REFED: A Subject Real-time Dynamic Labeled EEG-fNIRS Synchronized Recorded Emotion Dataset

Xiaojun Ning[1,2], **Jing Wang**[1,3,*], **Zhiyang Feng**[1,2], **Tianzuo Xin**[1,2], **Shuo Zhang**[1,2], **Shaoqi Zhang**[1,2], **Zheng Lian**[4], **Yi Ding**[5], **Youfang Lin**[1,2], **Ziyu Jia**[6,7,8,*]

[1] School of Computer Science & Technology, Beijing Jiaotong University, Beijing, China.
[2] Beijing Key Laboratory of Traffic Data Mining and Embodied Intelligence, Beijing, China.
[3] Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education, Beijing, China.
[4] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.
[5] College of Computing & Data Science, Nanyang Technological University, Singapore.
[6] Beijing Key Laboratory of Brainnetome and Brain-Computer Interface, Institute of Automation, Chinese Academy of Sciences, Beijing, China.
[7] Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China.
[8] Shanghai Key Laboratory of Data Science, Shanghai, China.

{ningxj, wj, fengzy, tianzuoxin, shuo_zhang, zhangsq, yflin}@bjtu.edu.cn, lianzheng2016@ia.ac.cn, ding.yi@ntu.edu.sg, jia.ziyu@outlook.com

Abstract

Affective brain-computer interfaces (aBCIs) play a crucial role in personalized human–computer interaction and neurofeedback modulation. To develop practical and effective aBCI paradigms and to investigate the spatial-temporal dynamics of brain activity under emotional inducement, portable electroencephalography (EEG) signals have been widely adopted. To further enhance spatial-temporal perception, functional near-infrared spectroscopy (fNIRS) has attracted increasing interest in the aBCI field and has been explored in combination with EEG. However, existing datasets typically provide only static fixation labels, overlooking the dynamic changes in subjects' emotions. Notably, some studies have attempted to collect continuously annotated emotional data, but they have recorded only peripheral physiological signals without directly observing brain activity, limiting insight into underlying neural states under different emotions. To address these challenges, we present the Real-time labeled EEG-fNIRS Dataset (REFED). To the best of our knowledge, this is the first EEG-fNIRS dataset with real-time dynamic emotional annotations. REFED simultaneously records brain signals from both EEG and fNIRS modalities while providing continuous, real-time annotations of valence and arousal. The results of the data analysis demonstrate the effectiveness of emotion inducement and the reliability of real-time annotation. This dataset offers the possibility for studying the neurovascular coupling mechanism under emotional evolution and for developing dynamic, robust affective BCIs.

1 Introduction

Emotion is not only an important internal mechanism that modulates cognition, behavior, and decision-making, but also plays a pivotal role in brain–computer interface (BCI) systems by bridging internal user states and external interactions [1, 2]. With advances in brain signal modeling and artificial

* Corresponding authors: Jing Wang and Ziyu Jia.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Track on Datasets and Benchmarks.

intelligence, affective BCIs (aBCIs) have emerged as a prominent research focus. The goal of aBCIs is to perceive users' emotional states in real-time and adapt interaction strategies, neurofeedback mechanisms, or personalized interventions accordingly. These capabilities hold great promise across a range of applications, including intelligent human–computer interaction, mental health care, and immersive user experiences [3, 4].

To develop effective and practical affective BCIs and to investigate the spatial-temporal dynamics of brain states under various emotions and emotional transitions, researchers have widely adopted emotion-inducement paradigms. These typically involve using emotional stimuli such as video clips to evoke affective responses while synchronously recording physiological signals [5, 6, 7, 8]. However, a significant gap remains: **To the best of our knowledge, no affective BCI dataset simultaneously records multimodal brain signals and provides real-time dynamic emotion annotation.**

On one hand, to directly observe brain activity, early studies focus on using electroencephalography (EEG) to capture neural electrical responses. Researchers decode subjects' emotional states by analyzing the spatial-temporal patterns of EEG signals [9, 10, 11, 12]. Many widely used datasets, such as DEAP [5], MAHNOB-HCI [13], SEED [14, 15], and DREAMER [6], are all affective BCI datasets that are dominated by EEG signals. However, due to the limited spatial resolution of EEG caused by volume conduction effects [16], recent studies explore the use of functional near-infrared spectroscopy (fNIRS) as a complementary modality for emotion recognition [17]. In addition, a few efforts have attempted to simultaneously record EEG and fNIRS to investigate the activity patterns of multimodal brain signals under different emotions [18]. Nevertheless, *all these brain-signal-based affective BCI datasets lack real-time, dynamic emotional annotations, limiting their applicability to dynamic affective decoding tasks in the real world.*

On the other hand, to provide emotion annotations, most emotion-inducing materials are designed to target specific emotional states. Some datasets directly use the targeted emotion of the stimulus as the annotation [19], while other datasets record the combined emotions produced by each subject under each piece of evoked material as the annotation through a form of subject self-assessment [20]. Both approaches assign a single, fixed emotional label to each stimulus clip. However, emotional experiences are inherently dynamic, especially during the viewing of affective videos [21]. Using a static label as the ground truth in training emotion recognition models introduces many errors and can negatively impact the effectiveness of model training and usage. Recently, a few datasets, such as CASE [22] and CEAP-360VR [21] have attempted to record real-time emotional changes during stimulus presentation. However, these datasets only captured peripheral physiological signals (e.g., ECG, EDA, skin temperature), without directly observing brain activity. In other words, *existing real-time dynamic annotated emotion datasets lack direct measurements of the brain.*

To address the above research gaps, we present the **REFED**, a **R**eal-time labeled **E**EG-**f**NIRS **E**motion **D**ataset. To the best of our knowledge, this is the first subject real-time dynamic labeled EEG-fNIRS synchronized recorded emotion dataset for affective brain-computer interfaces, aiming to more dynamically and comprehensively capture the individual's neural response during emotional stimuli. We collected multimodal brain signals from 32 healthy participants across 15 emotion-eliciting video clips. By combining EEG and fNIRS, we recorded brain activity from both electrophysiological and hemodynamic perspectives, enabling emotion modeling with high temporal and spatial resolution. Unlike existing affective BCI datasets, our dataset uniquely provides time-aligned dynamic emotion annotations based on participants' subjective experiences. During video watching, participants continuously rated their emotional states using a visual feedback tool, producing real-time dynamic emotion annotations. This provides an unprecedented foundation for studying the neural and vascular dynamic mechanisms underlying emotional evolution. Further details and access to the dataset can be found at `https://refed-dataset.github.io/`.

2 Related Work

To develop efficient and reliable affective brain-computer interfaces and explore the emotional secrets of the human brain, many researchers employ physiological signals elicited by emotional stimuli to reveal the neural and cognitive patterns associated with different emotional states [9, 23, 24].

On the one hand, to comprehensively capture the spatial-temporal dynamics and neurophysiological patterns of the brain in response to different emotions, most early affective BCI datasets primarily adopt EEG as the main data modality. The DEAP [5] and MAHNOB-HCI [13] datasets are among the earliest to explore neural electrical activity under audiovisual emotional elicitation of different emotions. The SEED series datasets [14, 15] provide EEG, EOG, and other electrophysiological signals from Chinese participants elicited by emotional movie clips. These datasets lay the technical foundation for standardized emotional data collection. As interest grows in emotion recognition algorithms under cross-session and cross-subject conditions, subsequent datasets such as DREAMER [6], AMIGOS [7], MPED [25], and FACED [26] datasets further diversify participant populations and experimental content. In recent years, to investigate brain activity with higher spatial resolution, fNIRS has gradually attracted attention in the affective BCI field. fNIRS directly measures cortical blood flow, offering a different perspective on brain activity. Recently, the TVED-fNIRS dataset [17] records fNIRS data from participants exposed to different emotional video stimuli. Studies by Si et al. [18] and the FEAD dataset [8] also combine EEG and fNIRS to uncover the spatial-temporal dynamic features of the brain during emotional processing. However, these datasets only provide traditional static annotations, without capturing the dynamic emotional changes of the participants, limiting the exploration of the brain's dynamic emotional processes.

On the other hand, to provide emotion annotations, current datasets generally adopt two main labeling strategies. The first is discrete emotion categories, such as happiness, sadness, anger, relaxation, and fear. These labels may be self-reported by participants or derived directly from the targeted emotion of the stimuli. Datasets like SEED [14, 15] and THU-EP [27] follow this approach. The second strategy is based on Russell's circumplex model [28] of emotion, using valence-arousal values, where valence indicates the positivity or negativity of an emotion, and arousal indicates the emotional energy intensity. This approach typically involves participants reporting their feelings using the self-assessment scale. It is widely used in datasets such as DEAP [5], MAHNOB-HCI [13], and DREAMER [6]. However, both labeling strategies are static, assigning a fixed emotion label to each stimulus segment. In reality, participants' emotional states often change dynamically during the presentation of stimuli. Different episodes trigger continuous emotion changes, and static labels limit the performance of emotion recognition models and hinder the study of fine-grained emotional neural mechanisms. Recently, datasets such as CASE [22] and CEAP-360VR [21] have adopted real-time joystick-based annotation to record dynamic emotional responses, but they only collect peripheral physiological signals without recording brain signals, lacking direct observation of brain activity.

In summary, current affective BCI research still faces gaps in both recording brain modalities and emotion annotation strategies. There is no publicly available emotion dataset with dynamic annotations based on multimodal brain signals. The dataset proposed in this paper combines the above two features and presents the dynamically annotated EEG-fNIRS multimodal brain signal dataset for emotion analysis. This offers the possibility to explore the spatial-temporal features of the brain in dynamic emotional changes and to design more practical and reliable affective brain-computer interfaces.

3 Recording Details

3.1 Overall Recording Protocol

The complete experimental protocol of our dataset collection is shown in Figure 1. All processes lasted about 1.5 hours per subject, and the specific experimental flow is as follows:

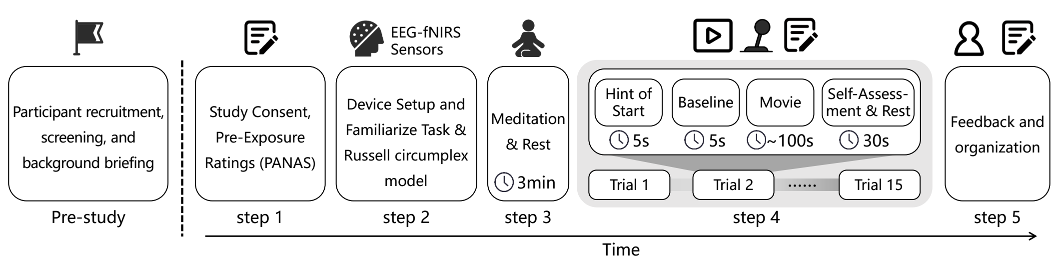

Figure 1: Experimental protocol.

1) **Obtain informed consent.** Prior to the study, informed consent was obtained from all subjects. Participants were thoroughly briefed on the experimental procedures, equipment, and any potential risks or discomforts associated with the study. We also collected basic demographic and individual difference information, including gender, age, etc. To assess participants' recent emotional states at the time of data collection, we administered the Positive and Negative Affect Schedule (PANAS) questionnaire [29]. PANAS evaluates emotional state at the time of data collection by asking participants to rate the intensity of 10 positive and 10 negative emotions they have experienced over a recent period.

2) **Device setup and system familiarization.** Subsequently, participants were fitted with an EEG cap. Conductive gel was applied to ensure that the impedance of all EEG channels was reduced below 50 kΩ. fNIRS optodes were then inserted and adjusted according to a predefined channel layout, and the light intensity was automatically calibrated to optimize signal quality. Next, participants received a detailed explanation of the theoretical basis of emotion assessment and were familiarized with the entire data acquisition procedure. The emotion assessment framework included key concepts such as arousal, valence, and dominance, as well as guidance on how to identify one's own emotional state. We introduced the full experimental procedure, which consisted of meditation, video viewing, real-time emotion annotation, and questionnaire completion. A brief demonstration program was used to help participants practice the full workflow, including providing continuous feedback on their emotional state using a joystick in the valence–arousal space. We confirmed that the subjects had fully understood and could operate proficiently based on their real-time annotations of sample videos of different affective tendencies in the demonstration system.

3) **Meditation break.** Participants were instructed to perform a 3-minute meditation, during which they closed their eyes, focused on their breath, and avoided any physical movement. This step helps participants relax and enter a calm, focused state, thereby enhancing the effectiveness of subsequent emotion elicitation.

4) **Video watching and recording.** In this step, we began the formal record. Participants watched a total of 15 video clips, each selected to induce different emotional states. The videos were categorized into five targeted emotion types: 1) high valence–high arousal (HVHA): high-energy positive emotions (e.g., excitement, happiness), 2) high valence–low arousal (HVLA): low-energy positive emotions (e.g., relaxation, serenity), 3) low valence–high arousal (LVHA): high-energy negative emotions (e.g., anger, fear), 4)low valence–low arousal (LVLA): low-energy negative emotions (e.g., sadness, boredom), 5) medium valence–medium arousal (MVMA): emotions with no apparent emotional tendency. Each video segment was preceded by a 5-second countdown to help participants prepare. Immediately afterward, 5 seconds of baseline data were recorded while participants were asked to sit still and fixate on a white cross on the screen. The video then played automatically, during which time participants were emotionally stimulated by the video. Meanwhile, participants were asked to use a joystick to continuously control a cursor on a 2D valence–arousal coordinate plane, reflecting their real-time emotional state (e.g., laughter moments, tear moments). After each clip, participants completed the Self-Assessment Manikin (SAM) scale [20] to provide an overall subjective evaluation of their emotional response to the video. The SAM self-assessment scale reflects the average emotion of watching the video clip, including valence, arousal, and dominance, which are recorded on a scale of 1 to 9, respectively. In addition, we also collect a separate familiarity rating for each clip.

5) **Feedback and organization.** At last, we communicate with the participant about the experimental process and feedback, and synthesize the SAM scale to understand the subjects' emotional evocations. Additionally, we examined the real-time emotional trajectories to ensure that the emotional responses were properly induced and annotated, thereby confirming the authenticity and reliability of the collected data.

3.2 Recording Device Settings

To simultaneously record EEG and fNIRS signals, we employed an ESI Neuroscan 64-channel EEG system and a Shimadzu LABNIRS fNIRS system. An EEG–fNIRS joint cap and a signal synchronization module developed by FiStar were also used to achieve joint acquisition and precise temporal alignment of the two modalities. The layout of the EEG and fNIRS channels used in the combined recording is illustrated in Figure 2. Meanwhile, to support real-time annotation of participants' emotional states and to automate the experimental workflow, we developed a custom real-time annotation and control system, as shown in Figure 3.

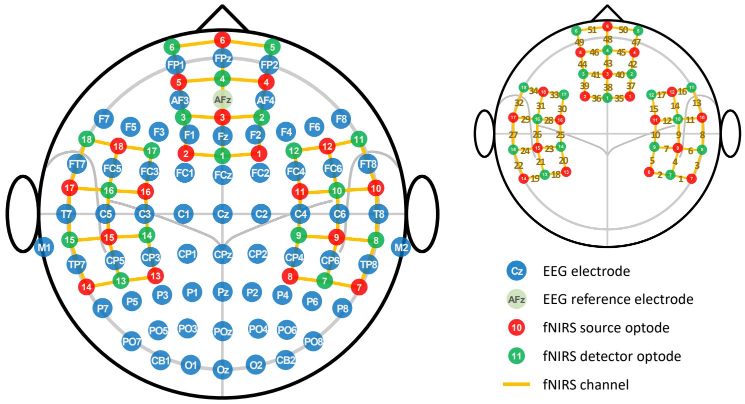

Figure 2: EEG-fNIRS layout. There are 64 channels of EEG and 51 channels of fNIRS in total.

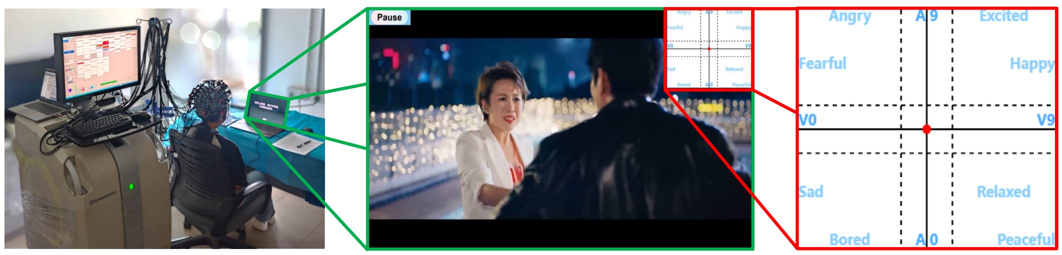

Figure 3: Experimental recording environment and the real-time labeling and control system.

The EEG data were recorded using a 64-channel wet-electrode system, with electrodes positioned according to the international 10–10 EEG system. The default reference electrode was placed at AFz, and the signals were sampled at 1000 Hz. During the recording, conductive gel (Greentek GT5) was applied to rapidly reduce electrode impedance. The fNIRS data were acquired using the Shimadzu LABNIRS system, which employs three near-infrared wavelengths (780, 805, and 830 nm) to detect blood flow changes in the cerebral cortex. LABNIRS calculates the change in concentration of oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) based on the difference in absorption spectra of these two hemoglobins. In our recordings, the fNIRS optodes were arranged to cover three major cortical regions: the prefrontal lobes, the left temporal lobes, and the right temporal lobes. There are a total of 18 emitters and 18 detectors positioned with a source–detector distance of 30 mm, forming 51 measurement channels. The sampling rate of the fNIRS system was 47.62 Hz (i.e., sampling every 21 ms). The prefrontal lobes are strongly associated with emotion generation and regulation, playing a central role in cognitive-affective processes [30]. The temporal lobes are primarily involved in language comprehension, emotional processing, and memory encoding [31]. These regions are the most critical areas of the cerebral cortex responsible for emotions.

The annotation and control system was developed using HTML and JavaScript to ensure broad compatibility across different computers. We ran the system on a separate computer to control the progress of the experiment, to prompt subjects for relevant information, and to automate video playback and automatic emotion annotation. Meanwhile, the computer is connected to an Xbox controller to allow participants to continuously annotate their emotional states during the video presentation. The emotion annotations were recorded in real-time and transmitted via serial communication to the acquisition system, enabling precise temporal alignment between the brain signals and the subjective emotional labels. During video play, the screen entered full-screen mode, with the video occupying the main area and a 2D coordinate system displayed in the top-right corner. The axes of the coordinate system represented valence and arousal, respectively. To reduce the cognitive load of annotation, we displayed examples of typical emotions in each quadrant of the coordinate plane as visual references. With this system, in a smooth situation, subjects only need to manipulate the right joystick of the controller to complete the real-time labeling, and all the experimental processes are carried out automatically.

3.3 Emotion Inducing Materials

We adopted Russell's circumplex model [28] of affect to assess emotional states. Under this model, emotions are characterized along two orthogonal dimensions: valence and arousal. Valence represents the degree of pleasantness, ranging from negative (low valence) to positive (high valence) emotions. Arousal reflects the level of physiological activation, ranging from low energy (low arousal) to high energy (high arousal). The two dimensions form four quadrants, i.e., high valence-high arousal (HVHA), high valence-low arousal (HVLA), low valence-high arousal (LVHA), and low valence-low arousal (LVLA), and each has its typical sentiment. In addition to these four quadrants, we also considered emotional states without a clear valence or arousal tendency, defined as medium valence–medium arousal (MVMA). This resulted in a total of five targeted emotion categories in our design. For each emotion category, we selected three videos, totaling 15 emotional video clips. Detailed information about these videos is provided in Appendix B.3. All participants watched the videos in the same order. Most of the video clips were sourced from publicly available and validated affective video libraries (including standard Chinese emotional film clips dataset [32, 33], Chinese positive emotion database (CPED) [34], and positive emotion database (PED) [35]), which ensured the basic effect of video induction. Due to the relatively heavy EEG cap worn during the experiment, we carefully controlled the average length of each video to approximately 100 seconds to avoid participant discomfort. Hence, the entire video-watching phase lasted about 40 minutes.

3.4 Recording Participants

We recruited 32 healthy participants (22 males, 10 females), aged between 18 to 34 years (M = 21.3, SD = 2.7), primarily consisting of undergraduate and graduate students. Individuals with a history of cardiovascular or cerebrovascular diseases, visual or auditory impairments, or cognitive or psychological disorders were excluded from the study. All participants were native Chinese speakers to eliminate the potential influence of language and cultural background on emotional elicitation. Before the recording session, participants were instructed to avoid caffeine and other neurostimulants within 12 hours, wash their hair in advance, and ensure adequate sleep the night before. All experiments were conducted during daytime hours in a quiet laboratory environment. To minimize interference with the quality of signal acquisition, fluorescent lighting was turned off during the sessions.

3.5 Dataset Annotations and Details

EEG-fNIRS data. The dataset contains physiological recordings from 32 participants during 15 emotion-inducing video clips, totaling 480 trials, approximately 820 minutes of data. For each trial, 5 seconds of baseline physiological data recorded before video onset are also included for calibration purposes. Each data segment is identified using a combination of `subject_ID` and `video_ID`. The original EEG and fNIRS signals are sampled at 1000 Hz and 47.62 Hz, respectively. For more details regarding data organization and usages, please refer to Appendix A.

Emotion annotations. For each participant's data segment, we recorded 1 Hz dynamic emotion labels (valence and arousal) as well as self-reported ratings using the SAM scale (Self-Assessment Manikin), which includes valence, arousal, dominance, and familiarity. The dynamic emotion annotations are represented as integers within the range of 1 to 255. Each video segment begins at the center of the two-dimensional coordinate system (coordinates: (128, 128)) and changes dynamically in response to the participant's emotional shifts. The SAM scale ratings are provided by the participants after each video and are in the range of 1 to 9.

Available supervised learning paradigms. When using this dataset for supervised learning, EEG-fNIRS data can be segmented into 1-second or longer intervals, where each segment serves as an input sample. As for the labels, a novel suggestion is to directly use the dynamic emotion label values (valence or arousal ranging from 1-255) as labels for regression tasks. This approach is more accurate than traditional classification tasks with hard boundaries. In the regression task, the model will be able to capture the subtle changes in emotion labels, offering greater potential for exploring the relationship between brain signals and emotional variations. Of course, more common valence-arousal classification tasks can also be performed. Since all labels begin from the neutral point (128,128), the proportion of neutral emotion data is relatively high. Here, we recommend three-class classification (low-middle-high) experiments for valence or arousal. We do not recommend conducting binary classification (low-high valence/arousal) experiments, as the neutral label point cannot be reasonably divided. Additionally, partitioning the 2D valence-arousal plane into 5 regions (i.e., 4 quadrants and the central region) based on the quadrants and central area is also a reasonable labeling scheme.

4 Experiments and Analysis

To validate the quality of the collected EEG-fNIRS data and real-time dynamic annotations, and to explore the patterns of neural and vascular activity of the brain under different emotions, we analyze the annotations, visualize the brain signal data, and evaluate the performance of the dataset in supervised learning.

4.1 Label Analysis

To verify the validity of the collected emotional annotations, we compute the average of all participants' real-time annotation trajectories and compare them with the results of the SAM scale.

Figure 4 shows the average valence-arousal trajectories of 32 participants while watching 15 emotional video clips. Overall, the videos elicit emotional responses effectively. Each video induces noticeable changes in the trajectories, and the real-time annotations are well consistent with the expected emotions. The trajectories of videos intended to elicit HVHA (i.e., happiness/pleasure), LVHA (i.e., fear/anger), and LVLA (i.e., sadness/boredom) emotions align best with expectations. The trajectories of MVMA (i.e., neutral) and HVLA (i.e., relaxed/calm) videos are relatively more ambiguous but still within an acceptable range. This ambiguity may be partially due to the choice of video materials for MVMA and HVLA, as participants show the greatest individual variation in their subjective perception of neutral and relaxed emotions. Some participants particularly enjoy food or scenery, which may lead to high arousal when watching such videos. Others may find neutral videos boring, resulting in lower arousal levels.

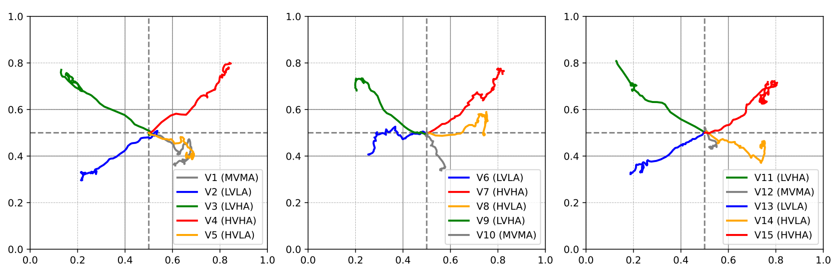

Figure 4: Average valence-arousal trajectories for participants watching 15 videos.

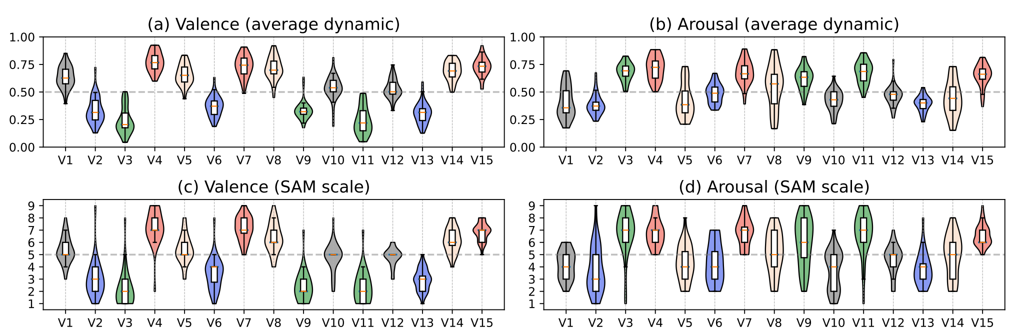

Figure 5: Valence/arousal distributions for participants watching 15 videos.

In addition, to evaluate the consistency between real-time annotations and participants' self-reported SAM scores, Figure 5 respectively shows violin plots of the average real-time annotations and SAM scores for each video. The SAM scores generally exhibit larger standard deviations, but both annotation methods reflect consistent emotional elicitation effects and align with the targeted emotions. To further demonstrate the reliability of the real-time annotations, we provide a joint distribution plot of all coordinate points collected during the annotation process in Appendix D.

4.2 Supervised Learning

To evaluate the effectiveness of our collected data, we conduct classification and regression experiments using the collected EEG and fNIRS signals as input and the dynamically annotated valence-arousal values as labels. Thanks to the dynamic and continuous valence-arousal annotations, we are able to implement regression tasks directly in emotion recognition models to predict arousal or valence levels. We deploy support vector machines with temporal and spectral domain features for supervised learning. For better optimization, we normalize the labels from the original [1, 255] range to the [0, 1] range. We apply 3-fold cross-validation for each participant, dividing each participant's 15 video segments into three groups, each containing one segment for each of the five target emotions. We report the average MAE and MSE across all 32 participants. We also perform three-class classification experiments, where the normalized labels are categorized into low (0 to 0.4), medium (0.4 to 0.6), and high (0.6 to 1.0) levels for valence and arousal. All other experimental settings are consistent with those of the regression task. Detailed experimental settings and complete experimental results are provided in Appendix E.

The experimental results of the supervised learning model TSMMF [36] are shown in Table 1. In our regression tasks, the models achieve valence prediction errors around 0.17 and arousal prediction errors of below 0.15. Figure 6 presents regression cases of predicted emotional trajectories, whose predicted values could follow the trends of the ground truth, demonstrating good fitting performance. In classification tasks, our dataset could achieve over 67% accuracy in arousal three-class classification performance and 62% valence three-class classification performance. Moreover, training with combined EEG and fNIRS modalities outperforms single-modality input, further validating the effectiveness of using multimodal brain signals. These also confirm the existence of the neurovascular coupling mechanism in emotional activities [37, 38], where EEG captures electrical brain activity and fNIRS reflects hemodynamic responses. These two modalities offer non-redundant insights, justifying their integration in our dataset.

Table 1: Performance of the REFED dataset under a supervised learning model (TSMMF [36]).

Modality	Valence - Classification		Arousal - Classification		Valence - Regression		Arousal - Regression	
	Accuracy ↑	F1-score ↑	Accuracy ↑	F1-score ↑	MAE ↓	MSE ↓	MAE ↓	MSE ↓
EEG	0.5961±0.1020	0.3965±0.0848	0.6527±0.1175	0.3720±0.0750	0.1822±0.0432	0.0588±0.0247	0.1542±0.0404	0.0402±0.0181
fNIRS	0.6199±0.1016	0.4485±0.1088	0.6645±0.1217	0.3956±0.0801	0.1716±0.0413	0.0542±0.0248	0.1453±0.0411	0.0376±0.0194
EEG+fNIRS	0.6269±0.1005	0.4611±0.1071	0.6701±0.1171	0.4060±0.0892	0.1705±0.0409	0.0531±0.0236	0.1445±0.0401	0.0369±0.0182

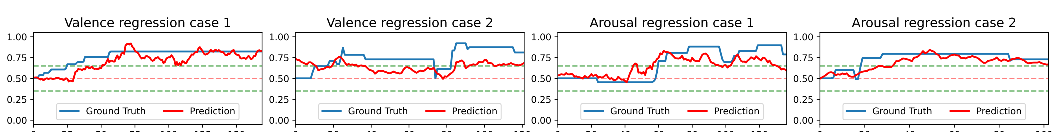

Figure 6: Some cases of valence/arousal regression tasks.

4.3 Visualization

To investigate neural and vascular patterns under different emotional states, we perform averaging analysis on the collected EEG and fNIRS data by differential entropy (DE) features. Figure 7 presents the average spatial activation of EEG and fNIRS (HbT) signals across different emotion categories.

As shown in Figure 7, EEG and fNIRS exhibit clear differences across the five emotional conditions. Overall, the HVHA (happiness/pleasure) emotion shows the highest overall activation levels and the richest brain activity, aligning with many neuroscience studies [39, 40]. Regarding the prefrontal lobes, LV (low-valence) emotions, particularly LVLA (sadness/boredom), show strong activation, while HV (high-valence) emotions reduce prefrontal activity, which aligns with Pessoa's cognitive–emotional

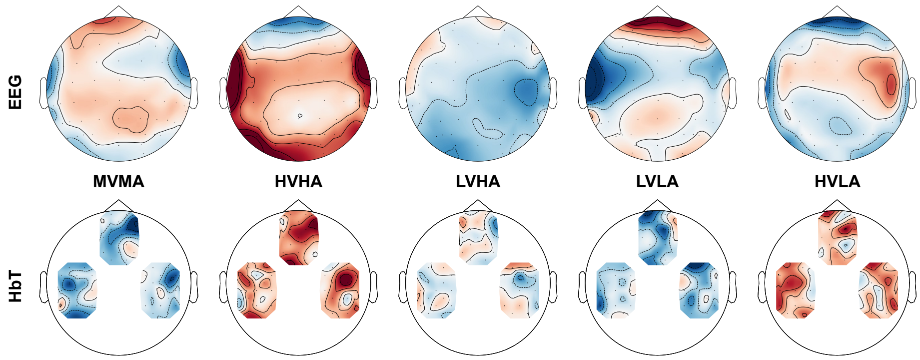

Figure 7: Average activation of EEG and fNIRS (HbT) under different emotions

interaction theory [41]. Specifically, MVMA (neutral) emotion elicits uniform activation in parietal and prefrontal regions, reflecting the brain's baseline state without strong emotional modulation. In contrast, HVHA (happiness/pleasure) emotion shows markedly distinct activation in temporal and occipital regions, potentially linked to the audiovisual impact of the video contents. LVLA (sadness/boredom) emotion exhibits the most salient prefrontal activation, along with some activity in the posterior parietal and occipital regions. This activation pattern may relate to subjects actively suppressing or reappraising negative emotions [42]. Finally, HVLA (relaxed/calm) emotion presents activation in the parietal and temporal regions, likely reflecting increased anterior cingulate activity during relaxation states [43]. Additionally, some discrepancies between electrophysiological and hemodynamic activity are also observed. For instance, with HVHA emotion, the parietal and temporal regions exhibit strong activation in both modalities, while the fNIRS additionally shows strong prefrontal activation. HVLA emotion also presents broader prefrontal activation in the fNIRS view. LVHA emotion reveals greater frontal and temporal activation in blood flow activity. Meanwhile, for LVLA emotions, frontal blood flow activity is much weaker than electrical activity. These findings demonstrate the objective existence of neurovascular coupling mechanisms, and two modalities offer non-redundant insights.

5 Ethical Considerations and Dataset Accessibility

The study for our dataset collection has been approved by the local Institutional Review Board. The subjects recruited for this experiment were healthy adults, and the experimental equipment involved was non-invasive, posing no negative physical or psychological risks. All participants provided informed consent for the recording of their physiological signals and scale data, and agreed that the data would be used solely for non-commercial scientific research purposes. To protect participant privacy, all personally identifiable information has been anonymized in the released dataset. The dataset poses no potential negative impact on society. Further details and access to the dataset are available at our website (`https://refed-dataset.github.io/`) following the `CC BY-NC-SA` license, and users can use it for research for non-commercial purposes.

6 Limitations

We propose the EEG-fNIRS emotion dataset with real-time dynamic annotations, offering new possibilities for developing novel affective BCIs. However, the dataset still has some limitations.

First, real-time annotation inevitably demands some cognitive effort from participants. During annotation, participants may become distracted from the video content, potentially affecting emotional induction and annotation quality. In fact, we evaluate multiple dynamic annotation strategies. One alternative is to have participants watch the video twice—first with full attention, and then again while recalling their emotional experience. However, our pilot study finds that participants struggle to recall emotional fluctuations. Another potential approach is to estimate emotions using facial expression recognition during the first viewing and correct annotations afterward. However, many participants display subtle facial expressions, and this method is both technically complex and prone to inaccuracies. Compared with retrospective labeling or facial-expression-based inference, real-time annotation more directly captures participants' subjective experiences and ensures temporal continuity, a strategy also adopted by prior continuous datasets such as CASE and CEAP-360VR. To mitigate potential burden, we implemented standardized training sessions, ensured participants were familiar with the annotation procedure, and controlled video duration with sufficient breaks to reduce fatigue. While hybrid annotation strategies may theoretically enhance reliability, they also introduce longer experimental sessions and additional participant demands. Overall, our chosen real-time paradigm reflects a balance between annotation accuracy and participant workload. How to perform dynamic emotion labeling more accurately and portably remains a direction worth exploring.

Furthermore, the unfamiliar laboratory environment, EEG-fNIRS equipment, and individual differences in culture and education level may also influence participants' emotional responses. To mitigate these effects, we implemented several measures such as turning off ambient lighting, maintaining stable temperature and humidity, and minimizing the duration of recording sessions. In the future, how to utilize more portable wearable devices to achieve affective brain-computer interfaces in more natural, daily environments remains a highly intriguing research topic.

7 Conclusion

In this study, we propose the REFED dataset, an affective brain-computer interface dataset integrating multimodal brain signals and real-time dynamic emotion annotations. It fills a gap in the study of neural mechanisms of emotion's dynamic evolution and the development of high-reliability aBCI models. By synchronizing the acquisition of EEG and fNIRS signals, REFED realizes the joint observation of neuroelectrical activity and hemodynamic response under emotional evocation, which provides unique data support for exploring emotion-related neurovascular coupling mechanisms. Meanwhile, the real-time valence and arousal annotation based on subjects' subjective reports realizes temporal alignment of brain signals with emotional state changes, which significantly improves the temporal modeling capability of the emotion recognition model. Experimental validation shows that the dataset meets the standard in terms of emotion evoked validity and labeling reliability, and the multimodal signal features show significant correlation with the dynamic labeling. The open sharing of REFED will promote the cross-modal neural representation parsing for the dynamic encoding of emotions in the following research directions, and lay an important foundation for the field of affective computation and brain-computer interfaces to move towards dynamic interaction paradigms with higher ecological validity.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 2024YJS045 & No. 2025JBMC030), the Excellent Youth Program of State Key Laboratory of Multimodal Artificial Intelligence Systems (No. MAIS2024311), and the Youth Science Fund Project of National Natural Science Foundation of China (No. 62201572 & No.62306317). This work was also sponsored by Beijing Nova Program (Grant No. 20250484804).

References

[1] Kübra Erat, Elif Bilge Şahin, Furkan Doğan, Nur Merdanoğlu, Ahmet Akcakaya, and Pınar Onay Durdu. Emotion recognition with eeg-based brain-computer interfaces: a systematic literature review. *Multimedia Tools and Applications*, 83(33):79647–79694, 2024.

[2] Ziyu Jia, Youfang Lin, Jing Wang, Zhiyang Feng, Xiangheng Xie, and Caijie Chen. Hetemotionnet: two-stream heterogeneous graph recurrent neural network for multi-modal emotion recognition. In *Proceedings of the 29th ACM international conference on multimedia*, pages 1047–1056, 2021.

[3] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. Human–ai collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1):46–57, 2023.

[4] Cheng Cheng, Wenzhe Liu, Xinying Wang, Lin Feng, and Ziyu Jia. Disd-net: A dynamic interactive network with self-distillation for cross-subject multi-modal emotion recognition. *IEEE Transactions on Multimedia*, 2025.

[5] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[6] Stamos Katsigiannis and Naeem 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*, 22(1):98–107, 2017.

[7] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE transactions on affective computing*, 12(2):479–493, 2018.

[8] Alireza Farrokhi Nia, Vanessa Tang, Valery Malyshau, Amit Barde, Gonzalo Maso Talou, and Mark Billinghurst. Fead: Introduction to the fnirs-eeg affective database-video stimuli. *IEEE Transactions on Affective Computing*, 2024.

[9] Ziyu Jia, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, and Jing Wang. Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2909–2917, 2020.

[10] Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2022.

[11] Cheng Cheng, Yong Zhang, Luyao Liu, Wenzhe Liu, and Lin Feng. Multi-domain encoding of spatiotemporal dynamics in eeg for emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 27(3):1342–1353, 2022.

[12] Xiyuan Jin, Jing Wang, Huaiyu Qin, Xiaojun Ning, Tianzuo Xin, and Youfang Lin. Group-wise relation mining for weakly-supervised fine-grained multimodal emotion recognition. *Neural Networks*, page 107543, 2025.

[13] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[14] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

[15] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, 49(3):1110–1122, 2018.

[16] Xianlun Tang, Jing Zhang, Yidan Qi, Ke Liu, Rui Li, and Huiming Wang. A spatial filter temporal graph convolutional network for decoding motor imagery eeg signals. *Expert Systems with Applications*, 238:121915, 2024.

[17] Xiaopeng Si, He Huang, Jiayue Yu, and Dong Ming. The fnirs-based emotion recognition by spatial transformer and wgan data augmentation towards developing a novel affective bci. *IEEE Transactions on Affective Computing*, 2024.

[18] Xiaopeng Si, He Huang, Jiayue Yu, and Dong Ming. Eeg microstates and fnirs metrics reveal the spatiotemporal joint neural processing features of human emotions. *IEEE Transactions on Affective Computing*, 2024.

[19] Ga-Young Choi, Jong-Gyu Shin, Ji-Yoon Lee, Jun-Seok Lee, In-Seok Heo, Ha-Yeong Yoon, Wansu Lim, Jin-Woo Jeong, Sang-Ho Kim, and Han-Jeong Hwang. Eeg dataset for the recognition of different emotions induced in voice-user interaction. *Scientific data*, 11(1):1084, 2024.

[20] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[21] Tong Xue, Abdallah El Ali, Tianyi Zhang, Gangyi Ding, and Pablo Cesar. Ceap-360vr: A continuous physiological and behavioral emotion annotation dataset for 360° vr videos. *IEEE Transactions on Multimedia*, 25:243–255, 2021.

[22] Karan Sharma, Claudio Castellini, Egon L Van Den Broek, Alin Albu-Schaeffer, and Friedhelm Schwenker. A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific data*, 6(1):196, 2019.

[23] Jing Wang, Zhiyang Feng, Xiaojun Ning, Youfang Lin, Badong Chen, and Ziyu Jia. Two-stream dynamic heterogeneous graph recurrent neural network for multi-label multi-modal emotion recognition. *IEEE Transactions on Affective Computing*, 2025.

[24] Ziyu Jia, Yucheng Liu, Haichao Wang, and Tianzi Jiang. Cross-modal knowledge distillation for enhanced unimodal emotion recognition. *IEEE Transactions on Affective Computing*, 2025.

[25] Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. Mped: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, 7:12177–12191, 2019.

[26] Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023.

[27] Xin Hu, Fei Wang, and Dan Zhang. Similar brains blend emotion in similar ways: Neural representations of individual difference in emotion profiles. *Neuroimage*, 247:118819, 2022.

[28] James A Russell, Anna Weiss, and Gerald A Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989.

[29] Vincent Tran. Positive affect negative affect scale (panas). In *Encyclopedia of behavioral medicine*, pages 1708–1709. Springer, 2020.

[30] Tor D Wager, Matthew L Davidson, Brent L Hughes, Martin A Lindquist, and Kevin N Ochsner. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, 59(6):1037–1050, 2008.

[31] Jennifer T Crinion, Matthew A Lambon-Ralph, Elizabeth A Warburton, David Howard, and Richard JS Wise. Temporal lobe regions engaged during normal speech comprehension. *Brain*, 126(5):1193–1201, 2003.

[32] Yan Ge, Guozhen Zhao, Yulin Zhang, Rebecca J Houston, and Jinjing Song. A standardised database of chinese emotional film clips. *Cognition and Emotion*, 33(5):976–990, 2019.

[33] Yong-Jin Liu, Minjing Yu, Guozhen Zhao, Jinjing Song, Yan Ge, and Yuanchun Shi. Real-time movie-induced discrete emotion recognition from eeg signals. *IEEE Transactions on Affective Computing*, 9(4):550–562, 2017.

[34] Yulin Zhang, Guozhen Zhao, Yezhi Shu, Yan Ge, Dan Zhang, Yong-Jin Liu, and Xianghong Sun. Cped: A chinese positive emotion database for emotion elicitation and analysis. *IEEE Transactions on Affective Computing*, 14(2):1417–1430, 2021.

[35] Xin Hu, Chu Zhuang, Fei Wang, Yong-Jin Liu, Chang-Hwan Im, and Dan Zhang. fnirs evidence for recognizably different positive emotions. *Frontiers in human neuroscience*, 13:120, 2019.

[36] Xiaopeng Si, Shuai Zhang, Zhuobin Yang, Jiayue Yu, and Dong Ming. A bidirectional cross-modal transformer representation learning model for eeg-fnirs multimodal affective bci. *Expert Systems with Applications*, 266:126081, 2025.

[37] Hasan Onur Keles, Randall L Barbour, and Ahmet Omurtag. Hemodynamic correlates of spontaneous neural activity measured by human whole-head resting state eeg+ fnirs. *Neuroimage*, 138:76–87, 2016.

[38] Michael K Yeung and Vivian W Chu. Viewing neurovascular coupling through the lens of combined eeg–fnirs: A systematic review of current methods. *Psychophysiology*, 59(6):e14054, 2022.

[39] Joshua M Carlson, Dan Foti, Lilianne R Mujica-Parodi, Eddie Harmon-Jones, and Greg Hajcak. Ventral striatal and medial prefrontal bold activation is correlated with reward-related electrocortical activity: a combined erp and fmri study. *Neuroimage*, 57(4):1608–1616, 2011.

[40] Richard J Davidson, Paul Ekman, Clifford D Saron, Joseph A Senulis, and Wallace V Friesen. Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I. *Journal of personality and social psychology*, 58(2):330, 1990.

[41] Luiz Pessoa. On the relationship between emotion and cognition. *Nature reviews neuroscience*, 9(2):148–158, 2008.

[42] Kevin N Ochsner and James J Gross. The cognitive control of emotion. *Trends in cognitive sciences*, 9(5):242–249, 2005.

[43] HD Critchley, RN Melmed, E Featherstone, CJ Mathias, and Raymond J Dolan. Brain activity during biofeedback relaxation: a functional neuroimaging investigation. *Brain*, 124(5):1003–1012, 2001.

[44] Qun He, Lufeng Feng, Guoqian Jiang, and Ping Xie. Multimodal multitask neural network for motor imagery classification with eeg and fnirs signals. *IEEE Sensors Journal*, 22(21):20695–20706, 2022.

[45] Ziyu Jia, Fengming Zhao, Yuzhe Guo, Hairong Chen, and Tianzi Jiang. Multi-level disentangling network for cross-subject emotion recognition based on multimodal physiological signals. In Kate Larson, editor, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 3069–3077. International Joint Conferences on Artificial Intelligence Organization, 8 2024. Main Track.

[46] Yunqiang Pei, Jialei Tang, Qihang Tang, Mingfeng Zha, Dongyu Xie, Guoqing Wang, Zhitao Liu, Ning Xie, Peng Wang, Yang Yang, and Hengtao Shen. Emotion recognition in hmds: A multi-task approach using physiological signals and occluded faces. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, page 5977–5986, New York, NY, USA, 2024. Association for Computing Machinery.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions and scopes of the paper which are mainly shown in Section 3, and 4 are same with the abstract and introduction.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 discuss the limitations of the proposed work.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Section 4.2 and 4.3 use experiments and visualizations proof the quality of the EEG-fNIRS data and dynamic annotations.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Section 4.2 and Appendix provide the information to reproduce the main experimental results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We host the dataset on the Hugging Face platform and provide a website to guide users in downloading the dataset and example codes.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All the data splits, hyperparameters, how they were chosen, type of optimizer, etc. are provided in Section 4.2 and the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Experiment in Section 4.2 include the standard error during 3-fold cross validation.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The detail of computer resources are provided in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research strictly adheres to the NeurIPS Code of Ethics, including considerations for data usage, human impact, and scientific integrity.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The potential positive societal impacts is discuss in Conclusion part. The paper does not have negative societal impacts which lead by intentional or unintentional misuse of the proposed technology.

    Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: All data have been anonymized to remove any personally identifiable information (PII), and no facial or voice data were collected; Participants provided informed consent for the collection and use of their data for research purposes; The data collection protocol was approved by the ethics committee; The dataset is distributed under a controlled license that requires users to agree to a Data Use Agreement, which prohibits misuse such as identification attempts or unethical profiling; A detailed data documentation is provided in the Appendix to ensure transparency and responsible reuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The models used in the paper are all open source, and the citations of the sources are provided in Section 4.2.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The collected dataset is equipped with a detailed documentation of the description, which is provided in the Appendix.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: The full text of instructions given to participants, as well as details about compensation are provided in section 6 and the appendix.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

Justification: The paper describes potential risks incurred by study participants and discloses these risks to subjects, as well as obtaining local institutional review board (IRB) approval.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: This research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

Appendix

A REFED Overview

The *REFED* (**R**eal-time Labeled **E**EG-**f**NIRS **E**motion **D**ataset) is an effective Brain-Computer Interface (aBCI) dataset integrating multimodal brain signals and real-time dynamic emotion annotation. REFED dataset fills a critical gap in the study of neural mechanisms of emotional dynamic evolution and the development of high-reliability aBCI models. By synchronizing the acquisition of EEG and fNIRS signals, REFED realizes the joint observation of neuroelectrical activity and hemodynamic response under emotional evocation, which provides unique data support for exploring emotion-related neurovascular coupling mechanisms. Meanwhile, the dynamic valence and arousal annotation based on participants' subjective reports realizes the temporal alignment of brain signals with emotional state changes, which significantly improves the temporal modeling capability of the emotion recognition model. Experimental validation shows that the dataset meets the standard in terms of emotion evoked validity and labeling reliability, and the multimodal signal features show significant correlation with the dynamic labeling. The open sharing of REFED will promote the cross-modal neural representation parsing for the dynamic encoding of emotions in the following research directions, and lay an important foundation for the field of affective computation and brain-computer interfaces to move towards dynamic interaction paradigms with higher ecological validity.

REFED Dataset Summary

Motivation

Summary: An affective BCI dataset integrating multimodal brain signals (EEG-fNIRS) and real-time emotion annotation (valence-arousal).
Example Use Case: Emotion recognition, Valence/Arousal classification, Valence/Arousal regression
Original Authors: X. Ning, J. Wang, Z. Feng, T. Xin, S. Zhang, S. Zhang, Z. Lian, Y. Ding, Y. Lin, Z. Jia

Metadata	
Hosting Platform	Hugging Face (https://huggingface.co/)
Keywords	Affective BCI, EEG, fNIRS, Real-time label, EEG-fNIRS
Format	.mat, .csv
Ethical Review Approval	IRB-CASIA
License	CC BY-NC-SA

Sensors	
EEG	ESI Neuroscan, 64 channels, 200Hz
fNIRS	Shimadzu LABNIRS, 51 channels, 47.6Hz

Annotations	
Dynamic Emotion	Valence and arousal
Self-Assessments	Valence, arousal, dominance, and familiarity
Other Data	PANAS scales

Participants	
Count	32
Gender	22 male, 10 female
Age	18~34 (M=21.3, SD=2.7)
Criteria	Healthy adults

Dataset Size	
Record Duration	about 820 minutes (Emotion Inducing) 40 minutes (Baselines Total)
Total Size	about 30 GB (Raw Data)

Figure 8: Summary of the REFED dataset.

A.1 Motivation

Our dataset aims to realize the following two innovations simultaneously:

1) **Multimodal brain signals for aBCI datasets**. While existing affective BCI (aBCI) datasets primarily rely on EEG, they are constrained by EEG's limited spatial resolution due to

volume conduction effects. By integrating EEG and fNIRS, our dataset captures both the fast-changing electrophysiological activity and the slower hemodynamic responses associated with emotional processing. This multimodal approach provides a more comprehensive and complementary view of brain activity, enhancing the interpretability and robustness of emotion decoding models. The spatial coverage of fNIRS, particularly over emotion-related brain regions such as the prefrontal and temporal cortices, enables deeper insights into the spatial patterns of affective neural responses.

2) **Real-time dynamic emotion annotations**. Traditional datasets typically assign static emotion labels to each stimulus, either based on the intended affective category of the video or post-hoc self-report summaries. However, emotional experiences during video viewing are inherently dynamic and continuously evolving. To better reflect the temporal progression of emotional states, we introduced a real-time continuous annotation protocol, allowing participants to report their emotional experiences on a moment-to-moment basis using a visual feedback interface and Xbox controller. This enables the training and evaluation of dynamic affective decoding models, which are more aligned with real-world human-computer interaction needs.

In summary, **our dataset fills a critical gap in affective BCI research by jointly providing multimodal brain signals and real-time dynamic emotion annotations**. This resource lays a solid foundation for advancing the understanding of emotion-related brain dynamics and developing practical aBCI systems capable of real-time emotional adaptation.

A.2 Comparison with Other Related Datasets

The REFED dataset introduces significant advancements in both the modality of brain signal acquisition and the granularity of data annotation for emotion-related datasets. By integrating EEG and fNIRS, REFED captures neural activity from both electrophysiological and hemodynamic perspectives, offering a more comprehensive and multimodal representation of emotional brain responses. In addition, unlike most existing datasets that rely on static, post-hoc labels, REFED incorporates real-time, dynamic emotion annotations collected continuously during stimulus presentation. These two innovations mark a notable step forward in the development of affective computing resources. As summarized in Table 2, we provide a comparative overview of representative affective BCI datasets, highlighting differences in their brain signal modalities and the type of emotional labels they offer, thereby underscoring the unique contributions of REFED.

Table 2: Comparison with other related datasets. Our REFED dataset provides brain signals in both EEG and fNIRS modalities, as well as dynamic emotion labeling and static emotion labeling.

Dataset	Brain Signals	Static Label	Dynamic Label	#Subjects	#Videos
DEAP [5]	EEG	SAM (Valence, Arousal, Dominance), Liking, Familiarity	×	32	40
MAHNOB-HCI [13]	EEG	SAM (Valence, Arousal, Dominance), Emotion, Predictability	×	27	20
SEED [14, 15]	EEG	Emotion (Targeted emotion)	×	15	15
DREAMER [6]	EEG	SAM (Valence, Arousal, Dominance)	×	23	18
AMIGOS [7]	EEG	PANAS, SAM (Valence, Arousal, Dominance), Liking, Familiarity, Personality, Emotion	×	40	16
MPED [25]	EEG	PANAS, SAM (Valence, Arousal), DES	×	23	28
THU-EP [27]	EEG	Emotion profiles	×	80	28
FACED [26]	EEG	SAM (Valence, Arousal), Emotion, Familiarity, Liking	×	123	28
TVED-fNIRS [17]	fNIRS	SAM (Valence, Arousal)	×	30	24
FEAD [8]	EEG+fNIRS	PANAS, SAM (Valence, Arousal, Dominance), Emotion, Familiarity	×	37	24
CASE [22]	× (only PPG)	×	Valence, Arousal	30	8
CEAP-360VR [21]	× (only PPG)	×	Valence, Arousal	32	8
REFED (Ours)	**EEG+fNIRS**	**PANAS, SAM (Valence, Arousal, Dominance), Familiarity**	**Valence, Arousal**	**32**	**15**

A.3 Dataset Organization

The REFED dataset mainly consists of EEG signals, fNIRS signals, real-time emotion annotations, and other related datasheets. The EEG signals, fNIRS signals, and emotion annotations are provided in MATLAB `.mat` format, and the other datasheets are provided in `.csv` format.

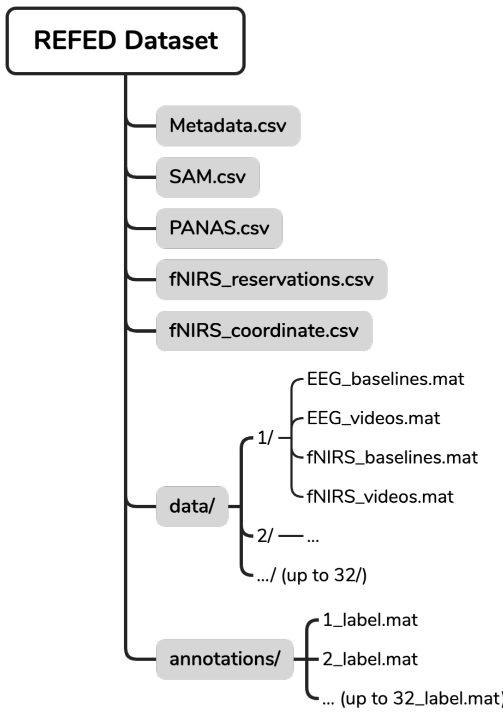

Figure 9: File organization of the REFED dataset.

`./Metadata.csv`
Subject demographic information (ID, gender, age, etc.).

`./SAM.csv`
Post-trial emotion scores based on the Self-Assessment Manikin (SAM) scale (valence and arousal).

`./PANAS.csv`
Positive and negative affect scores before and after the experiment (Positive and Negative Affect Schedule).

`./fNIRS_reservations.csv`
fNIRS acquisition logs including bad channel markers.

`./fNIRS_coordinate.csv`
3D coordinates of fNIRS channels for alignment and spatial analysis.

`./data/`
Each subject (1 to 32) has a corresponding subfolder under `./data/`, containing raw brain signals:

- EEG_baselines.mat: Resting EEG signal (baseline phase).
- EEG_videos.mat: Emotionally evoked phase EEG signaling (video stimulation).
- fNIRS_baselines.mat: Resting-state fNIRS signal (baseline phase).
- fNIRS_videos.mat: Emotionally evoked phase fNIRS signaling (video stimulation).
- EEG sampling rate: 1000 Hz; fNIRS sampling rate: 47.62 Hz.
- fNIRS signals type include: HbO, HbR, HbT, Abs 780 nm, Abs 805 nm, Abs 830 nm.
- Each `.mat` file stores multi-trial time series data (15 videos in total).

- EEG data shape: `channels × sampling_point`;
  fNIRS data shape: `signal_type × channels × sampling_point`.

`./annotations/`
Dynamic emotion annotations (valence and arousal) for each subject during video viewing:

- Each `*_label.mat` contains time-aligned joystick-based valence and arousal annotations.
- Annotations shape: `2 × time` (valence and arousal over time).

A.4 Dataset Usages

The REFED dataset provides 64-channel EEG data, 51-channel fNIRS data, static emotion annotations, and dynamic real-time emotion annotations under 15 different emotional video clips. It supports a wide range of research tasks, including emotion recognition, valence/arousal classification and regression, the discovery of patterns in EEG or fNIRS during emotional transitions, and the investigation of electrophysiological and hemodynamic mechanisms underlying emotional dynamics.

1) **Emotion Recognition Tasks**: Researchers can use the dataset to train models that classify discrete emotional states based on EEG, fNIRS, or their fusion. With well-defined emotion categories and multimodal brain signals, REFED enables both unimodal and multimodal emotion recognition approaches under different conditions of emotional valence and arousal.

2) **Valence / Arousal Classification Tasks**: The dataset allows for binary or multi-class classification tasks focused on emotional valence (positive vs. negative) and arousal (high vs. low). By leveraging both the static labels assigned to each video and the continuous annotations over time, researchers can build models that differentiate between different levels of affective intensity and polarity.

3) **Valence / Arousal Regression Tasks**: Using the real-time dynamic annotations, models can be trained to perform fine-grained regression of valence and arousal values on a continuous scale. This supports more nuanced modeling of emotional experience and is especially relevant for real-time affective decoding in applications like affective brain-computer interfaces (aBCIs) and neuroadaptive systems.

4) **Pattern Discovery in EEG or fNIRS During Emotional Shifts**: With synchronized EEG and fNIRS recordings, REFED enables the analysis of how electrical and hemodynamic brain signals change during transitions between different emotional states. Researchers can explore spatial-temporal patterns, connectivity dynamics, and cross-modal signal correlations during emotional fluctuations.

5) **Mechanistic Study of Electrophysiological and Hemodynamic Responses to Emotion**: The dataset provides a foundation for investigating the distinct and complementary roles of electrophysiological activity (captured by EEG) and hemodynamic responses (captured by fNIRS) in emotion processing. This includes analyzing temporal response profiles, regional activations, and how different neural substrates contribute to emotional regulation and perception.

By offering a rich, multimodal, and dynamically annotated dataset, REFED dataset serves as a valuable resource for advancing both practical applications in affective computing and theoretical understanding of the neural underpinnings of human emotion.

B Data Record Protocol

B.1 Dynamic Annotation and Control System

To streamline and standardize the data collection process for EEG-fNIRS during emotion elicitation experiments, we developed a real-time emotion annotation and control system based on HTML and JavaScript. This system automates key aspects of the experimental workflow, including procedural guidance, stimulus presentation, and the recording and transmission of emotional annotations. Designed to run on a main experimental computer, the system interfaces with an Xbox controller to enable participants to continuously report their emotional states in real time. By incorporating this system into the experiment, we ensure high consistency across sessions, minimize manual intervention, and significantly reduce the workload of the experimenter.

As illustrated in Figure 10, the experimental procedure begins with a brief meditation period, followed by the sequential presentation of 15 emotion-eliciting video clips. Each segment is preceded by an on-screen countdown to help participants prepare mentally and physically. In addition, prior to the formal experiment, the system can be switched into a training mode via a simple configuration file. In this mode, three sample video clips are presented, allowing participants to become familiar with the system interface and practice real-time emotion annotation, thereby improving the reliability and usability of the collected data.

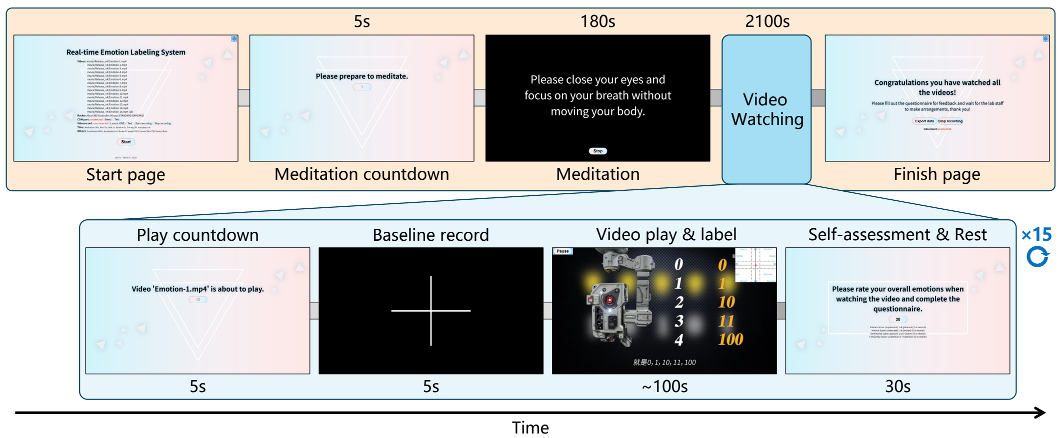

Figure 10: Control system running flow

Beyond the main experimental computer, a synchronization module developed by FiStar enables precise time alignment by transmitting dynamic emotion annotations to both the Neuroscan EEG and LABNIRS fNIRS systems. These systems concurrently acquire neural signals from a shared EEG-fNIRS cap, which integrates electrodes and optodes over emotion-relevant brain regions. The EEG and fNIRS data are recorded and processed on separate, dedicated PCs—referred to as the EEG recording PC and the fNIRS control PC, respectively. The overall hardware configuration, including device connections and the flow of data and annotation triggers across systems, is depicted in Figure 11. This tightly integrated setup ensures synchronized multimodal data acquisition, laying a robust foundation for subsequent analysis of emotion-related brain dynamics.

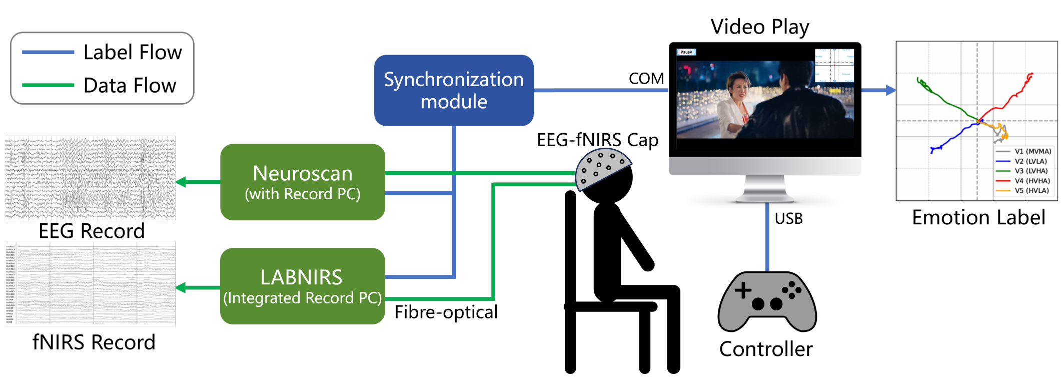

Figure 11: Device connection diagram

B.2 fNIRS Coordinates

Our fNIRS channel layout is carefully designed to balance spatial coverage and signal quality, comprising a total of 51 channels distributed across three functionally significant brain regions. These channels are configured using 18 light sources (emitters) and 18 detectors provided by the LABNIRS system. The selected regions—the prefrontal cortex, the left temporal lobe, and the right temporal lobe—are well-established in the literature as key areas involved in emotional perception, generation, and regulation. This configuration enables the system to capture hemodynamic changes in these emotion-relevant brain areas with both spatial specificity and functional relevance. The LABNIRS system operates at a sampling rate of 47.62 Hz, which is sufficient for monitoring the slow fluctuations in oxygenated and deoxygenated hemoglobin that reflect cortical activation patterns. Table 3 provides the three-dimensional (3D) coordinates of all 51 channels as defined by source detector positions.

Table 3: The coordinates of 51-channel fNIRS.

Channel	Source	Detector	3D Coordinates			Brain Region
			x	y	z	
CH_01	T7	R7	-77	36	64	Right temporal lobe
CH_02	T8	R7	-70	36	98	Right temporal lobe
CH_03	T7	R8	-81	22	43	Right temporal lobe
CH_04	T9	R7	-80	19	80	Right temporal lobe
CH_05	T8	R9	-67	18	116	Right temporal lobe
CH_06	T9	R8	-84	5	60	Right temporal lobe
CH_07	T9	R9	-77	2	98	Right temporal lobe
CH_08	T10	R8	-83	-8	38	Right temporal lobe
CH_09	T9	R10	-82	-13	76	Right temporal lobe
CH_10	T11	R9	-69	-18	112	Right temporal lobe
CH_11	T10	R10	-81	-25	54	Right temporal lobe
CH_12	T11	R10	-74	-32	90	Right temporal lobe
CH_13	T10	R11	-78	-36	35	Right temporal lobe
CH_14	T12	R10	-76	-44	68	Right temporal lobe
CH_15	T11	R12	-64	-52	100	Right temporal lobe
CH_16	T12	R11	-73	-55	48	Right temporal lobe
CH_17	T12	R12	-65	-64	78	Right temporal lobe
CH_18	T13	R13	70	38	100	Left temporal lobe
CH_19	T14	R13	78	37	64	Left temporal lobe
CH_20	T13	R14	67	20	119	Left temporal lobe
CH_21	T15	R13	80	21	81	Left temporal lobe
CH_22	T14	R15	82	22	43	Left temporal lobe
CH_23	T15	R14	77	3	100	Left temporal lobe
CH_24	T15	R15	84	6	60	Left temporal lobe
CH_25	T16	R14	69	-17	114	Left temporal lobe
CH_26	T15	R16	82	-12	76	Left temporal lobe
CH_27	T17	R15	83	-7	38	Left temporal lobe
CH_28	T16	R16	74	-32	91	Left temporal lobe
CH_29	T17	R16	81	-25	55	Left temporal lobe
CH_30	T16	R17	62	-51	101	Left temporal lobe
CH_31	T18	R16	76	-44	69	Left temporal lobe
CH_32	T17	R18	78	-36	35	Left temporal lobe
CH_33	T18	R17	64	-63	79	Left temporal lobe
CH_34	T18	R18	73	-55	49	Left temporal lobe
CH_35	T1	R1	-17	-76	121	Frontal lobe
CH_36	T2	R1	16	-75	121	Frontal lobe
CH_37	T1	R2	-34	-86	100	Frontal lobe
CH_38	T3	R1	0	-88	106	Frontal lobe
CH_39	T2	R3	32	-85	100	Frontal lobe
CH_40	T3	R2	-17	-98	85	Frontal lobe
CH_41	T3	R3	16	-98	85	Frontal lobe
CH_42	T4	R2	-33	-102	65	Frontal lobe
CH_43	T3	R4	-1	-104	69	Frontal lobe
CH_44	T5	R3	32	-101	65	Frontal lobe
CH_45	T4	R4	-17	-107	49	Frontal lobe
CH_46	T5	R4	16	-107	49	Frontal lobe
CH_47	T4	R5	-29	-108	33	Frontal lobe
CH_48	T6	R4	0	-112	38	Frontal lobe
CH_49	T5	R6	29	-107	33	Frontal lobe
CH_50	T6	R5	-13	-112	22	Frontal lobe
CH_51	T6	R6	14	-112	22	Frontal lobe

B.3 Emotion Inducting Videos

We selected a total of 15 emotional video clips to elicit participants' affective responses. These clips were carefully chosen, with the vast majority sourced from publicly available and previously validated emotion-eliciting video databases. The 15 videos were organized into three blocks, each containing five clips that correspond to five distinct target emotion categories. These categories were defined based on the widely used valence–arousal model and include:

1) High Valence–High Arousal (HVHA): High-energy positive emotions such as excitement and happiness;
2) High Valence–Low Arousal (HVLA): Low-energy positive emotions such as relaxation and serenity;
3) Low Valence–High Arousal (LVHA): High-energy negative emotions such as anger and fear;
4) Low Valence–Low Arousal (LVLA): Low-energy negative emotions such as sadness and boredom;
5) Medium Valence–Medium Arousal (MVMA): Neutral states or those without a clear affective polarity.

To ensure consistent emotion elicitation across participants, all subjects viewed the video clips in the same fixed sequence. Additionally, to help participants become familiar with the experimental interface and the real-time emotion annotation procedure, we selected three additional practice clips—each representing a different target emotion category—for use during the training session before the formal experiment.

Table 4 provides detailed information on the source of each emotional video clip as well as the duration of the edited segments used in the experiment. The majority of the videos were sourced from the publicly available video libraries PED (positive emotion database), CPED (Chinese positive emotion database), and SCEFD (standard Chinese emotional film clips dataset). These carefully curated stimuli form the core of our emotion-elicitation paradigm and provide a balanced and diverse set of emotional experiences for data collection.

Table 4: Summary of fifteen emotional induction video clips and three practice video clips.

id	Source	Description	Clip Length	Targeted emotion
p1	Bilibili[1]	Full understanding of how computers work	0:01:50	MVMA (Neutral)
p2	movie	Train to Busan	0:02:29	LVLA (Sad)
p3	PED[2]	My Neighbor Totoro	0:00:36	HVHA (Happy)
	Total		0:04:55	
	Average		0:01:38	
1	Bilibili[3]	How Computer Chips Work	0:02:16	MVMA (Neutral)
2	SCEFD[4]	Changjiang Qihao	0:02:19	LVLA (Sad)
3	SCEFD[4]	Inner Senses	0:01:32	LVHA (Fear)
4	CPED[5]	King of Destruction	0:01:21	HVHA (Happy)
5	CPED[5]	7 Strangest Places in the World	0:02:17	HVLA (Relax)
6	PED[2]	Gangs of New York	0:01:35	LVLA (Sad)
7	CPED[5]	Iron Fist of Shame	0:02:04	HVHA (Happy)
8	CPED[5]	Food those tantalizing moments	0:01:49	HVLA (Relax)
9	PED[2]	The Pianist	0:01:12	LVHA (Fear)
10	SCEFD[4]	Raise the Red Lantern	0:01:02	MVMA (Neutral)
11	SCEFD[4]	The Chrysalis	0:01:03	LVHA (Fear)
12	SCEFD[4]	Black Coal Thin Ice	0:01:05	MVMA (Neutral)
13	SCEFD[4]	Bodyguards and Assassins	0:01:53	LVLA (Sad)
14	CPED[5]	Epic Landscape Clips	0:01:45	HVLA (Relax)
15	CPED[5]	The Chinatown Detective	0:02:52	HVHA (Happy)
	Total		0:26:05	
	Average		0:01:44	

[1] `https://www.bilibili.com/video/BV1tM4y1B7qP`
[2] PED: positive emotion database [35]
[3] `https://www.bilibili.com/video/BV1gM411e74r`
[4] SCEFD: standard Chinese emotional film clips dataset [32, 33]
[5] CPED: Chinese positive emotion database [34]

B.4 Emotion Scales

Throughout the study, we utilized both the Positive and Negative Affect Schedule (PANAS) and SAM emotion scales. The PANAS scale was completed before the collection to assess the subject's recent basic mood state. The SAM scale was completed after the completion of each video during the collection process to assess the subject's effectiveness in responding to the mood evoked by that video.

B.4.1 PANAS Scale

The Positive and Negative Affect Schedule (PANAS) is a widely used self-report questionnaire designed to assess an individual's experiences of positive and negative emotions over a specific period. It consists of two subscales: Positive Affect (PA) and Negative Affect (NA), each comprising 10 adjective items. Participants rate the extent to which they have felt each emotion on a 5-point Likert scale. PANAS has demonstrated strong reliability and validity, making it a popular tool in psychological, neuroscientific, and clinical research.

Please mark "√" in the corresponding position.

Positive	Few or no	A little	Moderately	Quite a bit	Extremely
Interested	○	○	○	○	○
Excited	○	○	○	○	○
Strong	○	○	○	○	○
Enthusiastic	○	○	○	○	○
Proud	○	○	○	○	○
Alert	○	○	○	○	○
Inspired	○	○	○	○	○
Determined	○	○	○	○	○
Attentive	○	○	○	○	○
Active	○	○	○	○	○

Negative	Few or no	A little	Moderately	Quite a bit	Extremely
Distressed	○	○	○	○	○
Upset	○	○	○	○	○
Guilty	○	○	○	○	○
Scared	○	○	○	○	○
Hostile	○	○	○	○	○
Irritable	○	○	○	○	○
Ashamed	○	○	○	○	○
Nervous	○	○	○	○	○
Jittery	○	○	○	○	○
Afraid	○	○	○	○	○

Figure 12: PANAS Scale.

B.4.2 SAM Scale

The Self-Assessment Manikin (SAM) is a non-verbal, pictorial tool used to measure individuals' subjective emotional responses to stimuli. It assesses three core dimensions of affective experience: valence, arousal, and dominance, each represented by a series of human-like figures illustrating gradations of emotional states. Participants intuitively select the figure that best matches their feelings. Owing to its language-free design, SAM is particularly suitable for cross-cultural studies, children, and populations with limited verbal abilities, and is widely used in affective neuroscience and emotion induction research. Similar to most previous studies, we also collect familiarity as an additional variable along with the SAM Scale.

Please mark "√" in the corresponding position.

Video-x	1	2	3	4	5	6	7	8	9
Valence	○	○	○	○	○	○	○	○	○
Arousal	○	○	○	○	○	○	○	○	○
Dominance	○	○	○	○	○	○	○	○	○
Familiarity	○	○	○	○	○	○	○	○	○

Figure 13: SAM scale for a video clip.

C Ethical Considerations and Participant Selection

The local Institutional Review Board approved the ethical conduct of this study. All procedures involving human participants adhered strictly to established ethical standards, with an emphasis on participant safety, privacy, and informed consent.

We recruited healthy adult volunteers with no history of neurological or psychiatric disorders to ensure data reliability and minimize confounding factors. Prior to participation, all subjects were thoroughly informed of the study's purpose, procedures, and data usage policies. Informed consent was obtained for the collection of physiological signals (EEG and fNIRS) and subjective emotional annotations. The experiment employed entirely non-invasive methods and posed no known physical or psychological risks. Emotional video stimuli were carefully selected to avoid causing distress, and participants were free to withdraw at any time without penalty.

All personally identifiable information was removed or anonymized before dataset release. Only essential demographic details (e.g., age range and gender) are retained for scientific analysis. The dataset is fully anonymized and privacy-compliant and publicly released on Hugging Face under the `CC BY-NC-SA` license, permitting non-commercial scientific research with required attribution and share-alike terms. While we recognize the theoretical risk of misuse, such as unauthorized commercial use or re-identification, we have implemented multiple safeguards. In addition to anonymization, the dataset is distributed through reputable academic platforms with clearly stated usage terms. Users must agree to these terms prior to accessing. We will also monitor dataset usage and encourage the research community to report any suspected violations. If misuse is detected, we reserve the right to restrict access or take further action as necessary. By combining ethical licensing, thorough anonymization, and responsible data sharing practices, we strive to ensure that the REFED dataset remains a safe, valuable, and ethically used resource for advancing research in affective computing and brain–computer interfaces.

D Distribution of Annotations

Figure 14 exhibits the detailed distribution of the valence-arousal dynamic labeling. All valence and arousal degrees are normalized to 0-1. All videos start playing with both valence and arousal from the center point. The blue color depth of each subplot shows the distribution density of the labels, the blue histogram on the top shows the distribution of the valence, and the green histogram on the right shows the distribution of arousal.

First, we observed that the annotations for nearly all videos were clustered within their intended target quadrants in the valence–arousal space. Importantly, participants were not given any prior indication of the emotional tendency of each video, yet their actual ratings aligned well with the expected emotion categories. This consistency supports the effectiveness of the emotion elicitation protocol employed in our experiment. However, there were a few deviations. Specifically, Video 8 and Video 14, both categorized as high valence–low arousal (HVLA), showed slightly elevated arousal levels in the actual annotations compared to their target quadrant. These two clips featured content such as scenic landscapes and gourmet food, which, while generally calming, may evoke heightened interest or excitement in certain individuals depending on personal preferences. This variability highlights the importance of collecting subjective self-assessments rather than relying solely on predefined labels. Even when exposed to the same emotional stimulus, individuals may experience diverse emotional responses. Such inter-subject variability reinforces the value of real-time self-reported emotion annotations in achieving a more accurate and individualized understanding of affective states.

E Supervised Learning Experiments

To evaluate the learnability of EEG–fNIRS signals and the valence–arousal annotations in the REFED dataset, we conducted supervised training experiments. Specifically, we assessed model performance on both regression and three-class classification tasks along the valence and arousal dimensions.

Emotion Labels. For regression tasks, all valence and arousal labels were normalized to the [0, 1] range, and mean absolute error (MAE) and mean squared error (MSE) were used as evaluation metrics. For classification tasks, the normalized values were categorized as follows: values within [0.0, 0.4) were labeled as low, [0.4, 0.6] as medium, and (0.6, 1.0] as high. Accuracy and weighted

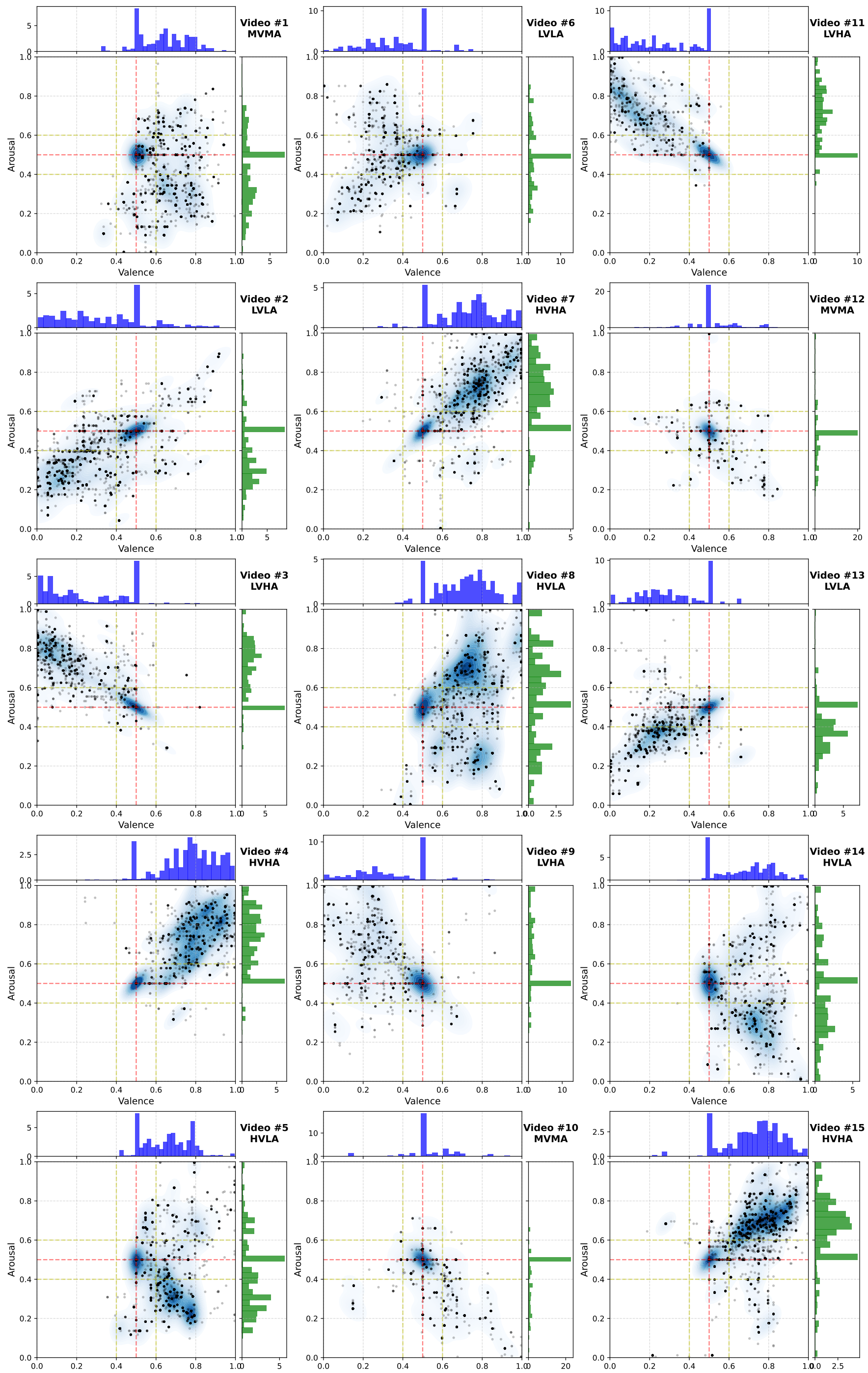

Figure 14: Distribution of detailed dynamic emotion labels (valence and arousal) of 15 videos.

F1 score were used as evaluation metrics for classification. All tasks are divided into samples of 1 second for data segmentation.

Baselines. We evaluated three representative models: Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), M2NN [44], MDNet [45], FERPO [46], and TSMMF [36]. SVM served as a baseline traditional machine learning method. For SVM, the temporal and frequency features from both modalities were flattened and concatenated before input. The MLP is the most basic deep learning model. We also used temporal and frequency features as input to the MLP. The

M2NN [44] is a Multimodal Multitask Neural Network, originally developed for joint EEG–fNIRS classification, directly consumes raw EEG and fNIRS data as input. MDNet [45] proposes a Multi-level Disentangling Network to address emotion recognition based on multimodal physiological signals. FERPO [46] is also a deep learning emotion recognition method that integrates multimodal representations. TSMMF [36] proposes a Temporal–Spatial Multimodal Fusion model, which leverages the bidirectional Cross-Modal Transformer to fuse EEG and fNIRS multimodal brain signals. We tuned the parameters of these deep learning methods to align with the sampling frequencies of our dataset.

Feature Extraction. We extracted temporal and frequency domain features from both EEG and fNIRS modalities for SVM and MLP models. For EEG, we derived differential entropy (DE) features [14, 15] across all channels in five standard frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31 Hz), and gamma (31–50 Hz), which form the feature matrix $\boldsymbol{H}_{\text{EEG}} \in \mathbb{R}^{64\times5}$. Assuming EEG signals x follow a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, the DE feature is calculated as:

$$\text{DE} = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\frac{(x-\mu)^2}{2\sigma^2} \ln\left(\frac{1}{\sqrt{2\pi\sigma}} \exp\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{2}\ln(2\pi e\sigma^2) \tag{1}$$

For fNIRS, we used the total hemoglobin (HbT) signal and extracted 5 time-domain features: mean, maximum, minimum, standard deviation, and slope. Additionally, we computed DE features in the 0.01–0.7 Hz frequency range to provide frequency-domain features. Hence, these fNIRS features form the feature matrix $\boldsymbol{H}_{\text{fNIRS}} \in \mathbb{R}^{51\times6}$.

Environments. All experiments were conducted on a workstation equipped with an Intel Core i5-13600K CPU, NVIDIA GeForce RTX 4070 Ti Super 16GB GPU, and 64 GB RAM. The software environment includes Python 3.13.2, Scikit-learn 1.6.1, scipy 1.15.2, mne 1.9.0, PyTorch 2.7.0, and CUDA 12.8.

Model parameters. SVM hyperparameter search over kernel types (linear, rbf, and poly) and C values (0.001 to 100). For deep learning methods, we used the Adam optimizer with grid search over batch size (16 to 512) and learning rate (1e-6 to 1e-2).

Results. Table 5 and Tables 6 summarize the performance of the REFED dataset on valence–arousal regression and three-class classification tasks using six representative models. Overall, all models achieved promising results, with best classification accuracies exceeding 0.6 and mean absolute errors (MAEs) around 0.17 for valence and 0.14 for arousal in regression tasks. In general, the combination of EEG and fNIRS consistently yielded the best performance across nearly all settings,

Table 5: Regression performance of the REFED dataset under supervised learning models.

Model	Modality	Valence		Arousal	
		MAE ↓	MSE ↓	MAE ↓	MSE ↓
SVM	EEG	0.1933±0.0415	0.0603±0.0241	0.1673±0.0486	0.0497±0.0316
	fNIRS	0.1867±0.0410	0.0595±0.0257	0.1598±0.0484	0.0436±0.0259
	EEG+fNIRS	0.1871±0.0411	0.0586±0.0248	0.1601±0.0477	0.0435±0.0230
MLP	EEG	0.1911±0.0431	0.0601±0.0245	0.1643±0.0445	0.0432±0.0204
	fNIRS	0.1908±0.0485	0.0618±0.0279	0.1570±0.0452	0.0406±0.0205
	EEG+fNIRS	0.1909±0.0501	0.0616±0.0282	0.1571±0.0444	0.0407±0.0200
M2NN [44]	EEG	0.1899±0.0442	0.0610±0.0255	0.1581±0.0440	0.0410±0.0188
	fNIRS	0.1934±0.0439	0.0601±0.0240	0.1624±0.0444	0.0424±0.0196
	EEG+fNIRS	0.1893±0.0429	0.0590±0.0235	0.1585±0.0446	0.0407±0.0191
MDNet [45]	EEG	0.1876±0.0437	0.0607±0.0248	0.1547±0.0430	0.0404±0.0181
	fNIRS	0.1793±0.0433	0.0592±0.0270	0.1479±0.0391	0.0391±0.0184
	EEG+fNIRS	0.1768±0.0429	0.0579±0.0265	0.1469±0.0381	0.0389±0.0178
FERPO [46]	EEG	0.1929±0.0440	0.0600±0.0249	0.1618±0.0440	0.0422±0.0205
	fNIRS	0.1730±0.0452	0.0546±0.0264	0.1487±0.0419	0.0386±0.0195
	EEG+fNIRS	0.1738±0.0434	0.0540±0.0245	0.1465±0.0419	0.0384±0.0193
TSMMF [36]	EEG	0.1822±0.0432	0.0588±0.0247	0.1542±0.0404	0.0402±0.0181
	fNIRS	0.1716±0.0413	0.0542±0.0248	0.1453±0.0411	0.0376±0.0194
	EEG+fNIRS	0.1705±0.0409	0.0531±0.0236	0.1445±0.0401	0.0369±0.0182

Table 6: Classification performance of the REFED dataset under supervised learning models.

Model	Modality	Valence		Arousal	
		Accuracy ↑	F1-score ↑	Accuracy ↑	F1-score ↑
SVM	EEG	0.5240±0.1354	0.4023±0.1408	0.6210±0.1503	0.5029±0.1666
	fNIRS	0.5730±0.1128	0.4798±0.1233	0.6428±0.1271	0.5466±0.1350
	EEG+fNIRS	0.5738±0.1147	0.4827±0.1231	0.6405±0.1301	0.5426±0.1392
MLP	EEG	0.5066±0.0836	0.4225±0.0693	0.4794±0.0966	0.4239±0.0799
	fNIRS	0.5478±0.0952	0.4929±0.0983	0.5456±0.0926	0.4987±0.0978
	EEG+fNIRS	0.5488±0.0934	0.4922±0.0963	0.5506±0.0889	0.5057±0.0934
M2NN [44]	EEG	0.4874±0.1049	0.4050±0.0866	0.4521±0.0886	0.3961±0.0627
	fNIRS	0.4939±0.0980	0.4294±0.0832	0.4840±0.0937	0.4290±0.0853
	EEG+fNIRS	0.5200±0.0896	0.4464±0.0850	0.5162±0.0966	0.4546±0.0981
MDNet [45]	EEG	0.5593±0.1034	0.3115±0.0618	0.6362±0.1214	0.3072±0.0566
	fNIRS	0.6053±0.1050	0.4456±0.1047	0.6647±0.1176	0.4239±0.0657
	EEG+fNIRS	0.6049±0.1011	0.4401±0.1010	0.6629±0.1164	0.3969±0.0650
FERPO [46]	EEG	0.5440±0.1236	0.2631±0.0644	0.6264±0.1376	0.2789±0.0593
	fNIRS	0.6299±0.0988	0.4539±0.1134	0.6765±0.1086	0.4231±0.0849
	EEG+fNIRS	0.6273±0.0989	0.4555±0.1115	0.6682±0.1111	0.4074±0.0798
TSMMF [36]	EEG	0.5961±0.1020	0.3965±0.0848	0.6527±0.1175	0.3720±0.0750
	fNIRS	0.6199±0.1016	0.4485±0.1088	0.6645±0.1217	0.3956±0.0801
	EEG+fNIRS	0.6269±0.1005	0.4611±0.1071	0.6701±0.1171	0.4060±0.0892

highlighting the advantage of multimodal brain signals in enhancing emotion recognition accuracy. Interestingly, the fNIRS modality alone often outperformed EEG alone, suggesting that fNIRS is a valuable and potentially underutilized signal source for affective computing. Overall, the TSMMF model achieved the best overall performance, while SVM and MLP performed the worst under certain settings. Although M2NN was originally designed for joint EEG–fNIRS decoding, its architecture was primarily tailored for motor imagery tasks, which may limit its adaptability to affective decoding scenarios. In contrast, MDNet and FERPO are designed for multimodal emotion recognition, but they may lack sufficient modeling capability for fNIRS signals. The TSMMF model, being specifically designed for EEG-fNIRS multimodal emotion recognition, can better capture the spatial-temporal features of brain activity.

F Brain Topography Visualization

Figure 15 presents average EEG–fNIRS topographic maps across all participants for the five target emotional categories. For each emotion, we display four complementary views: EEG, HbT, HbO, and HbR, enabling a multimodal visualization of neural and hemodynamic responses under distinct emotional states. Different emotional states elicit distinct patterns of brain activation. The HVHA emotion shows the most strongest brain activity, with both EEG and fNIRS views showing the highest levels of activation. The prefrontal cortex presents distinct patterns across emotional states. In the EEG view, prefrontal activity markedly decreases during high-valence emotions and increases during low-valence emotions. In contrast, the fNIRS view reveals stronger prefrontal hemodynamic responses under high-valence emotions. The similarities and differences between these two modalities jointly reflect the brain's neurovascular coupling mechanism and demonstrate their complementary roles in understanding emotional processing.

Figure 16 illustrates the average EEG–fNIRS topographic maps grouped by gender, providing comparative insights into potential sex-related differences in neural and hemodynamic activity patterns during emotion elicitation. From the EEG perspective, male participants exhibit significantly greater brain activity under HVLA emotion, while female participants show stronger both occipital and frontal activation under LVLA emotion. From the fNIRS perspective, female participants present more notable activation in the prefrontal cortex and bilateral temporal lobes under LVHA emotion, while exhibiting reduced blood flow activity under LVLA emotion. Overall, male and female participants exhibit both commonalities and differences in emotion-induced responses, reflecting potential neurophysiological mechanism differences in emotion processing across genders.

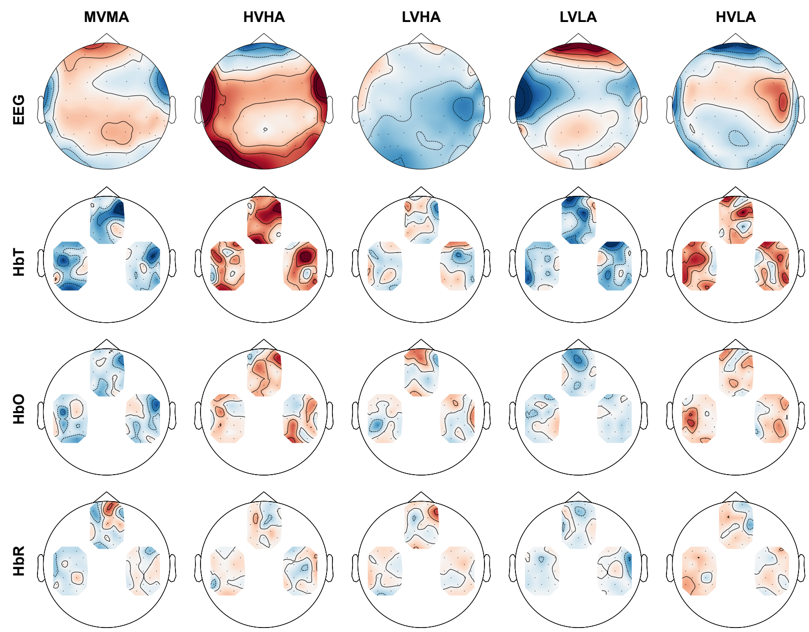

Figure 15: Average activation of EEG and fNIRS under different targeted emotions.

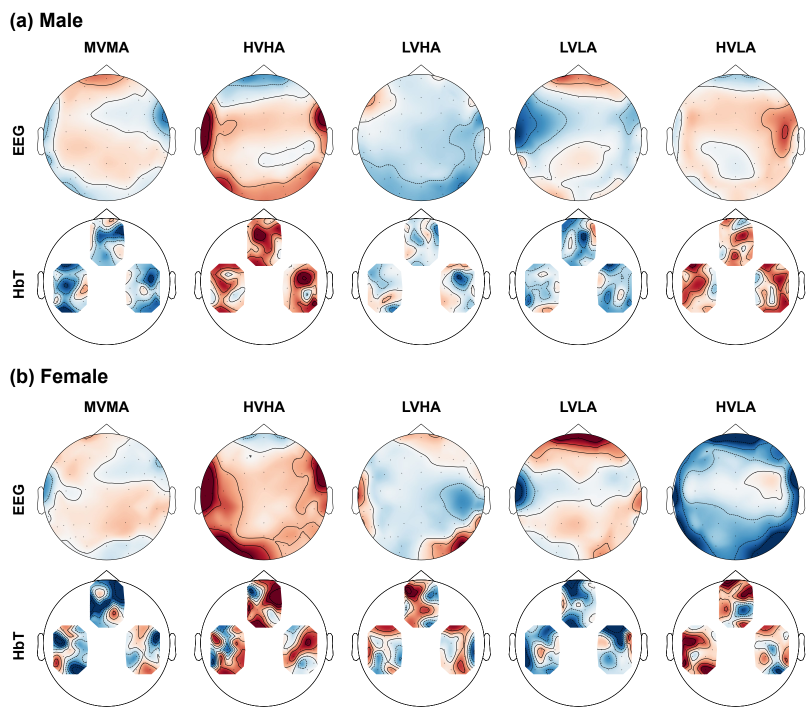

Figure 16: Average activation of EEG and fNIRS in different emotions by gender.