Neural Networks 116 (2019) 257–268

Contents lists available at ScienceDirect

Neural Networks

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

An unsupervised EEG decoding system for human emotion recognition ☆

Zhen Liang [a,b,*], Shigeyuki Oba [a], Shin Ishii [a,c]

[a] Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
[b] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, 518060, China
[c] ATR Cognitive Mechanisms Laboratories, Kyoto 619-0288, Japan

ARTICLE INFO

Article history:
Received 16 September 2018
Received in revised form 9 February 2019
Accepted 1 April 2019
Available online 25 April 2019

Keywords:
Electroencephalography
Brain activity
Emotion recognition
Hypergraph
Decoding model

ABSTRACT

Emotion plays a vital role in human health and many aspects of life, including relationships, behaviors and decision-making. An intelligent emotion recognition system may provide a flexible method to monitor emotion changes in daily life and send warning information when unusual/unhealthy emotional states occur. Here, we proposed a novel unsupervised learning-based emotion recognition system in an attempt to decode emotional states from electroencephalography (EEG) signals. Four dimensions of human emotions were examined: arousal, valence, dominance and liking. To better characterize the trials in terms of EEG features, we used hypergraph theory. Emotion recognition was realized through hypergraph partitioning, which divided the EEG-based hypergraph into a specific number of clusters, with each cluster indicating one of the emotion classes and vertices (trials) in the same cluster sharing similar emotion properties. Comparison of the proposed unsupervised learning-based emotion recognition system with other recognition systems using a well-known public emotion database clearly demonstrated the validity of the proposed system.

1. Introduction

Human emotion is recognized as a wide research topic covering different fields, including psychology, neuroscience, health science, and engineering. A successful model of human emotion would be beneficial for building an emotion recognition system and developing applications in emotion understanding and management. In the past decades, interdisciplinary research combining biometrics, computer science and human emotion has emerged and attracted attention from researchers in many fields. Various computation-based emotion recognition/detection systems have been proposed. Using facial information, Cruz et al. proposed a video-based temporal feature extraction and sampling approach to address the emotion recognition problem with AVEC datasets (Cruz, Bhanu, & Thakoor, 2014). Eyben et al. focused on selecting an effective acoustic parameter set for voice analysis, and this set was proven to be correlated with affective changes (Eyben et al., 2016). Several multifeature fusion techniques have also been proposed for efficient application to both visual and audio modalities (Chen, Chen, Chi, & Fu, 2017; Kim & Provost, 2017). However, instead of relying on the affective characterizations of multimedia, the affective characterization of brain signals could be considered a more direct method to index changes in human emotion with higher resolution.

Since electroencephalography (EEG) signals reflect activities of cortical neuron ensembles, EEG features extracted from specific brain regions have the potential to be connected to emotion dynamics (Alarcao & Fonseca, 2017). Koelstra et al. constructed a public emotion database (Database for Emotion Analysis using Physiological Signals, DEAP) of simultaneously recorded EEG signals under various spontaneous affective states (Koelstra et al., 2012), and it is the most widely used database for studying emotions with EEG signals. DEAP provides a platform for researchers to perform EEG-based studies for spontaneous affective state estimation and presents a benchmark for comparing methods for emotion class recognition under different emotion dimensions using the EEG modality. In recent publications, a number of EEG-based emotion detection algorithms have been proposed and verified using this database. Shahnaz, Masud, and Hasan (2016) proposed an emotion recognition method that performed a discrete wavelet transform (DWT) on selected empirical mode

☆ This study was supported by the New Energy and Industrial Technology Development Organization (NEDO), Post-K Project from Ministry of Education, Sports, Science and Technology (MEXT), and JSPS KAKENHI Grant Number JP19H04180 and 17H06310.

* Corresponding author at: Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan.
E-mail addresses: jane-l@sys.i.kyoto-u.ac.jp (Z. Liang), oba@i.kyoto-u.ac.jp (S. Oba), ishii@i.kyoto-u.ac.jp (S. Ishii).

https://doi.org/10.1016/j.neunet.2019.04.003
0893-6080/

decomposition (EMD) results and then utilized a support vector machine (SVM) to solve a two-class classification problem. Liu, Meng, Nandi, and Li (2016) extensively characterized frequency domain-based, time domain-based, wavelet-based and multielectrode features in EEG signals and further reduced the feature dimensionality by the maximum relevance minimum redundancy (mRMR) feature selection method. Emotion classification in terms of arousal and valence has been realized using the K-nearest neighbor (KNN) and random forest (RF) approaches. Atkinson and Campos also used the mRMR feature selection method in their model [2016], and various kernel classifiers were adopted to improve the classification accuracy. Yin, Wang, Liu, Zhang, and Zhang (2017) developed a new EEG feature selection approach, transfer recursive feature elimination (T-RFE), to build a subject-generic emotion classifier without requiring a large dataset from a single subject. In that study, a subject-based threshold was used to define low/high emotion classes, with promising performance. To further improve the accuracy of two-class emotion classification (low/high), researchers extended a brain–computer interface (BCI)-based emotion model with deep-learning structures (Lin, Li, & Sun, 2017; Yin, Zhao, Wang, Yang, & Zhang, 2017). On the other hand, to increase the number of labeled samples, Zhuang et al. (2017) segmented a single trial (1 min) into 12 short sections (5 s). Based on the segmented 480 samples for each subject, an EMD-based feature extraction and emotion recognition method was developed.

However, all of the abovementioned emotion recognition approaches are based on supervised learning. A number of training data with labeled emotional classes were required to learn a "reasonable" inferred function to map the input variables to the output labels. The learned inferred function was then tested on new unlabeled data, and the accuracy was evaluated. However, it would be very time consuming and unrealistic to manually label each training sample, especially for large data sets, and there is also a risk that the class label could be incorrect or biased, which is called "label noise" (Luo, Peng, Huang, Alahi, & Fei-Fei, 2017). Unsupervised learning would provide a more natural method to decode human emotions through EEG analysis and is better aligned with the mechanism of human cognition because it acquires useful information from distributed data by exploring the hidden structure therein, even without any associated teachers (Barlow, 1989). The key challenge for unsupervised decoding is to understand the hidden complex relationships among the EEG signals underlying various emotion classes. To address this problem, graph theory has provided sophisticated solutions for interpreting different data structures and proven useful in solving complex problems (Diestel, 2017). Compared to a simple graph representation (Tutte, 1998), the hypergraph representation proposed in Berge (1989) is capable of presenting more general types of relations and revealing more complex hidden structures than signal connections. While a simple graph representation can measure only the relationships between any two of the vertices, leading to a possible loss of information (Ducournau, Rital, Bretto, & Laget, 2009), in a hypergraph, an edge can connect an unlimited number of vertices. In other words, the complex relationships among any number of vertices connected by an edge and among different edges can be directly examined. One of the most important studies in the development of hypergraphs was presented in Zhou, Hung, and Scholkopf (2007), in which the concept of a hypergraph Laplacian was defined to resolve the hypergraph partition problem. This process was also termed a spectral hypergraph partitioning approach. Based on Zhou et al.'s study, the spectral hypergraph partitioning approach has been widely and successfully used in various classification applications, such as multimedia representation and processing (Hong, Chen, Wang, & Tang, 2016; Zhu, Shen, Xie, & Cheng, 2017).

In the brain research field, few studies have applied the hypergraph to image processing. Hu, Wei, and He (2014) utilized the hypergraph to solve the image segmentation problem for a series of computed tomography (CT) brain images. Liu, Gao, Yap, and Shen (2017) employed the hypergraph to construct higher-order relationships among multiple modalities (e.g., magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF) data) and further tested this method for automatic brain disease (Alzheimer's disease (AD) and mild cognitive impairment (MCI)) diagnosis. These previous studies regarded the target problem as image processing and then used the formulation of hypergraphs. To the best of our knowledge, there is no study that uses hypergraphs to extract relevant features for the pattern recognition of general brain signals.

To address the limitations of the traditional EEG decoding methods and explore complex and nonpairwise relationships among EEG signals under different emotional classes, in this study, we first introduce the hypergraph into the feature extraction of general brain signals and develop an unsupervised learning-based EEG decoding system for emotion recognition. In the constructed hypergraph, one vertex indicates one trial of EEG signals, and the relationships among the vertices are measured based on the extracted EEG characteristics in three domains (frequency, time, and wavelet). Subsequently, trials that shared similar EEG patterns were grouped together by applying hypergraph partitioning. Under the assumption that an emotion class can produce a set of similar EEG patterns, a single emotion class was assigned to multiple trials in the same group, enabling us to perform trial-wise decoding of the emotion class. In the experiments, the proposed decoding system was implemented in a purely unsupervised manner and categorized the emotion classes into low and high classes. The robustness of the decoding performance was fully demonstrated under four emotion dimensions (arousal, valence, dominance, and liking).

The main contributions of this study can be summarized as follows:

(1) A novel unsupervised learning-based framework for decoding EEG signals for human emotion recognition was proposed.
(2) A hypergraph was introduced to solve the EEG decoding problem; in this structure, the hidden complex relationships among EEG signals from different trials were well defined. To our knowledge, this is the first introduction of the hypergraph to represent the relationships among EEG trials and solve EEG decoding applications.
(3) This research demonstrated the possibility of using hypergraph-based EEG decoding for human emotion recognition and offered a new approach to solving other decoding problems in an unsupervised manner.

2. Database

The DEAP database (Koelstra et al., 2012) is a multimodal dataset for studying human affective states. For this database, a total of 32 subjects were invited, and 40 music videos (duration of 60 s) with different strongest emotions (close to the corners of the emotion quadrants) were selected as the triggering stimuli. During the experiments, the videos were played in a randomized sequence and were presented on a 17-inch screen at a resolution of 800 × 600. For each subject, 40 trials with different videos were conducted. Each trial was composed of 5 s of fixation and

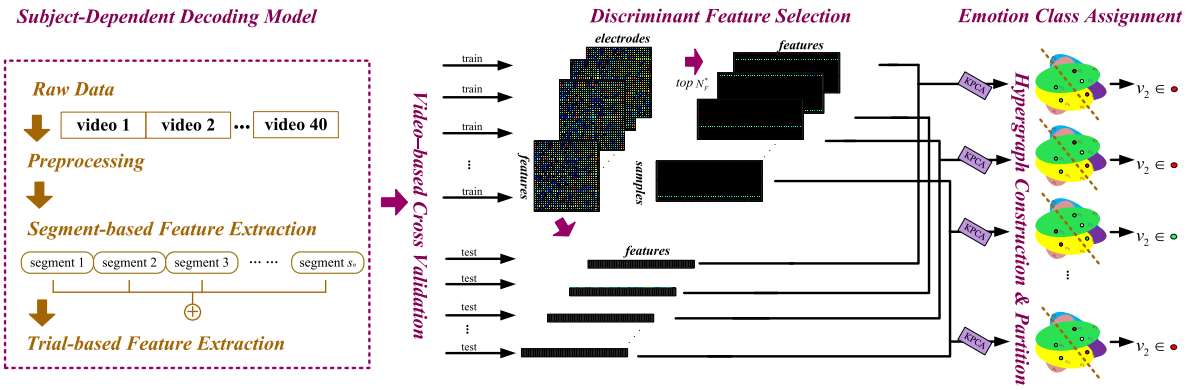


Fig. 1. The proposed unsupervised learning-based EEG decoding system for emotion recognition.

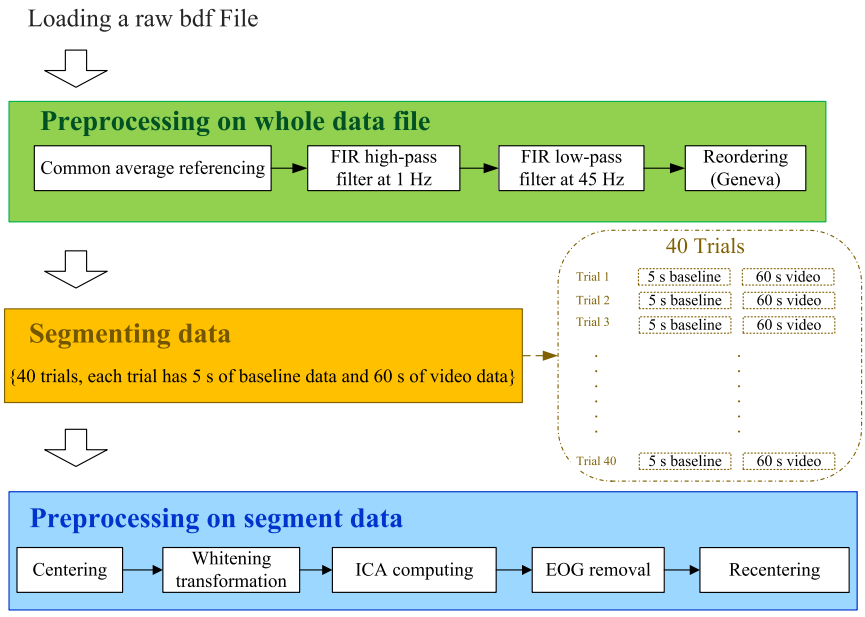


Fig. 2. Preprocessing flowchart for one subject's raw EEG signals.

60 s of video playing. After finishing each trial, the subjects were requested to give subjective feedback on their emotions while watching the video. Four emotion dimensions were considered for affective evaluation: arousal, valence, dominance, and liking. With the use of the Self-Assessment Manikin (SAM) system (Morris, 1995), subjective feedback with a range from 1 to 9 was given for each dimension of emotion, where 9 indicated an extremely strong emotion and 1 indicated an extremely weak emotion. Simultaneously, EEG signals were recorded at a sampling rate (fs) of 512 Hz from 32 active AgCl electrode sites according to the international 10–20 system placement (Jasper, 1958).

3. Methodology

The proposed unsupervised learning-based EEG decoding system is illustrated in Fig. 1. The collected raw EEG signals were first preprocessed for noise removal, and then the segment-based and trial-based features were extracted. After discriminant feature selection and extraction, a hypergraph was constructed to describe the hidden relationships among the trials in terms of the extracted features and further partitioned into different clusters. Consequently, trials in a single cluster were regarded as sharing similar patterns and assigned to the same emotion class.

3.1. Preprocessing

For the collected raw EEG signals, an overall preprocessing for noise removal was first performed, as shown in Fig. 2. The explanations of each step are provided below:

(1) Loading: loading a bdf file into MATLAB.

(2) Common average referencing: re-referencing the data by subtracting the average of all the collected electrodes from each single electrode.

(3) Finite impulse response (FIR) high-pass filter at 1 Hz: removing DC components at the low frequencies.

(4) FIR low-pass filter at 45 Hz: removing the other artifact noises at the high frequencies.

(5) Reordering: DEAP data were collected at two different locations (termed Twente and Geneva) with different channel orderings. In the present study, the EEG electrode locations for all the data were arranged in the Geneva order (refer to the project page: https://sites.google.com/site/janezhenliang/eeg-emotion).

According to the event marks, the collected EEG data of each subject were then segmented into 40 trials. Each trial included 5 s of baseline (termed baseline data below) and 60 s of video playing

(termed video data below). Further preprocessing was conducted for each trial as follows:

(6) Centering: aligning each channel to a zero mean.

(7) Whitening transformation: conversion to a data matrix that has an identity covariance matrix. Any correlations in the data were removed.

(8) ICA computing: running independent component analysis (ICA) in EEGLAB (Delorme & Makeig, 2004).

(9) Electrooculography (EOG) removal: removing the independent components with a lower fractal dimension (FD), based on the theory presented in Gomez-Herrero et al. (2006). Then, the EEG signals were reconstructed based on the remaining ICA components.

(10) Recentering: realigning the cleaned EEG data to a zero mean in each channel.

3.2. Segment-based feature extraction

To better monitor emotion changes over time, we examined EEG features over different timescales in comparison with single trial-based features. The video data were segmented into a number of short segments with the same time length T_s. There was no overlap between any two segments. As presented in Fig. 3, the segment-based features from 32 channels were characterized in three domains (frequency, time, and wavelet). The effect of the timescale on the EEG decoding is examined in Section 4.

3.2.1. Individual alpha frequency (IAF) detection

Conventionally, EEG frequency-domain features, e.g., the band powers for different frequency bands, are extracted under a fixed edge of the frequency bands (Jenke, Peer, & Buss, 2014). However, the rhythmic patterns of an EEG series can differ between subjects and between different mental states of the same subject (Klimesch, 1999). To study the variability in EEG rhythms across subjects, before feature extraction, the individual alpha frequency (IAF) for each subject in each trial was calculated as the normalized weighted sum of spectral estimates

$$\mathrm{IAF} = \frac{\sum_{i=1}^{n} p_{f_i} \times f_i}{\sum_{i=1}^{n} p_{f_i}}, \tag{1}$$

where p_{f_i} is the power spectrum at frequency f_i. Here, the interval for calculating the center of gravity was set to [7.5,12.5], and the frequency resolution was 0.25 Hz. Thus, n was equal to 21, and the frequency f_i $(i = 1, \ldots, n)$ was equal to {7.50 Hz, 7.75 Hz, …, 12.5 Hz}. For the calculated IAF values, the edges of the frequency bands can be dynamically defined (for details, refer to the project page). Note that all the IAF computations were performed on the baseline data and thus did not involve any emotion factor. Next, various EEG features were extracted from each short segment of video data.

3.2.2. Frequency-domain feature extraction

The power spectral density estimation algorithm (Welch, 1967) was first applied to compute the spectral power distribution using a Hamming window with 50% overlap. Subsequently, the average EEG power spectra on each subject-wise and trial-wise frequency sub-band (defined by the calculated IAF values) were extracted from each EEG channel. For clarity, the extracted band powers were named as follows: theta $\{\theta\}$, alpha1 $\{\alpha - 1\}$, alpha2 $\{\alpha - 2\}$, alpha3 $\{\alpha - 3\}$, beta1 $\{\beta - 1\}$, beta2 $\{\beta - 2\}$, beta3 $\{\beta - 3\}$, gamma1 $\{\gamma - 1\}$, gamma2 $\{\gamma - 2\}$, and gamm3 $\{\gamma - 3\}$.

In addition to the band powers, we also extracted the peak frequency in each frequency band according to the center-of-gravity calculation, similar to Eq. (1). The peak frequency is another critical characteristic in the frequency domain; it is the discrete frequency with the largest magnitude within a specific frequency band and has been shown to be correlated to the memory capacity (Moran et al., 2010) and cognitive processes (Haegens, Cousijn, Wallis, Harrison, & Nobre, 2014). In this study, the extracted peak frequencies were denoted as the theta peak (ITF), alpha peak (PAF), beta peak (IBF), and gamma peak (IGF), corresponding to the theta θ, alpha $\{\alpha - 1, \alpha - 2, \alpha - 3\}$, beta $\{\beta - 1, \beta - 2, \beta - 3\}$, and gamma $\{\gamma - 1, \gamma - 2, \gamma - 3\}$ bands, respectively. Thus, for each short segment of video data, the extracted frequency-domain features constituted a total of 14 features × 32 channels.

3.2.3. Time-domain feature extraction

Inspired by Jenke et al. (2014) and Liu et al. (2016)'s works, we also extracted EEG features in the time domain and identified the general characteristics of the time-series EEG data ($\boldsymbol{S}(t), t = 1, \ldots, T$).

(a) statistical features (7 features)

- power:

$$\mu_{\boldsymbol{S}} = \frac{1}{T}\sum_{t=1}^{T} |\boldsymbol{S}(t)|^2; \tag{2}$$

- mean:

$$P_{\boldsymbol{S}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{S}(t); \tag{3}$$

- standard deviation:

$$\sigma_{\boldsymbol{S}} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T} |\boldsymbol{S}(t) - \mu_{\boldsymbol{s}}|^2}; \tag{4}$$

- 1st difference:

$$\delta_{\boldsymbol{S}} = \frac{1}{T-1}\sum_{t=1}^{T-1} |\boldsymbol{S}(t+1) - \boldsymbol{S}(t)|; \tag{5}$$

- normalized 1st difference:

$$\tilde{\delta}_{\boldsymbol{S}} = \frac{\frac{1}{T-1}\sum_{t=1}^{T-1} \left|\tilde{\boldsymbol{S}}(t+1) - \tilde{\boldsymbol{S}}(t)\right|}{\sigma_{\boldsymbol{S}}}; \tag{6}$$

- 2nd difference:

$$\gamma_{\boldsymbol{S}} = \frac{1}{T-2}\sum_{t=1}^{T-2} |\boldsymbol{S}(t+2) - \boldsymbol{S}(t)|; \tag{7}$$

- normalized 2nd difference:

$$\tilde{\gamma}_{\boldsymbol{S}} = \frac{\frac{1}{T-2}\sum_{t=1}^{T-2} \left|\tilde{\boldsymbol{S}}(t+2) - \tilde{\boldsymbol{S}}(t)\right|}{\delta_{\boldsymbol{S}}}. \tag{8}$$

(b) Hjorth features (3 features)

Hjorth (1970) proposed three parameters to characterize EEG patterns in terms of the amplitude, timescale and complexity, which were defined as

- activity:

$$A_{\boldsymbol{S}} = \frac{1}{T-1}\sum_{t=1}^{T} |\boldsymbol{S}(t) - \mu_{\boldsymbol{s}}|^2; \tag{9}$$

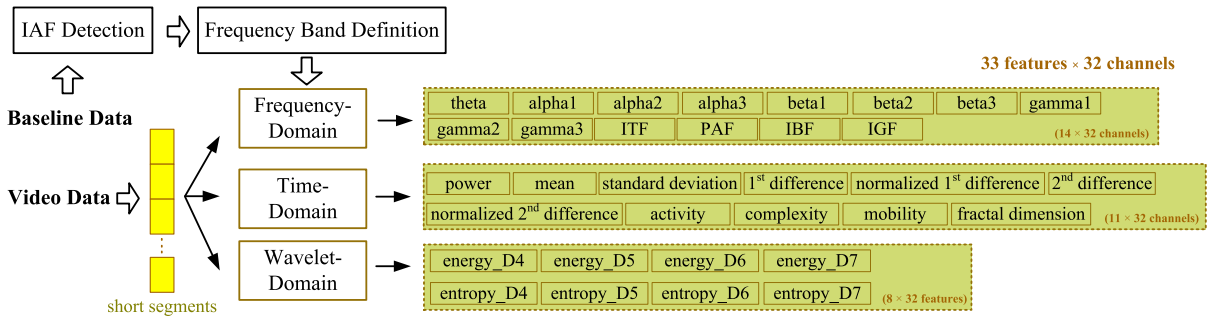


Fig. 3. Flowchart of segment-based feature extraction. Here, baseline data were used for individual alpha frequency (IAF) detection to define the boundary of each frequency band. Video data were segmented into short segments, which were used to extract and select features for further application in the modeling for emotion recognition.

- mobility:

$$M_{\boldsymbol{S}} = \sqrt{\frac{\frac{1}{T-1}\sum_{t=1}^{T-1}|\delta_{\boldsymbol{S}}(t)-\mu_\delta|^2}{A_{\boldsymbol{S}}}},\ \text{where } \mu_\delta = \frac{1}{T-1}\sum_{t=1}^{T-1}\delta_{\boldsymbol{S}}(t); \tag{10}$$

- complexity:

$$C_{\boldsymbol{S}} = \sqrt{\frac{\frac{1}{T-2}\sum_{t=1}^{T-2}\left|\gamma_{\boldsymbol{S}}(t)-\mu_\gamma\right|^2}{\frac{1}{T-1}\sum_{t=1}^{T-1}|\delta_{\boldsymbol{S}}(t)-\mu_\delta|^2}}\Big/ M_{\boldsymbol{S}},$$
$$\text{where } \mu_\gamma = \frac{1}{T-2}\sum_{t=1}^{T-2}\gamma_{\boldsymbol{S}}(t). \tag{11}$$

(c) fractal dimension (FD) (1 feature)

To characterize the shape information of EEG time-series data, the FD value was measured according to Sevcik's method, which was proven to be quite reliable in the presence of noise (Ansari-Asl, Chanel, & Pun, 2007; Sevcik, 1998). To obtain a more robust estimation of the FD value from EEG data (Gomez-Herrero et al., 2006), we first segmented the input into 10 short segments $\{\boldsymbol{S}_i, i = 1, \ldots, 10\}$ and then calculated the corresponding $FD_{\boldsymbol{S}_i}(i = 1, \ldots, 10)$ value for each $\boldsymbol{S}_i$. The final FD feature of $\boldsymbol{S}$ (t) was the average of all the obtained $FD_{\boldsymbol{S}_i}$ and was denoted as

$$FD = \frac{1}{10}\sum_{i=1}^{10} FD_{\boldsymbol{S}_i}, \tag{12}$$

where $FD_{\boldsymbol{S}_i} = 1 + \frac{\ln(L)}{\ln(2\times(N-1))}$. Here, N was the number of points of the waveform and was equal to $\frac{T}{10}$. L was the total length of the waveform and was determined by calculating the sum of the Euclidean distances of all coordinates (x_j, y_j) in $\boldsymbol{S}_i$ as follows:

$$L = \sum_{j=1}^{N-1}\sqrt{x_j^{*2} + y_j^{*2}}, \tag{13}$$

where $x_j^* = \frac{x_j}{x_{\max}}$ and $y_j^* = \frac{(y_j - y_{\max})}{(y_{\max} - y_{\min})}$. $x_{\max}$ was the maximum of x_j, and $y_{\min}$ and $y_{\max}$ were the minimum and maximum of y_j, respectively. In summary, the extracted time-domain EEG features from a short segment constituted a total of 11 features × 32 channels.

3.2.4. Wavelet-domain feature extraction

In addition to time-domain and frequency-domain EEG feature analyses, the wavelet transform is another effective tool for crafting specific forms of time-frequency representations of continuous EEG signals (Alomari, Awada, Samaha, & Alkamha, 2014; Gandhi, Panigrahi, & Anand, 2011). Gandhi et al.'s work (2011) demonstrated that the Coiflet wavelet of order 1 ('Coif1') would be the best mother wavelet for EEG signal analysis and classification. It was found that this wavelet achieved higher classification accuracy and required less computational time than other commonly used mother wavelets, such as Db, Harr, and Bior. In this study, EEG signals were decomposed by using the mother wavelet 'Coif1' up to level 7, and the energy and Shannon entropy properties were extracted from the detail coefficients at decomposition levels from 4 to 7 (for details, refer to the project page). For a short segment, the extracted EEG features in the wavelet domain constituted 8 features × 32 channels in total.

3.3. Feature normalization

Now, for each short segment (length of T_s), we had a feature vector $\mathbf{F}_{l,h}^{video}$, where l refers to the EEG channel and ranges from 1 to 32, and h indicates the feature type and ranges from 1 to 33 (14 frequency-domain + 11 time-domain + 8 wavelet-domain). The feature dimensionality was 1056 (33 features ×32 channels). Feature normalization was conducted for each feature dimension to make the distribution similar to a standardized normal distribution with a zero mean and unit variance as follows:

$$\widetilde{\mathbf{FS}}_{l,h} = \frac{\mathbf{F}_{l,h}^{video} - \mu_{l,h}}{\sigma_{l,h}}, \tag{14}$$

where $\mu_{l,h}$ and $\sigma_{l,h}$ correspond to the mean and standard deviation of $\mathbf{F}_{l,h}^{video}$, respectively. Note that to avoid information leaks in the performance validation (refer to Section 4), the test data were strictly protected in the training phase of the analysis, and the normalization parameters $\mu_{l,h}$ and $\sigma_{l,h}$ were calculated based only on the training data.

3.4. Trial-based feature extraction

For one video of a single subject (i.e., a single trial), a trial-based EEG feature vector was then formed by averaging all the extracted segment-based feature vectors over the trial as follows

$$\widetilde{\mathbf{FT}}_{l,h} = \frac{\sum_{s=1}^{s_n}\widetilde{\mathbf{FS}}_{l,h}^{(s)}}{s_n}, \tag{15}$$

where $\widetilde{\mathbf{FS}}_{l,h}^{(s)}$ was the extracted features at the sth short segment and s_n was the number of short segments in the video data. The feature dimensionality of $\widetilde{\mathbf{FT}}_{l,h}$ was the same as that of $\widetilde{\mathbf{FS}}_{l,h}$ (33 features ×32 channels = 1056).

Since the decoding model was subject-dependent, the obtained feature dimensionality was much greater than the number of trials (1056 features ≫ 40 trials). To avoid the complication of

dimensionality, an unsupervised learning-based feature selection and extraction approach was adopted to filter the most relevant features and generate a mapping to a nonlinear subspace for emotion modeling, as described below.

3.5. Unsupervised feature selection and extraction

Yang, Shen, Ma, Hung, and Zhou (2011) proposed an unsupervised discriminative feature selection (UDFS) algorithm by considering discriminative information, a manifold structure, and feature correlation. In UDFS, a joint structure including discriminative analysis and $\ell_{2,1}$ -norm minimization was developed with orthogonal constraints, and the feature discriminants were ranked in a batch manner. For clarity, let X represent $\widetilde{\mathbf{FT}}_{l,h}$ here, where $X = \{x_1, x_2, \ldots, x_n\}$ and $x_i \in \mathbb{R}^d$ (n is the number of samples, and d refers to the feature dimensionality). Inspired by the definition of discriminant analysis in the work of Fukunaga (1990) and studies of the local discriminative information analysis (Sugiyama, 2006; Yang, Xu, Nie, Yan, & Zhuang, 2010), the local discriminative information of the ith class (formed by data point x_i with its k nearest neighbors) is defined as

$$\begin{cases} S_t^{(i)} = \tilde{X}_i \tilde{X}_i^T \\ S_b^{(i)} = \tilde{X}_i G_{(i)} G_{(i)}^T \tilde{X}_i^T \end{cases}, \tag{16}$$

where $\tilde{X}_i$ is the centralized X_i and $G_{(i)}$ is the output label matrix for the ith class. $S_t^{(i)}$ and $S_b^{(i)}$ are the intraclass variance and interclass variance of the ith class, respectively. Furthermore, a local discriminant score is defined as

$$DS_i = Tr[(S_t^{(i)} + \lambda I)^{-1} S_b^{(i)}]. \tag{17}$$

A larger DS_i value indicates that the information is more discriminative. However, in the calculation of Eqs. (16) and (17), no output label G was available to be assigned in the unsupervised case. To estimate the G values in this case, it was assumed that there exists a linear relationship between the input data and the output class, given as $G = W^T X$. Thus, finding G becomes a search for the W. Here, $W \in \mathbb{R}^{d \times c}$ is a sparse feature weight matrix indicating the relationship between data point x_i and a class j ($j = 1, \ldots, c$) in terms of feature type p ($p = 1, \ldots, d$). If a weak connection exists between the data point and a class (j) in terms of the feature type (p), then element $w_{j,p}$ should be close to zero; conversely, if a strong connection exists, then $w_{j,p}$ should be very large. In other words, if $\sum_{j=1}^{c} w_{j,p}$ is very small, we do not regard feature type p as a discriminant feature for all the classes. To determine an optimal W under a consideration of a local discriminant score DS_i for X, an objective function is defined as

$$\min_{W^T W = I} Tr\left(W^T M W\right) + \rho \|W\|_{2,1}, \tag{18}$$

where ρ is a regularization parameter and M is a middle term related to the discriminant calculation (more details are available on the project page). Here, $\ell_{2,1}$-norm was leveraged to consider feature correlations across the whole feature space and to ensure that W was sparse in rows. In the implementation of determining the optimal W, only the samples X and the number of classes c were required (Roffo, Melzi, Castellani, & Vinciarelli, 2017). The solved W in Eq. (18) could be treated as a sparse feature selection matrix representing the combination coefficients of the most discriminant features, where one row indicates one feature type. If the values in the rows in W were close to zero, the corresponding features were not sufficiently discriminating. In other words, through sorting the $\left\|w^i\right\|_2$ ($W = [w^1, \ldots, w^d]^T$) in descending order, the corresponding features could be ranked from high to low in terms of feature discrimination. In this study, we optimized W for $\widetilde{\mathbf{FT}}_{l,h} \in \mathbb{R}^{1056}$. By calculating the corresponding $\left\|w^i\right\|_2$, a ranking list $R^T \in \mathbb{R}^{1056}$ was obtained; this list indicates the discrimination level of the extracted 1056 features. Only the top N_F^* features in R^T were retained to form the lower feature space $\mathbf{FA}^d \in \mathbb{R}^d$ ($d = N_F^* \ll 1056$).

Moreover, to make the representation of $\mathbf{FA}^d$ more compact, a nonlinear mapping was conducted to map $\mathbf{FA}^d$ to $\mathbf{FA}^g \in \mathbb{R}^g$ ($g < d$) through a kernel principal component analysis (KPCA) (Schölkopf, Smola, & Müller, 1998), with 95% of the variance retained. Consequently, the EEG features were reduced from 1056 to g dimensions and then denoted as $\mathbf{FA}^g$. Notably, in the implementation of cross-validation, N_F^* and the KPCA relationship were only determined by the training data and were then applied to the test data without any additional feature modifications.

3.6. Hypergraph construction and partitioning

In this section, we explain the hypergraph for describing complex relationships among different EEG trials in terms of the extracted EEG characteristics. As the proposed decoding system is specific to each subject, a hypergraph was constructed for each subject. For one subject, each trial (i.e., video) was treated as one individual vertex in the hypergraph. The relationships among the vertices were represented as hyperedges, in which one hyperedge connected a number of vertices sharing similar properties. Based on the constructed hypergraph, emotion recognition was realized using a spectral hypergraph partitioning that divided the hypergraph into a specific number of clusters that corresponded to different emotion classes. The preliminaries on the hypergraph are available on our project page.

3.6.1. Hypergraph construction

A hypergraph is given as $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_{|V|}\}$refers to the set of vertices and $E = \{e_1, e_2, \ldots, e_{|E|}\}$ indicates the set of hyperedges. In the hypergraph construction, each trial was treated as a vertex, and thus the total number of vertices was equal to the number of trials, denoted as N_T (in our case, $N_T = 40$). The similarity between any two vertices, v_{b_i} and v_{b_j}, was defined as

$$\mathfrak{X}_{v_{b_i}, v_{b_j}} = \exp(-dist_{v_{b_i}, v_{b_j}}), \tag{19}$$

where

$$dist_{v_{b_i}, v_{b_j}} = \sqrt{\sum_{f_t=1}^{g} (\mathbf{FA}(f_t)_{(b_i)}^g - \mathbf{FA}(f_t)_{(b_j)}^g)^2}. \tag{20}$$

$\mathbf{FA}_{(b_i)}^g$ and $\mathbf{FA}_{(b_j)}^g$ were the corresponding EEG characteristics of v_{b_i} and v_{b_j} and g refers to the dimensionality of the extracted EEG space. We defined the similarity matrix $\mathfrak{X} = \{\mathfrak{X}_{v_{b_i}, v_{b_j}}\}$, $\mathfrak{X} \in \mathbb{R}^{N_T \times N_T}$, in which the row and column index the vertices such that each element of $\mathfrak{X}$ showed similarity in [0, 1] between a couple of vertices. As mentioned above, a hyperedge can connect an arbitrary number of vertices that share similar patterns. In this study, we treated each vertex as a centroid and formed one hyperedge by the centroid vertex with its κ nearest vertices. For example, v_{b_i} was the centroid vertex, and the selected nearest neighbors $v_{b_\tau^*}$ were the vertices with the highest similarities to v_{b_i} in $\mathfrak{X}\left(v_{b_i}, v_{b_\tau^*}\right)$, $\tau = 1, \ldots, \kappa$. In the constructed hypergraph, the number of hyperedges was equal to the number of vertices, and every hyperedge size was equal to $\kappa + 1$.

Inspired by Huang, Liu, Zhang, and Metaxas's work (2010), we adopted a probabilistic incidence matrix $\mathbf{H}$, which was defined as

$$h\left(v_{b_i}, e_{b_j}\right) = \begin{cases} \mathfrak{X}(v_{b_i}, e_{b_j}), & \text{if} v_{b_i} \in e_{b_j} \\ 0, & \text{if} v_{b_i} \notin e_{b_j} \end{cases}, \tag{21}$$

where $\mathcal{X}(v_{b_i}, e_{b_j})$ is the similarity between vertex v_{b_i} and hyperedge e_{b_j}'s centroid vertex v_{b_j}, as defined in Eq. (19). Compared to the conventional incidence matrix, the probabilistic incidence matrix $\mathbf{H}$ not only reflects whether vertex v_{b_i} is connected to hyperedge e_{b_j} but also indicates the strength of the connectivity. The corresponding hyperedge weight $w_g(e_{b_j})$ of hyperedge e_{b_j} is a summation of all the similarities of the connected vertices, determined as $\sum_{v_{b_i} \in e_{b_j}} h\left(v_{b_i}, e_{b_j}\right)$. Then, we calculated the corresponding weight matrix $\mathbf{W}_g$, vertex degree matrix $\mathbf{D}_v$, and hyperedge degree matrix $\mathbf{D}_e$.

3.6.2. Hypergraph partitioning

Next, the decoding problem for emotion recognition is solved through hypergraph partitioning in an unsupervised manner. The commonly used hypergraph partitioning algorithms can be categorized into two types: (1) transforming a hypergraph to a simple graph and then using a spectral clustering algorithm to partition the vertices and (2) generating a hypergraph Laplacian. The second type is the most widely adopted one. As presented in Zhou et al.'s study (2007), hypergraph Laplacian-based partitioning is an NP-complete problem, where the cost function is given as

$$\Omega(f) = f^T \Delta f, \tag{22}$$

where $\Delta \in \mathbb{R}^{N_T \times N_T}$ is the hypergraph Laplacian, defined as

$$\Delta = \mathbf{I} - \boldsymbol{\Theta}. \tag{23}$$

Here, $\boldsymbol{\Theta} = \mathbf{D_v}^{-(1/2)}\mathbf{H}\mathbf{W}_g\mathbf{D_e}^{-1}\mathbf{H^T}\mathbf{D_v}^{-(1/2)}$ and $\mathbf{I} \in \mathbb{R}^{N_T \times N_T}$ is the identify matrix. Then, the optimized solution of the cost function $\Omega(f)$ is several eigenvectors with the smallest nonzero eigenvalues of Δ. Likewise, the optimal partitioning of the hypergraph structure is to the first several eigenvectors with the smallest nonzero eigenvalues in Δ that form an eigenspace for the subsequent vertex classification. In practice, we took the first l_H eigenvectors ($l_H < N_T$) to form a new eigenspace Δ' ($\Delta' \in \mathbb{R}^{N_T \times l_H}$) and adopted K-means to cluster Δ' into the number of emotion classes.

4. Experimental results and discussion

In this section, we present the performance of our proposed unsupervised learning-based decoding system on the DEAP database and compare the results with those in the existing literature. To fully examine the proposed unsupervised system, the decoding performance was verified for emotion class recognition under four different emotion dimensions (arousal, valence, dominance, and liking).

4.1. Evaluation process

Similar to Koelstra et al.'s work (2012), the decoding model was also evaluated using the accuracy P_{acc} and the F1-Score P_f, which are defined as

$$P_{acc} = \frac{n_{TN} + n_{TP}}{n_{TN} + n_{FN} + n_{TP} + n_{FP}} \times 100\%, \tag{24}$$

where n_{TN} and n_{TP} are the numbers of correctly predicted samples in two classes (low and high classes, in our case) and n_{FN} and n_{FP} are the corresponding misclassified sample numbers for the low and high classes.

$$P_f = \frac{2 \times P_{pre} \times P_{sen}}{P_{pre} + P_{sen}} \times 100\%, \tag{25}$$

where P_{pre} refers to the precision, given as $P_{pre} = \frac{n_{TP}}{n_{TP}+n_{FP}}$, and P_{sen} is the sensitivity, defined by $P_{sen} = \frac{n_{TP}}{n_{TP}+n_{FN}}$. To cross-compare the decoding results obtained in this study with those in other studies that used the DEAP database, the collected 9-point subjective feedback in the DEAP database was first discretized into two classes using a fixed threshold of 5 for each emotion dimension. Then, each emotion dimension had two emotion classes (low and high). In the actual implementation, the proposed decoding system classified the trials (videos) into low and high emotion classes based on EEG signals for each emotion dimension. Considering the inter-subject variability, the proposed unsupervised decoding system was tested in a subject-dependent manner; thus, for each subject, the system was evaluated in a video-based leave-one-out cross-validation manner. To further explain the modeling and evaluation processes, an example of emotion recognition on Subject 1 is illustrated below.

(1) Training and test data formation

The total collected 40 trials (corresponding to 40 videos) were randomly split into two groups: training data (39 videos) and test data (1 video).

(2) Feature extraction and selection

To avoid information leaks and achieve an unbiased modeling, the feature extraction and selection processes were conducted on the training data first, and the obtained parameters were then employed in the test data's feature extraction and selection procedure.

(3) Hypergraph construction

A hypergraph was constructed to describe the complex relationships within and between the training and test data.

(4) Hypergraph partitioning

The formed hypergraph was divided into two clusters through optimizing the calculated hypergraph Laplacian defined in Eq. (23).

(5) Emotion assignment

The emotion class for the test data was predicted based on the majority voting rule. As shown in Fig. 4, if most of the training data from the same cluster belonged to the low class, then the test data would be assigned to the low class (a), whereas if most of the training data in the same cluster belonged to the high class, then the test data would be classified as the high class (b).

(6) Repeated steps (1) to (5) until each video was treated as test data once

(7) Final validation result

The final validation result was to combine all the test results in step (5), and then the accuracy P_{acc} and the F1-score P_f were calculated as defined in Eqs. (24) and (25), respectively.

4.2. Timescale effect on decoding performance

We changed the time length of the short segments T_s to 1 s, 2 s, 3 s, 4 s, 5 s, 6 s, 10 s, 20 s, 30 s, and 60 s and evaluated the effect of the timescale on the EEG decoding for emotion recognition. The corresponding numbers of short segments s_n in one trial were 60, 30, 20, 15, 12, 10, 6, 3, 2 and 1. Based on the obtained segment-based features $\widetilde{\mathbf{FS}}_{l,h}^{(s)}, s = 1, \ldots, s_n$, the corresponding trial-based features $\widetilde{\mathbf{FT}}_{l,h}$ were formed and further processed for hypergraph construction and partitioning. The case in which T_s was equal to 60 s was the specific condition in which the trial-based features $\widetilde{\mathbf{FT}}_{l,h}$ were the same as the segment-based features $\widetilde{\mathbf{FS}}_{l,h}^{(s)}$. The validation performances under the four emotion dimensions are reported in Fig. 5, where N_F^*, κ, and l_H were empirically determined to be 31, 2, and 3, respectively. The results indicate that

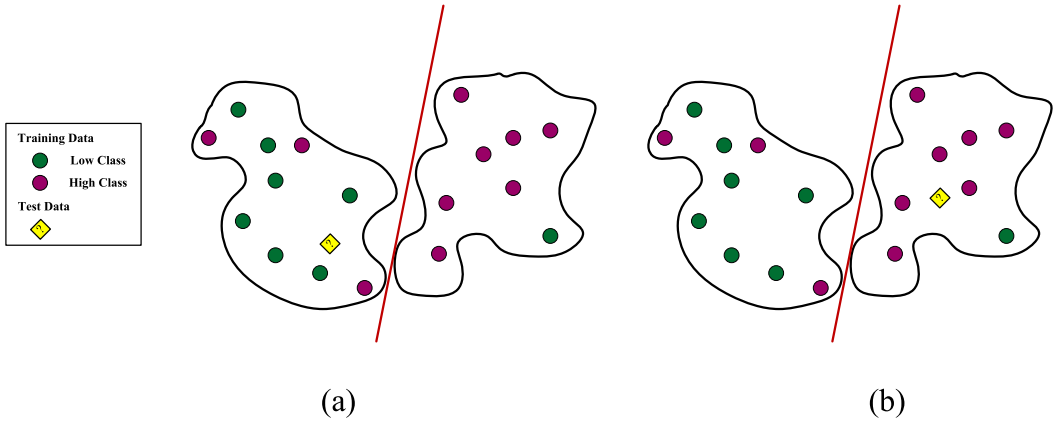


Fig. 4. Example of emotion assignment to the test data: (a) low class; (b) high class.

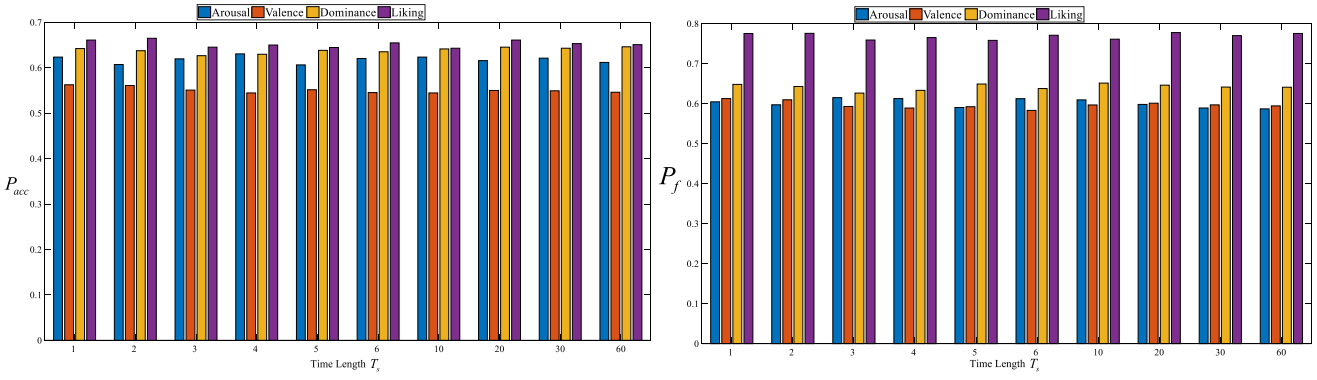


Fig. 5. Decoding performance when T_s was set to various values of the time length.

the best decoding performance for two-class emotion recognition was achieved for the liking dimension, where P_{acc} and P_f ranged from 64.30∼66.48 and 75.81∼77.76, respectively. The second and third best decoding performances were observed for the dominance dimension (P_{acc}: 62.66∼64.61 and P_f: 62.63∼65.12) and arousal dimension (P_{acc}: 60.63∼63.05 and P_f: 58.68∼61.48), whereas the poorest decoding performance was for the valence dimension (P_{acc}: 54.45∼56.25 and P_f: 58.30∼61.25). This result demonstrates that this decoding approach was less sensitive to the timescale than we expected. One possible reason may be that the timescale effect in feature representation was weakened in the process of conversion from the segment-based feature vector to the trial-based feature vector.

4.3. Performance comparison with existing literature

As is known, different validation procedures for feature extraction and modeling might lead to very different results. Before comparing our proposed decoding model to those in the existing literature, we first reviewed the existing studies using the DEAP database and briefly summarized the following aspects in Table 1

- **Supervised?:** whether supervised learning was used in the feature extraction and/or modeling;
- **Subject-dependent?:** whether the built model was subject-based;
- **Channels used:** how many channels were used for the EEG feature extraction;
- **Only EEG used?:** whether the decoding model used only the features extracted from EEG signals or also used features extracted from other peripheral signals;
- **Model:** what type of model was used;
- **Data size:** how many samples were used in the training and testing parts;
- **CV manner:** what type of cross-validation (CV) was adopted for the performance evaluation: Leave-One-Out CV (LOOCV) or N-fold CV;
- **Video-based CV?:** whether the sample(s) in the test data was/were from a different trial(s) (video(s)) than the samples used in the training data. If yes, the performance validation was a video-based CV. Usually, the test performance obtained in a non-video-based CV would be much better than that obtained in a video-based CV as the collected EEG signals are fairly consistent within the same video.

Notably, all the existing methodologies were supervised learning-based. Conversely, our proposed EEG decoding system for emotion recognition was a pure unsupervised learning-based pipeline: unsupervised feature extraction and selection and unsupervised modeling. The performance comparisons are reported in Table 2. The results reveal the validity of decoding emotions via unsupervised learning using EEG responses, even though the performances were generally worse than those with supervised learning.

4.4. Discussion

To better evaluate the proposed unsupervised decoding system, we further examined the pipeline with different parameter settings. Here, we used $T_s = 10$ s as an example.

Instead of presenting the overall performance across the subjects and videos, we checked the decoding performance for an

Table 1
A brief summary of the existing studies using the DEAP database.

Paper	Supervised?	Subject-Dependent?	Channels used	Only EEG used?	Model	Data size	CV manner	Video-Based CV?
Koelstra et al. (2012)	Yes	No	32	Yes	Gaussian naive Bayes classifier	32 subjects × 40 videos	LOOCV	Yes
Liu and Sourina (2012)	Yes	Yes	28	Yes	SVM	21 subjects × 40 videos	5-fold CV	Yes
Bahari and Janghorbani (2013)	Yes	Yes	32	Yes	Supervised-based K nearest neighbor	32 subjects × 40 videos	LOOCV	Yes
Naser and Saha (2013)	Yes	No	32	Yes	SVM	32 subjects × 40 videos	LOOCV	Yes
Yoon and Chung (2013)	Yes	No	32	Yes	Bayes classifier	32 subjects × 40 videos	LOOCV	Yes
Wang and Shang (2013)	Yes	No	32	No	Deep belief networks	32 subjects × 40 videos × 10 segments (segment length=1 s)	LOOCV	Yes
Torres-Valencia, Garcia-Arias, Lopez, and Orozco-Gutierrez (2014)	Yes	No	32	Yes	Hidden Markov models	32 subjects × 40 videos	10-fold CV	Yes
Chen, Hu, Moore, Zhang, and Ma (2015)	Yes	Yes	32	Yes	C4.5 decision tree	32 subjects × 40 videos	10-fold CV	Yes
Li et al. (2015)	Yes	Yes	32	Yes	SVM	32 subjects × 40 videos × 60 segments (segment length=1 s)	10-fold CV	No
Atkinson and Campos (2016)	Yes	No	14	Yes	SVM	32 subjects × 40 videos	8-fold CV	Yes
Liu et al. (2016)	Yes	No	32	Yes	Random forest	32 subjects × 40 videos	10-fold CV	Yes
Shahnaz et al. (2016)	Yes	No	32	Yes	SVM	32 subjects × 40 videos	LOOCV	Yes
Yin, Zhao et al. (2017)	Yes	Yes	32	No	Deep learning model	32 subjects × 40 videos	10-fold CV	Yes
Yin, Wang et al. (2017)	Yes	No	32	Yes	Least square LSSVM	32 subjects × 40 videos	10-fold CV	Yes
Lin et al. (2017)	Yes	No	32	No	Deep convolution neural network	32 subjects × 40 videos × 10 segments	10-fold CV	No
Zhuang et al. (2017)	Yes	Yes	8	Yes	SVM	32 subjects × 40 videos × 12 segments	LOOCV	Yes
Our method	**No**	**Yes**	**32**	**Yes**	**Hypergraph partitioning**	**32 subjects × 40 videos**	**LOOCV**	**Yes**

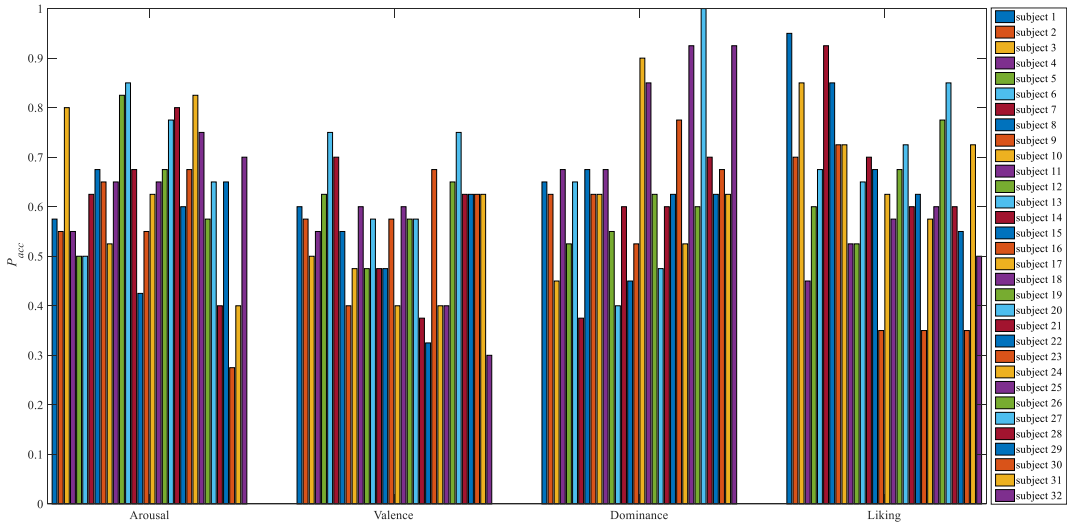
Table 2
Comparisons with other emotion recognition results in the existing literature.

Methodology		Arousal		Valence		Dominance		Liking	
		P_{acc}	P_f	P_{acc}	P_f	P_{acc}	P_f	P_{acc}	P_f
Koelstra et al. (2012)		62.00	58.30	57.60	56.30	–	–	55.40	50.20
Liu and Sourina (2012)		76.51	–	50.80	–	–	–	–	–
Bahari and Janghorbani (2013)		64.56	–	58.05	–	–	–	67.42	–
Naser and Saha (2013)		66.20	–	64.30	–	68.90	–	70.20	–
Yoon and Chung (2013)		70.10	74.90	70.90	74.70	–	–	–	–
Wang and Shang (2013)		51.20	–	60.90	–	–	–	68.40	–
Torres-Valencia et al. (2014)		55.00	–	58.75	–	–	–	–	–
Chen et al. (2015)		69.09	68.96	67.89	67.83	–	–	–	–
Li et al. (2015)		64.20	–	58.40	–	65.80	–	66.90	–
Atkinson and Campos (2016)		73.06	–	73.14	–	–	–	–	–
Liu et al. (2016)		71.20	–	69.90	–	–	–	–	–
Shahnaz et al. (2016)		66.51	76.68	64.71	74.94	66.88	76.67	70.52	81.94
Yin, Zhao et al. (2017)		77.19	69.01	76.17	72.43	–	–	–	–
Yin, Wang et al. (2017)		78.67	75.26	78.75	80.77	–	–	–	–
Lin et al. (2017)		87.30	78.24	85.50	80.06	–	–	–	–
Zhuang et al. (2017)		71.99	–	69.10	–	–	–	–	–
Our method	$T_s = 1$ s	62.34	60.44	56.25	61.25	64.22	64.80	66.09	77.52
	$T_s = 10$ s	62.34	60.93	54.45	59.66	64.14	65.12	64.30	76.10
	$T_s = 60$ s	61.17	58.68	54.61	59.43	64.61	64.10	65.08	77.53

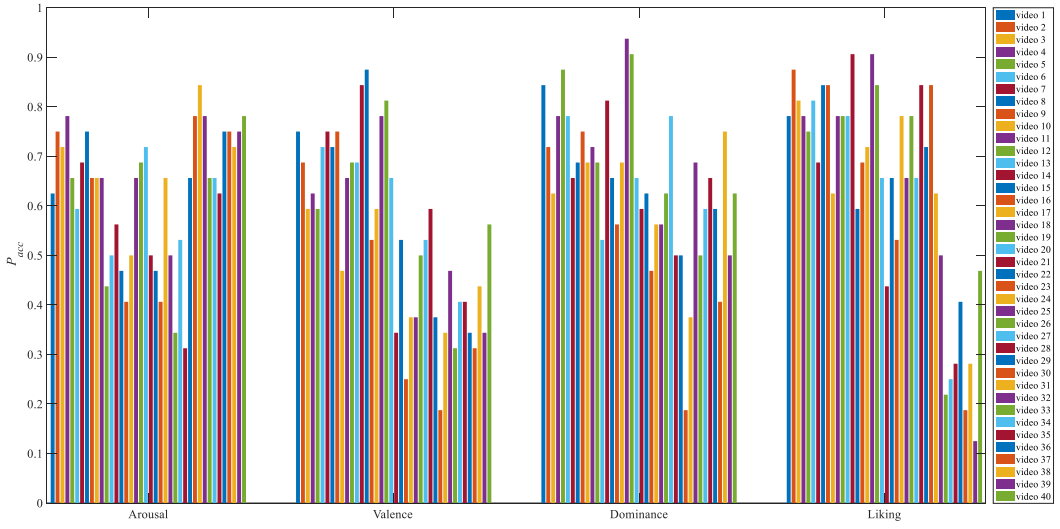
individual subject and video and presented the results in Fig. 6. As shown in (a), compared to other subjects, such as subject 27 (arousal: 65.00; valence: 75.00; dominance: 100; liking: 85.00), subject 16 always had poorer recognition accuracy (arousal: 55.00; valence: 57.50; dominance: 52.50; liking: 35.00). We expected that subject 27 would have more a consistent scoring criterion, in which the reported feedbacks could well reflect the emotion changes and the changes in the brain signals. To compare the decoding performances between different videos, we averaged all the subjects' results for each video. The results in (b) showed that, for example, video 7 could be considered a good stimulus for triggering human emotions (arousal: 68.75; valence: 75.00; dominance: 65.62; liking: 68.75). However, video 23 might have been a poor stimulus that failed to evoke emotion in any dimension (arousal: 40.62; valence: 25.00; dominance: 46.88; liking: 53.12).

We next examined the effect of the hyperedge size ($\kappa + 1$) on the relationship representation among the trials. Instead of fixing the hyperedge size to 3 as in the original setting, we implemented hypergraph constructions with different hyperedge sizes ranging from 3 to 30 ($\kappa = 2, \ldots, 29$) and presented the decoding performances in Fig. 7. We observed that the optimal hyperedge size was equal to 6 ($\kappa = 5$), regardless of emotion dimension. That is, one hyperedge was suggested to be formed by a centroid with its 5 nearest neighbors. On the other hand, we considered an emotion dimension dependent selection of the optimum hyperedge size. This selection found that the optimum hyperedge size was still 6 for emotion recognition for the arousal and liking dimensions but changed to 30 and 21 for the valence and dominance dimensions. The decoding results of the selected optimum hyperedge sizes in the two conditions are presented in Table 3. Likewise, in the hypergraph partitioning, instead of fixing the number of the remaining eigenvectors l_H to 3 as in the original setting, the effect of the l_H value was also verified. Here, we adjusted the l_H value from 1 to 20 when the hyperedge size was set to 6. The corresponding decoding performances with the optimal l_H were as follows: arousal ($l_H = 2$; P_{acc}: 63.67 and P_f: 61.26), valence ($l_H = 2$; P_{acc}: 56.33 and P_f: 60.72), dominance ($l_H = 19$; P_{acc}: 64.69 and P_f: 64.44), and liking ($l_H = 5$; P_{acc}: 66.64 and P_f: 77.59).

In addition to solving the decoding problem through hypergraph theory, we also evaluated the performance when a simple graph was adopted. Instead of forming a hyperedge with an arbitrary number of vertices, the edge in the simple graph was only allowed to connect to two vertices. Similarly, the spectral graph partitioning approach was applied to divide the simple graph into two clusters. The corresponding decoding performance is reported in Table 4. Compared to a simple graph structure, a



(a)



(b)

Fig. 6. Decoding performance for each subject and video.

Table 3
Decoding performance with the optimized hyperedge size.

Optimal hyperedge size	Arousal		Valence		Dominance		Liking	
	P_{acc}	P_f	P_{acc}	P_f	P_{acc}	P_f	P_{acc}	P_f
Independent of emotion dimension	6 63.52	62.02	55.31	59.88	64.22	65.51	65.86	76.95
Dependent on emotion dimension	6 63.52	62.02	30 57.19	60.77	21 65.16	66.21	6 65.86	76.95

Table 4
EEG decoding with a simple graph.

	Arousal	Valence	Dominance	Liking
P_{acc}	61.02	52.58	63.67	65.62
P_f	58.75	57.00	63.68	77.49

complex geometrical structure (hypergraph) would generally be a more suitable framework for describing the relationships in EEG trials.

To study the significance of the applied unsupervised feature selection and extraction approach in the proposed decoding pipeline, we conducted a Monte Carlo permutation test to examine whether any overfitting issues occurred during the feature selection and extraction procedure. Here, instead of using the ground truth labeled by the subjects, we labeled the data with random classes (the ratios of low and high classes in each dimension of emotion were the same as the original ground truth labeled by the subjects) and performed decoding for two-class emotion recognition. The corresponding results for each

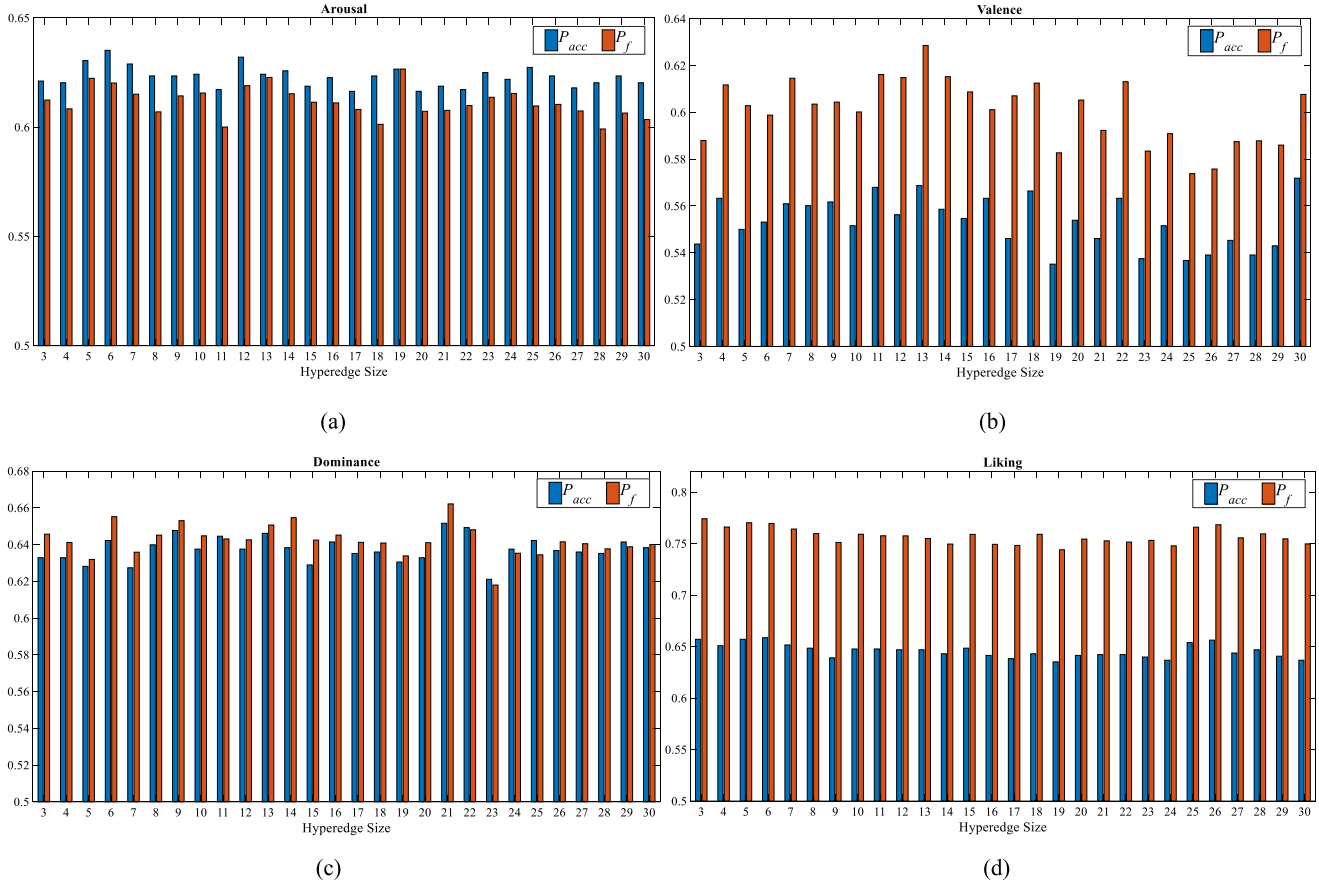


Fig. 7. Decoding performance when the hyperedge size was varied: (a) arousal; (b) valence; (c) dominance; and (d) liking.

dimension of emotion were as follows: arousal (P_{acc}: 59.77 and P_f: 58.08), valence (P_{acc}: 51.80 and P_f: 57.34), dominance (P_{acc}: 62.11 and P_f: 62.18), and liking (P_{acc}: 63.75 and P_f: 75.03). These results demonstrated that the emotion recognition performance decreased when the ground truth was randomly assigned. However, due to the unbalanced sample distribution of low and high classes, the random permutation results were not close to simple binary chance. Notably, the ratios of low and high classes (low/high) in each dimension of emotion were as follows: arousal (0.41/0.59), valence (0.43/0.57), dominance (0.38/0.62), and liking (0.33/0.67). More details are available on the project page.

5. Conclusion

In conclusion, this study proposed an unsupervised learning-based EEG decoding system for human emotion recognition. EEG features were characterized in terms of three domains (frequency, time, and wavelet) and with locations throughout the whole brain (channels). This system can be interpreted as an integration of the dynamic and diverse characteristics of a brain. After discriminant feature selection and extraction, a hypergraph structure was first introduced to describe the complex relationships among the EEG trials and was then partitioned into two clusters through a hypergraph Laplacian. The possibility of EEG-based decoding with unsupervised learning was demonstrated for the application of human emotion recognition. Our work in this study thus provides an attractive approach for the processing and decoding of EEG signals.

References

Alarcao, S. M., & Fonseca, M. J. (2017). Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, http://dx.doi.org/10.1109/TAFFC.2017.2714671.

Alomari, M. H., Awada, E. A., Samaha, A., & Alkamha, K. (2014). Wavelet-based feature extraction for the analysis of EEG signals associated with imagined firsts and feet movement. *Computer and Information Science*, *7*(2), 17–27.

Ansari-Asl, K., Chanel, G., & Pun, T. (2007). A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In *15th European signal processing conference* (pp. 1241–1245).

Atkinson, J., & Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications*, *47*, 35–41.

Bahari, F., & Janghorbani, A. (2013). EEG-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier. In *2013 20th Iranian conference on biomedical engineering* (pp. 228–233).

Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, *1*(3), 295–311.

Berge, C. (1989). *Hypergraphs*. North-Holland Mathematical Library.

Chen, J., Chen, Z., Chi, Z., & Fu, H. (2017). Facial expression recognition in video with multiple feature fusion. *IEEE Transactions on Affective Computing*, http://dx.doi.org/10.1109/TAFFC.2016.2593719.

Chen, J., Hu, B., Moore, P., Zhang, X., & Ma, X. (2015). Electroencephalogram-based emotion assessment system using ontology and data mining techniques. *Applied Soft Computing*, *30*, 663–674.

Cruz, A. C., Bhanu, B., & Thakoor, N. S. (2014). Vision and attention theory based sampling for continuous facial emotion recognition. *IEEE Transactions on Affective Computing*, *5*(4), 418–431.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Diestel, R. (2017). *Graph theory* (5th ed. 2017 edition). Springer.

Ducournau, A., Rital, S., Bretto, A., & Laget, B. (2009). A multilevel spectral hypergraph partitioning approach for color image segmentation. In *2009 IEEE international conference on signal and image processing applications* (pp. 419–424).

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., et al. (2016). The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, *7*(2), 190–202.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). Professional, Inc., San Diego, USA: Academic Press.

Gandhi, T., Panigrahi, B. K., & Anand, S. (2011). A comparative study of wavelet families for EEG signal classification. *Neurocomputing*, *74*(17), 3051–3057.

Gomez-Herrero, G., Clercq, W. D., Anwar, H., Kara, O., Egiazarian, K., Huffel, S. V., et al. (2006). Automatic removal of ocular artifacts in the EEG without an EOG reference channel. In *Proceedings of the 7th nordic signal processing symposium* (pp. 130–133).

Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *NeuroImage*, *92*, 46–55.

Hjorth, B. (1970). EEG Analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, *29*(3), 306–310.

Hong, C., Chen, X., Wang, X., & Tang, C. (2016). Hypergraph regularized autoencoder for image-based 3D human pose recovery. *Signal Processing*, *124*, 132–140.

Hu, J., Wei, X., & He, H. (2014). Brain image segmentation based on hypergraph modeling. In *2014 IEEE 12th international conference on dependable, autonomic and secure computing* (pp. 327–332).

Huang, Y., Liu, Q., Zhang, S., & Metaxas, D. N. (2010). Image retrieval via probabilistic hypergraph ranking. In *2010 IEEE conference on computer vision and pattern recognition* (pp. 3376–3383).

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Jenke, R., Peer, A., & Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, *5*(3), 327–339.

Kim, Y., & Provost, E. M. (2017). ISLA: temporal segmentation and labeling for audio-visual emotion recognition. *IEEE Transactions on Affective Computing*, http://dx.doi.org/10.1109/TAFFC.2017.2702653.

Klimesch, W. (1999). EEG Alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis using physiological signal. *IEEE Transactions on Affective Computing*, *3*(1), 18–31.

Li, X., Zhang, P., Song, D., Yu, G., Hou, Y., & Hu, B. (2015). EEG based emotion identification using unsupervised deep feature learning. In *SIGIR2015 workshop on neuro-physiological methods in IR research, ID: 44132*, (pp. 1–2).

Lin, W., Li, C., & Sun, S. (2017). Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal. In *International conference on image and graphics* (pp. 385–394).

Liu, M., Gao, Y., Yap, P. T., & Shen, D. (2017). Multi-hypergraph learning for incomplete multi-modality data. *IEEE Journal of Biomedical and Health Informatics*, 1–11. http://dx.doi.org/10.1109/JBHI.2017.2732287.

Liu, J., Meng, H., Nandi, A., & Li, M. (2016). Emotion detection from EEG recordings. In *2016 12th international conference on natural computation, fuzzy systems and knowledge discovery* (pp. 1722–1727).

Liu, Y., & Sourina, O. (2012). EEG-based valence level recognition for real-time applications. In *2012 international conference on cyberworlds* (pp. 53–60).

Luo, Z., Peng, B., Huang, D. A., Alahi, A., & Fei-Fei, L. (2017). Unsupervised learning of long-term motion dynamics for videos. In *The conference on computer vision and pattern recognition* (pp. 2203–2212).

Moran, R. J., Campo, P., Maestu, F., Reilly, R. B., Dolan, R. J., & Strange, B. A. (2010). Peak frequency in the theta and alpha bands correlates with human working memory capacity. *Frontiers in Human Neuroscience*, *4*(200), 1–12.

Morris, J. D. (1995). SAM: the self-assessment manikin an efficient cross-cultural measurement of emotional response. *Journal of Advertising Research*, *35*(8), 63–68.

Naser, D. S., & Saha, G. (2013). Recognition of emotions induced by music videos using DT-CWPT. In *2013 Indian conference on medical informatics and telemedicine* (pp. 53–57).

Roffo, G., Melzi, S., Castellani, U., & Vinciarelli, A. (2017). Infinite latent feature selection: A probabilistic latent graph-based ranking approach. In *2017 IEEE international conference on computer vision* (pp. 1398–1406).

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299–1319.

Sevcik, C. (1998). A procedure to estimate the fractal dimension of waveforms. *Complexity International*, *5*, source: http://www.complexity.org.au/ci/vol05/sevcik/sevcik.html.

Shahnaz, C., Masud, S. B., & Hasan, S. M. S. (2016). Emotion recognition based on wavelet analysis of empirical mode decomposed EEG signals responsive to music videos. In *2016 IEEE region 10 conference* (pp. 424–427).

Sugiyama, M. (2006). Local fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd international conference on machine learning* (pp. 905–912).

Torres-Valencia, C. A., Garcia-Arias, H. F., Lopez, M. A. A., & Orozco-Gutierrez, A. A. (2014). Comparative analysis of physiological signals and Electroencephalogram (EEG) for multimodal emotion recognition using generative models. In *2014 19th symposium on image, signal processing and artificial vision* (pp. 1–5). http://dx.doi.org/10.1109/STSIVA.2014.7010181.

Tutte, W. T. (1998). *Graph theory as I have known it*. Clarendon Press.

Wang, D., & Shang, Y. (2013). Modeling physiological data with deep belief networks. *International Journal of Information and Education Technology (IJIET)*, *3*(5), 505–511.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, *15*(2), 70–73.

Yang, Y., Shen, H. T., Ma, Z., Hung, Z., & Zhou, X. (2011). l2, 1-norm regularized discriminative feature selection for unsupervised learning. In *International joint conference on artificial intelligence* (pp. 1589–1594).

Yang, Y., Xu, D., Nie, P., Yan, S., & Zhuang, Y. (2010). Image clustering using local discriminant models and global integration. *IEEE Transactions on Image Processing*, *19*(10), 2761–2773.

Yin, Z., Wang, Y., Liu, L., Zhang, W., & Zhang, J. (2017). Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Frontiers in Neurorobotics*, *11*(19), 1–16. http://dx.doi.org/10.3389/fnbot.2017.00019.

Yin, Z., Zhao, M., Wang, Y., Yang, J., & Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods and Programs in Biomedicines*, *140*, 93–110.

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.

Zhou, D., Hung, J., & Scholkopf, B. (2007). Learning with hypergraphs: clustering, classification, and embedding. In B. Scholkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems, vol. 19* (pp. 1601–1608). Cambridge, MA: MIT Press.

Zhu, L., Shen, J., Xie, L., & Cheng, Z. (2017). Unsupervised topic hypergraph hashing for efficient mobile image retrieval. *IEEE Transactions on Cybernetics*, *47*(11), 3941–3954.

Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H., & Yan, B. (2017). Emotion recognition from EEG signals using multidimensional information in EMD domain. *BioMed Research International*, 8317357, 1–9.