scientific data

OPEN

DATA DESCRIPTOR

A Multimodal Dataset for Mixed Emotion Recognition

Pei Yang[1,5], Niqi Liu[1,5], Xinge Liu[1,5], Yezhi Shu[1], Wenqi Ji[1], Ziqi Ren[1], Jenny Sheng[1], Minjing Yu[2], Ran Yi[3], Dan Zhang[4] & Yong-Jin Liu[1]✉

Mixed emotions have attracted increasing interest recently, but existing datasets rarely focus on mixed emotion recognition from multimodal signals, hindering the affective computing of mixed emotions. On this basis, we present a multimodal dataset with four kinds of signals recorded while watching mixed and non-mixed emotion videos. To ensure effective emotion induction, we first implemented a rule-based video filtering step to select the videos that could elicit stronger positive, negative, and mixed emotions. Then, an experiment with 80 participants was conducted, in which the data of EEG, GSR, PPG, and frontal face videos were recorded while they watched the selected video clips. We also recorded the subjective emotional rating on PANAS, VAD, and amusement-disgust dimensions. In total, the dataset consists of multimodal signal data and self-assessment data from 73 participants. We also present technical validations for emotion induction and mixed emotion classification from physiological signals and face videos. The average accuracy of the 3-class classification (i.e., positive, negative, and mixed) can reach 80.96% when using SVM and features from all modalities, which indicates the possibility of identifying mixed emotional states.

Background & Summary

Affective computing plays an increasingly important role, especially in the era of Emotional Intelligence[1,2]. Among various affective computing tasks, emotion recognition is a key topic to achieving emotional intelligence[3]. Emotion recognition aims to enable machines to recognize people's emotional states automatically. Similar to other machine learning problems, data is a prerequisite for this kind of machine intelligence. Consequently, several datasets have been proposed to promote the research of emotion recognition from physiological and behavioral signals. For example, the DEAP dataset[4] collected Electroencephalogram (EEG), Galvanic Skin Response (GSR), blood volume pressure, respiration rate, skin temperature (ST), and Electrooculogram (EOG) signals of 32 participants. To contribute to affect recognition and implicit tagging research, Soleymani *et al.* constructed a multimodal database MAHNOB-HCI[5] consisting of recorded face videos, audio signals, eye gaze data, and psychological signals (i.e. EEG, ECG, GSR, respiration amplitude, ST). Recently, Park *et al.* created a multimodal sensor dataset K-EmoCon[6] for continuous emotion recognition in naturalistic conversations. Bota *et al.* released G-REx[7] to support research on group emotion analysis based on (photoplethysmography (PPG) and electrodermal activity (EDA) in real-world settings. Other datasets for emotion analysis include DECAF[8] (a database similar to DEAP, using Magnetoencephalogram (MEG) instead of EEG), AMIGOS[9], etc.

Although these existing datasets greatly promote the research of emotion recognition, they either focus on discrete emotion classification or focus on emotion classification in valence-arousal (VA) space. Mixed emotions are an important topic in emotion analysis and have received increasing attention[10–16]. Mixed emotion refers to the emotional state that is characterized by the co-occurrence of two or more emotional feelings[10], e.g., experiencing both positive and negative emotions. Although there are many opinions about whether human beings can feel mixed emotions[11–13], more and more evidence supports the fact that people can experience mixed emotional feelings[10,13,17]. One crucial issue for mixed emotion analysis is how to identify mixed emotional states. The most popular method used in previous works is the subjective reports, which measures mixed emotions through subject self-reports[17,18]. However, the limitation of subjective methods is that they may be subject to biases derived from memory[19]. In contrast, objective methods are less susceptible to subjective factors. Consequently, it

[1]Tsinghua University, Department of Computer Science and Technology, Beijing, 100084, China. [2]Tianjin University, College of Intelligence and Computing, Tianjin, 300350, China. [3]Shanghai Jiao Tong University, Department of Computer Science and Engineering, Shanghai, 200240, China. [4]Tsinghua University, Department of Psychology, Beijing, 100084, China. [5]These authors contributed equally: Pei Yang, Niqi Liu, Xinge Liu. ✉e-mail: liuyongjin@tsinghua.edu.cn

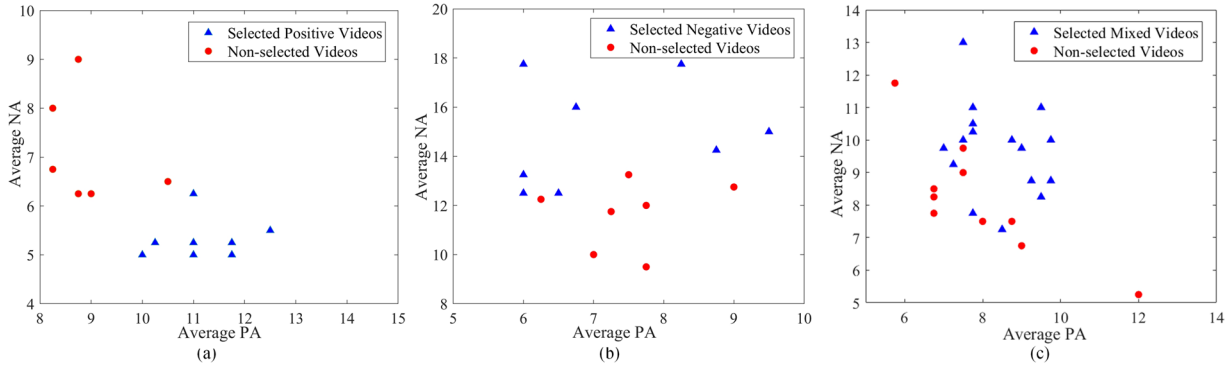

Fig. 1 Scatter plots of Average NA versus Average PA for selected and non-selected videos. (**a**) Scatter plot for selected and non-selected positive videos. (**b**) Scatter plot for selected and non-selected negative videos. (**c**) Scatter plot for selected and non-selected mixed videos.

is necessary to explore objective assessment methods based on expressive (e.g., facial expression) and physiological signals. Dataset is a key issue in the objective assessment of mixed emotions as none of the existing datasets involve mixed emotions. Therefore, it is urgent to establish a dataset for mixed emotion research.

Reliable emotion induction is a main challenge for the construction of mixed emotion datasets. Various kinds of stimuli, including self-selected music[20], motion pictures[21,22], video clips[23], etc., have been used to inducing mixed emotions. On the one hand, video clips are more effective at inducing emotions because they are more reflective of real life[24]. On the other hand, films usually have a high degree of ecological validity[25]. Due to the aforementioned advantages of video clips, we use the Stanford film library[23] as a stimuli source and use a rule-based stimuli filtering strategy to select video clips that can reliably elicit positive, negative, and mixed emotions.

Similar to previous datasets such as DEAP[4] and MAHNOB-HCI[5], the proposed dataset also includes multimodal signals. While the participants watched stimuli videos aimed at inducing mixed, positive, and negative emotions, we collected EEG, GSR, PPG, and frontal face videos using three portable devices. Finally, we established the multimodal dataset that contains physiological and face video data of 73 participants. To the best of our knowledge, the proposed dataset is currently the only available dataset for mixed emotion recognition, and it may help advance research in mixed emotion analysis.

Methods

Stimuli selection. In order to elicit emotions more effectively, especially for mixed emotions, we choose the film library of Stanford proposed by Samson *et al.*[23] as a candidate stimuli source (application for access can be submitted at https://spl.stanford.edu/film-clip-library-request-form), and conduct a further selection among these video clips through a rule-based video filtering step and experts evaluation.

To select video clips that can effectively elicit emotions from the Stanford film library, we first implemented a rule-based step using subjective rating scores from the library. The first rule was a language constraint that the video clips should not contain dialogues that may be relevant to the comprehension of the video content to avoid potential side-effects (e.g., for positive emotion, clips involving jokes telling were excluded). Then, we made rules for each target emotion based on emotion intensity to ensure the effectiveness of the video clips for inducing emotion. Specifically, we empirically selected the top 50% videos of Mixed Feelings (MF)[23] as candidate videos with mixed emotions. This collection is calculated with $I(\mathrm{MF}) = minimum\ (I(\mathrm{PA}), I(\mathrm{NA}))$, where $I(\mathrm{PA})$ and $I(\mathrm{NA})$ (both ranges from 5 (not at all) to 25 (extremely strong)) denoted the intensities of positive affect and negative affect respectively. For positive (negative) emotion, we considered video clips whose $I(\mathrm{PA})$ ($I(\mathrm{NA})$ for negative) in the top 60% and MF in the last 60% as candidate videos. By applying the above filtering rules, we finally got 26, 15, and 15 candidate excerpts for mixed, positive, and negative emotions, respectively.

After rule-based selection, we employed four experienced experts, all of whom have at least 5 years of research experience in emotion analysis, to further evaluate the candidate video clips. It should be noted that the four experienced experts will not participate in data collection experiments as subjects. The experts were asked to watch each candidate video and then to report their emotional states induced by the video through a short PANAS[26]. A video clip was selected if at least three of the four experts gave the same evaluation on it. Specifically, an expert rates a candidate positive video as positive if $I(\mathrm{PA}) - I(\mathrm{NA}) > 3$, where $I(\mathrm{PA})$ and $I(\mathrm{NA})$ are the expert's own ratings for the current candidate video. Similarly, an expert rates a candidate negative video as negative if his or her $I(\mathrm{NA}) - I(\mathrm{PA}) > 3$. Different from positive and negative emotions, we use $I(\mathrm{MF}) > 5$ as the criterion for defining mixed emotion video. We sort all candidate videos for mixed emotion after synthesizing the evaluation of four experts in descending $I(\mathrm{MF})$ order and ultimately select the top 16 video clips as the final mixed emotion stimulus. We present the average expert PA and NA scores on all candidate videos in Fig. 1. The results show that the selected positive videos and negative videos are generally located in high-value areas of PA and NA, respectively. For mixed emotion videos (see Fig. 1(c)), compared to non-selected videos, the selected mixed videos are more likely to be in the center of the Figure, which means that the selected videos have closer PA and NA values.

Emotion	Video ID	Video clip	Emotion	Video ID	Video clip
Positive	0	pos_babycontrolscheers.avi	Mixed	16	mix_bungeejumpaccidentmiscalculationl.avi
	1	pos_babydancingtornb.avi		17	mix_boogieboardbackfire.avi
	2	pos_babydancingtotechno.avi		18	mix_breakdanceheadbutt.avi
	3	pos_babyshiccupandlaugh.avi		19	mix_kidonskateboardfalls.avi
	4	pos_beatboxbabydance.avi		20	mix_pentrickelectricity.avi
	5	pos_pandasneezesalot.avi		21	mix_stiltscrashintocar.avi
	6	pos_singingdog.avi		22	mix_karatekickwrongtarget.avi
	7	pos_thirstybabydrink.avi		23	mix_boycrashesintopole.avi
Negative	8	neg_armbentfromskateboard.avi		24	mix_horsegrabsgirl.avi
	9	neg_boybreakswristbiking.avi		25	mix_manhitbynunchuck.avi
	10	neg_brokeankleskating.avi		26	mix_cranedrops.avi
	11	neg_bullhurtsman.avi		27	mix_tripleflipfaceplant.avi
	12	neg_bullwrongtarget.avi		28	mix_guybreaksglasscopyingbutt.avi
	13	neg_crocbitesman.avi		29	mix_guyongymnasticsparallelbars.avi
	14	neg_kidbikesofftruck.avi		30	mix_bikesplitafterjumpofframp.avi
	15	neg_manbreakslegfighting.avi		31	mix_painfulslingshotfail.avi

Table 1. Emotion inducing video clips selected from Stanford film library[23]. We selected 32 video clips, 8 for positive emotion, 8 for negative emotion, and 16 for mixed emotion.

Fig. 2 Photos of devices used for signal data collection. (**a**) Wireless dry electrode EEG device DSI-24. (**b**) Intelligent wristband ES1 used for GSR and PPG data acquisition.

As a result, we selected 32 video clips, including 8 for positive emotions, 8 for negative emotions, and 16 for mixed emotions. We list the names of the selected 32 video clips in Table 1.

Ethics statement. This experiment was reviewed by the Institution Review Board (IRB) of Tsinghua University (Project No. 20220110), including the research plans, recruitment strategy, data management procedure, privacy strategy, protection of participants, and informed consent. Participants were informed about the purpose and procedure of data collection, the rights and welfare of the participants, potential risks, and the protocol for the protection of privacy. Participants were informed that the videos and images of their faces obtained during this study will only be used for academic research and may be published in academic journals or books.

Participants. 80 healthy college students were recruited for this study. All participants took part in the experiment voluntarily. They had normal or corrected-to-normal vision and no history of psychiatric or psychological disorders by self-report. Participants received monetary compensation for their participation. The 80 participants included 48 females (60.0%) and 32 males (40.0%) with ages ranging from 18 to 35 (mean age (M) = 23.06, standard deviation (SD) = 3.37), and all participants were right-handed.

Signal acquisition. During the experiment, we collect Electroencephalogram (EEG), Galvanic Skin Response (GSR), Photoplethysmography (PPG)), and frontal face videos. We subsequently give a brief introduction to each of these four data modalities. *EEG*. Electroencephalogram is widely used for the recording of brain activities through electrodes. It has the advantages of being high temporal resolution, low cost, and non-intrusive, and has been proven effective for emotion recognition[27,28]. We use DSI-24 (WearableSensing Inc., USA, as shown in Fig. 2(a)), a wireless dry electrode EEG acquisition system, to record EEG signals. DSI-24 collects EEG signals from 21 channels at a sampling rate of 300Hz, and the channels corresponding to the international 10-20 system are Fp1, Fp2, Fz, F3, F4, F7, F8, Cz, C3, C4, T3, T4, Pz, P3, P4, T5, T6, O1, O2, A1, A2 (Fig. 3). The EEG signals are recorded using DSI-Streamer 1.08.44 (https://wearablesensing.com/dsi-streamer/), which is a data acquisition software for all DSI systems. During the EEG collection process, the default impedance settings were used, which are 0.1-1M Ω and 1-10M Ω for excellent and acceptable signal quality, respectively, to ensure the quality of the collected signals.

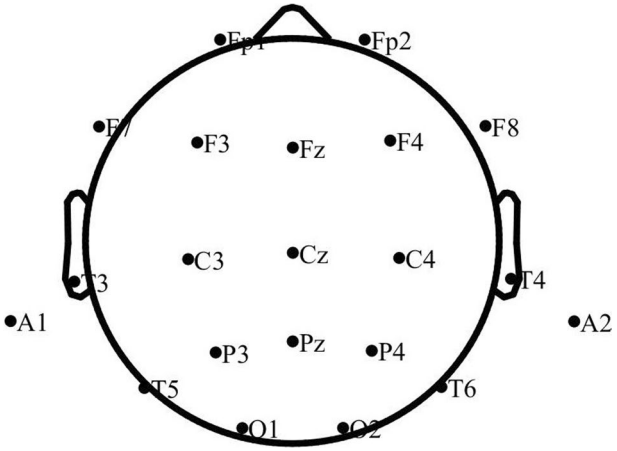

Fig. 3 Schematic diagram of channel locations according to international 10-20 system.

GSR. Galvanic skin response, which measures the electrical conductance of the skin, is a type of peripheral bio-signal from the automatic nervous system (ANS). It has been widely collected in previous emotion recognition databases, such as DEAP[4], MANHOB-HCI[5], and has been reported to be related to different emotion experiences[29,30]. In this study, we collect GSR signals by using a customized designed intelligent wristband (the Ergosensing wristband (ES1), as shown in Fig. 2(b)), which has been used in CPED dataset[31]. The AC excitation source frequency for ES1 is 24Hz, with a detection range of 0.01-100 μS. The sampling rate for GSR is 4Hz.

PPG. Photoplethysmography detects volumetric changes in blood using low-intensity infrared light, it has the advantages of low cost and non-invasive so that it can be collected at the surface of the skin. PPG has also been widely used in emotion recognition[32,33]. In this study, we record PPG at a sampling rate of 100Hz by using the same intelligent wristband ES1 used in the GSR collection, which includes one green LED and three photodiodes. We utilized the same intelligent wristband for both GSR and PPG signal collection throughout the entire experiment.

Video. Facial expression analysis is an important and popular[34–36] way to recognize emotion. In this study, we record the frontal face video of subjects using a built-in camera (on the DELL Latitude 5420 experimental PC) with a resolution of 640 × 480 and a frame rate of 30fps.

The experiment program was written with PsychoPy 2021.2.3[37] under Python 3.8. It communicates with wristband devices using the MQTT protocol (EMQ X 4.3.10 open source version is used in implementation, https://packages.emqx.net/emqx-ce/v4.3.10/emqx-windows-4.3.10.zip) for real-time GSR and PPG signal collection and records the frontal face videos using opencv-python 4.5. We use a PC (DELL Latitude 5420, i5-1135G7, 2.40GHz) to display the stimuli. The size of the screen is 14 inches (1920 × 1080), and each of the stimuli videos is displayed in full-screen mode while preserving the original aspect ratio. The built-in speaker is used, and the volume is adjusted to a comfortable level before the formal experiment. Figure 4 presents example samples of each signal (i.e., EEG, GSR, PPG, and Face video). Note that the EEG sample in Fig. 4 does not include channel data from ear electrodes A1, A2, and reference electrode Pz, and the duration of the EEG, GSR, and PPG samples are 5 seconds, 5 minutes, and 20 seconds, respectively.

Experimental protocol. All participants were given written informed consent upon arrival and were asked to read and voluntarily sign it. After that, they were informed about the experiment content, experiment protocol, meaning of affective scales, and instructions for completing the self-assessment form. After placing and checking the sensors, the experimenter ran the main program and a form was first presented on the screen to collect the name, age, gender, and other basic information of the participants. Then, the experiment started by the participant's pressing the 'OK' button.

The experiment procedure mainly consisted of a practice stage, a baseline recording stage, and 4 blocks. Figure 5 shows the timing diagram of the experiment, which began with a practice stage containing a single practice trial to make participants familiar with the procedure of one trial. After practice, the participant was asked to look at the black screen and remain relaxed to collect a three-minute recording of the resting state. Next, 32 film clips were presented in 4 blocks, each containing 8 trials with one video clip in each trial. Note that the film clips were divided into 4 blocks according to their original emotion labels (i.e., *positive, negative, mixed*) to make the labels of film clips the same in each block. A set of arithmetic operations and a 1-minute break were arranged between two consecutive blocks to eliminate the effect of the previous block. The presentation order of the blocks followed the Latin square design to eliminate any possible influence that block presentation order might have. Each trial consisted of the following concrete steps:

1. The display of one video clip for about 20-30 seconds.
2. Self-report for the emotional adjectives (10-item short positive affect (PA) and negative affect (NA) schedules (PANAS)[26]).

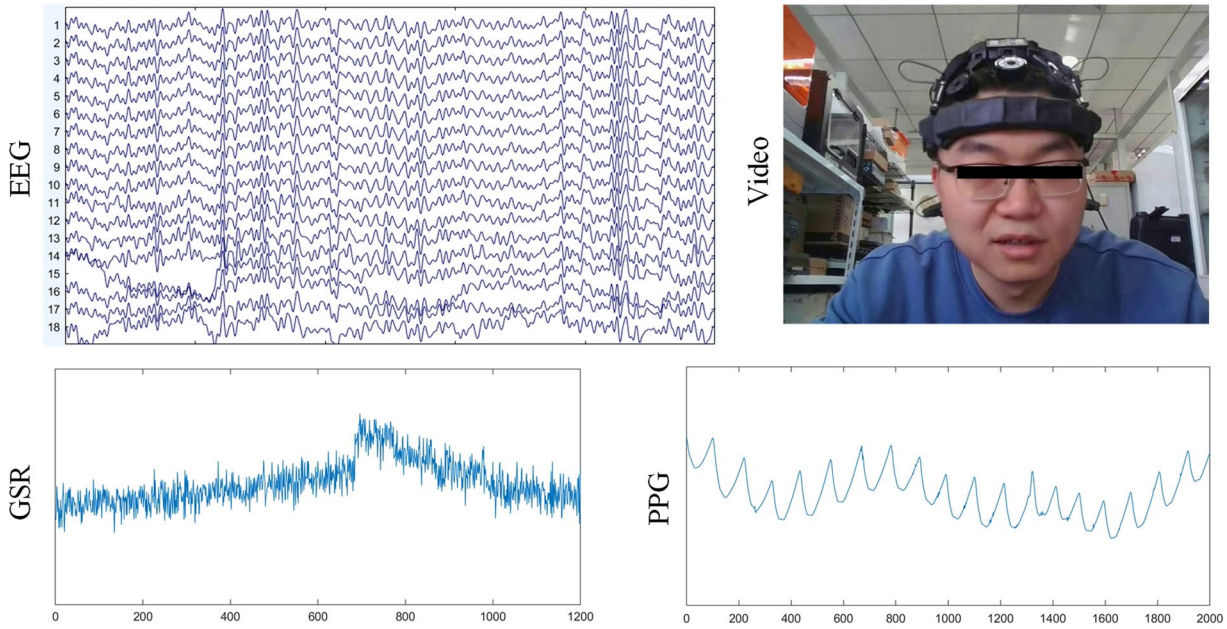

Fig. 4 Example samples of EEG, GSR, PPG, and Face video.

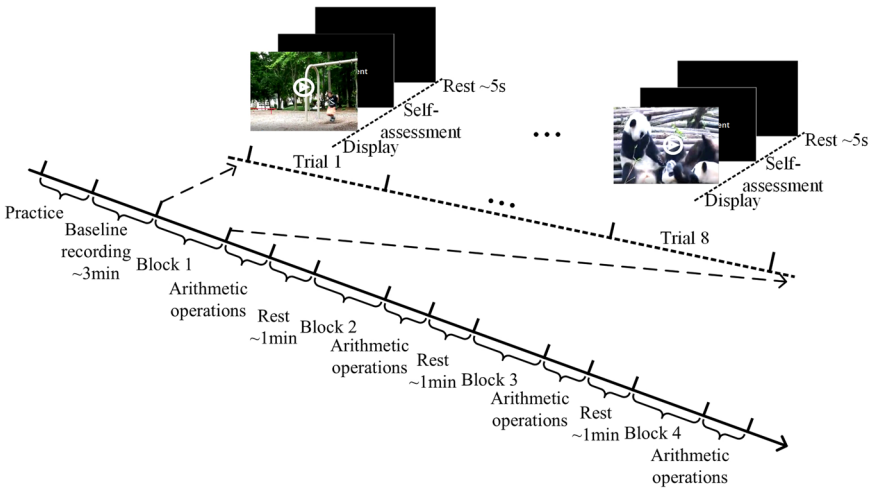

Fig. 5 Timing diagram of the experiment. The experiment procedure mainly consisted of a practice stage, a baseline recording stage, and 4 blocks (each containing 8 trials). A set of arithmetic operations and a 1-minute break were arranged to eliminate the effect of the previous block on the current one.

3. Self-report for arousal, valence, and dominance.
4. Self-report for two discrete emotions, namely amusement and disgust. We collected the self-rating scores of these two emotions since the positive and negative emotions in Stanford film library[23] mainly refer to amusement and disgust.
5. A 5-second break before the next trial.

Data Records

The present multimodal dataset is available on *Zenodo* (https://doi.org/10.5281/zenodo.11194571) upon request[38]. Interested researchers are invited to submit an access request via *Zenodo* to download the dataset. The proposed dataset includes two versions of data, raw data and aligned data (i.e., data from different modalities were aligned in time), both were gathered from 80 participants. Seven participants' data were discarded due to the deficiency in physiological signal data. Next, we introduce the structure of the proposed dataset in detail.

Raw data. The compressed file *Raw_data.zip* contains the available raw data. It includes 73 sub-folders named with numbers (i.e., the participant ID), each of which corresponds to a participant. For each participant, there are two sub-folders named with participant ID *'n'* and *'subn'*. Sub-folder *'n'* contains files of self-assessment, raw

Values of the first two items	Comment
$(-1, 0)$	-1 is a virtual video ID, and it means the start of recording when the value of the second item equals 0.
$(k, k{+}10)$, $k = 0, 1, \cdots, 31$	k denotes the k-th video, and $(k, k{+}10)$ for the first two items means start playing the k-th video clip.
$(k, k{+}100)$, $k = 0, 1, \cdots, 31$	k denotes the k-th video, and $(k, k{+}100)$ for the first two items means stop playing the k-th video clip.
(40, 50)	Begin of the resting stage. The virtual video ID 40 corresponds to the rest stage.
(40, 140)	End of the resting stage.
(50, 60)	Begin of the practice stage. The virtual video ID 50 corresponds to the practice stage.
(50, 150)	End of the practice stage.

Table 2. The values of the first items and the corresponding meaning.

signal data of GSR, PPG, and frontal face video, etc., and files of EEG recordings are in *'subn'*. The files in folder *'n'* are as follows:

1. *1_XXX.mp4* (*'XXX'* denotes a suffix): This file contains the frontal face video recording that records the complete facial videos of a participant throughout the experiment.
2. *camera.csv*: This file contains trigger information during the experiment. Each row in this file contains three items representing a trigger message. The first two items form a trigger, and the third item represents the corresponding timestamp. The values of the first two items and the corresponding meaning are presented in Table 2.
3. *raw_gsr.csv*: This CSV file contains the raw GSR data. For each line, the first item (if not null) represents the timestamp. The second item denotes the GSR data. The third item stores trigger information. Similar to *camera.csv*, we use $k + 10$ and $k + 100$ to indicate the start and stop of the k-th video, respectively.
4. *raw_ppg.csv*: This CSV file contains the raw PPG data. The organization structure of PPG data is the same as GSR data in *raw_gsr.csv*. The only difference between *raw_ppg.csv* and *raw_gsr.csv* is the number of lines per second (100 lines per second for PPG and 4 lines for GSR due to different sampling rates).
5. *Emotions.csv*: This file contains the rating scores for positive emotion *amusement* and negative emotion *disgust*. There are a total of 32 lines corresponding to 32 stimuli video excerpts. Each line contains a video ID defined in Table 1, two rating scores ranging from 1 to 5 for *amusement* and *disgust*, respectively.
6. *Arousal_Valence.csv*: This file includes 32 lines and each line contains a video ID and three evaluation scores, ranging from 1 to 9, for valence, arousal, and dominance, respectively.
7. *Panas.csv*-This file contains the PANAS evaluation scores. It also includes 32 lines and each line contains a video ID and ten rating scores ranging from 1 to 9 for the ten affective terms of short PANAS[26].

There are three EEG recording files in folder *'subn'*, all of which contain raw EEG data collected by DSI-Streamer in different file formats. The three files contain the same content with the different file formats, and more details are listed below:

1. *1_raw.csv*: As presented in Fig. 6, data in the first fourteen rows are basic information such as sampling frequency, device name, etc. The fifteen and sixteen lines are descriptions of each column of recorded data, including the channel index, the corresponding electrode, etc. From line 17 to the end of the file are the collected data, including the point in time when data is being recorded, 21-channel EEG data, triggers, etc.
2. *1_.dsi*: A DSI format file, which can be viewed and processed by using DSI-Streamer (https://wearablesensing.com/dsi-streamer/).
3. *1_raw.edf*: This file can be processed using EEGLab tool kit (https://sccn.ucsd.edu/eeglab/index.php).

Aligned data. In addition to raw data, we also provide aligned data in *Aligned_data.zip*. It contains time-aligned physiological signals and video clips corresponding to each trial for all participants. The *Aligned_data.zip* file contains 73 sub-folders, each named after the ID of its corresponding participant. There are a total of 33 files in a sub-folder, and the included files are as follows:

1. *k.mp4*, $(k = 0, \cdots, 31)$-The trimmed frontal face video clip corresponds to a trial with the k-th video excerpt. The length of the face video is aligned with other signals, namely EEG, GSR, and PPG.
2. *datas.mat*-This file contains the aligned physiological signals (i.e., EEG, GSR, and PPG) for all 32 trials, each corresponding to a stimuli video clip. The signal data for EEG, GSR, and PPG and their corresponding sampling rates are saved in variables *eeg_datas*, *gsr_datas*, *ppg_datas*, *fs_eeg*, *fs_gsr*, and *fs_ppg*, respectively. The second to last row of data in all three signal variables (i.e., *eeg_datas*, *gsr_datas* and *ppg_datas*) stores the IDs of stimuli video clips corresponding to the collected signals, and the values in the last row are time information. What needs to be declared is that we use time intervals relative to the start of each trial instead of timestamps, as we aligned all signal data based on event triggers. Specifically, non-negative integers in the last row represent the time in seconds from the start of a trial, e.g., 0 indicates the beginning of a new trial. We only assign time information for the first sampled data point of each second, and fill the time information for other data points with -1. Take *eeg_datas* as an example, it has a size of $20 \times N$, where the first 18 rows correspond to the selected 18 EEG channels. The 19th row stores the video ID that the EEG sample corresponds to, and the time information is saved in the last row.

1	# Mains_F	60									
2	# Sample_	300									
3	# Filter_D	43.3									
4	# Sensor_	uV									
5	# Headse	DSI-24 SN:2680									
6	# Data_Lc	DSI-Streamer-v.1.08.44									
7	# Date =	######									
8	# Time =	15:21:00									
9	# Patient_	1									
10	# Record_	Startdate 25-May-2022 EEG_XX X X									
11	# Filter =	Non-Filtered									
12	# Comments =										
13	# Referen	Pz									
14	# Trigger	Wireless									
15	# Channe	ch_1	ch_2	ch_3	ch_4	ch_5	ch_6	ch_7	ch_8	ch_9	...
16	Time	P3	C3	F3	Fz	F4	C4	P4	Cz	CM	...
17	0.0033	256.5	-1133.1	778.2	1015.2	-8.4	-249	-782.4	924.6	4594.2	
18	0.0067	253.5	-1132.2	777.9	1016.7	-10.8	-249.9	-777.6	926.4	4742.4	
19	0.01	251.4	-1133.1	779.7	1018.2	-7.5	-253.2	-780.9	924.9	4816.5	...
20	0.0133	248.4	-1136.4	777.3	1017.6	-7.8	-250.5	-781.2	917.1	4668.3	

Fig. 6 Screenshot of EEG recording file *1_raw.csv*.

It should be noted that signal data in *datas.mat* is preprocessed. Specifically, we performed independent component analysis (ICA) and filtering on the EEG signal using a Matlab toolbox EEGLab. We first excluded channels A1 and A2 and re-referenced the selected channels to channel Pz. After that, we filtered the signals using two filters, i.e. a band-pass filter from 1Hz to 50Hz and a 50Hz notch filter to remove various distractions such as power frequency interference. Then, we performed a baseline removal step to eliminate the baseline drift. We conducted ICA to further remove possible artifacts introduced by eye movement, muscle movements, etc. More specifically, we used the builtin function *runica*() of EEGLab to perform ICA and *pop_icflag*() for automatic component filtering, and the threshold confidence values for muscle and eye categories used in *pop_icflag*() were both set to (0.9, 1), which means only with confidence beyond 90% did we consider the component as an eye or muscle category. Finally, we reconstructed the EEG signals using the selected components and obtained the 18-channel signal for further feature extraction. For GSR signals, we used a Butterworth filter to filter out the high-frequency noise content from the continuously recorded raw GSR signal following a previous study[39]. More specifically, a 1.0 Hz 3rd-order low pass Butterworth filter was utilized to preprocess the original GSR signals. For PPG signals, we also chose a Butterworth filter, a 3d-order bandpass Butterworth with 0.6 Hz as the lower cutoff frequency and 5.0 Hz as the higher cutoff frequency to eliminate noise while retaining the original signal as much as possible according to Zhang's work[31]. Please note that the preprocessing steps of both GSR and PPG were conducted on the original signals corresponding to each trial.

Feature files. The file *Features.zip* contains the features used in section *Emotion Classification*. There are a total of 73 *mat* files, and the file names correspond to the IDs of the participants. Each *mat* file contains two variables *feas* and *vids* that store feature vectors extracted from signals and corresponding IDs of emotion induction videos. Specifically, each line of *feas* is a 913-dimension feature vector, including 90 EEG features, 28 GSR features, 27 PPG features, and 768 video features. The detailed introduction of these features is presented in Section *Feature Extraction*.

Technical Validation

Analysis of Self-Assessment Ratings. *Self-Assessment for PANAS score.* In this section, we employed the analytical methodology from the "Difference Within Film Clips" part of Saganowski *et al.*'s work[40] to examine the validity of PANAS questionnaire. Specifically, for each video type condition, we separately calculated the sum of the scores for the 5 positive items(hereinafter referred to as positive score) and the 5 negative items (hereinafter referred to as negative score) in the PANAS questionnaire. We first tested the normality of the data. The results of the Shapiro-Wilk test indicated that the data did not follow a normal distribution under most conditions (**p**s $<$ 0.01). Since we were comparing the differences between positive and negative scores, we did not use the reviewer-recommended repeated measures ANOVA. Consequently, to appropriately compare the differences between the positive and negative scores under each condition, we used the Wilcoxon Signed-Rank Test, a non-parametric method alternative to the paired sample t-test. We also calculated the rank biserial correlation to quantify the effect size. For positive videos, the results showed that the positive scores were significantly higher than the negative scores ($W = 3133$, $p < 0.001$), the rank biserial correlation was 0.983. For negative videos, the negative scores were significantly higher than the positive scores ($W = 32.5$, $p < 0.001$), and the rank biserial correlation was -0.979. For mixed video type condition, considering that the stimuli inherently contain both positive and negative emotions, we also calculated the total PANAS score to better measure the questionnaire's validity in assessing mixed emotions. We found that the total PANAS score showed significant differences compared to both the positive($W = 2249$, $p < 0.001$) and negative scores($W = 755$, $p < 0.001$). The rank biserial correlation values were 0.498 and -0.498, respectively. These results indicate that the PANAS questionnaire has a certain degree of discriminative power across different video types, demonstrating the validity of the PANAS questionnaire.

	Positive videos		Negative videos		Mixed videos		χ^2	df	ϵ^2
	M	SD	M	SD	M	SD			
Positive score	8.80	2.60	7.08	1.61	7.21	1.59	25.2***	2	0.106
Negative score	5.22	0.55	11.92	4.35	8.36	2.61	158.4***	2	0.663

Table 3. Results of One-way analysis of variance (Kruskal-Wallis) for differences between conditions in various self-reports. M=Mean, SD=Standard Deviation, *** indicates p<0.001.

		PS		NS	
		W	p	W	p
Positive videos	Negative videos	−6.40	<0.001	15.34	<0.001
Positive videos	Mixed videos	−5.86	<0.001	14.14	<0.001
Negative videos	Mixed videos	0.49	0.937	−7.92	<0.001

Table 4. Pairwise comparison results for differences between conditions in various self-reports. PS and NS represent positive scores and negative scores.

We further validate the self-reports scores across video type conditions. Given the non-normal distribution of the data (**p**s < 0.01), the nonparametric Kruskal-Wallis test was utilized instead of the one-way ANOVA. Pairwise comparisons were calculated using Dwass-Steel-Critchlow-Fligner method. The results were listed in Tables 3 and 4. For the positive score, the results showed that the scores for the positive video were significantly higher than those for the negative and mixed video ($\chi^2 = 25.2$, $p < 0.001$). For the negative score, the scores for the negative video were significantly higher than those for the positive and mixed video ($\chi^2 = 158.4$, $p < 0.001$).

Self-Assessment for Amusement and Disgust. We first present the statistical results for positive emotion *amusement* and negative emotion *disgust* of all participants for each stimuli video using box plots in Fig. 7. As defined in Table 1, videos with ID in [0,7], [8,15], and [16,31] are positive, negative, and mixed videos, respectively. The average rating scores in Fig. 7 indicate that, for positive videos, the average rating score of *amusement* is much higher than the average score of *disgust*. For negative videos, the results are just the opposite. In contrast, the average scores of *amusement* and *disgust* are very close for mixed videos (see Fig. 7). From another perspective, the average rating scores of *amusement* of positive videos are much higher than those of negative and mixed ones. In addition, the average rating scores of *disgust* of negative videos are greater than those of positive and mixed videos. All the above results indicate that the selected stimuli video excerpts have reliable emotion-inducing abilities and can successfully induce the target emotions as expected.

Self-Assessment for Valence, Arousal, and Dominance. We next analyze the ratings of valence, arousal, and dominance of the stimuli. We use self-assessment data from all 80 participants due to their availability. For each video clip, we compute the average rating score of all participants for valence, arousal, and dominance. We denote the valence, arousal, and dominance scores assessed by the i-th participant for the j-th video clip as v_{ij}, a_{ij}, d_{ij} ($i = 1, 2, \cdots, 80$ and $j = 1, 2, \cdots, 32$) respectively. The mean valence, arousal, and dominance scores of the j-th video can be obtained as $\bar{v}_j = \frac{\sum_{i=1}^{N} v_{ij}}{N}$, $\bar{a}_j = \frac{\sum_{i=1}^{N} a_{ij}}{N}$ and $\bar{d}_j = \frac{\sum_{i=1}^{N} d_{ij}}{N}$, where $N = 80$ denotes the number of participants. We then visualize the mean valence, arousal, and dominance scores of each video clip in the valence-arousal-dominance (VAD) space and the corresponding projection in the valence-arousal (VA) plane as shown in Fig. 8.

We can see from Fig. 8(a) that video clips corresponding to different emotions (i.e., positive, negative, and mixed emotions) are located in different regions of VAD space, and these regions can be relatively clearly separated from each other. As shown in Fig. 8(b), the valence scores of positive, negative, and mixed emotion videos are roughly satisfying > 5.5, < 3.5, and in [3.5,5.5] respectively. Although one of the mixed emotion videos is located in the region corresponding to negative emotion, the boundary between negative and mixed emotional points is relatively clear. In addition, Fig. 8(b) also indicates that negative videos generally have a high level of arousal compared to positive and mixed-emotion videos.

Physiological Signal Quality Validation. In this section, we will verify the quality of collected physiological signals. Specifically, we further conduct signal-to-noise ratio (SNR) analysis for EEG, GSR, and PPG signals and heart rate (HR) estimation for PPG.

SNR. To validate the signal quality of EEG, GSR, and PPG, we first performed SNR analysis on them. Specifically, we calculated the SNR of the signal corresponding to each trial for each subject. For EEG signals, 60 Hz is reported as a frequency boundary for measuring brain activities[41–43]. In implementation, we use 60 Hz as the desired upper frequency for signal separation and 60 Hz-100Hz for noise. As the cardiac-related components have negligible frequency components above 15Hz[44], we separated noise above 15Hz and signal below 15Hz from PPG data using 3rd order Butterworth filter. For GSR, we also separated noise and signal using 3rd order Butterworth filter with a cutoff frequency of 1Hz[45–47]. Then the SNR is calculated as $\text{SNR} = 10\log_{10}\left(P_{signal}/P_{noise}\right)$.

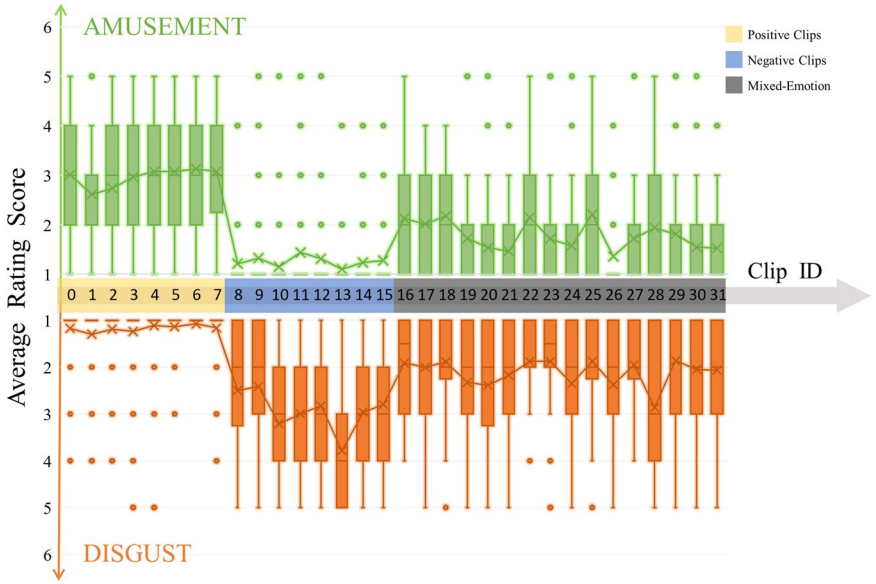

Fig. 7 Distribution of rating scores for amusement and disgust of all participants for each stimuli video clip. The positive axis represents amusement scores while the negative axis represents disgust scores, both of which are measured through 5-point scales. The tips of the top whisker and bottom whisker mark the maximum and minimum rating scores. Boxes represent quartiles of the rating score distributions. Crosses mark the mean values, and the circle dots are outliers.

For EEG signals, the Welch algorithm[48] is utilized to compute the power spectral density, thereby calculating P_{signal} and P_{noise}. For GSR and PPG signals, SNR $= 10\log_{10}(\mathrm{MS}(signal)/\mathrm{MS}(noise))$, where P_{signal} and P_{noise} are the power *signal* and the *noise* respectively, MS() denotes mean square amplitude. For each subject, we calculate the SNR for each channel of the EEG signal corresponding to each trial, resulting in a total of $Channels \times Trials \times Subjects = 18 \times 32 \times 73 = 42048$ SNR values. The experimental results for EEG signals showed that the average SNR of all trials for each electrode (i.e., a single channel) of each subject ranges from 2.98 dB to 30.81 dB, with standard deviation from 0.18dB to 15.84 dB. The total proportions of SNR values below 5dB and 10 dB are 543/42048=1.29% and 5404/42048=12.85%, respectively. For results of PPG and GSR, the average SNR of the PPG signal for each subject is greater than 60dB, while the average SNR of the GSR signal is greater than 50dB, in line with the results reported by Saganowski *et al.*[40], in which an average SNR around 30dB is reported (26.66dB-37.74dB).

HR. We further performed heart rate estimation using HeartPy (https://python-heart-rate-analysis-toolkit.readthedocs.io), which is a Toolkit that supports PPG based heart rate analysis. The HR estimation is conducted on the preprocessed PPG signals as introduced in Section *Aligned data*. The box plot in Fig. 9 presents the statistical results of heart rate estimation for each subject. The results demonstrated that the heart rate of most subjects ranged from 60 to 100 per minute, and there is no abnormal HR below 40 or above 120 according to the observation in[49].

Mixed Emotion Analysis From Physiological Signals and Face Videos. Our dataset targets the analysis of mixed emotion, which is supposed to vary from pure positive or negative emotion, so we conduct a classification task for these three emotion kinds (i.e., positive, negative, and mixed) to prove that the signals collected and processed have discrimination validity among these three classes. We use the aligned data for the classification task in implementation.

Signal Preprocessing. Signal preprocessing is crucial and can remove artifacts that may be introduced during signal acquisition from original physiological signals. In this study, we conducted signal preprocessing during the aligned data generation stage, including ICA for EEG, Butterworth filtering for GSR and PPG, etc. We did not perform any further preprocessing for each signal before the feature extraction step. For more detailed preprocessing information, please refer to Section *Aligned data*.

Feature Extraction. We extracted typical features from preprocessed physiological signals and face videos following previous works[50–52]. More specifically, for EEG signals, we extracted differential entropy (DE) features in each EEG channel from 5 different frequency bands, which are δ band (1-3Hz), θ band (4-7Hz), α band (8-13Hz), β band (14-30Hz) and γ band (31-50Hz).

For GSR signals, we extracted statistical features from both the time domain and frequency domain. Specifically, the median, mean, standard deviation, minimum, maximum, ratio of minimum, and the ratio of maximum of the raw GSR signal and its first-order and second-order derivatives were extracted as time domain

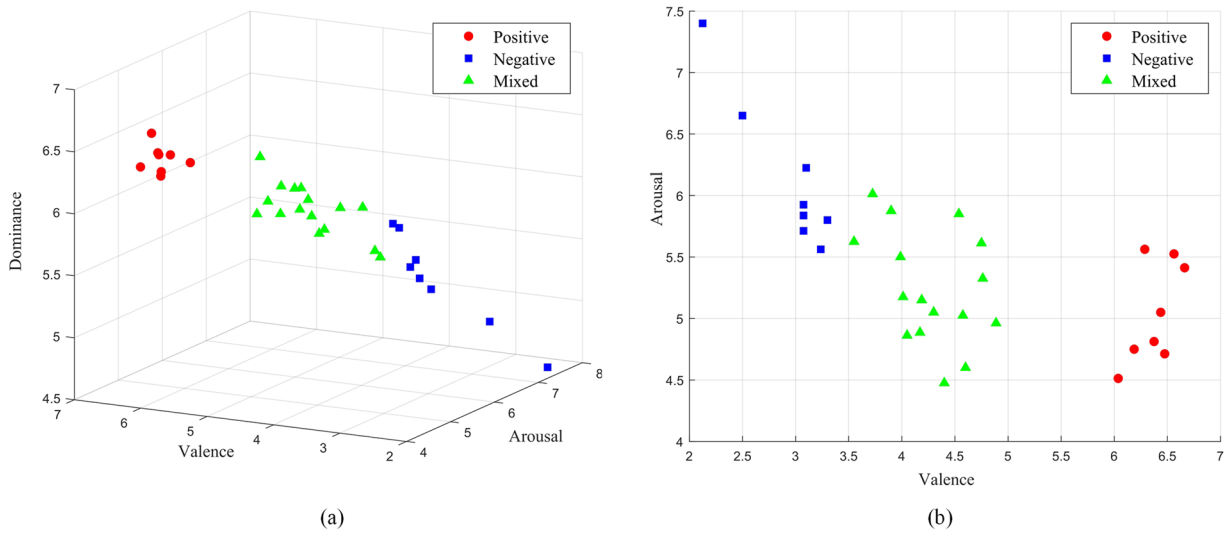

Fig. 8 Scatter plot of the average valence, arousal, and dominance rating scores of each video clip. (**a**) Scatter plot in valence-arousal-dominance space, (**b**) Projection of (a) onto the valence-arousal plane.

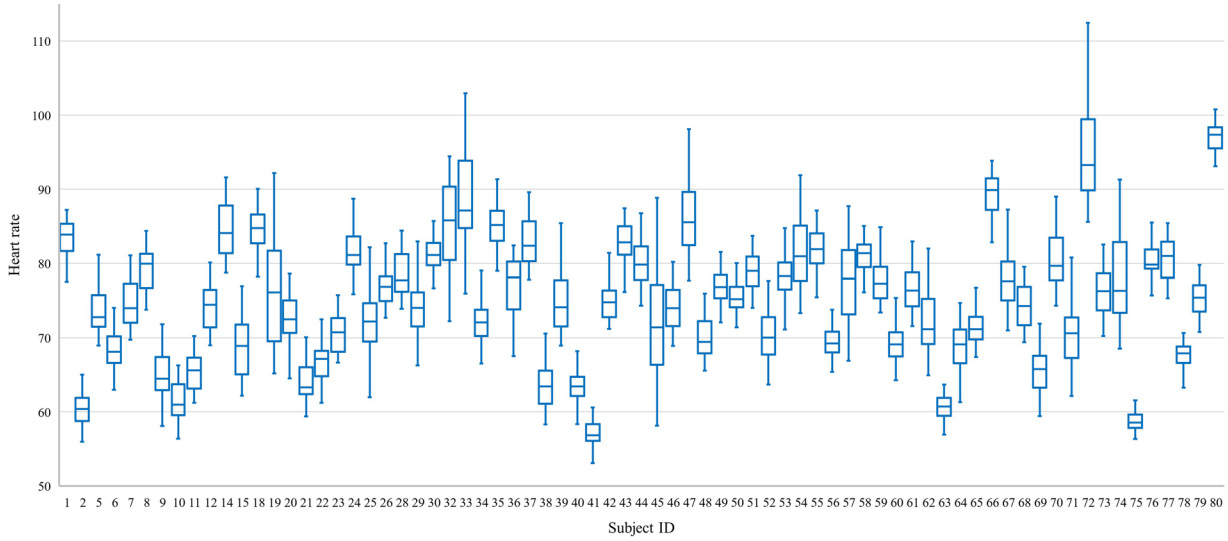

Fig. 9 Distribution of the estimated heart rate for each subject.

statistical features according to the feature extraction in the work of Udovičić *et al.*[51]. To extract frequency domain features, we first transformed the GSR signals from the time domain into the frequency domain using the Fast Fourier Transform (FFT), and then we extracted the median, mean, standard deviation, maximum, minimum, and range of the signal according to Udovičić *et al.*[51]. In addition, we extracted PSD of frequency band [0, 2] Hz using Welch's power spectral density. Consequently, we totally extracted 28 features (21 time domain features and 7 frequency domain features) from the GSR signals.

For the PPG signals, we also extracted time domain and frequency domain features. We extracted the same statistical time domain features as GSR signals, which means that 21 features were extracted from the preprocessed PPG signal. For the frequency domain, the same features except PSD were extracted. Finally, 27 time and frequency domain features were extracted from PPG signals, and these PPG features were also used in Udovičić's study[51]. It should be noted that the size of the sliding window for both GSR and PPG feature extraction was 5 seconds with an 80% overlap between two consecutive windows according to the work of Zhang *et al.*[31].

Local binary patterns from three orthogonal planes (LBP-TOP) is a widely used feature in facial expression analysis from video[52,53]. It extends the conventional LBP, which is designed for describing 2D textures of static images, to spatial-temporal space for dynamic texture description. Finally, we extracted a 768-dimension feature vector from each one-second face video.

To reduce the individual differences in physiological and video signals of each participant, we normalize the extracted features using Min-Max scaling as done in Udovičić's work[51]. More specifically, suppose

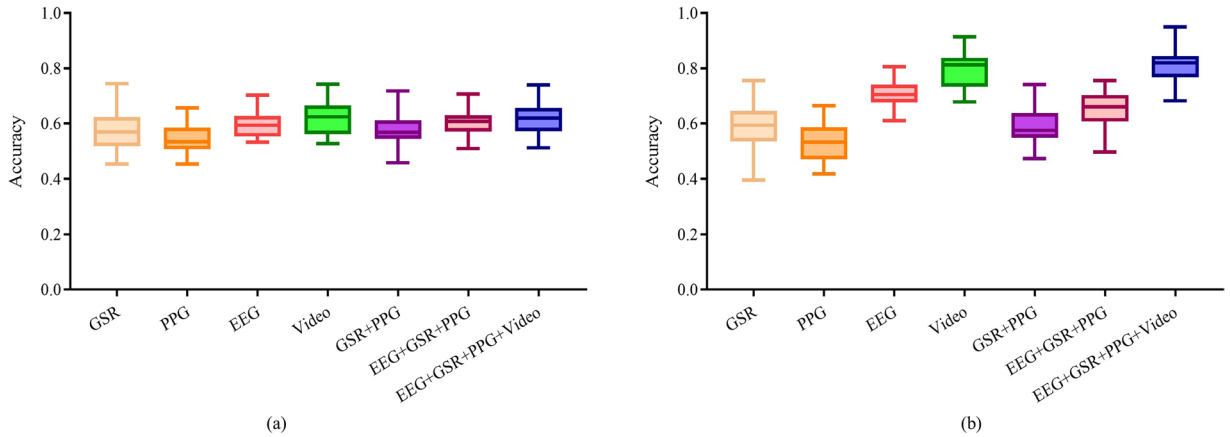

Fig. 10 Boxplot graph for the 3-class classification (i.e., positive, negative, and mixed) results of different combinations of features. (**a**) Results of using RF as the classifier. (**b**) Results of using SVM as the classifier.

$X_i = (X_{i1}, \cdots, X_{im})$ be a m-dimensional feature instance, then we can get the corresponding normalized feature $X_i^{'} = (X_{i1}^{'}, \cdots, X_{im}^{'})$ as follows:

$$X_{ij}^{'} = \frac{X_{ij} - X_{jmin}}{X_{jmax} - X_{jmin}} \tag{1}$$

where X_{ij} denotes the jth dimension of the ith feature instance, X_{jmin} and X_{jmax} are the minimum value and maximum value of the jth dimension of features from the same trial respectively.

Emotion Classification. To verify the feasibility of mixed emotion classification from physiological signals and face videos, we conducted experiments using two typical classifiers (i.e., support vector machine (SVM) and random forest (RF)) for positive, negative, and mixed emotion classification. We validated the classification performance in a participant-dependent protocol. The physiological signals and face video of each trial were divided into two parts according to 4:1, and the first and the second parts of all trials formed the original data of the train set and test set, respectively. We then extracted features introduced in 0.11.2 to form the training set and test set. Note that the emotion label of the corresponding stimuli clip was used as labels for training and test samples. More specifically, we adopted the 'RBF' kernel function for SVM and optimized parameter C using a grid-search strategy from 10^{-10} to 10^{10} in log-space. For the RF classifier, the max depth was set to 20 and the number of estimators was from 50 to 1000 with a step of 50.

The experimental results are presented in Fig. 10. We tested seven feature combinations, including four single modality features (i.e., EEG, GSR, PPG, Video) and three multiple modalities features (i.e., GSR+PPG, GSR+PPG+EEG, GSR+PPG+EEG+ Video). Results in Fig. 10 show that SVM and all features (i.e., EEG+GSR+PPG+Video) obtained the best accuracy. Besides, EEG performs better than other physiological signals: it achieves not only higher classification accuracy but also a smaller standard deviation. We also presented the confusion matrices of classification results obtained by using SVM and all features in Fig. 11 (Due to space limitations, we only presented the results of the first 28 subjects in the main text. The confusion matrices for all subjects can be found at https://github.com/ypthu/Multimodal-dataset-for-mixed-emotion-recognition/blob/main/Confusion-matrices-for-all-subjects.png). The results in Fig. 11 also indicate that the classification results are quite different among different participants. For example, it is relatively difficult to correctly identify positive emotions for participants 2, 11, and 21. In contrast, it is difficult to identify negative emotions for participant 22. Although there are individual differences, the overall results indicate that it is possible to identify mixed emotional states using physiological signals or facial video signals.

Usage Notes

Due to copyright issues, we did not include the stimuli video clips in the repository of data collection. However, the name list of the used video clips and the corresponding access application channel were provided in *Readme.md* under the *videos* folder of this repository (https://github.com/ypthu/Multimodal-dataset-for-mixed-emotion-recognition-Data-collection). Interested researchers can contact us if they encounter any problems in acquiring these stimuli videos.

Interested researchers can replicate the mixed emotion classification experiment either using the raw data or aligned data included in the Zenodo repository (https://doi.org/10.5281/zenodo.11194571)[38]. Face video files in this repository are in *mp4* format, and users can use any video player software to open them. GSR, PPG, and self-assessment data are saved in *csv* files, which can be opened by any spreadsheet or workbook software or imported into programming tools (e.g., python) for further analysis. For EEG recording data, we provide two formats in addition to *csv*, namely *dsi* and *edf*, which can be opened using DSI-streamer software (https://wearablesensing.com/dsi-streamer/) and EEGLab tool kit (https://sccn.ucsd.edu/eeglab/index.php)

Fig. 11 Confusion matrix of classification results obtained by using SVM and all features (i.e., features from GSR, PPG, EEG, and Video). Labels 0, 1, and 2 correspond to positive, negative, and mixed emotions. The number under each confusion matrix denotes the participant ID.

respectively. Furthermore, we provide aligned physiological data in *mat* files for each participant, and these data can be directly imported into Matlab, Python, and other programming tools for further processing.

Accessing data. To access the dataset, applicants need to sign a *Data Use Agreement (DUA)*, which can be obtained at *DUA.pdf*. The signed DUA need to be emailed to liuyongjin@tsinghua.edu.cn. The applicants are required to provide basic information, including their name, affiliation, and a detailed explanation of the purpose for which the dataset is being requested in the application email. It should be noted that the dataset can be

only used for academic research, the user may not use the dataset for any commercial purposes (including using screenshots from the dataset in advertisements; selling data from the dataset, etc.). The dataset repository on Zenodo can be found at https://doi.org/10.5281/zenodo.11194571.

Code availability

The data recording script is written with PsychoPy 2021.2.3 under Python 3.8. The source code and related resource files (e.g., pictures) have been uploaded, and a *Readme* is also provided for more details. The repository is available at: https://github.com/ypthu/Multimodal-dataset-for-mixed-emotion-recognition-Data-collection.

Codes for mixed emotion classification introduced in technical validation are implemented using Matlab and Python and are available at https://github.com/ypthu/Multimodal-dataset-for-mixed-emotion-recognition. The source files in this repository are mainly for raw data formatting, preprocessing, feature extraction, and emotion classification. More details for these source files can be found in *README.md* in this repository.

Received: 12 June 2023; Accepted: 23 July 2024;
Published online: 05 August 2024

References

1. Salovey, P., Mayer, J. & Caruso, D. Emotional intelligence: Theory, findings, and implications. *Psychological inquiry* **15**, 197–215 (2004).
2. Seyitoğlu, F. & Ivanov, S. Robots and emotional intelligence: A thematic analysis. *Technology in Society* **77**, 102512 (2024).
3. Picard, R. W., Vyzas, E. & Healey, J. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence* **23**, 1175–1191 (2001).
4. Koelstra, S. *et al*. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* **3**, 18–31 (2011).
5. Soleymani, M., Lichtenauer, J., Pun, T. & Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing* **3**, 42–55 (2011).
6. Park, C. Y. *et al*. K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data* **7**, 293 (2020).
7. Bota, P., Brito, J., Fred, A., Cesar, P. & Silva, H. A real-world dataset of group emotion experiences based on physiological data. *Scientific Data* **11**, 1–17 (2024).
8. Abadi, M. K. *et al*. Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing* **6**, 209–222 (2015).
9. Miranda-Correa, J. A., Abadi, M. K., Sebe, N. & Patras, I. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing* **12**, 479–493 (2018).
10. Larsen, J. T. & McGraw, A. P. Further evidence for mixed emotions. *Journal of personality and social psychology* **100**, 1095 (2011).
11. Russell, J. A. & Barrett, L. F. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology* **76**, 805 (1999).
12. Cacioppo, J. T. & Berntson, G. G. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological bulletin* **115**, 401 (1994).
13. Cohen, A. S., St-Hilaire, A., Aakre, J. M. & Docherty, N. M. Understanding anhedonia in schizophrenia through lexical analysis of natural speech. *Cognition and emotion* **23**, 569–586 (2009).
14. Zhou, K., Sisman, B., Rana, R., Schuller, B. W. & Li, H. Speech synthesis with mixed emotions. *IEEE Transactions on Affective Computing* (2022).
15. Oh, V. Y. & Tong, E. M. Specificity in the study of mixed emotions: A theoretical framework. *Personality and Social Psychology Review* **26**, 283–314 (2022).
16. Lange, J. & Zickfeld, J. H. Comparing implications of distinct emotion, network, and dimensional approaches for co-occurring emotions. *Emotion* (2023).
17. Williams, P. & Aaker, J. L. Can mixed emotions peacefully coexist? *Journal of consumer research* **28**, 636–649 (2002).
18. Larsen, J. T., McGraw, A. P. & Cacioppo, J. T. Can people feel happy and sad at the same time? *Journal of personality and social psychology* **81**, 684 (2001).
19. Aaker, J., Drolet, A. & Griffin, D. Recalling mixed emotions. *Journal of Consumer Research* **35**, 268–278 (2008).
20. Weth, K., Raab, M. H. & Carbon, C.-C. Investigating emotional responses to self-selected sad music via self-report and automated facial analysis. *Musicae Scientiae* **19**, 412–432 (2015).
21. Carrera, P. & Oceja, L. Drawing mixed emotions: Sequential or simultaneous experiences? *Cognition and emotion* **21**, 422–441 (2007).
22. Cohen, A. S., Callaway, D. A., Mitchell, K. R., Larsen, J. T. & Strauss, G. P. A temporal examination of co-activated emotion valence networks in schizophrenia and schizotypy. *Schizophrenia research* **170**, 322–329 (2016).
23. Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A. & Gross, J. J. Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition and emotion* **30**, 827–856 (2016).
24. Uhrig, M. K. *et al*. Emotion elicitation: A comparison of pictures and films. *Frontiers in psychology* **7**, 180 (2016).
25. Gross, J. J. & Levenson, R. W. Emotion elicitation using films. *Cognition & emotion* **9**, 87–108 (1995).
26. Mackinnon, A. *et al*. A short form of the positive and negative affect schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual differences* **27**, 405–416 (1999).
27. Petrantonakis, P. C. & Hadjileontiadis, L. J. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on affective computing* **1**, 81–97 (2010).
28. Alarcao, S. M. & Fonseca, M. J. Emotions recognition using eeg signals: A survey. *IEEE Transactions on Affective Computing* **10**, 374–393 (2017).
29. Nourbakhsh, N., Wang, Y., Chen, F. & Calvo, R. A. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th australian computer-human interaction conference*, 420–423 (2012).
30. Liu, M., Fan, D., Zhang, X. & Gong, X. Human emotion recognition based on galvanic skin response signal feature selection and svm. In *2016 international conference on smart city and systems engineering (ICSCSE)*, 157–160 (IEEE, 2016).
31. Zhang, Y. *et al*. Cped: a chinese positive emotion database for emotion elicitation and analysis. *IEEE Transactions on Affective Computing* (2021).
32. Li, F., Yang, L., Shi, H. & Liu, C. Differences in photoplethysmography morphological features and feature time series between two opposite emotions: Happiness and sadness. *Artery Research* **18**, 7–13 (2017).
33. Zhang, X. *et al*. Photoplethysmogram-based cognitive load assessment using multi-feature fusion model. *ACM Transactions on Applied Perception (TAP)* **16**, 1–17 (2019).

34. Liliana, D. Y. Emotion recognition from facial expression using deep convolutional neural network. In *Journal of physics: conference series*, vol. 1193, 012004 (IOP Publishing, 2019).
35. Kessous, L., Castellano, G. & Caridakis, G. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces* **3**, 33–48 (2010).
36. Ioannou, S. V. *et al*. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks* **18**, 423–435 (2005).
37. Peirce, J. *et al*. Psychopy2: Experiments in behavior made easy. *Behavior research methods* **51**, 195–203 (2019).
38. Yang, P. *et al*. A multimodal dataset for mixed emotion recognition. *zenodo* https://doi.org/10.5281/zenodo.8002281 (2022).
39. Anusha, A. *et al*. Electrodermal activity based pre-surgery stress detection using a wrist wearable. *IEEE journal of biomedical and health informatics* **24**, 92–100 (2019).
40. Saganowski, S. *et al*. Emognition dataset: emotion recognition with self-reports, facial expressions, and physiology using wearables. *Scientific data* **9**, 158 (2022).
41. Xu, J., Ren, F. & Bao, Y. Eeg emotion classification based on baseline strategy. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, 43–46 (IEEE, 2018).
42. Murugappan, M. & Murugappan, S. Human emotion recognition through short time electroencephalogram (eeg) signals using fast fourier transform (fft). In *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, 289–294 (IEEE, 2013).
43. Taran, S. & Bajaj, V. Emotion recognition from single-channel eeg signals using a two-stage correlation and instantaneous frequency-based filtering method. *Computer methods and programs in biomedicine* **173**, 157–165 (2019).
44. Patterson, J. A., McIlwraith, D. C. & Yang, G.-Z. A flexible, low noise reflective ppg sensor platform for ear-worn heart rate monitoring. In *2009 sixth international workshop on wearable and implantable body sensor networks*, 286–291 (IEEE, 2009).
45. Chang, C.-Y., Chang, C.-W. & Lin, Y.-M. Application of support vector machine for emotion classification. In *2012 Sixth International Conference on Genetic and Evolutionary Computing*, 249–252 (IEEE, 2012).
46. Hashemi, M. Design and development of gsr biofeedback device. *European Journal of Engineering and Formal Sciences* **4**, 42–51 (2021).
47. Moser, M. K., Resch, B. & Ehrhart, M. An individual-oriented algorithm for stress detection in wearable sensor measurements. *IEEE Sensors Journal* (2023).
48. Welch, P. 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**, 70–73 (1967).
49. Sagie, A., Larson, M. G., Goldberg, R. J., Bengtson, J. R. & Levy, D. An improved method for adjusting the qt interval for heart rate (the framingham heart study). *The American journal of cardiology* **70**, 797–801 (1992).
50. Duan, R.-N., Zhu, J.-Y. & Lu, B.-L. Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 81–84 (IEEE, 2013).
51. Udovičić, G., Derek, J., Russo, M. & Sikora, M. Wearable emotion recognition system based on gsr and ppg signals. In *Proceedings of the 2nd international workshop on multimedia for personal health and health care*, 53–59 (2017).
52. Zhao, G. & Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence* **29**, 915–928 (2007).
53. Yan, W.-J. *et al*. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one* **9**, e86041 (2014).

Acknowledgements

This work was supported by the Natural Science Foundation of China (U2336214, 62332019).

Author contributions

P.Y., N.Q.L., and X.G.L. designed the experiment, collected the data, processed the data, contributed to technical validation, and prepared the first draft of the manuscript. Y.Z.S. participated in the experiment design and stimuli selection. W.Q.J. and Z.Q.R. collected the data and verified the dataset. J.S., M.J.Y., and R.Y. supervised the data collection and preparation of the first draft of the manuscript. D.Z. and Y.J.L. originated the concept for this study and supervised the experiment design and data collection. All authors contributed to manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.-J.L.