[12]	Real-time subject-dependent algorithm with the most stable features.	Limit number of channels were selected for the experiment and different types of features were extracted and compared.	✓			✓
[13]	Obtaining a detailed understanding of the affective EEG patterns using RQA method.	Non-linear approach for detecting instantaneous changes in the emotional EEG induced by visual stimuli.	✓			✓
[14]	Presenting a feature extraction technique in which the mirror neuron system concept was adapted to efficiently foster emotion induction by the process of imitation.	Testing four different classifiers in order to accomplish efficient emotion recognition. Maximum classification rate was achieved.		✓		✓
[16]	An end-to-end classifier with no need for a feature selection/extraction method.	Correct features of each class will be automatically learned with a deep neural network.			✓	✓
[19]	Identifying the features that can best discriminate the emotions.	Using three-dimensional emotion model that better can classify than two-dimensional model.			✓	✓
[20]	A new channel selection method (synchronization likelihood).	Reducing the number of EEG channels.		✓		✓
[21]	Searching emotion-specific features and evaluating the efficiency of different classifiers.	Achieving a user-independent emotion recognition system.	✓		✓	✓
[22]	Using ReliefF-based channel selection methods.	Channel reduction.		✓		✓
[23]	Adopting discriminative Graph regularized Extreme Learning Machine (GELM) to identify the stable patterns over time for emotion recognition.	Developing novel emotion EEG dataset as a subset of SEED. Also, systematic comparison of different feature extraction, feature selection, feature smoothing and pattern classification methods on DEAP.			✓	✓
[27]	Comparative study on several state-of-the-art domain adaptation techniques.	Testing the method on different well-known datasets and self-recorded datasets that was achieved in different conditions.		✓		✓
[30]	A deep learning framework based on a multiband feature matrix (MFM) and a capsule network (CapsNet).	Using three-dimensional emotion model that better can classify than two-dimensional model.			✓	✓

2.2. Preprocessing

The EEG spectrum ranges from 0.5 to 45 Hz. This band is not exclusive of EEG signals. The spectrum of some other electro-physiological signals such as Electrocardiogram (ECG) and Electromyogram (EMG) lies in this band. Furthermore, during EEG acquisition process, noise and interferences can appear. Therefore, it is necessary to preprocess the recorded signals to remove these perturbances. In the preprocessing step, the signals are down-sampled from 512 Hz to 128 Hz and a band pass frequency filter from 4.0 to 45.0 Hz applied to reduce ECG and EMG artifacts from the EEG signals. Notch filters are also applied to eliminate the effect of power line at 50 or 60 Hz [32]. In this paper, this stage was preformed using EEGLab toolbox in MATLAB.

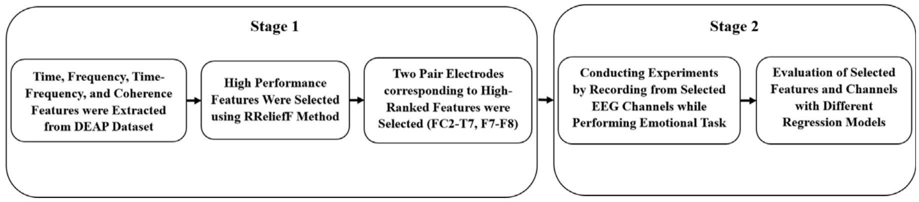

Fig. 1. Overview of the steps taken in this study for finding appropriate features and regression model for estimation of valence.

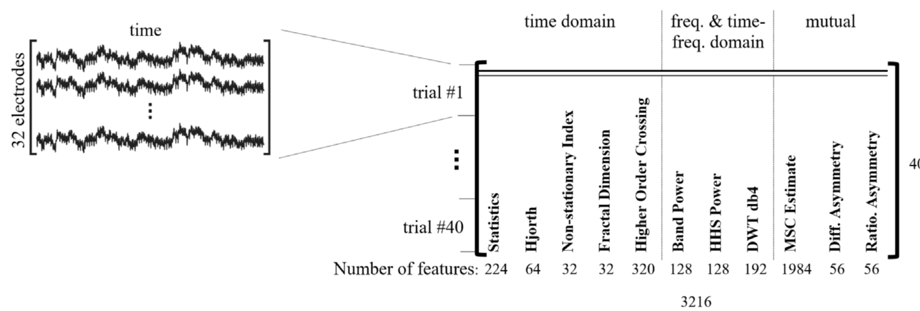

Fig. 2. Name and number of features for each feature group extracted from EEG signals of DEAP database recorded from 32 electrodes.

2.3. Feature extraction

There is a wide range of features relevant to emotions which are proposed in the literature. These features are in time, frequency and time–frequency domain. They are typically calculated from the recorded signals of a single electrode. However, a few features combining signals of more than one electrode can be found in the literature. In this study, various features from different categories were implemented so that we can compare them for recognition of valence, with exactly the same method. The extracted features in this research are as follows:

2.3.1. Time domain features

Although time domain features of EEG have received less attention in the literature, some studies suggest that they are more reliable than the frequency domain features [19]. In this study several time domain features were included, as listed below:

- Statistics of Signal,
- Hjorth Features,
- Non-Stationary Index (NSI),
- Fractal Dimension (FD),
- Higher Order Crossings (HOC).

2.3.2. Frequency domain features

Frequency domain features of the EEG signals have received more attentions in the previous researches on emotion recognition [19]. In this study, we use the band power spectrum features.

2.3.3. Time–frequency domain features

Since the brain signals are non-stationary, time–frequency analysis may better present the dynamic changes in the signals. From the category of these features, the following are used:

- Hilbert–Huang Spectrum (HHS),
- Discrete Wavelet Transform.

2.3.4. Features calculated from combinations of electrodes

Combinations of electrodes help us to achieve more effective features in emotion recognition. In this research, we use

- Magnitude Squared Coherence Estimate,
- Differential Asymmetry,
- Rational Asymmetry.

In Table 2, a summarized description of the mentioned features is available.

The main focus of this study is on reducing the number of electrodes to achieve an optimal system in operation. At first, we extract the features (mentioned in Section 2.3) of the 32-channel EEG data in DEAP database. The original EEG data that we got from DEAP dataset are pre-processed as explained in Section 2.2. The EEG spectrum is divided into five frequency bands: *delta* (from 0.5 to 4 Hz), *theta* (from 4 to 7 Hz), *alpha* (from 8 to 13 Hz), *beta* (from 14 to 30 Hz), and *gamma* (from 31 to 45 Hz) [32]. Since a bandpass filter from 4 to 45 Hz is used in pre-processing step, the results of delta frequency bands are not included. Therefore, the feature extraction in frequency domain is related to four frequency bands. A feature matrix, generated from the EEG data of 40 trials, leads to 40 rows or feature vectors. A temporal window is employed for EEG segmentation, and features are extracted from all 32 electrodes on each segment. Candra et al. [40] reported that the effective window size for arousal and valence recognition is from 3 to 10 and from 3 to 12 seconds, respectively, using the DEAP database. To increase the samples, the one minute signals are divided into 15 samples using a window of size 4 seconds in each trial. The resulting number of features for each feature group is indicated in Fig. 2. Except for MSCE and asymmetry features, other features are associated with one electrode.

2.4. Feature selection

Reducing the extracted features space, which lead to achieving the most relevant features, is one of the challenges in affective computing. In fact, feature selection removes the irrelevant features to avoid over-specification which increases the computational complexity of the model. Different methods have been used in the literature. Regression-based feature selection methods have been given little consideration in the field of emotion recognition, and the researches in this field are mainly focused on discrete emotional states. In this study, we assume the emotional states as continuous variables, and use regression-based methods to select the most informative features.

Feature Selection methods can generally be categorized into filter, wrapper and embedded methods [41,42]. While wrapper and embedded methods select features based on classification algorithm in the underlying model, filter methods are classifier-independent. Another advantage of filters is that they are less computationally expensive than wrappers and embedded methods, thus they are more suitable for big

Table 2
Summarized description of the features used in this study.

Features	Method/Description	Author	Year
Statistics	Power, Mean, Standard deviation, 1st difference, Normalized 1st difference, 2nd difference, Normalized 2nd difference.	Takahashi [33]	2003
Hjorth	Mobility:Measure of the signal mean frequency, Complexity:Measure of the deviation of the signal from the sine shape.	Ansari-asl et al. [20]	2007
NSI	Variation of the local average over time.	Kroupi et al. [34]	2011
FD	A ratio providing a statistical index of irregularity of a time series.	Liu et al. [35]	2013
HOC	Captures the oscillatory pattern of signal by counting the number of zero-crossings of the filtered EEG time series.	Petrantonakis et al. [14]	2010
Band power	Average power of frequency bands using STFT method.	Chanel et al. [36]	2007
Hilbert–Huang Spectrum	Average power of frequency bands using empirical mode decomposition and Hilbert transform.	Hadjidimitriou et al. [16]	2012
Discrete Wavelet Transform	Energy of detail coefficients, corresponding to band frequencies, extracted from $db4$ wavelet transform.	Murugappan et al. [37]	2010
Magnitude Squared Coherence	Represents the correspondence of two signals at each frequency.	Khosrowabadi et al.[38]	2010
Differential Asymmetry	Differences in band powers of symmetric electrodes on scalp.	Lin et al. [39]	2010
Rational Asymmetry	Ratio of band powers of symmetric electrodes on scalp.	Lin et al. [39]	2010

datasets. In this study, we apply RReliefF [31] which is an extension of the univariate technique Relief algorithm in which, a proxy of statistics is calculated for each feature to estimate feature relevance to the target value. These statistics are defined as a quality weight $W(A)$ for each feature A, which is updated based on the following equation:

$$W(A)^{[i+1]} = W(A)^{[i]} - \sum_{j=1}^{k} \mathrm{diff}(A, R_i, H_j)/m + \mathrm{diff}(A, R_i, M_j)/m, \qquad for \quad i = 1, \ldots, m \tag{1}$$

where m is the total number of instances in the dataset, and M and H are the nearest miss and hit of randomly selected instance R_i, respectively. The function diff() calculates the difference between instances as:

$$\mathrm{diff}(A, R_1, R_2) = \frac{|\mathrm{value}(A, R_1) - \mathrm{value}(A, R_2)|}{\max(A) - \min(A)}. \tag{2}$$

The Relief algorithm performs the feature selection based on k nearest hits/misses. In regression problems, an updated version of this algorithm, RReliefF, is proposed [43]. Since the predicted value (class) is continuous in such problems, the mentioned concept of the nearest neighbors is not useful. In RReliefF, a probability function is suggested using the relative distances between the predicted values (class) of the two instances. In this algorithm, $W(A)$ is reformulated as

$$W(A) = \frac{P_{\mathrm{diffC|diffA}} P_{\mathrm{diffA}}}{P_{\mathrm{diffC}}} - \frac{(1 - P_{\mathrm{diffC|diffA}}) P_{\mathrm{diffA}}}{1 - P_{\mathrm{diffC}}}, \tag{3}$$

where

$$P_{\mathrm{diffA}} = P\big(\text{different value of A|nearest instances}\big)$$
$$P_{\mathrm{diffC}} = P\big(\text{different prediction|nearest instances}\big)$$
$$P_{\mathrm{diffC|diffA}} = P\big(\text{diff. prediction|diff. value of A and nearest instances}\big) \tag{4}$$

2.5. Regression

The selected features in Section 2.4 are fed to regression models to recognize emotions. In this study, three conventional regression models including Linear Regression (LR), Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP) are implemented. Using different models lets us compare the results and select the best model.

LR model finds the line that best fits the data points. This model fits a line between a dependent variable y_i and independent variables $\{x_{i1}, x_{i2}, \ldots, x_{ip}\}$ as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon, \qquad for \quad i = 1, \ldots, m \tag{5}$$

where m is the total number of observations, p is the number of features, β_js are regression coefficients and ϵ is the error term. The parameters of this model can be computed using maximum likelihood method.

SVM Categorizes the data by finding the best hyperplane that separates data. In an n-dimensional space, the equation of the hyperplane is

$$w^T x + b = 0, \tag{6}$$

where x is the vector of points on the hyperplane, w^T is the gradient vector, and b is the y-intercept vector. If data can be linearly separable, hard margin SVM is used. In this model, two parallel hyperplanes can be selected which separate the two classes of data with a margin between them. In the cases that data are not linearly separable, soft margin SVM has been introduced. A model of SVM for regression, called

SVR, is proposed in [44] whose main idea is to accept a tolerance ϵ in SVM model, i.e.,

$$\begin{aligned} &\min \quad 1/2w^T w \\ &S.t. \quad |y_i - (w^T x_i + b)| < \epsilon, \end{aligned} \tag{7}$$

where $y_i \in \{-1, 1\}$. In this research, MATLAB Regression Learner is used to build the support vector regression with Gaussian kernel. The parameter ϵ was also selected using Eq. (8).

$$\epsilon = iqr(Y)/13.49 \tag{8}$$

MLP is a class of neural network with at least three layers of nodes, and it uses back-propagation for training [45]. We in the simulation part use the trainbr function which is based on the Bayesian Regularization back-propagation algorithm. To do this, we use the MATLAB neural network tool and the Fitting app menu, which creates a network with a single hidden layer and an arbitrary number of neurons. To determine number of hidden neurons, using trial and error method, first the network was trained using 5, 10, and 15 neurons in the hidden layer and the training error was calculated. The results of this study showed that network training using 10 neurons in the hidden layer has the least error compared to the other two cases. Due to the better performance of this number of neurons to ensure the optimal number of neurons in the hidden layer, the network was trained using 8, 9, 11 and 12 neurons in the hidden layer and the training error was calculated. The results show that the use of 10 neurons in the hidden layer leads to the least training error in the least repetition. Therefore, in this study, a neural network with a hidden layer and 10 neurons was used.

2.6. Emotion experiment design

After performing the explained emotion recognition steps on DEAP, the most significant features and consequently the most important electrodes are discovered. Next, we design an experiment to collect EEG data using only the selected electrodes. Since the literature review has shown that the videos are more affective than images and audios on emotions, we use emotional video clips. Furthermore, to consider the impact of native culture factors on emotions of participants (subjects), Iranian video clips were used. Since long films would make the participants tired, the maximum lengths of the films we used are one minute, to be short enough. The films also should be such that they elicit only a single emotional state.

To do the experiment, 5 males and 7 females of 20 to 30 years old native Iranian students from the University of Isfahan, who were self-reported healthy, were recruited as participants. The experiments were performed based on the principles of Helsinki. A consent form was signed by each participant. A Rayan Mindware System at a sampling rate of 250 Hz was used to record EEG signals from 4 active AgCl electrodes (2 bipolar selected channels) according to the international 10–20 system. These 4 electrodes have been high-lighted in Fig. 3, which shows the layout of EEG electrodes on the caps. Participants' distance from the screen was about 1 meter, and the size of the screen was 15 inches.

Participants were informed of the experiment protocol, such as watching the film clips attentively without diverting their attention from the screen, and of the meaning of different scales used for self-assessment. They reported their emotions while watching the films by a score, from 1 to 9 and keywords, e.g. fear and sadness. For each experiment, 10 video clips for each of positive, neutral and negative emotions were provided (30 clips in total). Each trial consists of a 15-second time-interval for fixation cross before the film presentation, and a 10 seconds time-interval for self-assessment after presentation. Fig. 4 shows the detailed structure. Self-assessment manikins with numbers 1 to 9 were used to select the emotion levels by the participants. Fig. 5 shows the used assessment interface. As seen in the figure, the valence scale ranges from unhappy or unpleasant to happy or pleasant. The arousal scale ranges from calm or bored to stimulated or excited.

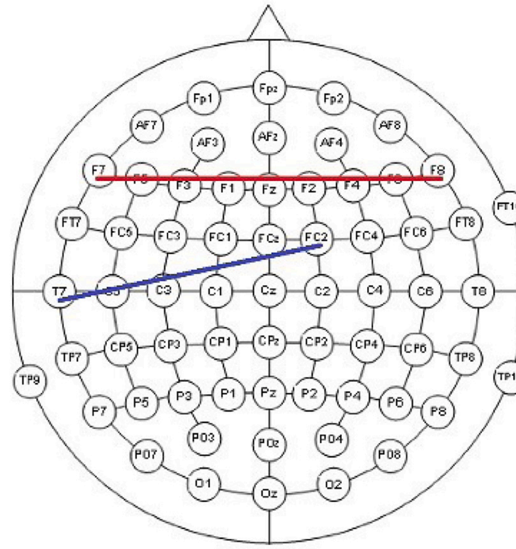

Fig. 3. Layout of EEG electrodes on the cap. The two bipolar channels used in this study are marked with red and blue color.

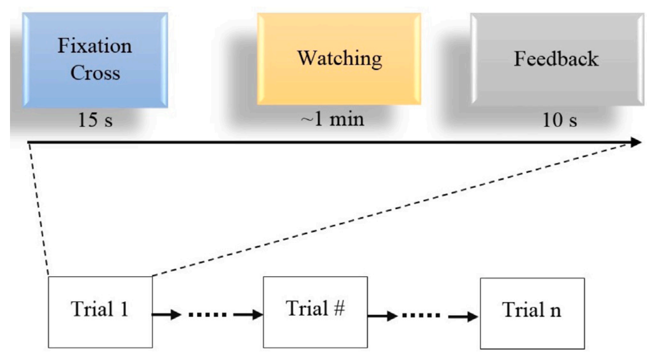

Fig. 4. Structure of the protocol of the designed experiment.

3. Experiment results

In this section, we explain the results of emotion recognition process performed in this study. In Section 3.1, the results of processing on DEAP dataset are discussed. The performance of our designed experiment is presented in Section 3.2.

3.1. Processing results on DEAP dataset

The number of extracted features from EEG signals in DEAP was 3216 features of different categories. A z-normalization was done to have a dataset with deviation one and zero mean. To perform the feature selection (and consequently channel selection) RRelief weights of features were computed. Using the weights, high-ranked features among the top-N features were computed as presented in Table 3. As seen in the table, the top 2% (2% of 3216 $\cong$ 64 features), 5%, 10%, 20% and 30% of the number of features are listed. In the table, the numbers in the parentheses, located under each feature name, show the total number of the features in that specific category. For example, among the extracted features, there are 32 features for HHS gamma that 6 ones of them are located in top-2% features. The results of evaluating different category of features show that the features related to gamma bands are more significant in emotion recognition. Similar results are reported in some previous studies, e.g. [11,22,26]. Specially, as seen in Table 3, MSCE gamma is the most significant

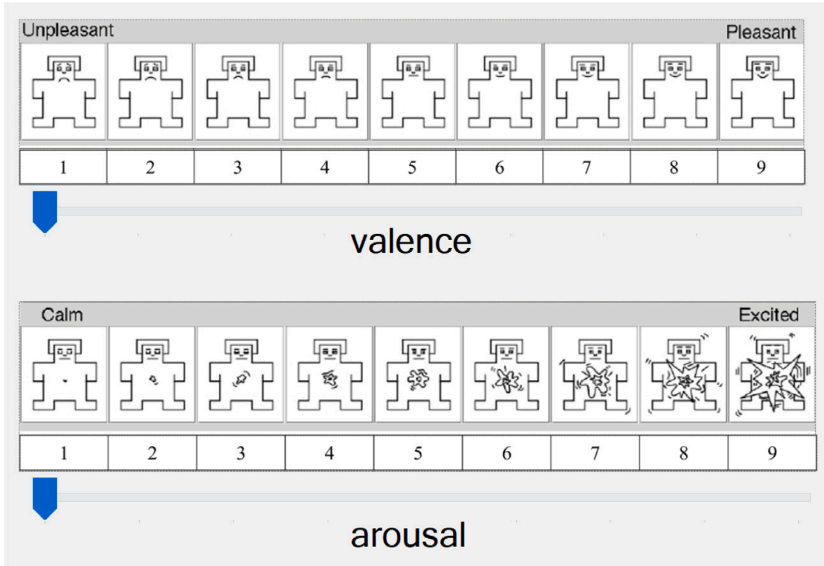

Fig. 5. User's emotion assessment setup after watching each trial of emotional stimuli (video clips).

Table 3
High-ranked features among the top-N features selected with RReliefF algorithm(the numbers in the parentheses of the first row show the total number of features in that specific category. numbers in the parentheses of the first column represent the percentage of top-N features among the total number of features).

Top-N Features	High-ranked Features						
	HHS gamma (32)	MSCE beta (496)	MSCE gamma (496)	LREE gamma (32)	Mobility (32)	2nd diff (32)	PSD gamma (32)
64 *features*(2%)	0.1(6 of 32)	0.015 (1)	0.6 (39)	0.03 (2)	0.05 (3)	0.05 (3)	0.07 (4)
160 *features*(5%)	0.075 (12)	0.08 (12)	0.45 (71)	0.03 (5)	0.05 (8)	0.05 (8)	0.07 (10)
320 *features*(10%)	0.05 (17)	0.06 (19)	0.4 (126)	0.025 (8)	0.03 (11)	0.04 (15)	0.06 (20)
640 *features*(20%)	0.04 (23)	0.1 (67)	0.28 (181)	0.02 (11)	0.028 (18)	0.03 (20)	0.05 (31)
960 *features*(30%)	0.025(24)	0.12 (115)	0.24 (233)	0.01 (11)	0.027 (26)	0.03 (29)	0.04 (32)

Table 4
Top-rated electrodes for each category of features(the most repeated electrodes among different classes are highlighted).

Feature	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6
HSS gamma	P7	P3	F8	T8	CP6	T7
LREE gamma	FC2	Pz	FC1	P3	Cz	FC6
PSD gamma	T8	F7	T7	Fp1	Cz	F3
MSCE gamma	FC2,Cz	AF3,F7	F7,FC2	F7,FC5	FC2,P4	Fp1,T8
MSCE beta	T7,O1	Fp1,F8	T7,CP5	P3,Pz	CP1,AF4	Fp1,FC6
Mobility	F8	P7	PO4	CP2	CP1	F4
2nd diff	F8	P7	CP5	P3	PO4	F4

feature. Furthermore, using a paired t-test method, the related results were statistically evaluated. This evaluation shows that the frequency features outperform the time domain features significantly ($p < 0.05$).

Since each feature was calculated for each channel or each pair of channels (in case of mutual features), RReliefF was also used to rank and find the most significant channels [22]. This was performed both for single electrode features that ranked highest in the previous step, and also for the paired electrode features in the high rank set. Table 4 presents the ranking of the channels and their related class of features. The most repeated electrodes, $FC2, F7, F8, T8, T7$, and $P7$, are highlighted in the table. As expected, the significant ones are more related to MSCE gamma features. Since our aim is reaching the optimal number of channels, we next compare the accuracy in emotion recognition using the significant features related to these 6 channels in order to select the minimum number of the significant ones. SVR was used to validate the feature (channel) selection results. To assess the performance of our proposed method, and to compare the results with the literature (in which the emotional states are discrete), it is required to classify the regression outcomes. To do this, we segmented the valence-arousal space to the two parts; namely, positive and negative emotions. If the valence rating was smaller (respectively, larger) than 5, then the class of emotion is recognized as negative (respectively,

Table 5
Mean accuracy rate of SVR classifier for selected single electrode features performed on DEAP dataset.

Selected features	Selected channels					
	P7	F8	T8	T7	FC2	F7
PSD gamma	56.05	57.28	57.13	58.28	57.87	57.33
HSS gamma	62.66	63.15	61.84	61.75	**63.16**	62.85
LREE gamma	60.06	59.6	59.54	60.64	60.78	59.79
Mobility	55.5	56.42	55.49	55.53	56.52	55.57
2nd diff	49.42	49.39	50.47	50.4	49.37	49.48

Table 6
Mean accuracy rates of SVR classifiers for MSCE features (calculated from combinations of electrodes).

	P7	F8	T8	T7	FC2	F7
P7		64.86	64.74	62.79	64.86	64.74
F8	65.94		60.84	64.62	64.13	64.68
T8	65.93	65.73		64.63	64.63	63.6
T7	67.18	65.98	66.02		65.12	64.7
FC2	65.84	67.21	66.08	67.05		61.95
F7	65.87	**67.45**	66.29	66.19	65.91	

Note: Right-top of the table shows the results for the beta band of the electrode pairs, and left-bottom shows the results for the gamma band.

positive). With this point of view, classification performance is evaluated through the classification accuracy rate for the two positive and negative emotional classes.

The mean accuracy rates of using SVR and classifying the results are shown in Tables 5 and 6 for single electrode features and for MSCE features (calculated from combinations of electrodes), respectively. In this assessment, the features related only to the 6 selected channels in the previous stage, are considered.

According to the results tables, the best accuracy of the single electrode features is 63.16% using HHS feature of $FC2$ channel. The accuracy of MSCE features is 67.45% related to $F7$-$F8$ channel. An overall review on the two tables shows that the accuracy rates of MSCE features are mostly higher than those of single electrode features. We also evaluated the combinations of 2, 3, 4, 5, and 6 electrodes, which achieve the best accuracies (%)/standard deviations (%) of 63.45/2.53, 64.23/3.27, 61.15/2.87, 62.03/4.21, and 61.35/3.50, respectively, using the HHS features of gamma band. Thus, using MSCE, which is extracted from the combination of two electrodes, promotes the accuracy in recognizing the emotions. This was also confirmed using the regression result. Using SVR as regressor and MSCE features, we achieved the best normalized Root Mean Square Error (RMSE) value of 0.25 for pair electrodes $F7$-$F8$. Normalized RMSE was used as an indication of percentage error in regression. To compute the normalized value, RMSE was divided by the range of outputs which was 8 for the case of 1 to 9 emotion labeling. Finally, using these results and a correlation computation among the features, $F7$-$F8$ and $FC2$-$T7$ were selected as the optimal channels.

3.1.1. Comparison with the literature

A comparison of the accuracy rates of different models in the literature are presented in Table 7. To be comparable, we selected the papers in which emotion recognition was performed using EEG signals in DEAP dataset. The method in this paper gains an average recognition accuracy of 67.45% on the DEAP dataset for two classes of valence dimension (positive and negative) with MSCE features using two EEG channels, F7–F8 and FC2-T7, for all 32 participants. The performance of this method is acceptable among the others as seen in the table. There are some papers such as [19,23] which have reached a higher accuracy, but with a larger number of channels. We reduced the number of channels as much as possible in return for a minor decrease in accuracy rate.

3.2. Experiment results

In this section, we discuss and evaluate the results of the designed experiment explained in Section 2.6. From the processing results of DEAP dataset, finally two bipolar channels $F7$-$F8$ and $T7$-$FC2$ were selected as the most significant channels in the emotion recognition. The EEG data gathered from the two bipolar selected channels were passed through a 4 to 45 Hz band-pass filter, and were next segmented to the same-length epochs using time-windows. The results in Section 3.1 showed that the following features are more significant in EEG-based emotion recognition among the set of features: PSD gamma, HHS gamma, REE gamma, Mobility, 2nd difference, MSCE beta and MSCE gamma. Therefore, we extracted only these features from the experiment. To recognize the emotion states, we used LR, SVR and MLP as regressors. We adopt a 5-fold cross-validation scheme with LR, SVR, and MLP. The normalized RMSE of these regressors for each extracted feature is presented in Fig. 6 and Table 8. Comparing the results shows that the frequency domain features, especially HHS gamma and MSCE gamma, perform better in emotion detection which their extraction method are as follows. We computed HHS for each epoch of signal, which is done via empirical mode decomposition (EMD) to arrive at intrinsic mode functions (IMFs) to represent the original signal:

$$\xi(t) = \sum_{k=1}^{N} IMF_k(t) + r_N(t). \tag{9}$$

By computing the Hilbert transform of each IMF_k, a conjugate pair IMF_k^H is formed, so that an analytic signal can be written as:

$$Z_k(t) = IMF_k(t) + jIMF_k^H(t) = A_k(t)e^{j\theta_k(t)}. \tag{10}$$

Due to their generating process, IMFs constitute narrow band components and, therefore, a meaningful instantaneous frequency can be estimated by the derivative of $\theta_k(t)$ which yields a time–frequency representation of amplitude.

$$f_k(t) = \frac{1}{2\pi}\frac{d\theta_k}{dt}. \tag{11}$$

Further, the squared amplitude is computed, and the average of each frequency band is calculated as features. MSCE represents the correspondence of two signals ξ_i and ξ_j at each frequency, taking values between 0 and 1. It is defined as:

$$C_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_i(f)P_j(f)}, \tag{12}$$

where $P_ij(f)$ is the cross power spectral density and P_i and P_j are the power spectral density of ξ_i and ξ_j, respectively. In order to reduce the large amount of features resulting from all possible combinations of electrodes, C_{ij} is averaged over the frequency band.

Thus, it seems that the gamma band of EEG signals involves more information about the emotions. Such results have been reported in some previous researches, e.g. [48–50]. Moreover, by examining these results, it can be concluded that the second channel, i.e., $FC2-T7$ plays more important roles in distinguishing between valence emotional states.

Next, we considered the combination of the two significant features, i.e., HHS and MSCE in gamma frequency band. The evaluation of the regression in this scenario was performed using three criteria: RMSE, R-squared, and Pearson correlation coefficient. Table 9 presents the results of these evaluations for LR, SVR and MLP, respectively. Comparing Tables 9 and 8 shows that the combination of the frequency features improves the emotion recognition by decreasing the detection error. The results show that an RMSE value of 0.22 was achieved using only 2 bipolar EEG channels. Fig. 7 demonstrates the correlation between the valence levels recognized by the algorithm and the valence as reported by the subjects. Overall a good correlation is observed for the negative valences and then the positives ones, while the estimated values for the middle valence range is almost flat. This suggests that the

Table 7
Comparison of accuracy achieved by various studies using EEG in the DEAP dataset.

Study	Year	Method		Advantages	Disadvantages	Study Results
		Classifiers	Extracted Features			
Chung et al. [46]	2012	Bayesian	Frequency domain (spectral powers)	Sufficient Participants (32), Simple implementation of the proposed classifier	Using large number of channels (32), Discrete values were considered for the valence/arousal axis	66.6%, 66.4% for valence and arousal (2 classes) with all 32 participants.
Koelstra et al. [25]	2012	Gaussian naive Bayes	Frequency domain (spectral powers), Mean, average and energy of signals, Other features related to Physiological Signals	Presented a novel semi-automatic stimuli selection method, Different category of features are extracted and compared, Sufficient Participants (32), Making a publicly available database for affective detection		62.0%, 57.6% for arousal and valence (2 classes) with all 32 participants.
Candra et al. [40]	2015	SVM	Wavelet	Finding the effective window size for arousal and valence for EEG signals		65% for valance and arousal, 32 channels, 32 participants
Zhang et al. [22]	2016	SVM	Frequency domain (spectral powers)	EEG channel reduction, Results achieved using sufficient number of participants	Discrete values were considered for the valence/arousal axis, Only a specific group of features are compared	58.51% for 4 emotional states (4 classes), 8 channels, with selected 16 participants
Chao et al. [30]	2019	A capsule network (CapsNet)	Frequency domain	Using three-dimensional emotion model that better can classify than two-dimensional model	Using large number of channels (32), Discrete values were considered for the valence/arousal axis	66.73% for valance and 68.2% arousal, 32 channels
Yan et al. [47]	2019	RT-ERM	Time and frequency domain	Using sufficient number of participants		62.12% for valance and 69.1% arousal, 32 channels, 32 participant, (4 class)
Zheng et al. [23]	2019	KNN, SVM, GELM	PSD, DE, DASM, RASM, ASM, DCAU	Developing novel emotion EEG dataset as a subset of SEED, Systematic comparison of different feature extraction, feature selection, feature smoothing and pattern classification methods on DEAP		69.67% for quadrants of VA space (4 classes) with all 32 participants.
Nawaz et al. [19]	2020	SVM, KNN, Decision Tree	Power, Entropy, FD, Statistical features, Wavelet energy	Using 3-dimensional emotion model that can classify better than 2-dimensional model		77.62% for valance and 78.96% for arousal, (2 classes), 14 channels, 32 participants
Our method		SVR, LR, MLP	Time, frequency and time–frequency domain, Features calculated from combinations of electrodes	Decreasing the number of channels, Emotional state as a continuous variable, Evaluating a comprehensive list of features	Only estimation of valence (not arousal), Better estimation of lower valences compared to higher values	67.45% for valence (2 classes), 2 channels, with all 32 participants.

algorithm is more sensitive to valence changes in extreme values, but has had difficulty in identifying differences in the middle range. This however might be correlated with nonuniform distribution of valence associated with the videos used for this study.

HHS is a time–frequency feature that extracts information from the dynamics of non-stationary EEG. Moreover, it is a non-linear method that is shown to be more resistance to noise [16]. MSCE is a frequency-domain feature that looks at the mutual information between two electrodes. Recently many studies have demonstrated the value of looking at functional connectivity between different regions of the brain in recognizing brain states. Higher performance of Gama band for these features seems relevant, as rhythms in gamma frequency

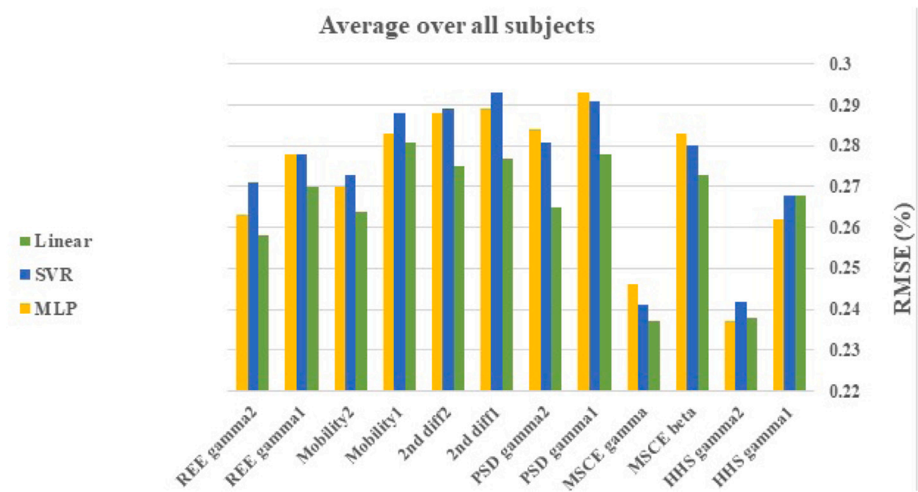

Fig. 6. RMSE of LR, SVR and MLP regressors for each extracted features of self-recorded signals.

Table 8
Average error (RMSE) in emotion recognition using Multi Linear, SVR and MLP regressors.

Feature	Multi-Linear	SVR	MLP
HHS gamma1	0.268 ± 0.03	0.268 ± 0.03	0.262 ± 0.04
HHS gamma2	$\mathbf{0.238 \pm 0.03}$	$\mathbf{0.242 \pm 0.04}$	$\mathbf{0.237 \pm 0.04}$
MSCE beta	0.273 ± 0.02	0.28 ± 0.03	0.283 ± 0.03
MSCE gamma	$\mathbf{0.237 \pm 0.04}$	$\mathbf{0.241 \pm 0.04}$	$\mathbf{0.246 \pm 0.03}$
PSD gamma1	0.278 ± 0.01	0.291 ± 0.02	0.293 ± 0.01
PSD gamma2	0.265 ± 0.01	0.281 ± 0.02	0.284 ± 0.01
2nd diff1	0.277 ± 0.01	0.293 ± 0.02	0.289 ± 0.02
2nd diff2	0.275 ± 0.01	0.289 ± 0.03	0.288 ± 0.02
Mobility1	0.281 ± 0.02	0.288 ± 0.03	0.283 ± 0.01
Mobility2	0.264 ± 0.03	0.273 ± 0.05	0.27 ± 0.05
REE gamma1	0.27 ± 0.02	0.278 ± 0.02	0.278 ± 0.02
REE gamma2	$\mathbf{0.258 \pm 0.02}$	$\mathbf{0.271 \pm 0.03}$	$\mathbf{0.263 \pm 0.03}$

Table 9
The amount of three criteria in emotion recognition using Multi Linear, SVR and MLP when combination of features are performed.

Regression Algorithm	RMSE	R Squard	Pearson correlation coefficient
Multi-Linear	0.25	0.3	0.49
SVR	0.24	0.34	0.56
MLP	0.22	0.48	0.67

range is correlated to large scale brain network activity and cognitive phenomena in the brain [50].

4. Conclusion

The aim of this study was to improve the performance of our method in practice by decreasing the number of electrodes. This reduces the computational complexity and facilitates user convenience in a realistic practical system. We evaluated a comprehensive list of features for emotion recognition to propose a light efficient algorithm for valence recognition, using a few number of electrodes. The evaluation was performed not only on the well-known DEAP dataset, but also on the data recorded in this study. The experiment was designed and features were selected based on the results from DEAP dataset. Among 32 channels, two bipolar channels, $F7$-$F8$ and $T7$-$FC2$, were selected. The best average classification accuracy was 67.45% for positive and negative valence with 32 participants using only 2 EEG channels. Some studies in the literature have achieved a higher accuracy rate on DEAP data, but by using more channels. In fact, in this paper, reducing the number of channels and improving the practicality of the method are achieved in return for acceptable decrease in accuracy of emotion classification.

Next, an experiment was designed using two selected bipolar channels, 12 participants and new culturally relevant video clips as stimuli. Since the emotional state was regarded as a continuous variable in this paper, regression based SVR was used to recognize the emotions. The results of this new dataset indicated that features obtained from frequency domain, specifically high frequency bands, perform better than other types of features. It was also observed that the combination of frequency features improves the emotional recognition error. Combination of MSCE and Hilbert–Huang Spectrum provided the best performance with 0.22 error (78% accuracy) which show the reliability and superior performance of our regression based method in comparison with the existing approaches.

Several limitations of this work must be taken into consideration. First, Feature and channel selection was performed only for the estimation of valence. The same approach can be used to evaluate features and channels for the other dimension of Russel model, i.e. arousal. The selected features and electrodes performed better in estimation of lower valences (negative emotions) compared to higher values (positive emotions). This is potentially due to experiment stimuli design or the imperfection of feature selection method. Another limitation of this study was that the final conducted experiments were performed with a 2 channel wearable EEG recording system. While the aim of this selection was to demonstrate, at the end, that detection of valence can be achieved with an instrument usable for out of the lab applications, performing the experiment with an EEG recording system with more electrodes could allow further investigation of the potential effect that culture and stimuli selection could have on the optimal channels and features. Future studies aim to surpass some limitations of this work, expanding the model to include arousal, and increasing the number of subjects to obtain more reliable and representative results. Multichannel recording and different feature selection algorithms may also be considered in future approaches.

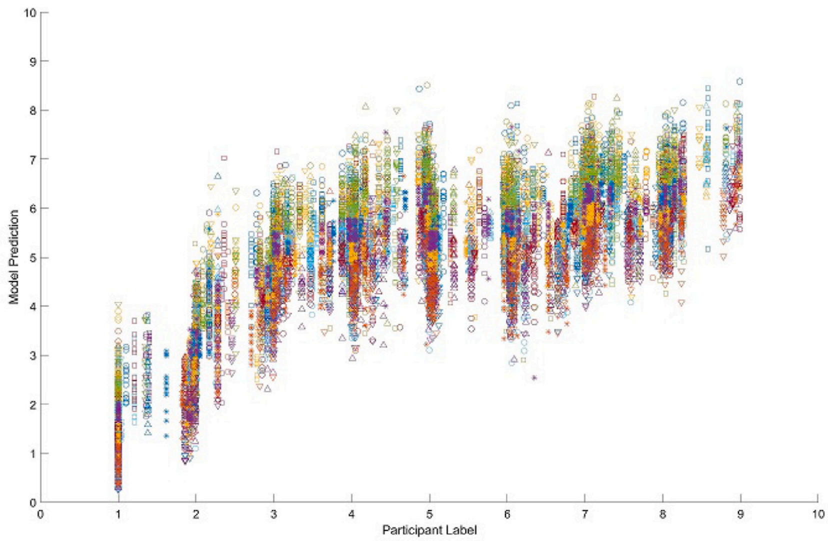

Fig. 7. Correlation between the valence levels recognized by the algorithm and the reported one by the subjects.

CRediT authorship contribution statement

Mahshad Javidan: Investigation, Formal analysis, Writing – original draft. **Mohammadreza Yazdchi:** Supervision, Methodology. **Zahra Baharlouei:** Validation, Writing – review & editing. **Amin Mahnam:** Conceptualization, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, H. He, Multisource transfer learning for cross-subject EEG emotion recognition, IEEE Trans. Cybern. 50 (7) (2020) 3281–3293.

[2] M. Egger, M. Ley, S. Hanke, Emotion recognition from physiological signal analysis: A review, J. Electron. Notes Theor. Comput. Sci. 343 (2019) 35–55.

[3] X. Huang, J. Kortelainen, G. Zhao, X. Li, A. Moilanen, T. Seppanen, M. Pietikainen, Multi-modal emotion analysis from facial expressions and electroencephalogram, J. Comput. Vis. Image Underst. 147 (2016) 114–124.

[4] X.-W. Wang, D. Nie, B.-L. Lu, Emotional state classification from EEG data using machine learning approach, J. Neurocomput. 129 (2014) 94–106.

[5] S.M. Alarcao, M.J. Fonseca, Emotions recognition using EEG signals: A survey, IEEE Trans. Affect. Comput. 10 (3) (2019) 374–393.

[6] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, L. Albera, Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources, IEEE Trans. Affect. Comput. 11 (2) (2020) 244–257.

[7] A. Martinez-Rodrigo, B.G.-M.R. Alcaraz, P. Gonzalez, A. Fernandez-Caballero, Multiscale entropy analysis for recognition of visually elicited negative stress from eeg recordings, Int. J. Neural Syst. 29 (2) (2019) 1.

[8] L. Piho, T. Tjahjadi, A mutual information based adaptive windowing of informative EEG for emotion recognition, IEEE Trans. Affect. Comput. 11 (4) (2020) 722–735.

[9] R. Yuvaraj, M. Murugappan, Hemispheric asymmetry non-linear analysis of EEG during emotional responses from idiopathic Parkinson's disease patients, Cogn. Neurodyn. 10 (3) (2016) 225–234.

[10] B. García-Martínez, A. Martinez-Rodrigo, R. Alcaraz, A. Fernández-Caballero, A review on nonlinear methods using electroencephalographic recordings for emotion recognition, IEEE Trans. Affect. Comput. 10 (1) (2019) 1.

[11] X.-W. Wang, D. Nie, Bao-LiangLu, Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (2014) 94–106.

[12] Z. Lan, O. Sourina, L. Wang, Y. Liu, Real-time EEG-based emotion monitoring using stable features, Vis. Comput. 32 (2016) 347–358.

[13] A. Goshvarpour, A. Abbasi, Dynamical analysis of emotional states from electroencephalogram signals, Biomed. Eng.: Appl. Basis Commun. 28 (2) (2016) 165.

[14] P.C. Petrantonakis, L.J. Hadjileontiadis, Emotion recognition from EEG using higher order crossings, IEEE Trans. Inf. Technol. Biomed. 14 (2) (2010) 186–197.

[15] H. Cai, X. Sha, X. Han, S. Wei, B. Hu, Pervasive EEG diagnosis of depression using deep belief network with three-electrodes EEG collector, in: Proc. IEEE Int. Conf. Bioinformatics and Biomedicine, BIBM, 2016, pp. 1239–1246.

[16] S. Hadjidimitriou, L. Hadjileontiadis, Toward an EEG-based recognition of music liking using time-frequency analysis, IEEE Trans. Biomed. Eng. 59 (12) (2012) 3498–3510.

[17] M.Z. Soroush, K. Maghooli, S.K. Setarehdan, A.M. Nasrabadi, A novel method of eeg-based emotion recognition using nonlinear features variability and Dempster–Shafer theory, Biomed. Eng.: Appl. Basis Commun. 30 (4) (2018) 1850026.

[18] K. Guo, H. Candra, H. Yu, H. Li, H.T. Nguyen, S.W. Su, EEGbased emotion classification using innovative features and combined SVM and HMM classifier, in: Proc. 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, EMBC, 2017, pp. 489–492.

[19] R. Nawaz, K.H. Cheah, H. Nisar, Y.V. Voon, Comparison of different feature extraction methods for EEG-based emotion recognition, Biocybern. Biomed. Eng. 40 (3) (2020) 910–926.

[20] K. Ansari-Asl, G. Chanel, T. Pun, A channel selection method for EEG classification in emotion assessment based on synchronization likelihood, in: 15th European Signal Processing Conference, 2007, pp. 1241–1245.

[21] X.W. Wang, D. Nie, B.-L. Lu, EEG-based emotion recognition using frequency domain features and support vector machines, in: International Conference on Neural Information Processing, ICONIP, 2011, pp. 734–743.

[22] J. Zhang, M. Chen, S. Zhao, S.H. aand Zhiguo Shi, Y. Cao, Relieff-based EEG sensor selection methods for emotion recognition, Sensors 16 (10) (2016) 1558.

[23] W.-L. Zheng, J.-Y. Zhu, B.-L. Lu, Identifying stable patterns over time for emotion recognition from EEG, IEEE Trans. Affect. Comput. 10 (3) (2019) 417–429.

[24] J.A. Russell, A circumplex model of affect, J. Personal. Soc. Psychol. 39 (6) (1980) 1161–1178.

[25] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, DEAP: A database for emotion analysis using physiological signals, IEEE Trans. Affect. Comput. 3 (1) (2012) 18–31.

[26] W. Zheng, B. Lu, Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, IEEE Trans. Auton. Mental Dev. 7 (3) (2015) 162–175.

[27] Z. Lan, O. Sourina, L. Wang, R. Scherer, G.R. Muller-Putz, Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets, IEEE Trans. Cogn. Dev. Syst. 11 (1) (2019) 85–94.

[28] S. Hatamikia, A.M. Nasrabadi, Common spatial pattern method for channel reduction in eegbased emotion recognition, Modares J. Electr. Eng. 13 (1) (2013) 31–43.

[29] O. Sourina, Y. Liu, A fractal-based algorithm of emotion recognition from EEG using arousal-valence model, in: International Conference on Bio-Inspired Systems and Signal Processing, BIOSIGNALS, 2011, pp. 209–215.

[30] H. Chao, L. Dong, Y. Liu, B. Lu, Emotion recognition from multiband EEG signals using capsnet, Sensors 19 (9) (2019) 2212.

[31] M. Robnik-Sikonja, I. Kononenko, An adaptation of Relief for attribute estimation in regression in Machine Learning, in: Proceedings of the Fourteenth International Conference, Vol. 5, ICML, 1997, pp. 296–304.
[32] S. Sanei, Signal transforms and joint time–frequency analysis, in: Adaptive Processing of Brain Signals, John Wiley & Sons, Ltd, ISBN: 9781118622162, 2013, pp. 72–89.
[33] K. Takahashi, A. Tsukaguchi, Remarks on emotion recognition from multi-modal bio-potential signals, in: SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483), Vol. 2, 2003, pp. 1654–1659.
[34] E. Kroupi, A. Yazdani, T. Ebrahimi, EEG Correlates of different emotional states elicited during watching music videos, Affect. Comput. Intell. Interact. 6975 (2011) 457–466.
[35] Y. Liu, O. Sourina, Real-time fractal-based valence level recognition from EEG, in: Transactions on Computational Science XVIII, 2013, pp. 101–120.
[36] G. Chanel, A. Karim, P. Thierry, Valence-arousal evaluation using physiological signals in an emotion recall paradigm, in: 2007 IEEE International Conference on Systems, Man and Cybernetics, 2007, pp. 2662–2667.
[37] M. Murugappan, N. Ramachandran, Y. Sazali, Classification of human emotion from EEG using discrete wavelet transform, J. Biomed. Sci. Eng. 3 (04) (2010) 390.
[38] R. Khosrowabadi, H.C. Quek, A. Wahab, K.K. Ang, EEG-based emotion recognition using self-organizing map for boundary detection, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 4242–4245.
[39] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, J.-H. Chen, EEG-Based emotion recognition in music listening, IEEE Trans. Biomed. Eng. 57 (7) (2010) 1798–1806.
[40] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H.T. Nguyen, S.W. Su, Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2015, pp. 7250–7253.
[41] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28.
[42] Y. Saeys, I.n. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics (ISSN: 1367-4803) 23 (19) (2007) 2507–2517.
[43] M. Robnik-Sikonja, I. Kononenko, An adaptation of Relief for attribute estimation in regression, in: Machine Learning: Proceedings of the Fourteenth International Conference, Vol. 5, ICML'97, 1997, pp. 296–304.
[44] H. Drucker, C.J. Burges, L. Kaufman, A. Smola, V. Vapoik, Support vector regression machines, in: Advances in Neural Information Processing Systems, 1997, pp. 155–161.
[45] N.K. Kasabov, Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering, The MIT Press, 1996.
[46] S. Chung, H. Yoon, Affective classification using Bayesian classifier and supervised learning, in: 2012 12th International Conference on Control, Automation and Systems, 2012, pp. 1768–1771.
[47] J. Yan, S. Chen, S. Deng, A EEG-based emotion recognition model with rhythm and time characteristics, Brain Informatics 6 (7) (2019) 1.
[48] M. Li, B.-L. Lu, Emotion classification based on gamma-band EEG, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 1223–1226.
[49] B. Guntekin, E. Bar, Event-related beta oscillations are affected by emotional eliciting stimuli, Neurosci. Lett. 483 (3) (2010) 173–178.
[50] N. Martini, D. Menicucci, L. Sebastiani, R. Bedini, A. Pingitore, N. Vanello, M. Milanesi, L. Landini, A. Gemignani, The dynamics of EEG gamma responses to unpleasant visual stimuli: From local activity to functional connectivity, NeuroImage 60 (2) (2012) 922–932.