Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition

TENGFEI SONG[1,2], WENMING ZHENG[1], (SENIOR MEMBER, IEEE), CHENG LU[1,2], YUAN ZONG[1], XILEI ZHANG[1] AND ZHEN CUI [3], (Member, IEEE)

[1]Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 210096, China (e-mail: wenming_zheng@seu.edu.cn)
[2]School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: songtf@seu.edu.cn)
[3]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhen.cui@njust.edu.cn)

Corresponding author: Wenming Zheng (e-mail: wenming_zheng@seu.edu.cn).

This work was supported in part by the National Basic Research Program of China under Grant 2015CB351704, in part by the National Natural Science Foundation of China under Grant 61572009 and Grant 61772276, in part by the Key Research and Development Program of Jiangsu Province, China under Grant BE2016616.

ABSTRACT To explore human emotions, in this paper, we design and build a Multi-Modal Physiological Emotion Database (MPED), which collects four modal physiological signals, i.e., electroencephalogram (EEG), galvanic skin response (GSR), respiration (RSP) and electrocardiogram (ECG). To alleviate the influence of culture dependent elicitation materials and evoke desired human emotions, we specifically collect an emotion elicitation material database selected from more than 1500 video clips. By considerable amount of strict man-made labelling, we elaborately choose 28 videos as standardised elicitation samples, which are assessed by psychological methods. The physiological signals of participants were synchronously recorded when they watched these standardised video clips that described six discrete emotions and neutral emotion. With three types of classification protocols, different feature extraction methods and classifiers (SVM and KNN) were used to recognize the physiological responses of different emotions, which presented the baseline results. Simultaneously, we present a novel attention-LSTM (A-LSTM) which strengthens the effectiveness of useful sequences to extract more discriminative features. Additionally, correlations between the EEG signals and the participants' ratings are investigated. The database has been made publicly available to encourage other researchers to use it to evaluate their own emotion estimation methods.

INDEX TERMS Discrete emotion recognition, physiological signals, EEG, affective computing, machine learning, video-induced emotion, LSTM

I. INTRODUCTION

AFFECTIVE computing research had emerged as an interdiscipline, including neuroscience, pedagogy, cognitive science, psychology and computer science, and had received more attentions in recent years. As one of the important parts in affective computing, emotion recognition had received increasing attention during the past several years. Facial expression [1] [2] [3] and speech [4] are two of the popular modalities of recognizing human emotions. However, both modalities are non-physiological signals and may not directly reflect the intrinsic mental states of human beings. In contrast to the non-physiological signals, physiological signals, such as electroencephalogram (EEG) [5], electrocardiogram (ECG) [6], galvanic skin response (GSR) [7] and respiration (RSP) [8], are more likely to reflect the real emotional states because people are easier to hide the non-physiological response. Therefore, there have emerged more applications, such as driving fatigue detection [9] and assessment of workload [10], all of which are based on emotion recognition using physiological signals.

To build a physiological emotion database, a key problem to be considered is the partition of emotional space, which is commonly divided into two categories, i.e., dimensional emotion model and discrete emotion model. Dimensional emotion model is expressed using multiple dimensions or scales to categorize emotions. Russell [11] proposed the

valence-arousal scale dimensional model and Mehrabian [12] proposed the pleasure-arousal-dominance scale model. Besides, Plutchik [13] proposed the emotion wheel, which is dimensional model as well. Discrete emotion model consists of a constant number of basic emotions, like the tree structure of emotions [14] and the six basic emotions (i.e., joy, sadness, surprise, fear, anger and disgust) [15]. In terms of dimensional emotion model, it is easier to mark different emotion via numerical values in the valence-arousal coordinate space. On the contrary, a particular feeling is difficult to be named in dimensional emotion model. The performance of dimensional model will be degraded dramatically [16], when emotions are close in the valence-arousal space. Although researcher attempted map valence and arousal classes into basic emotion, it is still different from discrete emotions, which are induced from target materials. The same category discrete emotions could be in different levels of valence and arousal [17].

In dealing with physiological emotion database for discrete emotion recognition, it is notable that constructing elicitation materials plays an important role. Lang et al. [18] built the international affective picture system (IAPS) based on colorized photographs, which are extensively used for studies of emotion elicitation. Although IAPS has been assessed by psychological methods, the cultural dependency may affect the performance on discrete emotion elicitation. Under this consideration, more studies [19] [20] [21] used elicitation materials selected by themselves to induce various emotions. But the psychological assessment is insufficient and it is hard to guarantee the effectiveness of elicitation materials.

In this paper, we build a Multi-Modal Physiological Emotion Database (MPED) to study discrete emotion recognition from multi-modal physiological signals, i.e., EEG, GSR, RSP and ECG. As a noninvasive signal, EEG describes the ongoing brain activity and has been applied in cognitive neuroscience to evaluate the regulation and processing of emotion. ECG reflects the emotional activity, especially the differentiate between positive and negative emotions. GSR describes the skin's ability to conduct electricity, which is effective to capture emotion states, especially for arousal difference. Former studies have demonstrated that the magnitude of GSR is linearly associated in arousal. RSP measures the breathing activity of subjects. Generally, negative emotions will induce the irregularity in RSP pattern. Compared with fNIRS [22] and fMRI [23], our modalities are easier to be acquired. The physiological signals of 23 participants were recorded when they watched different emotional videos. To alleviate the influence of culture dependent elicitation materials, 28 videos describing seven different emotion states (i.e., joy, funny, anger, disgust, fear, sad and neutrality) were selected as elicitation materials from over 1500 video clips and assessed by considerable amount of strict man-made labelling. Besides, we evaluated the rated scores and the result validated the effectiveness of our elicitation materials. The summary of database content was presented in Table 1.

For emotion database based on physiological signals, the extracted features are commonly discrete sequences. There have emerged some deep learning methods, like Long Short-Term Memory (LSTM), to solving sequential classification tasks [24]. LSTM uses the memory cell to deal with vanishing gradient problem and to capture long-term temporal dependence. LSTM has been applied for emotion recognition using physiological signals and the results are comparable to state-of-the-art methods [25] [26]. Although LSTM performs well on emotion recognition task using physiological signals, some sequences are not that contributing for emotion recognition which may limit the effectiveness of classification model. Besides, for these physiological signals, especially EEG, the individual differences are significant. Attention model provides a way to dynamically select features according to different input data, which is more appropriate to deal with the problem on individual differences. Additionally, a residual connection is helpful to enhance good features and suppress noises via soft masks. For LSTM, the input data, hidden state and memory cell are three parts of great importance to extract discriminative features. Under this consideration, we introduced attention mechanism to constrain the input data, hidden states and memory cell of LSTM for each iterative process so that the useful features were selected.

Extensive experiments with different feature extraction methods on MPED are conducted to provide the baseline result and prove the effectiveness of the proposed attention-LSTM (A-LSTM). Besides, we used spearman correlation coefficients to explore the correlation between EEG energy distribution in scalp and different emotion states.

TABLE 1. Summary of Database Content.

Elicitation material	
Number of videos	170
No. of ratings per video	16-22
Contents of rating	PANAS, SAM and DES
Rating values	Discrete scale of 1-5 for PANAS; Discrete scale of 1-9 for SAM and DES
Experiment	
Number of participants	23
Number of video	28
Emotion categories	Joy, funny, disgust, anger, fear, sad and neutrality
Recorded signals	62-Channel EEG (1000Hz), respiration, galvanic skin response, electrocardiogram
Self-report	
Rating scales	Arousal, Valence, and DES
Rating values	Discrete scale of 1-9

II. RELATED WORK

A. PHYSIOLOGICAL EMOTION DATABASES

Recently, researchers have built many physiological emotion databases and we summarized in Table 2, where all databases

TABLE 2. The summary of physiological emotion databases reviewed in experiment environment.

Database	Stimuli	Modality	Sub.	Feature extraction method and classifier	Emotional states	Len.	Emotional Model
[27]	IAPS, IADS	EEG	5	Fourier analysis, PCA and FDA	Valence and arousal	-	Dimensional model
[21]	Music	ECG, RSP, SC, EMG	3	FFT, SSE, HRV, BRV features, SVM and MLP.	Arousal and valence	14400 s	Dimensional model
DEAP [28]	Video	EEG, EOG, EMG, GSR, RSP, BP, ST	32	Spectral power features, Gaussion naive Bayes classifier.	Valence, arousal and liking	2520 s	Dimensional model
MAHNOB-HCI [29]	Video	EEG, GSR, ECG, RSP, ST, eye gaze	27	Spectral power features, SVM and Adaboost	Valence and arousal	-	Dimensional model
DEAMER [30]	Video	EEG, ECG	23	PSD features, KNN, LDA and SVM	Valence, arousal and dominance	1080 s	Dimensional model
ENTERFACE׳06 [22]	IAPS	EEG, fNIRS	5	STFT features and TBM	Positive, negative and calm	-	Discrete model
[31]	Music	EEG	26	PSD, DASM, RASM features, SVM and MLP.	Joy, anger, sadness, pleasure	480 s	Discrete model
[32]	Music	EEG	9	SPG, HHS, ZAM features, KNN, QDA and SVM	Liking and disliking	1125 s	Discrete model
SEED [20]	Video	EEG	15	PSD, DE, DASM, RASM, DCAM features, SVM, KNN and DBN	Positive, negative and neutral	3394 s	Discrete model
[33]	Video	EEG	30	STFT features of five frequency bands, SLDA and SVM	Joy, amusement, tenderness,anger, sadness, fear, disgust and neutrality	1271 s	Discrete model
HR-EEG4EMO [19]	Video	EEG	27	PSI, HOC, SCF, FD features and SVM.	Positive and negative	2015 s	Discrete model
RCLS [34]	Video	EEG	14	PSD, DE, HOC, FD, Wavelet, Hjorth features, GRSLR, GSCCA, GraphSC, CCA, SVM and RF.	Positive, negative and neutral	1628 s	Discrete model
Our MPED	Video	EEG, ECG, RSP, GSR	23	PSD, STFT, HHS, HOC, Hjorth features, KNN and SVM	Joy, funny, anger, fear, disgust, disgust and neutrality	5684 s	Discrete model

"Sub." denotes the number of subjects. "Len." denotes the length of physiology signals for each subject. IAPS and IADS denote the International Affective Picture System and the International Affective Digital Sounds. The feature extraction methods include Short-time Fourier Transform (STFT), Fast Fourier Transform (FFT), Subband Spectral Entropy (SSE), Heart Rate Variability (HRV), Breathing Rate Variability (BRV), Power Spectral Density (PSD), Differential Entropy (DE), Differential Asymmetry(DASM), Differential Caudality (DCAM), Rational Asymmetry (RASM), Hilbert-Huang Spectrum (HHS), Zhao Atlas Marks Distribution (ZAM), Phase Synchronization Index (PSI), High Order Crossing (HOC), Fractal Dimension (FD) and Spectral Crest Factor (SCF). The classifiers include Fisher's Discriminant Analysis (FDA), Support Vector Machine (SVM), K Nearest Neighbors (KNN), Transferable Belief Model (TBM), Multilayer Perceptron (MLP), Quadratic Discriminant Analysis (QDA), Deep Belief Network (DBN), Sparse Linear Discriminant Analysis (SLDA), Graph Regularized Sparse Linear regression (GRSLR), Group Sparse Canonical Correlation Analysis (GSCCA), Graph regularized Sparse Coding (GraphSC), Canonical Correlation Analysis (CCA) and Random Forest (RF).

contain EEG signals.

The studies of Bos [27] and Savran [22] used IAPS to induce different emotions. Bos [27] recorded EEG signals around the frontal and parietal lobes when participant watching images and videos. The limited number of electrodes were applied for emotion recognition. Savran [22] presented a database to detect and estimate emotion using fNIRS and EEG signals from five participants. The fusion of both modalities signals was considered. Kim et al. [21] investigated music-induced emotion recognition using different physiological signals. Lin et al. [31] evoked different emotions based on music and explored the correlation between brain activity and emotional states. Hadjidimitriou et al. [32] recorded EEG signals of nine subjects during music listening and adopted three different time-frequency analysis methods to investigate correlations between EEG signals in five frequency bands with two emotional states, i.e., like and dislike. Pantic et al. published DEAP [28] and MAHNOB-HCI [29] to explore human emotions from multimodal signals in dimensional model. SEED [20] consists of EEG signals of 15 participants during video watching and a deep belief networks was applied for emotion recognition. RCLS [34] is made up of 14 subjects' EEG signals during video watching and a new classification method, i.e., graph regularized sparse linear regression, was proposed for EEG based emotion recognition. Liu et al. [33] published a elicitation material database for emotion elicitation and built a system to recognize different emotional states based on EEG signals in a real-time. Katsigiannis et al. [30] recorded 23 subjects' EEG signals elicited by audio-visual stimuli using low-cost devices and the fusion of EEG and ECG signals was evaluated. Becker et al. [19] published a database including 257-channel EEG data and reconstructed brain activity on the cortical surface. The classification result demonstrated

the effectiveness of source reconstruction. Although there have been so many databases on discrete emotion recognition existing, multi-modal physiological emotion database is extremely deficient. Fusing multiple modalities provides an effective way to improve classification results by exploiting the complementary nature of different modalities. Therefore, it is essential to explore discrete emotions using multi-modal physiological signals.

B. THE FRAMEWORK FOR EMOTION RECOGNITION USING PHYSIOLOGICAL SIGNALS

Generally, discrete emotion recognition using physiological signals can be commonly divided into two step, i.e., feature extraction and classification.

In feature extraction stage, different time domain, frequency domain and time-frequency domain [35] features are commonly applied. Hjorth [36] proposed a time domain feature to describe the acitivity, mobility and complexity of time series. Petrantonakis et al. [37] presented the HOC feature to reflect the oscillatory pattern of time sequence. The PSD [27] was the most popular to extract features from frequency domain. Beside, some time-frequency methods had been applied for feature extraction, like short-time Fourier transform (STFT) spectrum [31] and Hilbert-Huang transform [35] spectrum. For the last classification process, there were many classical classifiers, like support vector machine (SVM) [38] and k-NearestNeighbor (KNN) [39]. To present the baseline results on our database, these feature extraction methods and classifiers were evaluated in this paper.

C. DEEP LEARNING METHODS

Recently, deep learning methods, especially convolutional neural network (CNN) and recurrent neural network (RNN), have been most popular among classification tasks [40]. More studies fucus on emotion recognition using physiological signals by deep learning methods. Zheng et al. presented deep belief networks (DBN) [20] to evaluate emotions based on EEG signals. Song et al. provide dynamical graph convolutional neural networks (DGCNN) [26] to build a graph connection based on training data for EEG emotion recognition. As an extension of RNN, Long short-term memory (LSTM) uses the memory cell to deal with vanishing gradient problem and to capture long-term temporal dependence. LSTM has been applied to solving many difficult problems, such as language modeling [41], protein secondary structure prediction [42] and translation [43]. Soleymani et al. had applied LSTM for EEG emotion analysis. In [25], Zhang et al. proposed Spatial-Temporal Recurrent Neural Network (STRNN) for emotion analysis and gain comparable results. More deep classification methods were applied for emotion recognition and the result with higher accuracy proved the advantage of deep learning methods.

III. CONSTRUCTION OF MULTI-MODAL PHYSIOLOGICAL EMOTIONAL DATABASE

A. EMOTION ELICITATION

1) Collection of elicitation materials

Twenty-four Chinese volunteers (ten males and fourteen females) with average age of 21.46 years (range = 18-24, SD = 1.87) selected about 1500 Chinese video clips for preliminary screening, which were shown after the year of 2005, including film clips, TV News and TV shows, and so on. All of these video clips last 2.5 to 5 minutes and contain the complete content to elicit the target emotion. The preliminary screened video clips consist of eleven emotional states, i.e., joy, happiness, romance, warmth, love, funny, passion, sadness, anger, fear, disgust and neutrality. To select the satisfactory elicitation materials, two specialists (1 male and 1 female) at the elicitation of emotion and nine research assistants (5 males and 4 females) majoring in psychology evaluated the preliminary screened video clips. 170 video clips with high evaluated scores by research assistants were selected for further evaluation.

162 graduate and undergraduate students (86 males and 76 females) with average of 23.21 years (range = 18-29, SD = 1.65) participated our Chinese emotion video evaluation experiment. We divided participants into seventeen groups and each group consisted of 7-11 people. The participants of each group watched the video clips via a projector and the voice was set comfortable to be hear. For each group, 20 video clips were shown in a random order and there was enough break time between two video clips to avoid the interference from the close clips. After watching each video clip, the participants finished three questionnaires, i.e., the positive and negative affective scheme (PANA) [44], self-assessment manikin (SAM) [45] and differential emotion scale (DES) [46], according to their true feelings. During the rating process of the video clips, participants finished the evaluation forms without any communication and were told that they could drop out of the rating whenever they wanted in case that they made the wrong decision. Finally, each of these video clips was evaluated by at least 16 participants.

Three psychological questionnaires, i.e., PANAS, SAM and DES, related to various emotions were applied for elicitation materials assessment. PANAS is a 5-point scale (1 = ¡°not at all¡±, 5= ¡°extremely¡±) containing 20-item mood words, i.e., 10-item words for positive affective subscale and 10-item words for negative affective subscale. PANAS has been proved reliable, valid, and efficient for reflecting two primary dimensions of mood, i.e., positive and negative affects. The SAM uses a non-verbal, graphic representation ro assess arousal, valence and dominance. The psychological study of [45] has proved the effectiveness of SAM. Different from [45], we adopted a 9-point scale (1 = “not at all”, 9= “extremely”). The DES was used for assessing the different component of emotions, which consisted of ten basic emotions, 3-item words for each emotion. The same as SAM, the DES was based on a 9-point scale (1 = “not at all”, 9= “extremely”).

TABLE 3. Details of the presented standardised Chinese emotion elicitation material database.

Emotion	Video Name	length (sec)	Clip Content	Video Name	length (sec)	Clip Content
Joy	Naked Wedding	146	A man and a woman get married	Where Are We Going, Dad?	164	Many children eat fruits together
	Perfect Two	216	The father and son live together happily	Where Are We Going, Dad?	122	The fathers eat with children together
Funny	Tonight 80's Talk Show	284	A man talks about changes of his wife	The Mermaid	201	A man tells police that he found a mermaid
	iPartment	150	A person's head stuck in a door	Top Funny Comedian	178	Two people battle for top one in DEYUNSHE
Anger	The Flowers Of War	170	A woman was raped by invaders	City of Life and Death	280	Thousands of people were killed in Nanjing
	City of Life and Death	150	A woman was raped and killed by invaders	City News	189	A woman maltreats aged people
Sadness	Take the wrong car	232	A daughter says goodbye to her father	Soulmate	182	The best friend of a girl died in the hospital
	Aftershock	196	A mother lost her daughter for saving her son	Aftershock	198	The daughter lost her leg after the earthquake
Fear	Hungry Ghost Ritual	295	Many ghosts kill a lot of people	Bunshinsaba Vs Sadako	179	A girl walks alone in the teaching building
	Seeing Ghosts	245	A woman can see the person who was dead	ChangChen Ghost Stories	141	A woman nightmares at night
Disgust	Dr.Qin	241	The forensic check the scene of crime	Detective Lei	180	Detective Lei eats worms
	Dr.Qin	154	The forensic check the scene of crime	Survivor Games	182	A group of people eat worms
Neutrality	Vacuum cleaner	197	A man shows how to use vacuum cleaner	Financial basic knowledge	173	The lesson about financial basic knowledge
	Operating Systems	172	The lesson about introduction of operating systems	Machine Learning	227	The lesson about machine learning

TABLE 4. The average score of each of the seven emotion categories along each of the five sub-dimensions of DES scale.

Emotion Category	DES sub-dimensions				
	$Dim.1$ $(Pleasure)$	$Dim.2$ $(Angry)$	$Dim.3$ (Sad)	$Dim.4$ $(Fear)$	$Dim.5$ $(Disgust)$
Joy	6.38	1.17	1.26	1.12	1.45
Funny	7.32	1.13	1.05	1.09	1.2
Anger	1.12	6.23	6.04	4.29	7.02
Fear	1.24	2.67	2.33	6.8	5.32
Disgust	1.66	3.08	2.46	4.94	6.54
Sad	1.55	2.28	5.85	2.09	2.11
Neutrality	2.07	1.57	1.22	1.12	2.86

2) Selection of elicitation materials

Based on scores of PANAS, SAM and DES, three widely used scales to evaluate emotional information from different perspectives [44] [45] [46], we conducted k-mean algorithm to produce seven distributed clusters which correspond to emotion categories of joy, funny, anger, sad, disgust, fear and neutrality, respectively. Then, five sub-dimensions of DES, namely pleasure, angry, sad, disgust and fear were used to evaluate and to base the selection of top four videos ranking in corresponding sub-dimensions for each category. Finally, 28 video clips were used for eliciting the target emotions. The descriptive details of these videos were shown in Table 3, and their average scores along each of the five sub-dimensions of DES scale (i.e., pleasure, angry, fear, disgust and sad) were summarized in Table 4.

Notably, joy and funny videos were usually splitted into two emotion categories according to the intensity of pleasantness [47]. We followed this approach and, as demonstrated in the second column of Table 4, used four videos rating middle higher in the pleasantness sub-dimension to elicit JOY while used another four rating highest in this dimension to elicit FUNNY. T-test (conducted by matlab script $ttest2.m$) confirmed that the selected funny videos were rated, by the participants, as eliciting significantly more pleasantness than joy videos ([t(153)=2.67, P=0.009]). Despite clear distinction between positive and negative emotions, it is well known that dimensional ratings within negative emotion categories are to a large extent overlapped [48] [17]. For example, the subjective ratings in sub-dimensions of $Angry$ and $Disgust$ for selected angry videos were quite close and held no significant difference ([t(146)=-1.90, P=0.13]). Due to well-established neural circuits in preference for fear processing [49] [50], fear emotion can be more easily distinguished from other negative emotions [51]. This statement was also confirmed by data here showing that, for the selected fear videos, the intensity rating in the sub-dimension of $Fear$

was significantly higher than that of $Disgust$ ([t(140)=3.09, P=0.002]), or any other sub-dimensions (P<0.001). For the selected videos of neutrality, however, ratings in all the five sub-dimensions were generally low. All the evidences above validated the approach we used here for video selections.

B. MULTI-MODAL PHYSIOLOGICAL SIGNALS SAMPLING

1) Scene configuration and experimental setting

The videos were displayed by a projector and the participants were told to adjust the voice in a suitable level to make sure that they are comfortable with the recording experiment. According to the international 10-20 system, the 62-channel EEG signals were recorded using ESI NeuroScan System at sampling rate of 1000 Hz and the location of EEG electrodes on a cap is shown in Fig. 1 (a). Before EEG signals recording, participants were told to clear the scalp and exfoliate so as to record more accurate signals. The EEG signals on the position of left\right mastoids were recorded with signals of left mastoid as the reference. The RSP, ECG and GSR were recorded by BIOPAC System at sample rate of 250 Hz, and the physiological signals were sent by wireless technology. GSR signals were recorded with finger electrodes connecting middle finger and index finger, which is shown in Fig. 1 (b). The sensor used to record RSP shown in Fig. 1 (c) was tied around chest, which was close to the diaphragm. To record ECG, we conducted limb lead connection with left wrist connected to positive electrode, right wrist connected to negative electrode and left ankle connected to ground electrode respectively. Fig. 1 (d) describes the positions of electrodes in recording ECG. We used the E-prime to play elicitation materials and send marks to ESI NeuroScan System and BIOPAC System simultaneously.

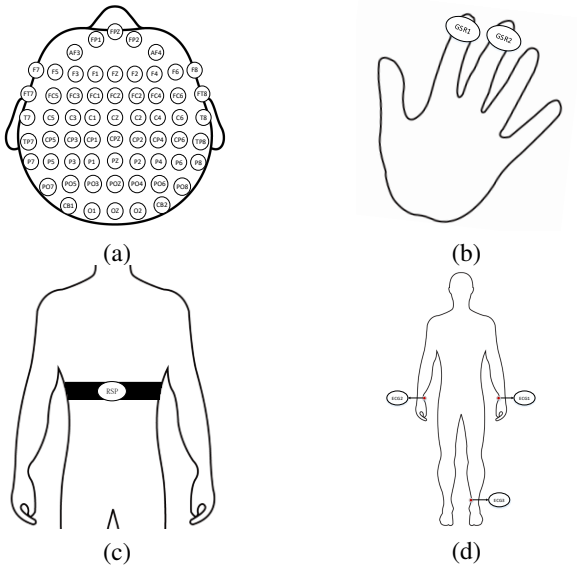


FIGURE 1. (a) The electrode placement for the international 10-20 system; (b) The electrode placement to record GSR signals;(c) Position of sensor to record RSP signals;(d) The electrode placement to record ECG signals.

2) Experiment Protocol

The research has been approved by the institutional review board of the Southeast University, and it adhered to the tenets of the Declaration of Helsinki. There are thirty healthy Chinese participants without psychiatric disorder and neurological illness according to participants' self-report, aged between 18 and 25 (mean age 20), participating in the experiment. Before the experiment, every participant signed an informed consent form, and then, they were instructed about the experiment protocol and self-assessment.

To avoid that long time experiment makes participants exhausted, we divided the experiment into two parts and the interval of two parts was at least 24 hours so that the participants have enough time to rest. The details of protocol used for eliciting various emotion were presented in Fig. 2. Each separated experiment consisted of a 120-second resting status, 14 trials and rating for self-assessment. In the 120-second resting status, participants were told to keep eyes closed and relax, during which psychological signals were recorded as well. Each trial was made up of a 10-second countdown hinting process, play of the Chinese video clip and then a 30-second resting process. The video was displayed in a given random order. The rating for self-assessment was conducted in the end after the participants took off all the sensors. Compared with the protocol, where every video is closely followed by self-assessment rating, the actual arrangement of self-assessment rating has been put at the end of experiment to prevent exhausting subjects through cutting their wearing time of experiment devices. Participants reviewed the watched videos and finished the SAM and DES for self-assessment.

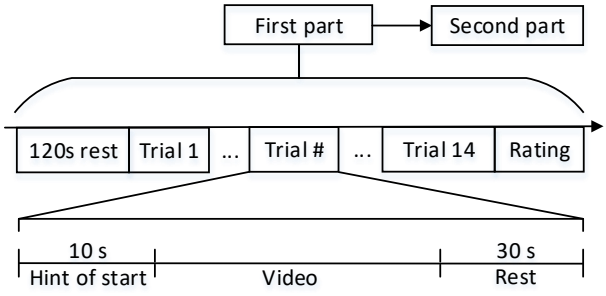


FIGURE 2. The protocol of video-induced emotion elicitation.

After all experiment processes were clear to the participant, the participant was asked to clear the scalp and use exfoliator to slough away lackluster skin on the location of corresponding electrode patch. After all the sensors were placed correctly and all the signals appeared no abnormal, the experiment started when participant pressed the key on the keyboard.

TABLE 5. Number of features from EEG, GSR, ECG and RSP signals.

Modality	Number of features per sample				
	δ	θ	α	β	γ
EEG(HOC)	62×20	62×20	62×20	62×20	62×20
EEG(Hjorth)	62×3	62×3	62×3	62×3	62×3
EEG(PSD)	62	62	62	62	62
EEG(HHS)	62	62	62	62	62
EEG(STFT)	62	62	62	62	62
GSR	4				
ECG	8+8+2				
RSP	2+2+2				

IV. FEATURE EXTRACTION FROM PHYSIOLOGICAL SIGNALS

After signals acquisition, a key process was feature extraction from these physiological signals. Effective features were easier to be distinguished by the classification model. In this section, we presented different feature extraction methods and a slide window of one second with no overlap was conducted to divide the physiological signals into many 1-second samples. The detailed number of features from a sample were summarized in Table 5, in which the EEG features were extracted from five frequency bands, i.e., δ, θ, α, β and γ, the GSR features contained the mean value, standard deviation, and the mean of first and second derivations, the ECG features contained eight energy mean values, eight SSE values and mean value and standard deviation of NN intervals, the RSP features contained two energy mean values, two SSE values and mean value and standard deviation of PP intervals.

A. EEG FEATURE EXTRACTION

Before EEG feature extraction, we conducted the independent component analysis (ICA) to remove electrooculography (EOG) artifacts. The raw EEG signals were decomposed to many independent components, which consist of EOG artifact and EEG signals. A classifier trained was used to identify the EOG artifacts and EEG signals, which were used to reconstruct the EEG signals without EOG artifacts. All the EEG signals were filtered into five frequency bands, i.e., delta (1-4Hz), theta (4-8Hz), alpha (8-14Hz), beta (14-31Hz) and gamma (31-50Hz).

1)$Frequency\ Domain\ Feature\ Extraction$: The most popular features for EEG-based emotion recognition are power spectral density (PSD) features [52] [27], which is the average energy from different frequency bands. The energy of frequency domain can be formulated as $P(f) = \mid \int_{\infty}^{+\infty} x(\tau)e^{-j2\pi f\tau}d\tau \mid^2$.

We used the fast fourier transform (FFT) to calculate the discrete fourier transform (DFT). A 1000 samples window was applied for the estimation of PSD from the five frequency bands respectively. The logarithms of the PSD from different frequency bands were used as the features.

2)$Time\ Domain\ Feature\ Extraction$: Hjorth [36] proposed a kind of time domain feature, which can be formulated as: $A_x = \frac{\sum_t (x(t)-\mu)^2}{T}$, $M_x = \sqrt{\frac{var(\dot{x}(t))}{var(x(t))}}$, $C_x = \frac{M(\dot{x}(t))}{M(x(t))}$, where A_x, M_x and C_x represent activity, mobility and complexity respectively. A_x denotes the variance of the input signal $x(t)$ and $\dot{x}(t)$ is the time derivative of $x(t)$. The activity, mobility and complexity of EEG signals were listed in order as the feature for further classification.

Petrantonakis and Hadjileontiadis developed higher order crossings (HOC) feature [37], which is used for describing the oscillatory pattern of time series. The time series $x(t)$ is processed by a sequence of high-pass filters: $\Im_k\{x(t)\} = \bigtriangledown^{k-1}x(t)$, in which $\bigtriangledown$ denotes the difference operator. $X_t(k)$ was used for evaluation of the number of zero-crossings by $X_t(k) = \begin{cases} 1, & \text{if } \Im_k\{x(t)\} \geq 0 \\ 0, & \text{if } \Im_k\{x(t)\} < 0 \end{cases}, k = 1, 2, ...; t = 1, ..., N$, and then HOC are calculated as $D_k = \sum_{t=2}^{N}[X_t(k) - X_{t-1}(k)]^2$.

In our experiment, we constructed the HOC-based feature vector F_{HOC} as $F_{HOC} = [D_1, D_2, ..., D_L], 1 < L \leq \mathcal{J}$, in which L represents the maximum order of F_{HOC} and $\mathcal{J}$ represents the maximum order of HOC.

3)$Time - Frequency\ Domain\ Feature\ Extraction$: Time-frequency spectrum (TFS) has been used for the analysis of EEG signals [32]. we conducted two method, i.e., short-time Fourier transform (STFT) [31] [53] and Hilbert-Huang transform (HHT) [35] [54] to evaluate the time-frequency features.

The time-frequency spectrum using STFT, which is a kind of linear decomposition of time series, are calculated by $TFS_{STFT}(t, f) = \mid \int_{\infty}^{+\infty} w(\tau - t)x(\tau)e^{-j2\pi f\tau}d\tau \mid^2$, in which $TFS_{STFT}(t, f)$ denotes the time-frequency distribution of the time series $x(t)$ and $w(\tau - t)$ represents the short-time analysis window.

The Hilebert-Huang Spectrum (HHS) consists of two parts, i.e., empirical mode decomposition (EMD) of a time series and HHT. The intrinsic mode functions (IMFs) are obtained via EMD: $x(t) = \sum_{i=1}^{K} IMF_i(t) + r_K(t)$, in which $r_K(t)$ represents the residue that is constant or monotonic signal. An analytic signal reconstructed by a conjugate pair (IMF and IMF_k^H) can be formulated as: $Z_k(t) = IMF_k(t) + jIMF_k^H(t) = A_k(t)e^{j\theta_k(t)}$, where $A_k(t)$ represents the instantaneous amplitude of $Z_k(t)$ and $\theta_k(t)$ denotes the instantaneous phase of $IMF_k(t)$. The instantaneous frequency can be evaluated by $f_i(t) = \frac{1}{2\pi}\frac{d\theta_i}{dt}$ and the original time series $x(t)$ can be obtained by $x(t) = \sum_{k=1}^{N} A_k(t)e^{j2\pi \int f_k(t)dt}$, where the squared amplitude $A_k^2(t)$ and instantaneous frequency $f_k(t)$ form the time-frequency spectrum based on HHT, i.e., $TFS_{HHT}(t, f)$.

B. ECG FEATURE EXTRACTION

ECG signals describe the changes of muscular contraction related with cardiovascular activity. To make ECG signals easier for classification, we extracted the frequency domain features and time domain features respectively.

Fast Fourier transform was conducted and the coefficients in frequency range 0-10 Hz were divided into eight nonover-

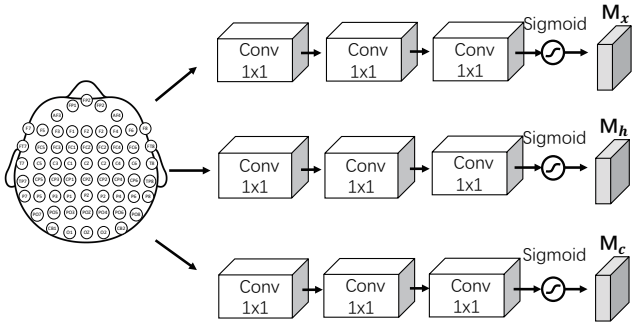


FIGURE 3. The illustration of proposed attention mechanism using EEG signals. We use three branches to output three masks for input data, hidden state and memory cell. The input data are extracted EEG features. Convolution kernel with size 1×1, Relu, and sigmoid function are used for the extraction of deep features.

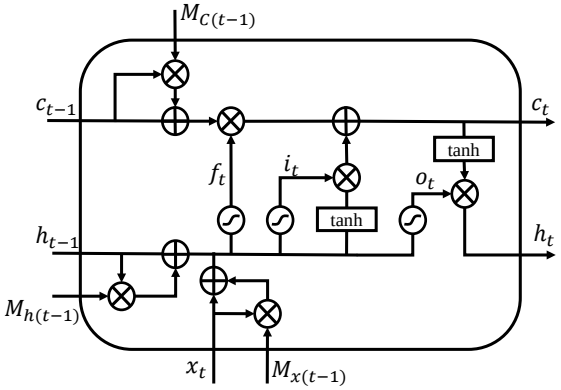


FIGURE 4. The A-LSTM block for a iterative process.

lapping subbands, which have equal bandwidth. The mean energy value and subband spectral entropy (SSE) of each subband were calculated as features. To evaluate SSE, we normalized the spectrum as $e_i = \frac{E_i}{\sum_{i=1}^{N} E_i}$, for $i = 1 \dots N$, in which E_i is the power of the ith frequency component of the spectrum and $\tilde{\mathbf{e}} = \{e_1 \dots e_N\}$ can be regarded as a probability mass function (PMK)-like form of the spectrum. For each subband, the SSE can be calculated by $H_{sub} = -\sum_{i=1}^{N} e_i \cdot log_2 e_i$, where N is the number of frequency components for each subband.

To extract time domain features, we first detect all the R peaks over the whole ECG signals during watching a video rather than 1-second signals and then calculate discrete Normal-to-normal (NN) interval values based on the detected R peaks, where each NN is located at the center position between two neighboring R peaks. Then, we apply the cubic spline interpolation approach to fit a cubic spline curve based on the discrete NN values so as to visualize the NN variations over the whole time. Finally, the cubic spline curve is sampled based on the same sampling rate of the ECG signals. In this case, we can finally obtain a set of sample points within 1 second and hence we are able to calculate both mean and the standard deviation of the sample points within 1 second as the time domain features of ECG signals within 1 second.

C. GSR FEATURE EXTRACTION

GSR describes the resistance of the skin via two electrodes on index and middle fingers. The resistance is affected by perspiration, which is related to various emotions. Former study [55] discovered that the value of GSR is linearly correlated to the level of arousal. To extract effective features, we used a low-passed filter with cutoff frequency of 0.2 Hz to remove noises. All the GSR signals were normalized and we calculated the mean value, standard deviation, and the mean of first and second derivations as features for emotion recognition.

D. RSP FEATURE EXTRACTION

RSP signals measure the changes of thoracic expansion, which contains less artifacts than other signals recorded by electrodes. For RSP signals, we commonly focused on breathing rate and the intensity. To extract intensity related features, we used similar methods with ECG features. The Fourier coefficients within range 0-0.8Hz were divided into two nonoverlapping subbands with equal bandwidth and the energy mean values and SSE values of each subband were evaluated. We evaluated the peaks of breathing and the peak-to-peak (PP) intervals. The mean value and the standard deviation of all PP intervals were calculated as the features related to breathing rate.

V. AN ATTENTION-LSTM METHOD

In this section, we will introduce the proposed A-LSTM, which selects sequences of interest to extract more discriminative features. The attention mechanism is motivated by the property of human perception. Humans tend to focus part of visual space to catch useful information which is needed, rather than perceiving the whole scene at once. The focused information will guide the next decision making.

Taking an EEG sample as example, the illustration of the proposed attention mechanism based on EEG signals are shown in Fig. 3. The proposed attention mechanism consists of three branches to produce three masks, which are used to select the information of input data, hidden states and memory cells. To avoid the interaction between different channels, we adopt convolution kernel with size 1×1 to extract deep features. The input of three attention branches are extracted EEG features, i.e., $x \in \mathbb{R}^{T \times K}$, where T is the number of EEG channel and K is the number of features in a EEG channel. We use the convolution kernel with size 1×1 to project EEG features to high level dimensions and this operation is conducted in each EEG channel. The 1×1 convolution operation are followed by a Relu activity function, which is a kind of non-linear transformation setting negative values to zeros. The third 1×1 convolutional operation in each branch is used for dimensionality reduction to output the desired size. Sigmoid activation functions are adopted

to normalize the output range to [0, 1]. Three masks, i.e., $M_x \in \mathbb{R}^{T\times K}$, $M_h \in \mathbb{R}^{T\times D_h}$ and $M_c \in \mathbb{R}^{T\times D_c}$, select the sequences of input data x, hidden state h and memory cell c, respectively, where D_h is the dimension of hidden state and D_c is the dimension of memory cell.

For an EEG sample, x_t represents the features in t-th EEG channel. The attention process can be described as

$$\begin{aligned}\mathcal{X}_t &= x_t + x_t \circ M_{x(t-1)}\\ \mathcal{H}_{t-1} &= h_{t-1} + h_{t-1} \circ M_{h(t-1)}\\ \mathcal{C}_{t-1} &= c_{t-1} + c_{t-1} \circ M_{c(t-1)}.\end{aligned} \tag{1}$$

In equations (1), we use a residual connection to capture the focused information. Compared with the connection that just calculates the elementwise production between attention mask and original information, the residual connection just strength the information of interest. The selected input data $\mathcal{X}_t$, hidden state $\mathcal{H}_{t-1}$ and memory cell $\mathcal{C}_{t-1}$ were processed by the following equation:

$$\begin{aligned}i_t &= \sigma(W_{xi}\mathcal{X}_t + W_{hi}\mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i)\\ f_t &= \sigma(W_{xf}\mathcal{X}_t + W_{hf}\mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f)\\ c_t &= f_t \circ \mathcal{C}_{t-1} + i_t \circ tanh(W_{xc}\mathcal{X}_t + W_{hc}\mathcal{H}_{t-1} + b_c)\\ o_t &= \sigma(W_{xo}\mathcal{X}_t + W_{ho}\mathcal{H}_{t-1} + W_{co} \circ c_t + b_o)\\ h_t &= o_t \circ tanh(c_t).\end{aligned} \tag{2}$$

In equations (2), t is the step number to iterate in A-LSTM and $\circ$ denotes elementwise product. b_i, b_f, b_c and b_o are four biases. As a key part of LSTM related methods, memory cell c_t can be regarded as a accumulator of the hidden states information. The memory cell c_t is controlled by a "input gate" i_t and a "forgotten gate" f_t. If i_t is activated based on selected input data $\mathcal{X}_t$, hidden state $\mathcal{H}_{t-1}$ and memory cell $\mathcal{C}_{t-1}$, the useful information is saved. If f_t is activated, the past selected memory cell $\mathcal{C}_{t-1}$ can be forgotten. The "output gate" o_t is evaluated by fusing $\mathcal{X}_t$, $\mathcal{H}_{t-1}$ and c_t via sigmoid activity function σ. Finally, o_t controls c_t that is processed by tanh activity function and the hidden state h_t is achieved based on o_t and c_t. The attention-LSTM block for a iterative process was presented in Fig. 4. To predict different emotion states, the hidden states of all iteration processes are projected to desired dimension. During the training process, we adopt the cross entropy function to measure the similarities between predicted label vectors and real label vectors.

The processes of other physiological signals, i.e., ECG, GSR and RSP, were similar with EEG signals and the extracted features was iterated by A-LSTM. For the fusion of four modal signals, we used four A-LSTM networks and the hidden states of four A-LSTM are concatenated into a vector, which is then projected to desired dimension to output the predicted label.

VI. EVALUATION

A. CLASSIFICATION PROTOCOL

Different emotions were induced by video clips, which elicited more intense emotion in the latter part. With this consideration, the last 120-second physiological signals during watching videos were used for emotion recognition. For each subject, there were physiological signals of 28 trials existing and all the physiological signals were divided into one-second samples with no overlap. The physiological signals of 7 trials consisting of seven emotions were served as testing data and the physiological signals of the rest trials were served as training data. Additionally, we presented three different kind of protocols for subject-dependent emotion recognition using physiological signals.

Protocol one : To explore the differences among discrete emotions and balance the training data, we conducted eight classification strategies (i.e., joy-N-anger, joy-N-fear, joy-N-disgust, joy-N-sad, funny-N-anger, funny-N-fear, funny-N-disgust and funny-N-sad) to distinguish positive, negative and neutral emotions.

Protocol two : To investigate the influences of unbalanced training data, we conducted positive-negative-neutral classification protocol and the data of negative emotions (anger, sad, disgust and fear) were larger than positive emotions (joy and funny) and neutrality, which was challenging with these unbalanced training data and testing data.

Protocol three : A seven emotions classification protocol was presented for multi-class emotion recognition, i.e., joy, funny, anger, fear, disgust, sad and neutrality.

TABLE 6. Time consumption of different classification methods when predicting 30 testing samples.

Method	Time consumption (s)	Method	Time consumption (s)
SVM	0.0284	STRNN	0.0451
KNN	0.0748	LSTM	0.1296
DBN	0.0099	A-LSTM	0.1878
DGCNN	0.0165	-	-

B. IMPLEMENTATION DETAILS

In this paper, we conduct various classification methods, i.e., SVM, KNN, DBN, STRNN, DGCNN, LSTM and A-LSTM. SVM is implemented by libsvm toolbox with linear kernel. KNN is conducted using Matlab with default value K=1. The DBN consists of three hidden layers and the dimension of each hidden layer is set to 500. The learning rate of DBN is set to 0.01. For STRNN, the numbers of the input, hidden, and output nodes are set to be 5, 30, and 30, respectively. For DGCNN, the dimensions of graph filtering layers are set to 64 and the learning rate is set to 0.01. For LSTM and A-LSTM, the learning rates are set to 0.01. The dimensions of hidden states and memory cells are all set to 128. For A-LSTM, the dimensions of convolutional layers in attention module are set to 32. SVM and KNN are applied with Intel(R) Core(TM) i7-4790K CPU. DBN, STRNN, DGCNN, LSTM and A-LSTM are applied with TITAN Xp. The time consumption of these classification methods when predicting 30 testing samples are presented in Table 6.

TABLE 7. The average accuracies and standard deviations (%) of subject dependent discrete emotion recognition in protocol one among the various methods using different single modal signals.

Modality	Method	joy-N-anger	joy-N-fear	joy-N-disgust	joy-N-sad	funny-N-anger	funny-N-fear	funny-N-disgust	funny-N-sad
EEG (HOC)	SVM	64.81 / 16.25	67.37 / 16.08	50.23 / 15.71	63.36 / 14.77	73.18 / 20.27	78.62 / 17.54	56.36 / 14.89	71.41 / 15.53
	KNN	47.75 / 17.11	61.46 / 17.00	47.52 / 14.29	58.45 / 18.37	58.51 / 20.67	71.61 / 15.41	53.82 / 16.03	70.93 / 17.08
	LSTM	74.09 / 14.28	75.47 / 15.42	65.41 / 14.32	70.01 / 13.49	75.07 / 15.83	78.89 / 15.63	70.39 / 12.83	77.60 / 14.96
	A-LSTM	**75.40/14.63**	**77.16/15.50**	**68.10/13.23**	**71.03/13.52**	**80.47/13.80**	**81.16/11.78**	**74.14/12.83**	**81.01/13.43**
EEG (Hjorth)	SVM	55.76 / 16.05	49.94 / 20.14	44.01 / 13.20	49.82 / 14.53	61.82 / 16.23	58.89 / 15.58	52.29 / 18.16	54.99 / 13.05
	KNN	45.51 / 14.34	39.74 / 18.04	37.92 / 13.89	40.43 / 13.34	59.44 / 18.32	50.86 / 17.21	48.19 / 13.68	53.58 / 15.58
	LSTM	65.65 / 15.38	60.07 / 16.66	55.17 / 13.43	63.44 / 14.81	71.85 / 16.38	67.66 / 13.16	64.17 / 12.95	68.26 / 12.53
	A-LSTM	**66.96/15.20**	**61.88/13.15**	**57.84/11.73**	**64.48/12.19**	**76.50/17.51**	**69.63/11.73**	**66.64/13.41**	**70.04/12.18**
EEG (PSD)	SVM	63.79 / 14.34	52.46 / 14.50	48.95 / 15.95	52.48 / 14.95	68.43 / 16.80	66.83 / 17.46	59.93 / 11.39	61.27 / 18.24
	KNN	46.40 / 10.85	40.22 / 11.68	38.37 / 9.68	41.22 / 9.48	55.30 / 13.25	51.53 / 12.96	50.93 / 10.67	51.63 / 12.70
	LSTM	68.44 / 14.62	62.10 / 14.61	59.13 / 12.42	63.13 / 14.67	75.65 / 17.87	72.09 / 17.98	70.87 /15.80	71.55 / 15.00
	A-LSTM	**73.09/15.61**	**65.85/13.63**	**62.04/12.25**	**64.35/16.79**	**79.47/16.04**	**75.28/18.35**	**73.80/13.74**	**72.52/14.73**
EEG (HHS)	SVM	61.43 / 18.26	57.72 / 14.31	49.36 / 19.99	54.19 / 18.29	68.45 / 15.96	68.12 / 18.60	59.13 / 13.59	63.31 / 16.70
	KNN	48.94 / 14.99	43.94 / 13.93	41.06 / 13.72	42.26 / 13.17	62.33 / 18.64	57.78 / 16.10	52.69 / 12.37	56.69 / 14.80
	LSTM	**77.44/12.00**	66.24 / 16.95	63.54 / 13.29	67.86 / 15.15	79.55/ 15.96	76.71 / 15.58	72.95 / 13.34	75.98 / 14.69
	A-LSTM	75.83 / 15.76	**67.62/16.72**	**65.53/12.75**	**70.01/12.13**	**82.34/15.04**	**78.76/15.86**	**74.21/12.96**	**75.99/13.86**
EEG (STFT)	SVM	62.53 / 17.66	55.98 / 16.69	49.30 / 17.77	54.32 / 14.93	68.99 / 17.63	65.10 / 15.51	59.61 / 13.59	65.29 / 17.99
	KNN	47.21 / 13.01	41.81 / 13.64	40.57 / 14.21	42.45 / 11.82	59.75 / 18.13	55.46 / 15.47	51.36 / 13.20	54.11 / 16.59
	LSTM	73.54 / 14.58	65.48 / 16.81	61.88 / 14.54	**68.51/15.04**	79.24 / 15.75	**77.40/14.97**	70.95 / 13.52	76.30 / 15.20
	A-LSTM	**74.42/13.93**	**66.50/13.77**	**63.62/11.93**	67.50 / 12.75	**83.07/13.93**	77.03 / 13.21	**72.78/11.83**	**78.49/14.19**
GSR	SVM	42.45 / 9.02	50.70 / 17.82	39.06 / 16.66	45.08 / 21.59	48.68 / 19.04	47.71 / 17.93	46.23 / 17.79	44.06 / 20.90
	KNN	42.43 / 9.07	45.37 / 14.82	39.58 / 13.53	41.73 / 16.61	47.56 / 14.43	47.39 / 14.67	43.76 / 15.24	46.25 / 14.28
	LSTM	61.09 / 15.04	64.59 / 15.77	56.24 / 13.42	59.49 / 18.01	63.86 / 11.66	**62.42/15.02**	61.82 / 14.99	59.26 / 17.98
	A-LSTM	**63.95/15.79**	**66.67/15.92**	**57.71/13.29**	**63.16/17.34**	**67.69/11.08**	62.25 / 15.53	**63.55/14.88**	**61.99/16.15**
RSP	SVM	31.36 / 6.97	31.90 / 13.14	32.67 / 12.43	35.52 / 10.01	34.37 / 5.13	37.09 / 12.84	35.11 / 6.32	34.34 / 11.39
	KNN	36.03 / 9.28	34.81 / 11.25	36.41 / 6.60	35.16 / 9.70	37.78 / 12.20	36.46 / 12.09	38.03 / 10.60	36.44 / 11.91
	LSTM	51.52 / 12.23	50.24 / 12.98	51.52 / 13.00	52.92 / 14.17	47.68 / 13.98	**50.17/14.87**	48.82 / 15.32	48.76 / 12.02
	A-LSTM	**52.97/12.45**	**50.86/12.61**	**51.59/13.16**	**53.09/13.90**	**50.40/14.20**	49.87 / 16.90	**50.04/13.30**	**49.46/12.93**
ECG	SVM	40.40 / 8.72	42.63 / 11.44	35.93 / 10.46	41.69 / 9.55	46.69 / 8.53	46.75 / 9.23	44.77 / 11.50	42.45 / 9.02
	KNN	42.57 / 8.03	41.82 / 8.90	36.55 / 5.91	38.57 / 9.28	46.36 / 9.64	44.71 / 9.28	43.14 / 7.45	42.42 / 9.07
	LSTM	47.84 / 6.60	49.15 / 8.55	43.83 / 6.69	48.54 / 9.56	53.16 / 8.83	53.08 / 8.07	51.73 / 9.21	50.79 / 8.11
	A-LSTM	**50.00/7.95**	**50.97/9.49**	**44.82/5.90**	**50.66/10.04**	**55.12/10.07**	**55.24/8.76**	**54.49/8.07**	**51.96/9.25**

C. COMPARISONS OF EVALUATION RESULTS

In Table 7 and Table 8, the emotion classification accuracies (standard deviations) in protocol one are presented. For single modal physiological signals, we can see that EEG signals achieved better classification results with any presented classification methods than GSR, RSP and ECG signals, which indicates that EEG signals are more effective to reflect physiological differences among positive, neutral and negative emotion states. According to Table 7, we can see that 'funny' is easier to be distinguished than 'joy' since classification strategies related to 'funny' have better performance than that related to 'joy'. Averagely, the accuracies using deep learning methods, i.e., LSTM and A-LSTM, are at least 10% higher than that using SVM and KNN. In all, the proposed A-LSTM achieved higher mean classification accuracies than SVM, KNN and LSTM, which demonstrates the efficacy of attention mechanism. Especially, HOC, HHS and STFT features achieve better performance than Hjorth and PSD features for EEG-based emotion recognition. To evaluate the fusion of multi-modal physiological signals in protocol one, the average emotion classification accuracies (standard deviations) of using single EEG signals and fusing EEG, GSR, RSP and ECG signals are displayed in Table 8. We average the classification results of eight classification strategies as the classification accuracies in Table 8. Four modal signals are fused with equal weights, which may limit the performance of different classification methods. Even so, A-LSTM and LSTM achieve higher accuracies than SVM and KNN when fusing four modal signals.

The results following protocol two, which is challenging due to the unbalanced training data, are shown in Table 9. With unbalanced training data and testing data, F1 scores are more significant. For emotion recognition using single model

TABLE 8. The average accuracies and standard deviations (%) of subject dependent discrete emotion recognition in protocol one using single modal signals and four modal signals.

Modality	Method	EEG (HOC)	EEG (Hjorth)	EEG (PSD)	EEG (HHS)	EEG (STFT)	GSR	RSP	ECG
Single modality	SVM	65.67 / 16.38	54.22 / 15.26	59.61 / 15.20	60.21 / 16.91	59.86 / 16.29	45.50 / 17.59	34.05 / 9.78	42.66 / 9.81
	KNN	58.76 / 16.99	47.67 / 15.49	46.95 / 11.41	50.71 / 14.72	49.31 / 14.35	44.26 / 14.08	36.39 / 9.28	40.02 / 8.45
	LSTM	73.37 / 14.60	64.53 / 14.41	67.87 / 15.37	72.87 / 14.49	72.09 / 14.94	61.10 / 15.24	50.20 / 13.57	49.77 / 8.20
	A-LSTM	**76.06/13.59**	66.75 / 13.39	70.88 / 15.00	73.79 / 14.39	72.93 / 13.19	**63.37/15.00**	**51.65/8.69**	**51.66/8.69**
Four modalities	SVM	48.73 / 18.52	39.21 / 14.75	35.24 / 9.39	42.61 / 15.29	44.62 / 15.01	-	-	-
	KNN	36.27 / 10.39	36.37 / 10.63	36.35 / 10.50	36.43 / 10.54	36.47/ 10.51	-	-	-
	LSTM	75.26 / 14.70	71.38 / 14.00	70.98 / 13.78	**78.79/13.48**	**77.99/13.32**	-	-	-
	A-LSTM	76.01 / 13.99	**71.81/13.52**	**72.02/13.45**	77.27 / 13.90	76.48 / 13.51	-	-	-

TABLE 9. The average accuracies and F1 scores (%) of subject dependent discrete emotion recognition in protocol two using single modal signals and four modal signals.

Modality	Method	EEG (HOC)	EEG (Hjorth)	EEG (PSD)	EEG (HHS)	EEG (STFT)	GSR	RSP	ECG
Single modality	SVM	67.18 / 57.24	61.69 / 51.70	51.14 / 24.24	61.83 / 54.20	57.06 / 24.43	58.70 / 28.56	41.87 / 24.09	57.53 / 26.66
	KNN	64.78 / 54.86	55.87 / 46.13	43.91 / 33.79	56.71 / 48.01	43.96 / 34.39	46.78 / 38.87	44.17 / 33.31	**45.75/37.52**
	LSTM	69.72 / 60.83	67.91 / 57.78	68.01 / 57.95	71.87 / 65.84	71.92 / 65.12	55.30 / 40.41	55.31 / 32.04	53.70 / 34.25
	A-LSTM	**72.27/67.12**	**68.45/60.48**	**70.46/65.93**	72.48 / 66.14	71.57 / 67.74	**60.24/44.61**	**56.17/34.51**	53.20 / 35.22
Four modalities	SVM	59.19/ 44.90	48.55 / 31.37	38.62 / 21.51	46.61 / 35.29	51.39/ 40.46	-	-	-
	KNN	44.52 / 33.70	44.39 / 33.35	44.01 / 33.33	44.64 / 33.68	44.94 / 33.84	-	-	-
	LSTM	65.82 / 58.68	67.45 / 60.46	61.51 / 51.24	71.44 / 65.26	70.51 / 65.24	-	-	-
	A-LSTM	65.63 / 59.67	67.21 / 60.33	60.69 / 52.63	**72.02/67.29**	**72.36/68.22**	-	-	-

TABLE 10. The average accuracies and standard deviations (%) of subject dependent discrete emotion recognition in protocol three using single modal signals and four modal signals.

Modality	Method	EEG (HOC)	EEG (Hjorth)	EEG (PSD)	EEG (HHS)	EEG (STFT)	GSR	RSP	ECG
Single modality	SVM	34.03 / 7.80	28.48 / 7.24	31.34 / 6.77	31.16 / 9.13	31.14 / 8.06	21.21 / 10.11	14.55 / 4.74	19.00 / 4.16
	KNN	29.18 / 8.06	23.72 / 7.87	22.35 / 4.62	25.07 / 8.62	23.65 / 8.05	18.63 / 5.66	15.54 / 4.20	18.63 / 3.72
	LSTM	37.46 / 10.11	33.62 / 8.23	30.09 / 8.31	39.71 / 8.66	38.55 / 8.43	30.07 / 6.73	23.34 / 5.28	20.74 / 3.44
	A-LSTM	**38.43/11.81**	33.72 / 6.69	26.42 / 10.14	39.21 / 8.71	38.74 / 7.75	**31.19/7.19**	**23.65/7.39**	**25.10/5.15**
Four modalities	SVM	25.39 / 9.33	18.53 / 6.94	15.03 / 3.77	18.69 / 8.11	21.21 / 8.84	-	-	-
	KNN	15.71 / 4.21	15.44 / 4.39	15.47 / 4.23	15.70 / 4.34	15.70 / 4.32	-	-	-
	LSTM	37.49 / 8.76	37.33 / 8.79	31.11 / 6.51	40.95 / 8.03	41.82 / 7.12	-	-	-
	A-LSTM	38.35 / 8.24	**38.52/8.04**	**33.52/7.29**	**41.88/7.67**	**42.10/8.09**	-	-	-

signals, A-LSTM achieves higher F1 scores than LSTM, SVM and KNN. With HOC, Hjorth and PSD as EEG features, the F1 scores fusing four modal signals are lower than that using single EEG signals. With HHS and STFT as EEG features, A-LSTM achieves the best F1 scores among different classification methods. Among different modal signals, EEG signals achieve the best F1 scores, which demonstrates the efficacy of EEG to deal with the unbalanced training samples. Additionally, we present confusion matrixes using different single model signals with A-LSTM in Fig 5. For EEG signals, the accuracy of each emotion category is comparable. However, GSR, ECG and RSP achieve a poor performance that almost testing data were recognized to negative category. Compared with the result using single EEG signals, the improvement of both accuracies and F1 scores fusing multi-modal signals is limited and the result with HOC and PSD as EEG features even decreases a lot due to the unbalanced training data.

Following protocol three, the classification results using different single modal physiological signals and fusing four modal physiological signals are shown in Table 10. The results show that again deep learning methods, i.e., LSTM and A-LSTM, have better performance than traditional methods, i.e., SVM and KNN. Averagely, EEG and GSR achieve higher accuracies for emotion recognition than ECG and GSR. For discrete emotion recognition using single GSR, RSP and ECG signals, A-LSTM achieves better performance than any other methods with the accuracies of 31.19%, 23.65% and 25.10%, respectively. With SVM and KNN as classifiers, the fusion of multi-modal signals achieves lower classification accuracies. However, deep learning methods, especially A-LSTM, improve the classification accuracies on

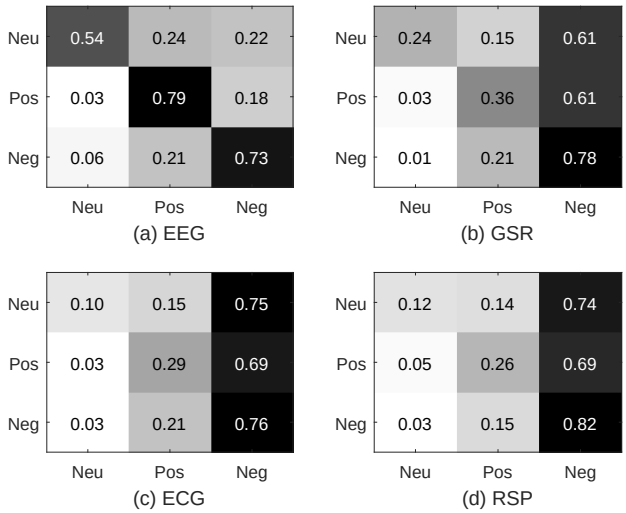


FIGURE 5. The confusion matrixes using different single model signals with A-LSTM in protocol two.

TABLE 11. Experiment results(%) in protocol one and protocol three (average accuracies and standard deviations), as well as in protocol two (average accuracies and F1 scores) using two types combination of EEG, ECG, RSP and GSR.

Protocol	Modality	SVM	KNN	LSTM	A-LSTM
Protocol One (acc/std)	EEG + GSR	59.32 / 16.32	49.35 / 14.33	74.94 / 14.34	**75.35/13.35**
	EEG + ECG	60.10 / 16.15	49.33 / 14.37	72.28 / 14.10	**72.70/13.40**
	EEG + RSP	40.54 / 13.58	36.46 / 10.74	72.10 / 14.60	**72.24/13.86**
	GSR + ECG	51.90 / 14.16	46.75 / 10.92	63.99 / 13.44	**65.19/13.99**
	GSR + RSP	31.71 / 8.51	36.46 / 10.80	63.61 / 15.82	**65.44/14.96**
	ECG + RSP	31.73 / 7.67	36.39 / 10.7	51.93 / 8.02	**54.83/9.47**
Protocol Two (acc/F1)	EEG + GSR	60.85 / 54.09	55.47 / 46.73	70.07 / 62.42	**71.33/65.67**
	EEG + ECG	60.63 / 54.19	55.57 / 46.71	71.51 / 63.90	**72.58/66.58**
	EEG + RSP	50.09 / 35.77	44.93 / 33.84	72.04 / 63.54	**72.94/66.52**
	GSR + ECG	58.03 / 35.23	47.26 / 40.46	57.10 / 42.89	**58.54/46.59**
	GSR + RSP	42.62 / 23.08	44.21 / 33.41	59.16 / 45.20	**60.66/48.23**
	ECG + RSP	39.45 / 21.54	44.14 / 33.53	53.45 / 35.89	**53.80/40.94**
Protocol Three (acc/std)	EEG + GSR	31.54 / 8.24	23.68 / 7.04	40.34 / 8.69	**40.65/7.96**
	EEG + ECG	31.09 / 7.64	23.74 / 7.06	38.66 / 8.04	**38.96/7.71**
	EEG + RSP	19.79 / 7.14	15.70 / 4.32	39.22 / 8.40	**39.79/8.25**
	GSR + ECG	23.42 / 7.62	20.63 / 4.07	32.73 / 6.57	**32.97/7.40**
	GSR + RSP	14.45 / 4.97	15.61 / 4.58	31.92 / 7.82	**34.20/8.40**
	ECG + RSP	13.64 / 3.94	15.65 / 4.26	25.29 / 4.28	**27.01/4.36**

emotion recognition. With HHS and STFT as EEG features, A-LSTM achieves the accuracies of 41.88% and 42.10% respectively, via fusing four modal physiological signals.

To investigate more combination modes of fusing multi-modal signals, we present the classification results using two types combination of EEG, ECG, RSP and GSR in Table 11. The results show that again A-LSTM has better performance than LSTM, SVM and KNN. Simultaneously, the results demonstrate that EEG is more effective to be fused for emotion recognition.

To demonstrate the efficacy of the proposed method,

TABLE 12. Experiment results(%) in protocol one and protocol three (average accuracies and standard deviations), as well as in protocol two (average accuracies and F1 scores) using single EEG (STFT) signals.

Method	Protocol One (acc/std)	Protocol Two (acc/F1)	Protocol Three (acc/std)
SVM	59.86 / 16.29	57.06 / 24.43	31.14 / 8.06
KNN	49.31 / 14.35	43.96 / 34.39	23.65 / 8.05
DBN	65.83 / 13.20	65.98 / 59.19	29.26 / 9.19
STRNN	65.38 / 13.20	66.84 / 60.57	35.64 / 9.57
DGCNN	71.13 / 15.77	68.02 / 61.11	36.92 / 12.78
LSTM	72.09 / 14.94	71.92 / 65.12	38.55 / 8.43
LSTM+X	71.09 / 14.50	71.56 / 65.30	38.31 / 8.74
LSTM+H	71.43 / 14.51	71.53 / 65.16	38.59 / 8.10
LSTM+C	71.78 / 13.80	71.71 / 65.67	38.13 / 7.83
A-LSTM	**72.93/13.19**	**71.57/67.74**	**38.74/7.75**

we conduct extensive experiments for EEG-based emotion recognition in Table 12. The proposed A-LSTM achieves better results than SVM, KNN, DBN, STRNN and DGCNN. Besides, the results on emotion recognition by fusing LSTM with each of three attention branches are presented in Table 12, where 'LSTM+X', 'LSTM+H', 'LSTM+C' indicate the fusion of LSTM with three attention branches for input data , hidden state and memory cell, respectively. In protocol one and protocol three, although the LSTM achieves higher recognition accuracies than that using each of three attention branches, the A-LSTM using three attention branches achieves the best recognition accuracies. In protocol two, the F1 scores using each of three attention branches is higher than LSTM and A-LSTM achieves the highest F1 score.

D. CORRELATIONS BETWEEN EEG SIGNALS AND RATINGS

In this part, the further investigation of correlations between the self-assessment rating and EEG signals is conducted. In the self-assessment process, the participants rated every videos with arousal, valence and SAM of 1-9 scales. In SAM, there are three-items words related to positive emotions (joy and funny) and twelve-items words related to negative emotions (anger, fear, disgust, sad), three-items word for each negative emotion. For each emotion, we calculate the average rating of three-items words as the subjective scores. The PSD features from five different frequency bands (i.e., delta band, theta band, alpha band, beta band and gamma band) are used for the measurement of power changes. In Fig. 6, the Spearman correlation coefficients between the energy changes and the subjective scores are used to measure the correlations between different EEG channels and various emotions.

For positive emotions, i.e., joy and funny, we can obviously found that the power in nearly all areas of the scalp are negatively correlated with the level of positive emotions, especially in theta band and alpha band. In beta band and gamma band, energy in prefrontal area of the scalp has strong negative correlations with the level of positive emotions and

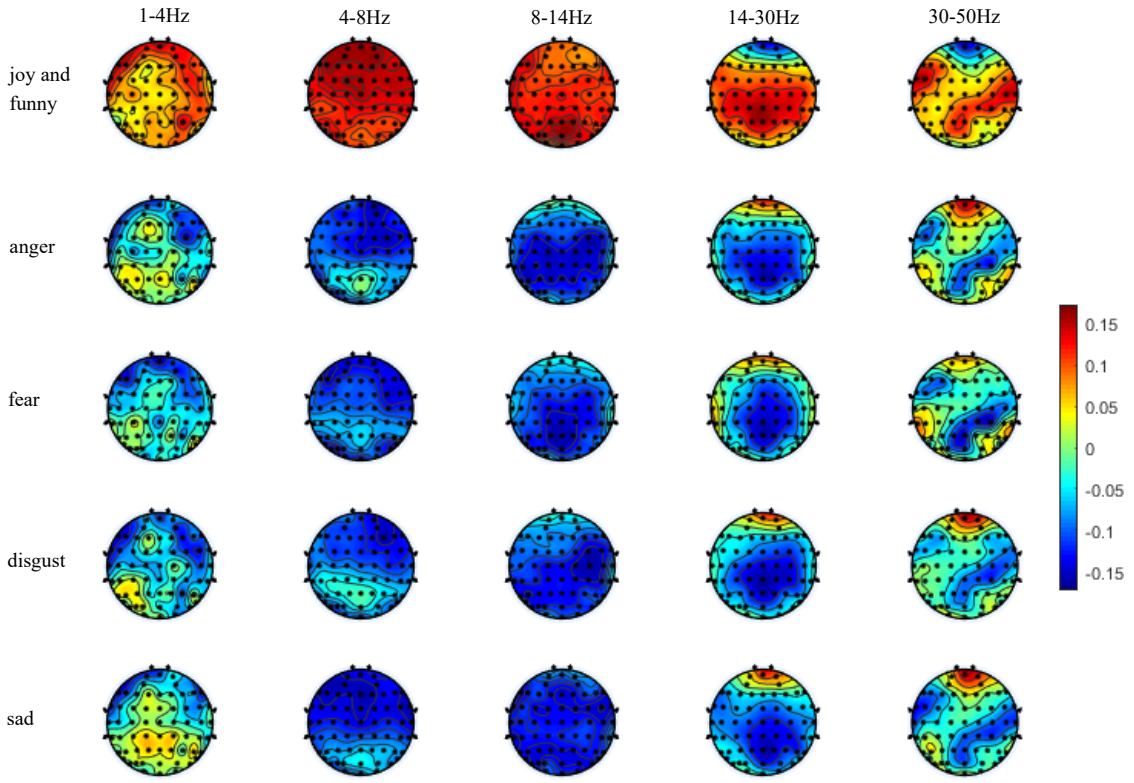


FIGURE 6. The average spearman correlation coefficients of joy, funny, anger, fear, disgust, sad and self-assessment ratings with power changes in five frequency bands of delta (1-4Hz), theta (4-8Hz), alpha (8-14Hz), beta (14-30Hz) and gamma (30-50Hz).

the posterior part in the scalp is positively correlated with the positive emotions. Negative emotions show the different correlation rule that nearly all areas in the scalp have the negative correlations with various negative emotion states in theta band and alpha band. In the higher frequency band, negative emotions have higher positive correlations with the energy in the prefrontal area. Compared with other negative emotions, fear emotion state indicates the lower positive correlations with the energy changes of prefrontal area in beta and gamma band and higher negative correlations with the energy changes of the posterior part in gamma band. Besides, fear emotion state has lower negative correlations with power changes of the posterior part in theta band than anger, disgust and sad. From the point of different frequency band, negative emotion states reflect the higher positive correlations with the energy changes of posterior part of the scalp in the low frequency band (delta band). The average spearman correlation coefficients among negative emotions have the similar pattern, and it satisfies that negative emotion categories are to a large extent overlapped [48] [17].

Former studies [20] [32] on emotion recognition demonstrated that EEG-based emotion classification accuracies using energy features in beta and gamma bands are higher than that in the lower frequency bands, which proved that these energy features in high frequency bands are more useful for emotion recognition. Besides, some works [56] [57] proved that the prefrontal part is highly related with human emotions. In terms of the average spearman correlation coefficients of our experiment results in beta and gamma bands, we can summarize that the energy in prefrontal part is positively correlated with the level of positive emotion states and negative correlated with the level of negative emotion states.

VII. CONCLUSIONS

In this paper, we presented a MPED for discrete emotion recognition. The MPED consists of multi-modal physiological signals of 23 participants. Each participant watched 28 video clips describing seven different emotions, i.e., joy, funny, anger, sad, disgust, fear and neutrality. The selection of these video clips were assessed by three psychological questionnaires (PANAS, SAM and DES), which guaranteed the effectiveness of these elicitation materials. Besides, T-test was conducted to evaluate the effectiveness of the selected elicitation materials and the results validated the selection approach we used here. People who assessed the elicitation materials share similar culture background with people participating emotion elicitation experiment such that their understanding will not diverge a lot. The self-assessment during the recording of physiological signals was put at the end of experiment to prevent exhausting participants through cutting their wearing time of experiment devices.

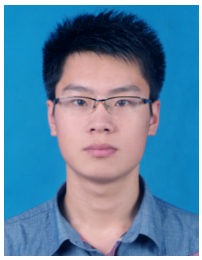
With different feature extraction and classification methods, we presented three protocols for subject-dependent discrete emotion classification. Averagely, the result demonstrated the efficacy of the proposed A-LSTM, which achieved better result than LSTM, SVM and KNN. In protocol one and protocol three, the fusion of multi-modal physiological signals using A-LSTM and LSTM apparently improved the classification accuracies. In protocol two, although fusing multi-modal physiological signals presented a feasible way to improve the classification accuracies on discrete emotion recognition, the improvement was limited and it was crucial to develop more algorithms to deal with the challenging task. Besides, we fuse multi-modal physiological signals with equal weights, which may limit the performance on attention process and fusing multi-modal signals. Therefore, there still remains much room for improvement.

Finally, significant correlates were found between the participant ratings and EEG signals. The spearman correlation coefficients analysis of this database demonstrated that the energy in prefrontal part is positively correlated with the level of positive emotion states and negative correlated with the level of negative emotion states. The difference of coefficients distribution between positive emotions and negative emotions was apparent to be distinguished and The coefficient distributions among negative emotions were similar, which validated that negative emotion categories are to a large extent overlapped.

REFERENCES

[1] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 36, no. 1, pp. 96–105, 2006.
[2] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (kcca)," IEEE transactions on neural networks, vol. 17, no. 1, pp. 233–238, 2006.
[3] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2528–2536, 2016.
[4] P. Song and W. Zheng, "Feature selection based transfer subspace learning for speech emotion recognition," IEEE Transactions on Affective Computing, 2018.
[5] W. Zheng, "Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis," IEEE Transactions on Cognitive and Developmental Systems, vol. 9, no. 3, pp. 281–290, 2017.
[6] J. Cai, G. Liu, and M. Hao, "The research on emotion recognition from ecg signal," in International Conference on Information Technology and Computer Science, 2009, pp. 497–500.
[7] G. Wu, G. Liu, and M. Hao, "The analysis of emotion recognition from gsr based on pso," in International Symposium on Intelligence Information Processing and Trusted Computing (IPTC), 2010, pp. 360–363.
[8] P. Philippot, G. Chapelle, and S. Blairy, "Respiratory feedback in the generation of emotion," Cognition & Emotion, vol. 16, no. 5, pp. 605–627, 2002.
[9] L.-C. Shi and B.-L. Lu, "Eeg-based vigilance estimation using extreme learning machines," Neurocomputing, vol. 102, pp. 135–143, 2013.
[10] C. A. Kothe and S. Makeig, "Estimation of task workload from eeg data: new and current tools and perspectives," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 6547–6551.
[11] J. A. Russell, "A circumplex model of affect." Journal of personality and social psychology, vol. 39, no. 6, p. 1161, 1980.
[12] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," Current Psychology, vol. 14, no. 4, pp. 261–292, 1996.
[13] A. Ben-Zeev, "The nature of emotions," Philosophical Studies, vol. 52, no. 3, pp. 393–409, 1987.
[14] W. G. Parrott, Emotions in social psychology: Essential readings. Psychology Press, 2001.
[15] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti et al., "Universals and cultural differences in the judgments of facial expressions of emotion." Journal of personality and social psychology, vol. 53, no. 4, p. 712, 1987.
[16] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," Psychological science, vol. 18, no. 12, pp. 1050–1057, 2007.
[17] S. Hamann, "Mapping discrete and dimensional emotions onto the brain: controversies and consensus," Trends in cognitive sciences, vol. 16, no. 9, pp. 458–466, 2012.
[18] P. J. Lang, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," Technical report, 2005.
[19] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, and L. Albera, "Emotion recognition based on high-resolution eeg recordings and reconstructed brain sources," IEEE Transactions on Affective Computing, 2017.
[20] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," IEEE Transactions on Autonomous Mental Development, vol. 7, no. 3, pp. 162–175, 2015.
[21] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," IEEE transactions on pattern analysis and machine intelligence, vol. 30, no. 12, pp. 2067–2083, 2008.
[22] A. Savran, K. Ciftci, G. Chanel, J. C. Mota, L. H. Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut, "Emotion detection in the loop from brain signals and facial images," Acta Horticulturae, vol. 671, no. 671, pp. 151–157, 2006.
[23] K. L. Phan, T. Wager, S. F. Taylor, and I. Liberzon, "Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in pet and fmri," Neuroimage, vol. 16, no. 2, pp. 331–348, 2002.
[24] A. Graves, Long Short-Term Memory. Springer Berlin Heidelberg, 2012.
[25] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial-temporal recurrent neural network for emotion recognition," IEEE Transactions on Cybernetics, vol. PP, no. 99, pp. 1–9, 2017.
[26] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," IEEE Transactions on Affective Computing, 2018.
[27] D. O. Bos et al., "Eeg-based emotion recognition," The Influence of Visual and Auditory Stimuli, vol. 56, no. 3, pp. 1–17, 2006.
[28] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18–31, 2012.
[29] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42–55, 2012.
[30] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," IEEE journal of biomedical and health informatics, vol. 22, no. 1, pp. 98–107, 2018.
[31] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "Eeg-based emotion recognition in music listening," IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1798–1806, 2010.
[32] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an eeg-based recognition of music liking using time-frequency analysis," IEEE Transactions on Biomedical Engineering, vol. 59, no. 12, pp. 3498–3510, 2012.
[33] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from eeg signals," IEEE Transactions on Affective Computing, 2017.
[34] Y. Li, W. Zheng, Z. Cui, Y. Zong, and S. Ge, "Eeg emotion recognition based on graph regularized sparse linear regression," Neural Processing Letters, pp. 1–17, 2018.
[35] N. E. Huang and Z. Wu, "A review on hilbert-huang transform: Method and its applications to geophysical studies," Reviews of geophysics, vol. 46, no. 2, 2008.

[36] B. Hjorth, "Eeg analysis based on time domain properties," Electroencephalography and clinical neurophysiology, vol. 29, no. 3, pp. 306–310, 1970.
[37] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from eeg using higher order crossings," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 186–197, 2010.
[38] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines. ACM, 2011.
[39] F. Sebastiani, "Machine learning in automated text categorization," Acm Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[41] M. Zaheer, A. Ahmed, and A. J. Smola, "Latent lstm allocation: Joint clustering and non-linear dynamic modeling of sequence data," in International Conference on Machine Learning, 2017, pp. 3967–3976.
[42] S. K. Sønderby and O. Winther, "Protein secondary structure prediction with long short term memory networks," arXiv preprint arXiv:1412.7828, 2014.
[43] T. Luong, I. Sutskever, Q. Le, O. Vinyals, and W. Zaremba, "Addressing the rare word problem in neural machine translation," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), vol. 1, 2015, pp. 11–19.
[44] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales." Journal of personality and social psychology, vol. 54, no. 6, p. 1063, 1988.
[45] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," Journal of behavior therapy and experimental psychiatry, vol. 25, no. 1, pp. 49–59, 1994.
[46] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," Cognition & emotion, vol. 9, no. 1, pp. 87–108, 1995.
[47] N. H. Frijda, The emotions. Cambridge University Press, 1986.
[48] M. Tamietto and B. De Gelder, "Neural bases of the non-conscious perception of emotional signals," Nature Reviews Neuroscience, vol. 11, no. 10, p. 697, 2010.
[49] P. Kirsch, C. Esslinger, Q. Chen, D. Mier, S. Lis, S. Siddhanti, H. Gruppe, V. S. Mattay, B. Gallhofer, and A. Meyer-Lindenberg, "Oxytocin modulates neural circuitry for social cognition and fear in humans," Journal of neuroscience, vol. 25, no. 49, pp. 11 489–11 493, 2005.
[50] M. J. Miserendino, C. B. Sananes, K. R. Melia, and M. Davis, "Blocking of acquisition but not expression of conditioned fear-potentiated startle by nmda antagonists in the amygdala," Nature, vol. 345, no. 6277, p. 716, 1990.
[51] C. Méndez-Bértolo, S. Moratti, R. Toledano, F. Lopez-Sosa, R. Martínez-Alvarez, Y. H. Mah, P. Vuilleumier, A. Gil-Nagel, and B. A. Strange, "A fast pathway for fear in human amygdala," Nature neuroscience, vol. 19, no. 8, p. 1041, 2016.
[52] T. Musha, Y. Terasaki, H. A. Haque, and G. A. Ivamitsky, "Feature extraction from eegs associated with emotions," Artificial Life and Robotics, vol. 1, no. 1, pp. 15–19, 1997.
[53] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music," Psychophysiology, vol. 44, no. 2, pp. 293–304, 2007.
[54] N. E. Huang, Hilbert-Huang transform and its applications. World Scientific, 2014, vol. 16.
[55] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," Psychophysiology, vol. 30, no. 3, pp. 261–273, 1993.
[56] R. N. Cardinal, J. A. Parkinson, J. Hall, and B. J. Everitt, "Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex," Neuroscience & Biobehavioral Reviews, vol. 26, no. 3, pp. 321–352, 2002.
[57] A. R. Damasio, "The somatic marker hypothesis and the possible functions of the prefrontal cortex," Phil. Trans. R. Soc. Lond. B, vol. 351, no. 1346, pp. 1413–1420, 1996.



TENGFEI SONG Tengfei Song received the B.S. degree in communication engineering from Hohai University, Jiangsu, China, in 2016. He is currently working towards his Ph.D. degree in the Department of information and communication engineering of Southeast University, China. His research interests include affective computing, computer vision, machine learning and pattern recognition.

WENMING ZHENG (SM'18) Wenming Zheng received the B.S. degree in computer science from Fuzhou University, Fuzhou, China, in 1997, the M.S. degree in computer science from Huaqiao University, Quanzhou, China, in 2001, and the Ph.D. degree in signal processing from Southeast University, Nanjing, China, in 2004. Since 2004, he has been with the Research Center for Learning Science, Southeast University. He is currently a Professor with the Key Laboratory of Child Development and Learning Science, Ministry of Education, Southeast University. His research interests include affective computing, pattern recognition, machine learning, and computer vision. He is an associated editor of IEEE Transactions on Affective Computing, an associated editor of Neurocomputing and also an editorial board member of The Visual Computer.

CHENG LU Cheng Lu received the B.S. and M.S.degree from the School of Computer Science and Technology, Anhui University, China, in 2013 and 2017, respectively. Currently, he is a Ph.D. candidate in the School of Information Science and Engineering, Southeast University, under the supervision of professor Wenming Zheng. His research interests include affective computing, machine learning and pattern recognition.

YUAN ZONG Yuan Zong received the BS and MS degrees in electronics engineering from Nanjing Normal University, Nanjing, China, in 2011 and 2014, respectively, and the PhD degree in Biomedical Engineering from Southeast University, Nanjing, China, in 2018. He is currently a Lecturer with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Sciences and Medical Engineering, Southeast University. From 2016 to 2017, he was a Visiting Student with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. His research interests include affective computing, pattern recognition, and computer vision.

XILEI ZHANG Xilei Zhang received the M.S. degree in basic psychology in 2014 from Faculty of Psychology, South-West University (SWU), Chongqing, China, and received the Ph.D. degree in cognitive neuroscience in 2018, in Institute of Psychology, Chinese Academy of Sciences (CAS), Beijing, China. Currently, he is a post-doctoral fellow, working with Professor Wenming Zheng in the Key Laboratory of Child Development and Learning Science, Ministry of Education, Southeast University. His research intersects cognitive neuroscience (psychophysics and neuroimaging) and machine learning methods to unravel the neural signatures of consciousness under the contexts of conscious and unconscious processing of visual fear.

ZHEN CUI Received the Ph.D. degree in computer science from Institute of Computing Technology (ICT), Chinese Academy of Science (CAS), Beijing, in Jun. 2014. He was a Research Fellow in the Department of Electrical and Computer Engineerning at National University of Singapore (NUS) from Sep. 2014 to Nov. 2015. He also spent half a year as a Research Assistant on Nanyang Technological University (NTU) from Jun. 2012 to Dec. 2012. Now he is a professor of Nanjing University of Science and Technology, China. His research interests cover computer vision, pattern recognition and machine learning, especially focusing on deep learning, manifold learning, sparse coding, face detection/alignment/recognition, object tracking, image super resolution, emotion analysis, etc.