Multimodal Physiological Analysis of Impact of Emotion on Cognitive Control in VR

Ming Li, Junjun Pan*, Yu Li, Yang Gao†, Hong Qin, Yang Shen

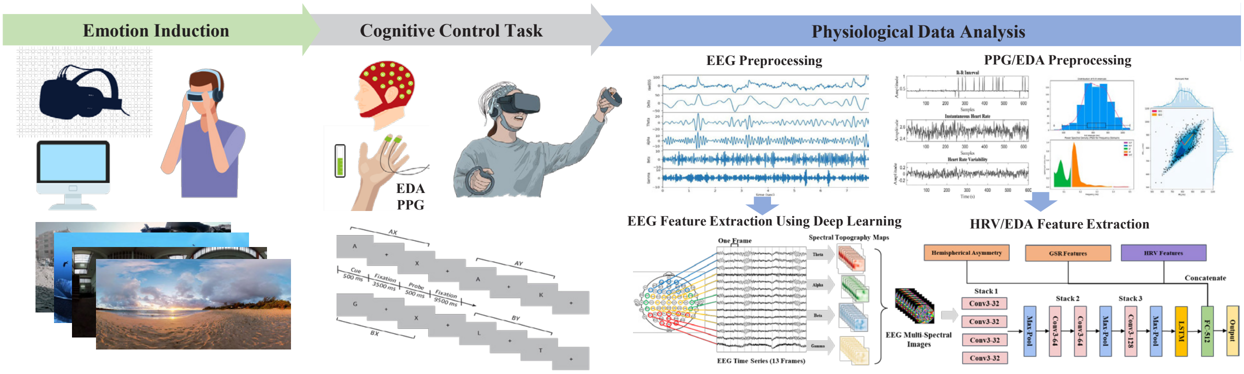

Fig. 1: The research framework: emotion induction materials on the left, the AX-CPT cognitive task in the middle, and the analysis and processing of multimodal physiological data on the right

Abstract—Cognitive control is often perplexing to elucidate and can be easily influenced by emotions. Understanding the individual cognitive control level is crucial for enhancing VR interaction and designing adaptive and self-correcting VR/AR applications. Emotions can reallocate processing resources and influence cognitive control performance. However, current research has primarily emphasized the impact of emotional valence on cognitive control tasks, neglecting emotional arousal. In this study, we comprehensively investigate the influence of emotions on cognitive control based on the arousal-valence model. A total of 26 participants are recruited, inducing emotions through VR videos with high ecological validity and then performing related cognitive control tasks. Leveraging physiological data including EEG, HRV, and EDA, we employ classification techniques such as SVM, KNN, and deep learning to categorize cognitive control levels. The experiment results demonstrate that high-arousal emotions significantly enhance users' cognitive control abilities. Utilizing complementary information among multi-modal physiological signal features, we achieve an accuracy of 84.52% in distinguishing between high and low cognitive control. Additionally, time-frequency analysis results confirm the existence of neural patterns related to cognitive control, contributing to a better understanding of the neural mechanisms underlying cognitive control in VR. Our research indicates that physiological signals measured from both the central and autonomic nervous systems can be employed for cognitive control classification, paving the way for novel approaches to improve VR/AR interactions.

Index Terms—Cognitive control, emotion, physiological signal, deep learning

1 Introduction

Virtual Reality (VR) is increasingly being incorporated into cognitive studies due to its immersive and interactive capabilities, which enhance information retrieval and cognitive abilities within virtual environments [1]. During various VR tasks and challenges, users typically rely on input devices such as controllers or gloves to select objectives, plan actions and continually make decisions. The successful execution of these tasks and decision-making processes requires users to possess strong cognitive control abilities. Cognitive control, also known as executive function, comprises a set of psychological processes that involve goal selection, regulation of thoughts and actions to achieve these goals, and sustained attention. It reflects our ability to regulate and pursue goal-oriented behavior. Accounting for users' cognitive abilities and limited cognitive resources, some VR-based applications dynamically adjust VR environments, games, or tasks based on users' psychological/physiological states and cognitive performance. This personalized approach enhances suitability for users, thereby increasing the immersion and engagement of the VR experience [2].

During the interactive process, the performance of the cognitive control task is influenced not only by the external environment but also by an individual's subjective feelings, particularly emotional states closely tied to cognition. Previous research has shown that emotional interference can greatly disrupt goal-directed behavior, leading to the reallocation of processing resources and affecting the performance of cognitive control tasks [3, 4]. Additionally, some studies suggest that cognitive control could be influenced by heightened arousal levels and that this impact varies among individuals [5]. As depicted in Fig.2, the effect of emotional arousal levels on individual performance generally follows Hebb's U-shaped curve model. In this model, excessively high or low levels of emotional arousal tend to hinder task performance, while a moderate level of emotional arousal stimulates users to perform optimally. In terms of emotional valence, positive emotions generally

- *∗: Corresponding author*
- *†: Corresponding author*
- *Ming Li, Junjun Pan, and Yang Gao are with Beihang University. E-mail: {minglee | pan_junjun | gaoyangvr}@buaa.edu.cn.*
- *Yu Li is with Beijing Normal University. E-mail: yu.li@mail.bnu.edu.cn.*
- *Hong Qin is with Stony Brook University. E-mail: qin@cs.stonybrook.edu.*
- *Yang Shen is with Beijing Normal University. E-mail: shenyang@bnu.edu.cn.*

Manuscript received 4 October 2023; revised 17 January 2024; accepted 24 January 2024. Date of publication 5 March 2024; date of current version 15 April 2024. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2024.3372101, provided by the authors. Digital Object Identifier no. 10.1109/TVCG.2024.3372101

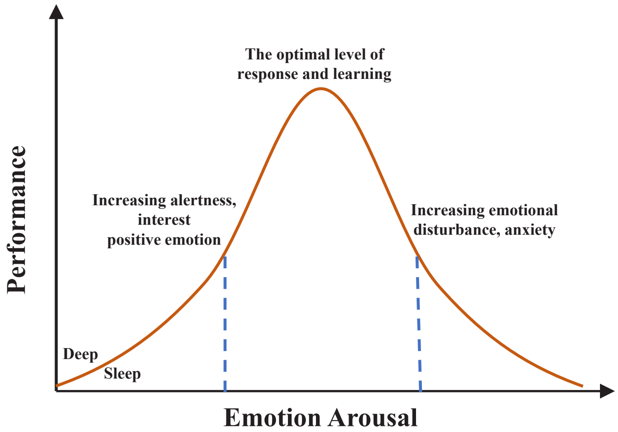

Fig. 2: Schematic illustration of the association between emotion arousal levels and cognitive performance.

enhance user performance, while negative emotions tend to hinder it [6]. However, there are also studies with contrasting conclusions. Mayer et al. argue that individuals have limited working memory capacity, and cognitive load restricts the acquisition and integration of information [7]. Both positive and negative emotions can increase cognitive load and decrease cognitive control levels. Consequently, there is no unified and comprehensive conclusion regarding the effects of arousal/valence emotions on cognitive control.

Hence, understanding the effects of emotions on cognitive control and the underlying physiological mechanisms is crucial. This understanding can help induce specific emotional states in users to enhance cognitive control levels and identify biological markers of cognitive control. In particular, VR has been demonstrated to be an ecologically valid paradigm for studying emotions [8]. Furthermore, due to its ability to induce high-arousal emotions, VR holds a unique advantage in investigating the impact of emotions on cognitive control [9]. However, most current research has primarily focused on examining the impact of emotional valence on cognitive control. As demonstrated in Pekrun's cognitive/motivational model of emotion effects [10] and Linnenbrink-Garcia's research on affect and engagement [11], it is not sufficient to solely distinguish between positive and negative emotional states. It is vital to consider the underlying level of activation or arousal. Therefore, building on Russell's arousal-valence model of emotion, we have embarked on a comprehensive exploration of the relationship between different types of emotions and cognitive control [12].

Based on our literature review, we formulate the following hypotheses: The high emotional arousal induced by VR videos significantly enhances user performance. Excessively high or low emotional arousal diminishes task performance. To verify these hypotheses, we have designed a VR-based emotion induction experiment using 8 VR videos as stimuli for emotion induction [13]. Subsequently, we conducted corresponding cognitive control tasks known as the AX-Continuous Performance Test (AX-CPT) to systematically investigate how emotions affect cognitive control. Additionally, we collected both subjective and objective data from participants to evaluate their cognitive control level. Subjective data included cognitive task performance (test accuracy and reaction time), reflecting an individual's cognitive control level. For objective data, considering the implicit nature of cognitive control, changes are difficult to observe through explicit expressions and behaviors. However, different cognitive control levels trigger a series of neurophysiological reactions in the human body, often accompanied by changes in the Central Nervous System (CNS) and Autonomic Nervous System (ANS) activity [14–17]. Electroencephalography (EEG) signals, Heart Rate Variability (HRV), and Electrodermal Activity (EDA) are effective markers of these two systems' responses, respectively [18]. Thus, we collected three types of physiological signals: EEG, HRV, and EDA. By continuously monitoring users' cognitive control levels using multimodal physiological data, we aim to identify objective and quantifiable biomarkers. The contributions of this study are as follows:

- We explore the relationship between emotions and cognitive control in VR, utilizing the valence-arousal emotion model. Our research indicates that heightened emotional arousal can indeed enhance cognitive control levels. These findings hold significant implications for improving user experiences and capabilities within VR environments, providing invaluable insights for the advancement of human-computer interaction.
- Compared to single-modality, our deep learning networks that utilize multimodal physiological signals show a substantial performance enhancement, attaining an accuracy of 84.52% in recognizing cognitive control levels. Additionally, information drawn from different modalities effectively complements each other.
- We identify stable neural patterns associated with cognitive control. Results from time-frequency analysis uncover higher cognitive control levels are associated with a decrease in α oscillatory energy and an increase in β oscillatory energy. This fundamental evidence aids our comprehension of the neural mechanisms involved in cognitive control and provides valuable guidance for designing adaptive VR/AR systems.

The remainder of the paper is structured as follows: Section 2 reviews the relevant background literature on the relationship between emotions and cognition in VR. Section 3 provides a detailed description of our experiment, including the design of cognitive control tasks and a comprehensive overview of the experimental process. In Section 4, we present and discuss the results of our study. In Section 5, we introduce a classification method based on multimodal physiological data. Section 6 is dedicated to the discussion, and Section 7 provides a summary of this paper.

2 Related Works

We first discuss the recent research on cognitive control in VR. Next, we delve into the influence of emotions on cognitive control and the underlying motivation for this study. Lastly, we provide an overview of the physiological patterns associated with cognitive control.

2.1 Cognitive Control in VR

Cognitive control refers to a set of higher-order cognitive processes and abilities enabling goal-directed and adaptable behavior in various situations. On the other hand, VR provides ecologically valid scenarios where users can exercise these cognitive behaviors and abilities. Unlike traditional laboratory environments, VR presents unique opportunities for experiential active learning. It allows precise control over stimulus levels and task difficulty and possesses the capability of objectively measuring users' cognitive control levels. As such, VR has proven to be not just a suitable, but an exceptional tool for both assessing and enhancing users' cognitive control skills.

M. Clements et al. conducted an insightful study on motion tracking in a VR environment to assess the effects on, and reasons for, improvements in participants' motor skills. Participants were engaged in simulated marksmanship training using a firearm game controller while receiving immediate feedback and scores. The results revealed significant enhancements in shot accuracy, precision, and reaction times during the cognitive training process, suggesting a notable improvement in users' cognitive control levels [19]. In another study, Jacoby et al. compared the effectiveness of a Virtual Reality Mall (VMall) with traditional Occupational Therapy (OT) in improving executive functions in patients with Traumatic Brain Injury (TBI). The study demonstrated that most participants exhibited an improvement in their cognitive control levels after treatment. Specifically, performance on executive function tests in the VR environment was superior, suggesting that VR can lead to better improvements in complex daily activities compared to traditional OT methods [20]. Cao et al. explored the impact of Cinematic Virtual Reality (CVR) on users' cognitive activities. They investigated whether the sense of immersion provided by VR could enhance the cognitive integration of information in CVR. This was done by comparing cognitive activities when watching 3D computer-generated animations in both non-CVR and CVR environments. The research showed that participants exhibited higher early visual attention in the CVR environment [21].

Furthermore, Luong et al. investigated the possible additional impact of wearing a Head-Mounted Display (HMD) on users' cognitive control compared to real-world environments. Participants took part in standardized cognitive control tasks (N-back tasks) with varying difficulty levels, both in virtual and real environments. Data was collected through self-report questionnaires, task performance measures, and behavioral and physiological indicators to evaluate cognitive states. Interestingly, no significant difference was found in task performance between the virtual and real environments [22]. This study highlights that the use of a VR HMD does not inherently impact users' cognitive control levels. However, the underlying reasons for the observed improvements in cognitive control levels within VR, and the neural mechanisms involved, remain less understood. It's important to note that cognitive control abilities are influenced not only by external factors such as the environment and devices but are primarily affected by subjective states, including individual emotions and personal experiences.

2.2 Relationship Between Cognitive Control and Emotion

Cognitive control represents our ability to guide and maintain goal-oriented behavior. Individual emotions can significantly influence various aspects of cognitive control, including attention, memory, and decision-making [23, 24]. It's tempting to view emotions as the adversary of cognitive control. Indeed, in moments of emotional intensity, our capacity to consciously direct our thoughts and actions may diminish. However, emotions can also play a beneficial role. For instance, experiencing anger can serve as a powerful motivator. Some theories propose that emotions and cognitive control are not opposing forces but rather integrated and harmonious components. Emotions can aid in resolving control dilemmas, fostering a shift in the entire system towards a more unified and situationally appropriate control state [25].

D. Parsons et al. conducted a study investigating the impact of different emotional stimuli on executive functions. They devised two conditions: one where participants completed emotional tasks in isolation, and a second where cognitive control and emotional tasks were performed concurrently. To realize this, they developed a virtual reality Stroop task (VRST), where Stroop stimuli were presented within a virtual environment. The experiment sought to assess neurocognitive and psychophysiological responses when executing these tasks in both low-fear (safe) and high-fear (ambush desert) environments. The findings suggested that an increase in cognitive workload was correlated with the more cognitively demanding Stroop condition. Further, it was implied that shifting attention from the environment to the Stroop stimuli may necessitate enhanced inhibitory control [26]. In a separate study, Storbeck et al. explored the influence of emotions on cognitive control levels, specifically spatial working memory capacity. They induced positive, negative, and neutral emotions in participants and subsequently had them engage in a spatial working memory span task to measure their working memory capacity. The study revealed that positive emotions improved working memory capacity, while negative emotions had no impact on performance. Interestingly, when compared to a neutral emotional state, negative emotions did not reduce working memory capacity, suggesting that they do not adversely affect executive control [27].

While these studies provide valuable insights into the impact of emotions on cognitive tasks, there remains considerable debate and controversy over their conclusions: Do all types of emotions either enhance or diminish cognitive control levels? Are the effects fleeting or more persistent? It's noteworthy that these studies predominantly focus on the valence (i.e., positivity or negativity) or arousal (excitement or calmness) of emotions, often overlooking the integration of the two-dimensional emotion model (Russell's Arousal-Valence model) with cognitive control. Consequently, interpreting these findings to elucidate the relationship between emotions and cognitive control can be both challenging and potentially misleading. Therefore, we propose an approach that leverages the Russell arousal-valence model of emotions and exploits the unique advantages of VR in inducing emotions. This would allow for a systematic and comprehensive analysis of the intricate relationship between emotions and cognitive control. To the best of my knowledge, there exists a paucity of research exploring this critical issue within VR environments.

2.3 Physiological Pattern of Cognitive Control

Cognitive control is a multifaceted construct that encompasses various higher-order brain functions, regulated by the central nervous system (CNS). Additionally, the autonomic nervous system (ANS), which governs internal bodily functions such as heart rate (HR), respiration (RESP), blood pressure, and digestion, plays a crucial role. The ANS is divided into the sympathetic nervous system, whose activation may lead to emotional excitement, and the parasympathetic nervous system, which when activated can induce relaxation and relieve anxiety [28]. These activities can directly influence cognitive control levels. Consequently, it's possible to objectively assess cognitive control functions by monitoring the psychophysiological changes in the CNS and ANS [29, 30].

EEG primarily originates from the electrical activity of the cerebral cortex, directly measuring the activity induced by the functioning of the CNS [31]. EEG signals are categorized into different brain regions and five frequency bands: δ (1-4Hz), θ (4-7Hz), α (8-12Hz), β (13-28Hz), and γ (29-50Hz). Notably, α activity is regarded as an effective and reliable indicator for cognitive functions. It mirrors the brain's resting state and is associated with processes such as attentional regulation, mental relaxation, and the inhibition of irrelevant sensory input. In a study by Mahjoory et al., EEG activity during resting state and cognitive control tasks were compared. They found that better performance was linked with increased α oscillations in the right hemisphere's frontal, temporal, and occipital regions [32]. Similarly, θ and β waves have also been found to correspond with cognitive control or executive functions [33–35]. Furthermore, Hinault et al. utilized EEG to investigate the spatiotemporal neural correlates of cognitive control. Participants performed a numerical Stroop task under two conditions: congruent stimuli (larger numbers displayed in larger fonts) and incongruent stimuli (larger numbers displayed in smaller fonts). Their results indicated that when processing incongruent stimuli, participants exhibited higher levels of cognitive control, with activation observed in the frontal and parietal regions of the brain [36].

The most commonly employed measurements for evaluating the activity of the ANS include ECG, Photoplethysmogram (PPG), , and Respiratory Rate. Both ECG and PPG signals can be used to determine heart rate (HR) and accurately mirror changes in inter-beat intervals (IBI) corresponding to heartbeats. This allows for the calculation of trends in HRV indices [37]. For instance, a study by Lee et al. assessed HRV data in individuals with Internet Gaming Disorder (IGD) and healthy controls during gameplay. Their findings indicated that individuals with IGD exhibited reduced high-frequency HRV during real-time gaming sessions, a phenomenon related to diminished goal-oriented cognitive control over the game.

In the realm of VR-based cognitive research, most studies have primarily focused on exploring specific patterns of physiological signals under varying cognitive states or cognitive loads [38]. However, it's important to note that cognitive control is not synonymous with cognitive load or cognitive states. In situations where cognitive load is high, attention may become dispersed, leading to suboptimal cognitive control performance. Moreover, there is a notable scarcity of research investigating the neurophysiological mechanisms underlying cognitive control or executive function performance during VR experiences. To date, no effective biomarkers have been identified. As a response to this gap in the research, we collected multimodal physiological data and cognitive control task performance metrics. These data are crucial for a more comprehensive understanding of cognitive control and emotions.

3 Methods

We have designed a VR emotion induction and cognitive control task to investigate the impact of emotions on cognitive control, specifically focusing on task accuracy and reaction time. Following this, we explore the psychophysiological mechanisms underlying this influence through the analysis of multimodal physiological data. In the following section, we offer a detailed description of the experiment.

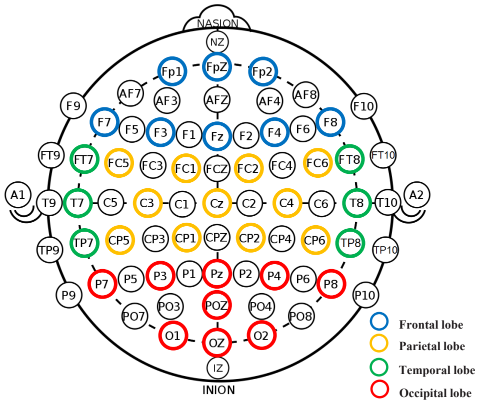

Fig. 3: The EEG cap layout comprises 34 electrodes grouped into 10 regions based on prior knowledge.

3.1 Experiment Apparatus

In this study, we used 1) An HTC Vive Pro HMD for displaying VR videos and AX-CPT tasks, with a resolution of 1200 × 1080 pixels per eye and a screen refresh rate of 90 Hz. We measured and adjusted the inter-pupillary distance for each participant before using VR. 2) An HP laptop computer with an Intel Core i9 processor, GTX 3080T graphics card, and 32 GB of RAM. 3) An ANT Neuro EEG signal acquisition system[1], capable of capturing signals from 128 electrodes on the scalp. The electrode placement follows the 10-20 international system. We only used 34 electrodes in this study. As depicted in Fig. 3, based on prior knowledge of neuroscience, we divided the 34 electrodes into 4 functional areas: the frontal lobe, parietal lobe, temporal lobe, and occipital lobe area. All electrodes were referenced to CPz and grounded to the forehead. Prior to the experiment, the non-abrasive gel was applied to the electrode cap, and electrode impedance should be kept below $5k\Omega$ before and during the experiment. The sampling rate was set to 1000 Hz. 4) A Shimmer3 GSR+ system[2] was used to capture PPG and EDA signals. Similar to ECG, we can also calculate HR and HRV based on PPG. The sensors were placed on the non-dominant wrist of the participant. EDA sensors were placed on the middle phalanges of the index and middle fingers. PPG sensors were placed on the tip of the thumb. All data were recorded at a sampling rate of 200 Hz. The hardware and sensor configurations are shown in Fig. 4.

Fig. 4: Hardware and the sensor setup. The HMD is affixed using lateral straps, ensuring that the upper strap is kept loose to avoid pressure on the central forehead electrode and minimize artifacts.

[1]https://www.ant-neuro.com/
[2]https://shimmersensing.com/

3.2 Participants and Ethics

The experiment was conducted in an indoor environment maintained at a temperature of 26 degrees Celsius. We recruited a total of 26 participants, aged between 21 and 45 (M=31.2, SD=3.16), including 12 females and 14 males. All participants are right-handed and have normal vision. None of the participants has a history of heart disease, neurological disorders, or other cognitive-related illnesses. The study received approval from the local ethics committee.

3.3 Emotion Induction Material and Evaluation

Currently, there is a dearth of VR-based emotion induction materials. Following the criteria used in previous studies, we selected 8 VR videos from Stanford's Immersive VR video datasets, YouTube[3,4], and Steam to induce 4 emotional states of Russell's arousal-valence model: Low Arousal/Low Valence (LALV), Low Arousal/High Valence (LAHV), High Arousal/Low Valence (HALV), and High Arousal/High Valence (HAHV) [39]. The specific video titles, descriptions, sources, and lengths are presented in Tab.1.

To evaluate emotions, we utilized the well-established 9-point Self-Assessment Manikin (SAM) scale prevalent in the field of affective computing. This scale gauges subjects' arousal, valence, and dominance [40]. Valence ranges from negative to positive, arousal varies from calm to excitement, and dominance spans from low control to high control.

3.4 Cognitive Task

In our experiment, we utilized the classic psychology paradigm: AX-CPT. This paradigm is a widely used cognitive control task to examine context processing and goal maintenance. It measures participants' executive control ability without losing attention [41], by asking participants to quickly and accurately identify specific letter combinations within a series of letter stimuli. We developed a VR version of the AX-CPT task in accordance with specific design and development guidelines for cognitive assessment in immersive VR [42]. Specifically, we used the Unity3D game engine for development and designed the interaction between the participants and the virtual scene using the SteamVR SDK. We employed controllers to ensure easy and ergonomic interaction, with the trigger on the Vive controller being used to drive the task.

The task commenced with the appearance of a cross-fixation point at the screen's center, which lasted for 400 ms. Following the disappearance of the fixation point, a cue stimulus (either the uppercase letter A or B) was displayed for 300 ms before disappearing. This was succeeded by a blank delay period of 3,000 ms, after which a probe stimulus (either the uppercase letter X or Y) was presented for 300 ms. Here, "B" represents any letter other than "A", and "Y" represents any letter other than "X"' Participants were required to make a judgment within 1800 ms. If an X followed the letter A, participants were instructed to pull the left trigger (i.e., AX). If a Y appeared immediately after the letter A or if an X or Y followed the letter B (i.e., AY, BX, or BY), they were instructed to press the right trigger. Participants were required to press the trigger quickly and accurately without aiming with the controller. After the trigger response, a blank jitter screen was presented for a random duration of 1,500 ms, 2,500 ms, or 3,500 ms, averaging to 2500 ms. The participants' reaction times were calculated within the period of the blank jitter screen presentation. Thus, the average duration for each trial was approximately 6,500 ms.

In the task, cues A and probe stimuli X were presented with high frequencies. The experiment comprised a total of 100 trials. Specifically, in line with the classic AX-CPT paradigm, it included 70% AX sequences, with AY, BX, and BY sequences each making up 10%. Participants developed a response bias towards A and X. Therefore, in AY, BX, and BY trials, participants needed to overcome the conflict between the response bias (dominant response) and the correct judgment (target). Thus, they might miss the target stimulus or make incorrect responses. Lastly, we recorded the task accuracy of all trials and the

[3]https://www.youtube.com/watch?v=ViLReDIvk_A
[4]https://www.youtube.com/watch?v=C0Rl4m38gOU

Table 1: Comprehensive list of 8 immersive VR clips in the experiment.

ID	Title	Descriptions	Sources	Lengths(s)
1	The Nepal Earthquake Aftermath	Short film on the effects of an earthquake in Nepal Short	Stanford Immersive VR Videos	240
2	War Zone	Journalistic clip of a war-torn city	Stanford Immersive VR Videos	183
3	The Real Run	Will you be able to resist the ruthless nun who haunts the place?	Youtube	174
4	The Conjuring 2	Un Film de James Wan avec Vera Farmiga et Patrick Wilson	Youtube	188
5	The Malaekahana Sunrise	Viewer sees the sun rising over the horizon at a beach	Stanford Immersive VR Videos	120
6	The Mountain Stillness	Atmospheric shots of Canadian snowy mountains	Stanford Immersive VR Videos	128
7	TheBlu: Whale Encounter	View underwater sea life and encounter a whale	Steam	100
8	TheBlu: Reef Migration	View Reef Migration on the ocean floor	Steam	350

Table 2: Extracted Features for Each Modality (Feature Dimension).

Modality	Extracted features
EEG (204)	Power Spectrum of θ , α , β and γ bands for 34 electrodes.The Hemispherical Asymmetry between 17 pairs of symmetrical electrodes in the 4 bands.
HRV (60)	HR, Standard deviation of IBIs, Root mean square of successive RR interval differences, Percentage of differences among adjacent IBIs more than 50ms, Mean of IBIs, 50 spectral power in the bands from [0-5] Hz component of the ECG signal, low frequency [0.01, 0.08] Hz, medium frequency [0.08,0.15] and high frequency [0.15, 0.5] Hz components of HRV spectral power, Poincaré analysis features (2 features). [43].
EDA (31)	Mean skin resistance and mean of derivative, mean differential for negative values only (mean decrease rate during decay time), the proportion of negative derivative samples, number of local minima in EDA signal, average rising time of EDA signal, spectral power in the [0-2.4] Hz band, zero crossing rate of skin conductance slow response (SCSR) [0-0.2] Hz, zero crossing rate of skin conductance very slow response (SCVSR) [0-0.08] Hz, mean SCSR, SCVSR peak magnitude.

average reaction time for each trial. The reaction time refers to the time frame from when participants see the probe stimulus (X or Y) in each trial to when they press the trigger button.

3.5 Experimental Procedures

All participants volunteered to participate in the study. Before the experiment, the experimenter briefed the participants about the purpose, procedures, potential risks of the study, and the significance of the subjective questionnaires. Subsequently, the participants practiced our AX-CPT task to ensure their understanding of the task requirements. Thereafter, the participants signed a written informed consent form, and they were informed that they could discontinue the experiment at any time if they encountered issues during the experiment. Finally, the experimenter fitted the participants with the EEG cap and the Shimmer3 GSR+ device, checked the signal quality, and then initiated the experiment.

The experiment consists of a total of 1 baseline stage and 8 sessions. It began with participants quietly seated in a chair for a 5-minute baseline period. Subsequently, the 8 VR videos were played in random order across 8 sessions, and 2 VR videos that targeted the same emotion were not shown consecutively. Each session consisted of 4 parts. First, they randomly viewed one of 8 VR videos, each with a duration of approximately 3 minutes and 30 seconds. After watching each video, participants completed the SAM questionnaire to assess emotional responses. The questionnaire was displayed on the VR screen to avoid the need for repeated HMD removals during the experiment. Subsequently, the AX-CPT task commenced, consisting of 100 trials and lasting a total of 650 seconds. At last, participants rested for 5 minutes to prevent mental fatigue. They then continued with this cycle 7 more times, completing a total of 8 video viewing and cognitive control tasks. The entire experiment took approximately 2.6 hours to complete.

3.6 Physiological Feature Extraction

We provide a detailed description of the methods used for extracting features from each of the bio-sensing modalities. All extracted features are listed in Tab.2.

3.6.1 EEG Preprocessing and Feature Extraction

Raw EEG data often contain noise such as body movements, VR displays, electrooculography (EOG), ECG, and electromyography (EMG) [44]. We utilized EEGLAB, an open-source toolbox in MATLAB that provides powerful algorithms for EEG preprocessing [45]. Initially, we downsampled the raw signals from 1000 Hz to 256 Hz using the "pop_resample.m" function in EEGLAB and re-referenced to the mastoid electrodes M1 and M2. Subsequently, we applied a band-pass filter ranging from 2 to 47 Hz to the signals using the FIR filter in EEGLAB. The EEG signals were then visually inspected, and segments with amplitudes exceeding ± 100 (μv) were manually removed. Finally, we employed Independent Component Analysis (ICA), a blind source separation (BSS) method, to remove eye-blink artifacts and non-cognitive-related artifacts from the signals. The EEG signals were decomposed into 34 independent components (ICs). Following a visual inspection by experts and using the SASICA plugin in EEGLAB, an average of 5.46 ± 0.84 components were removed per participant [46]. The preprocessed EEG data was then transformed into the frequency domain for feature extraction. For each EEG channel, we used the Fast Fourier Transform (FFT) on the time series for each trial to extract the power spectrum of the signal. We then calculated the sum of squared absolute values within the θ, α, β, and γ bands. Additionally, we computed hemispheric asymmetry features between 17 pairs of symmetrical electrodes in the 4 frequency bands. In the end, for each sample, we obtained 204 power spectrum features (34 channels $\times$ 4 bands and 17 symmetrical channels $\times$ 4 bands), as shown in Tab.2.

3.6.2 HRV Feature Extraction

For each session, we first applied a moving average filter with a window length of 0.25 seconds to filter the noise of the PPG signal. Then, we used a peak detection algorithm to identify the R-peaks [47]. The number of peaks per minute represents the Heart Rate (HR). The Inter-beat Intervals (IBI) were obtained by calculating the time differences between consecutive peaks. Finally, we computed the time-domain and frequency-domain features of the IBI. The names and descriptions of the HRV features are shown in Tab.2.

3.6.3 EDA Feature Extraction

EDA signals typically comprise a tonic component (Skin Conductance Level, SCL) and a phasic component (Skin Conductance Response, SCR). Following the method in [48], we utilized nonnegative deconvolution to break down EDA data. Initially, we estimated tonic SC activity by analyzing inter-impulse data through standard deconvolution of the raw EDA signals. Subsequently, the phasic SC data was derived by subtracting the tonic SC activity from the original SC data, and nonnegative deconvolution was applied to it. We then identified

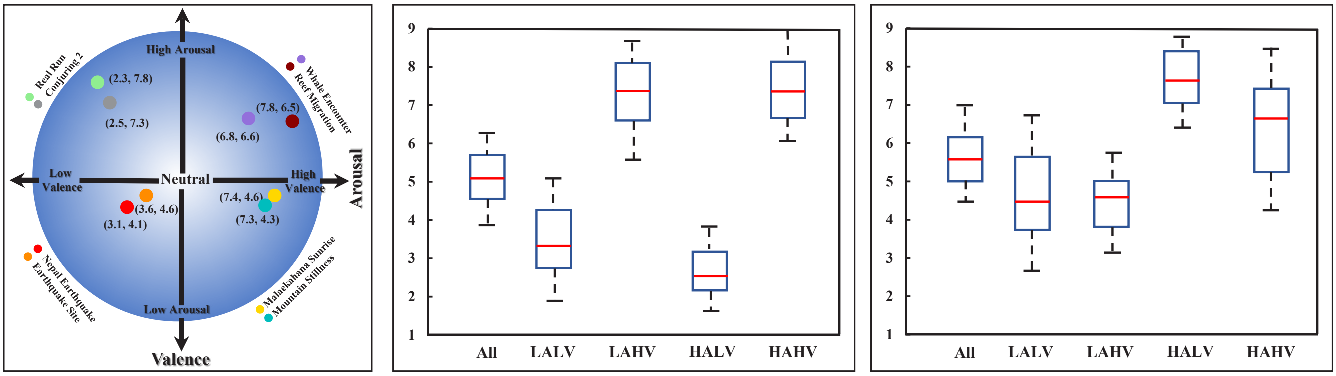

Fig. 5: Average ratings for valence and arousal by all participants under the 4 emotional conditions. The left diagram represents the distribution of the 8 VR videos in the valence-arousal model along with their corresponding scores. The middle and right diagrams represent the distribution of average Valence and Arousal ratings for the 4 emotion induction conditions (LALV, LAHV, HALV, HAHV).

single impulses and corresponding pore opening components through segmentation of the driver and remainder signals. The original SC data was subsequently reconstructed using the tonic and phasic components. Finally, we extracted 31 EDA features, as presented in Tab.2.

3.7 Statistical Analysis

To assess the normality of the data, we employed the Shapiro-Wilk test. Subsequently, if the data exhibited a normal distribution, we proceeded to conduct One-way Repeated Measures ANOVA to investigate any discrepancies in the average reaction time and accuracy of task completion across the 4 emotional states. A significance level of 0.05 was employed for all statistical analyses. Notably, all statistical analyses were performed using SPSS 22.0.

4 Data Analysis

In this section, we have conducted analyses on the self-reported data and cognitive control task scores collected during the experiment.

4.1 Analysis of Self-ratings

As indicated on the left panel of Fig. 5, the average ratings for valence and arousal under the 4 emotional conditions are displayed. The distribution of the 8 videos across the 4 quadrants can be observed, illustrating a wide range of emotional responses. In the middle and right panels, box plots further highlight the significant differences in valence and arousal ratings across these emotional states. To be specific, under LALV condition, the average valence rating is 3.33, paired with an arousal rating of 4.33. Under LAHV, the valence rating increases to an average of 7.36, while the arousal rating is slightly higher at 4.47. For HALV, the valence rating drops to an average of 2.41, with a notably higher arousal rating of 7.58. Finally, under HAHV, the valence rating is relatively high at an average of 7.30, coupled with an arousal rating of 6.54. Across all 4 conditions, the average valence rating stands at 5.09, while the average arousal rating is 5.73.

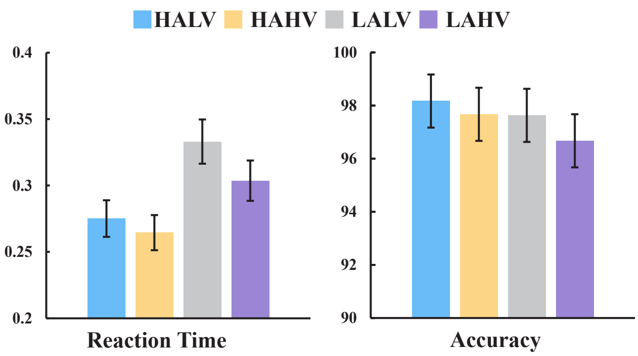

Fig. 6: Average reaction time and accuracy in cognitive control tasks under 4 emotional conditions.

4.2 Analysis of Cognitive Control Task Performance

In our AX-CPT experiment, we used a total of 70 AX stimuli, along with 10 each of AY, BX, and BY stimuli. As outlined in Fig. 6, the ANOVA results suggest that participants, under HAHV condition, exhibit the shortest average reaction time per trial (M = 0.264s, SD = 0.04). This is followed by HALV condition (M = 0.275s, SD = 0.035). Significant differences in average reaction times are noted between these two conditions and LALV condition (M = 0.333s, SD = 0.081), as well as LAHV condition (M = 0.303, SD = 0.0532). Specifically, the differences are highly significant between HAHV and both LALV and LAHV ($p < 0.001$), and also between HALV and both LALV and LAHV ($p < 0.01$ and $p < 0.05$, respectively). Moreover, substantial differences are identified between HA and LA emotions ($p < 0.01$), while no significant differences are found between HV and LV conditions ($p = 0.067$).

Furthermore, for task completion accuracy, no significant differences are found among the 4 emotional states (M = 98.17, SD = 1.33 for HALV; M = 97.67, SD = 1.63 for HAHV; M = 97.63, SD = 1.37 for LALV; M = 96.67, SD = 0.98 for LAHV). Similarly, no significant differences are found between HA and LA, as well as between HV and LV emotions.

5 Classification of Multimodal Physiological Data

In this section, we present the method and classification results based on multimodal physiological signals (EEG, HRV, and EDA). The primary focus of our experiment is to categorize levels of cognitive control into high and low and to identify biomarkers that can accurately reflect these levels. To achieve this, we use the performance results from the cognitive control tasks carried out under 4 different emotional conditions as our ground truth. We regard the multimodal physiological data gathered during tasks after high arousal stimulation as indicative of high cognitive control levels. Conversely, data obtained under low arousal stimulation represent low cognitive control levels. By adopting this approach, we ensure a balanced data representation for both categories.

To demonstrate the feasibility of identifying levels of cognitive control during VR experiences, we systematically evaluate the classification performance of three distinct classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and deep learning-based classification models. In the course of our experiment, we gather EEG, HRV, and EDA data from a total of 26 participants. Each participant engages in 8 experimental sessions, split evenly between high and low cognitive control levels — 4 sessions for each. We execute subject-dependent evaluations, wherein the training and testing data are sourced from different sessions of the same participant. The training dataset comprises data from six sessions (three sessions each from high and low cognitive control levels), while the testing data is derived from the remaining two sessions of the same participant. In the end, we conducted a total of 26 experiments and computed the average classification accuracy to measure the effectiveness of the classifiers.

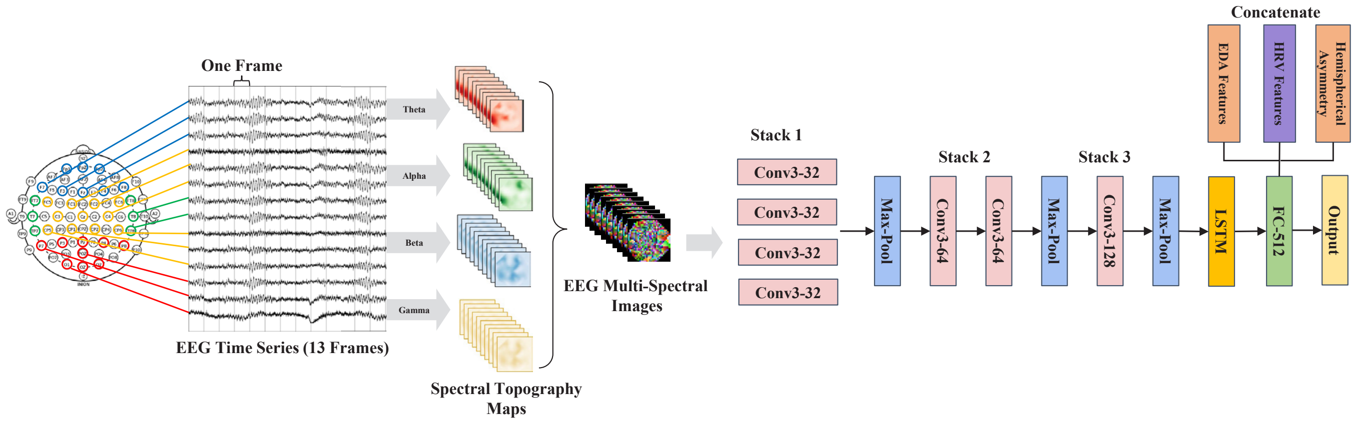

Fig. 7: Our deep learning model: (1)The multi-channel EEG data for each trial is split into 13 frames, each spanning 0.5 seconds; (2) For each frame, calculate the power spectrum in θ, α, β, and γ bands within each channel to generate EEG topographical maps; (3) Merge the topographical maps from different frequency bands and fed into a recurrent-convolutional network to extract features, then concatenate the extracted feature with manually extracted HRV and EDA features.

5.1 Traditional Classifiers

For SVM and KNN, the two classifiers use the features of three modalities of data summarized in Tab.2 as inputs. Before training, all features are rescaled between -1 and 1. KNN is a supervised classification algorithm that organizes feature space into either binary or multi-class clusters. In our study, we set K=5 as the baseline for comparison with other methods. The SVM projects input data into a higher-dimensional feature space using a kernel function, making the data more easily separable than in the original feature space. The SVM classifier is versatile, supporting a range of kernel functions. In our case, we employ the Radial Basis Function (RBF) kernel.

5.2 Baseline Models

We use EEGNet as the baseline and compare it with our DL method [49]. EEGNet is a compact convolutional neural network that utilizes depthwise and separable convolutions to learn interpretable features across various BCI tasks. Moreover, we employ the same implementation details as those mentioned in the original paper.

5.3 Deep-Learning Classifiers

For deep learning networks, we propose a unique method that transforms EEG signals into images. This approach eliminates the need for extensive training data, instead leveraging deep convolution models to extract feature representations from sequences of images.

5.3.1 EEG-images-based Deep-Learning Classifiers

For EEG, in the cognitive control task, each session consists of 100 trials, with each trial lasting for 6.5 seconds. Each trial is divided into 13 frames, each lasting 0.5 seconds, as depicted in Fig.7. For every such window, we transform the EEG signals into a sequence of topology-preserving multispectral images. These images are fed into a recurrent convolutional network, enabling the extraction of robust feature representations from the image sequence.

Specifically, given that EEG electrodes are distributed over the scalp in a 3-dimensional space, we first project them onto a 2-D surface using a topology-preserving Azimuthal Equidistant Projection (AEP), thus preserving the spatial structure between electrodes. Following this, for each frequency band, we apply the Clough-Tocher method to interpolate the power spectrum values across the scalp and estimate values between the electrodes on a 32×32 grid [50]. As a result, 4 topographical activity maps corresponding to the θ, α, β, and γ bands are obtained. These maps are subsequently merged to form a multispectral image. Ultimately, for each 0.5s window, the EEG image size is $13 \times 32 \times 32 \times 4$, and the size of the entire EEG trial is $13 \times 32 \times 32 \times 4$.

Based on these EEG image sequences, otherwise referred to as EEG movies, we employ a recurrent neural network to extract robust feature representations. As depicted in Fig.7, a ConvNet is first utilized to process the spatial and spectral variation in EEG. All ConvNets employ a small 3×3 receptive field, a stride of 1 pixel, and the ReLU activation function, with a padding of 1 pixel applied to preserve spatial resolution. The initial number of kernels is 32, and for deeper stacks of layers, the number of kernels within each convolutional layer is doubled. By stacking multiple convolutional layers in this manner, we achieve more effective receptive fields at higher dimensions while requiring fewer parameters. The window size of max-pooling layers is set to 2×2 with a stride of 2 pixels. Furthermore, for a trial comprising 13 frames, ConvNet parameters are shared across different frames, effectively reducing the number of parameters. Subsequently, the outputs from all ConvNets are reshaped into a continuous sequence of frames, and an LSTM is employed to extract temporal evolution features from the ConvNet sequence. The number of memory cells in the LSTM layer is set to 128. We then concatenate the extracted EEG features with the hemispherical asymmetry features between 17 pairs of symmetrical electrodes, HRV, and EDA features (as mentioned in Section 3.6). Finally, the concatenated features are fed to a fully connected layer and a softmax layer. Thus, we extract spatial, frequency, and temporal features from EEG, deriving robust EEG feature representations.

5.3.2 Implementation Details

In this paper, we have implemented this network framework using PyTorch. Owing to the weight sharing among different frames, which amplifies gradient differences in various layers, we have employed the Adam optimizer to minimize the cross-entropy loss function. We set the learning rate at 0.001, while the coefficients β_1 and β_2, which are used for computing the running averages of the gradient and squares respectively, are set to 0.9 and 0.99. The batch size is configured at 32. To mitigate overfitting, we have applied a dropout probability of 0.5 to all fully connected layers.

5.4 Classification Performance

First, we carry out an independent evaluation of the classification performance of traditional machine learning methods on single-modality data. Then, we assess the results derived from fusing multi-modal features and draw a comparison between the performance of recurrent convolutional networks and traditional methods.

5.4.1 Classification Results of Different Bands

In our endeavor to identify the brain frequency band most strongly associated with cognitive control, we perform a comparative analysis of EEG power spectral features across various frequency bands. As indicated in Tab.3, all bands are the confluence of classification performance across all 4 frequency bands. SVM records an accuracy

Table 3: The Average Accuracy and Standard Deviation of Classification Results for Different Bands.

Classifier	θ	α	β	γ	All Bands
SVM	57.86/9.73	**69.23/8.72**	67.66/11.37	60.94/10.93	73.68/7.70
KNN	56.39/9.89	**69.70/7.98**	65.41/12.37	60.30/10.44	72.28/8.34
DL	68.31/7.33	**79.65/6.64**	75.67/8.21	70.22/8.03	82.68/6.75

rate of 69.23%/8.72% in the α band, while KNN posts 69.70%/7.98%. Yet, DL demonstrates the highest accuracy at 79.65%/7.64%, outclassing KNN by a significant 9.95%, all the while maintaining a lower standard deviation of 1.34%. Moreover, regardless of the frequency band under examination, the DL method consistently delivers superior accuracy and lower standard deviations. These accuracy levels stand at 73.31%/7.33% for the θ band, 75.67%/8.21% for the β band, and 78.22%/8.03% for the γ band, outstripping SVM by substantial margins of 15.45%, 8.01%, and 17.28%, respectively.

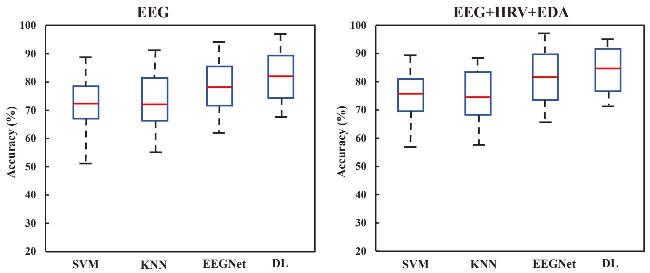

Fig. 8: Box plot of the accuracies with different classifiers. The left figure represents the classification accuracy using EEG data alone, while the right figure represents the classification accuracy using multimodal data.

5.4.2 Classification Results of Different Modalities

The accuracy for each modality is encapsulated in Tab.4. From our analysis, when using EEG, KNN achieves an accuracy rate of 72.28%±8.34% with F1-scores of 66.37%, while SVM notches an accuracy rate of 73.68%±7.70% with F1-scores of 66.53%. EEGNet also achieves an accuracy rate of 78.63%±7.59% with F1-scores of 73.66%. Remarkably, when employing recurrent convolutional networks for EEG classification, the accuracy peaks at 82.68%±6.75% with F1-scores of 78.67%. Furthermore, for HRV, SVM outperforms KNN by achieving a performance score of 71.24%±11.53% against KNN's 69.73%±10.26%, with F1-scores of 70.22% and 60.29%, respectively. Conversely, for EDA, KNN's performance (71.03%±10.92%, F1-scores of 64.88%) overshadows SVM's (70.88%±11.34%, F1-scores of 64.51%).

Additionally, we analyze the classification accuracy of multimodal feature fusion. In this case, the fusion approach for SVM and KNN involves the direct concatenation of three manually extracted modal features into a larger composite feature vector for input. In contrast, the DL method first employs a recurrent convolutional network to extract spatial-temporal-frequency features from EEG data. This is then concatenated with features derived from HRV and EDA, and the combined feature set is fed into a fully connected layer. Our results demonstrate that the classification accuracy using SVM and KNN reaches 76.54%±7.68%, F1-scores of 71.32% and 74.02%±8.21%, F1-scores of 68.40%, respectively, outperforming the results derived from any single modality. For the baseline model, EEGNet achieves 80.95%/±8.14%, with F1-scores of 75.89%, which is also higher than the single modality. Notably, the employment of recurrent convolutional networks peaks the accuracy at 84.52%±6.42%, with the F1-scores of 80.32%. To better illustrate the advantage of DL, the box plots in Fig.8 display the accuracy of the three classifiers. We utilize a one-way ANOVA to establish statistical significance. The results confirm that the classification performance of multimodal fusion significantly surpasses that of a single modality ($p< 0.01$).

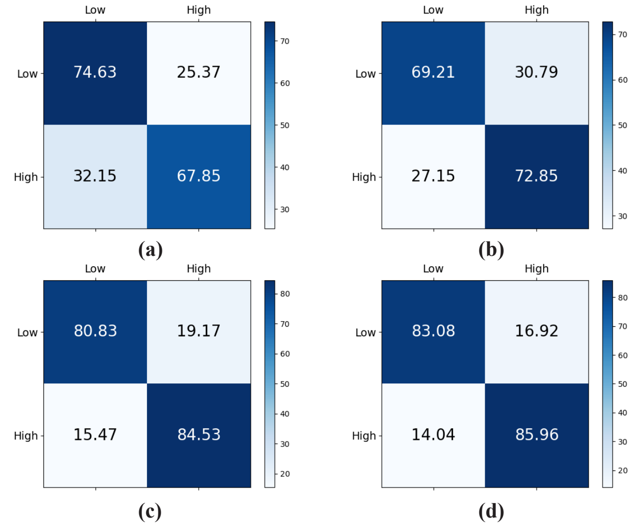

Fig. 9: The confusion matrices for different classifiers using different modalities. The numbers in the figure represent recognition accuracy. (a) SVM(HRV). (b) KNN(EDA). (c) DL(EEG). (d)DL(EEG+HRV+EDA). The content within parentheses indicates the modalities used.

Table 4: The Average Accuracy and F1-Scores of Classification Results for Different Modalities.

Modality	EEG	HRV	EDA	EEG+HRV+EDA
SVM	73.68/66.53	71.24/70.22	70.88/64.51	76.54/71.32
KNN	72.28/66.37	69.73/60.29	71.03/64.88	74.02/68.40
EEGNet	78.63/73.66	—-	—-	80.95/75.89
DL	**82.68/78.67**	—-	—-	**84.52/80.32**

5.4.3 Confusion Matrices

Fig.9 presents the confusion matrices for various classifiers, utilizing both single-modal and multi-modal fusion techniques. The matrices provide detailed insights into the strengths and weaknesses of identifying cognitive control levels using different modalities. In each matrix, every row signifies a target class, while every column signifies the predicted class output by the classifier. The results indicate that for the SVM classifier, the classification accuracy for recognizing low cognitive control levels using HRV data is 74.63%, while for high cognitive control levels, it stands at 67.85%. When using EDA data, the classification accuracy for low cognitive control levels is 69.21%, and for high cognitive control levels, it is 72.85%. Lastly, when we utilize DL methods, the classification accuracy for identifying low cognitive control levels using EEG data reaches 80.83%, and for high cognitive control levels, it reaches a peak of 84.53%. Finally, when we apply DL methods for multi-modal data classification, the classification accuracy remains highest for low cognitive control levels at 83.08%, and for high cognitive control levels, it peaks at 85.96%.

6 Discussion

In this experiment, we have conducted a systematic analysis of the impact of emotions on cognitive control within a VR environment, using the valence-arousal emotion model as our primary framework. Additionally, we have explored the underlying physiological mechanisms influencing this relationship. We have employed multi-modal physiological data to classify levels of cognitive control, identifying the most relevant frequency bands linked to cognitive control. In this section, we present a summary of our findings and juxtapose them with insights gleaned from neuroscience research.

6.1 Effect of Emotion on Cognitive Control

We elicit emotions in participants through VR videos and administer classical psychological cognitive control tasks. As depicted in Fig. 5, the distribution of the 8 videos and the outcomes illustrated in the box plots confirm the successful induction of the desired emotional states. This lays the foundation for exploring the influence of emotions on cognitive control levels.

The findings depicted in Fig. 6 undeniably demonstrate that various emotional states exert an influence on the performance of cognitive control tasks. When compared to a low arousal emotional state, users displayed significantly reduced reaction times during task execution under high arousal emotional states. Moreover, there is no discernible difference in the accuracy of participants' performance in the cognitive control task under both emotional conditions. This implies that participants are capable of effectively carrying out the task without being distracted and maintaining their focus irrespective of the condition.

6.2 The Complementary Characteristics Between Different Modalities.

Tab.4 unequivocally illustrates that the DL method consistently outperforms traditional classifiers, regardless of whether it involves EEG alone or the combined EEG+HRV+EDA. This highlights the effectiveness of deep learning techniques in adeptly extracting robust temporal, spatial, and frequency domain features from EEG data and discarding irrelevant features, ultimately leading to improved classification outcomes. Moreover, when we combine data from all three modalities, both traditional machine learning and DL methods surpass single-modality approaches. The performance of SVM improved from its highest single-modality accuracy of 73.68% to 76.54%, marking a gain of 2.86%. Similarly, KNN shows an enhancement, rising from its highest single-modality accuracy of 72.28% to 74.02%, representing an increase of 1.74%. On the other hand, EEGNet achieves an enhancement from 78.63% to 80.95%, with an increase of 2.32%. The DL method demonstrates a significant boost, escalating from 82.68% to 84.52%, indicating an increase of 1.84%. The outstanding performance of the DL method indicates that the model effectively learned the spatial-temporal-frequency features of EEG signals.

We delve deeper into the reasons behind this performance improvement and scrutinize the confusion matrices in Fig. 9 to reveal the strengths and weaknesses of each modality. It can be observed that high cognitive control levels are more readily identifiable, whereas low cognitive control levels pose a bigger challenge. Specifically, EEG and EDA possess advantages over HRV in classifying high cognitive control levels. The KNN classifier achieves an accuracy of 72.85% in recognizing high cognitive control levels, while EEG reaches 84.53%, and the multimodal approach tops at 85.96%. These accuracies are significantly higher (by 3.64%, 3.70%, and 2.88%, respectively) than the accuracies for recognizing low cognitive control levels. In contrast, HRV struggles with the identification of high cognitive control levels, achieving an accuracy of only 67.85%, which is 6.78% lower than its accuracy in recognizing low cognitive control levels. These findings suggest that the three modalities exhibit differing discriminative capabilities in recognizing cognitive control. The fusion of multimodal features amalgamates complementary information from each modality, effectively enhancing the overall performance. This carries significant implications for identifying individuals' cognitive control levels using objective physiological data in VR applications.

6.3 Neural Patterns of Cognitive Control

We further analyze stable neural patterns associated with cognitive control and plot the power spectrum features in our experiment. As shown in the red box section in Fig. 10, through time-frequency analysis, we can discern that under high cognitive control conditions, there is a reduction in oscillatory energy within the α band, while energy within the β band increases. Conversely, in situations of low cognitive control, the energy within the α band tends to increase. These findings align with previous neuroscience research, which has confirmed that an increase in α band oscillations is typically linked to states of relaxed attention and tranquility (i.e., low cognitive control state) [51].

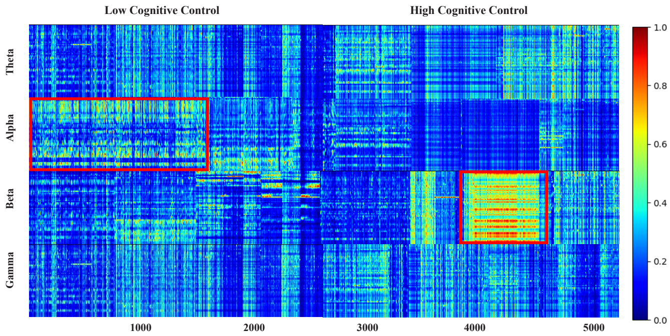

Fig. 10: The power spectrum map of one experiment, the horizontal axis represents time, and the vertical axis represents the power spectrum of 4 frequency bands.

Moreover, VR tasks tend to divert users' attention (increasing external attention). In states of low cognitive control, users' focus tends to be more internally directed, aligning with the observed increase in α band power [52]. The elevation in β band oscillations is associated with heightened attention. Consequently, high cognitive control is associated with an increase in β band power. These observations also account for the differing classification results in Section 5.4.1, where the classification accuracy for the α and β bands stands at 79.65% and 75.67%, respectively, both of which surpass the accuracy observed in the θ and γ bands. Hence, the stable neural patterns in these two bands offer potential neural markers associated with cognitive control. Utilizing these features allows real-time and objective assessment of users' cognitive control levels, enabling dynamic adjustments to the system, which provides a foundation for designing adaptive VR/AR systems [53–56].

7 Conclusion

In this study, we undertake a comprehensive investigation of the influence of emotions on cognitive control levels in VR, grounded in the valence-arousal emotional model. We also employ multimodal physiological signals to identify individuals' cognitive control levels. Our results highlight a significant enhancement in cognitive control levels when users experience high-arousing emotions. Based on the classification results using SVM, KNN, and deep learning, a classification accuracy of 84.52% is achieved in recognizing high and low levels of cognitive control. Furthermore, compared to single modality (EEG: 82.68%; HRV: 71.24%; EDA: 71.03%, the fusion of multimodal features markedly boosts classification accuracy 84.52%. Moreover, the modalities offer complementary information. Notably, results from time-frequency analysis reveal consistent neural patterns related to cognitive control, with higher cognitive control levels associated with diminished α oscillation energy and an increase in β oscillation energy. This foundational evidence enriches our comprehension of the neural mechanisms underpinning cognitive control. It enables us to gain insights into users' cognitive control levels without causing any disruption, facilitating dynamic adjustments of tasks within VR/AR applications for a more harmonious HCI.

Acknowledgments

This research is supported by the National Key R&D Program of China (No.2022ZD0115902, 2023YFC3604500), the National Natural Science Foundation of China (No.62307003), and the Beijing Science and Technology Plan Project (No.Z221100007722001, Z231100005923039).

References

[1] Tiffany Luong, Anatole Lecuyer, Nicolas Martin, and Ferran Argelaguet. A survey on affective and cognitive VR. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5154–5171, 2022. 1

[2] Hamdi Ben Abdessalem and Claude Frasson. Real-time brain assessment for adaptive virtual reality game: A neurofeedback approach. In *Brain*

Function Assessment in Learning: First International Conference, BFAL 2017, Patras, Greece, September 24-25, 2017, Proceedings 1, pages 133–143, 2017. 1

[3] Jennifer A Johnson and Robert J Zatorre. Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex*, 15(10):1609, 2005. 1

[4] Dolcos and F. Brain systems mediating cognitive interference by emotional distraction. *Journal of Neuroscience*, 26(7):2072–2079, 2006. 1

[5] Jan L Plass and Slava Kalyuga. Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, 31:339–359, 2019. 1

[6] Elisa Straub, Andrea Kiesel, and David Dignath. Cognitive control of emotional distraction–valence-specific or general? *Cognition and Emotion*, 34(4):807–821, 2020. 2

[7] Richard E Mayer. *The Cambridge handbook of multimedia learning*. Cambridge university press, 2005. 2

[8] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. Virtual reality for emotion elicitation – A Review. *IEEE Transactions on Affective Computing*, 14(4):2626–2645, 2023. 2

[9] Feng Tian, Minlei Hua, Wenrui Zhang, Yingjie Li, and Xiaoli Yang. Emotional arousal in 2D versus 3D virtual reality environments. *PloS one*, 16(9):e0256211, 2021. 2

[10] Reinhard Pekrun. The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied psychology*, 41(4):359–376, 1992. 2

[11] Lisa Linnenbrink-Garcia and Reinhard Pekrun. Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, 36(1):1–3, 2011. 2

[12] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980. 2

[13] Benjamin J Li, Jeremy N Bailenson, Adam Pines, Walter J Greenleaf, and Leanne M Williams. A public database of immersive vr videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures. *Frontiers in psychology*, 8:2116, 2017. 2

[14] Lynda Gerry, Barrett Ens, Adam Drogemuller, Bruce Thomas, and Mark Billinghurst. Levity: A virtual reality system that responds to cognitive load. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2018. 2

[15] Nicolette Stogios, Alexander Gdanski, Philip Gerretsen, Araba F Chintoh, Ariel Graff-Guerrero, Tarek K Rajji, Gary Remington, Margaret K Hahn, and Sri Mahavir Agarwal. Autonomic nervous system dysfunction in schizophrenia: impact on cognitive and metabolic health. *npj Schizophrenia*, 7(1):22, 2021. 2

[16] Erika Johannessen, Adam Szulewski, Nada Radulovic, Matthew White, Heather Braund, Dan Howes, Dirk Rodenburg, and Claire Davies. Psychophysiologic measures of cognitive load in physician team leaders during trauma resuscitation. *Computers in Human Behavior*, 111:106393, 2020. 2

[17] Megan K MacPherson, Defne Abur, and Cara E Stepp. Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load. *Journal of Voice*, 31(4):504–e1, 2017. 2

[18] Juan Sztajzel et al. Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss medical weekly*, 134(35-36):514–522, 2004. 2

[19] Jillian M Clements, Regis Kopper, David J Zielinski, Hrishikesh Rao, Marc A Sommer, Elayna Kirsch, Boyla O Mainsah, Leslie M Collins, and Lawrence G Appelbaum. Neurophysiology of visual-motor learning during a simulated marksmanship task in immersive virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 451–458, 2018. 2

[20] Michele Jacoby, Sara Averbuch, Yaron Sacher, Noomi Katz, Patrice L. Weiss, and Rachel Kizony. Effectiveness of executive functions training within a virtual supermarket for adults with traumatic brain injury: A pilot study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2):182–190, 2013. 2

[21] Ruochen Cao, Lena Zou-Williams, Andrew Cunningham, James Walsh, Mark Kohler, and Bruce H Thomas. Comparing the neuro-physiological effects of cinematic virtual reality with 2D monitors. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 729–738, 2021. 3

[22] Tiffany Luong, Nicolas Martin, Ferran Argelaguet, and Anatole Lécuyer. Studying the mental effort in virtual versus real environments. In *2019 IEEE Conference on virtual reality and 3D user interfaces (VR)*, pages 809–816, 2019. 3

[23] Gerald L Clore and Jeffrey R Huntsinger. How emotions inform judgment and regulate thought. *Trends in cognitive sciences*, 11(9):393–399, 2007. 3

[24] George Loewenstein, Jennifer S Lerner, et al. The role of affect in decision making. *Handbook of affective science*, 619(642):3, 2003. 3

[25] Jeremy R Gray. Integration of emotion and cognitive control. *Current directions in psychological science*, 13(2):46–48, 2004. 3

[26] Thomas D. Parsons and Christopher G. Courtney. Interactions between threat and executive control in a virtual reality stroop task. *IEEE Transactions on Affective Computing*, 9(1):66–75, 2018. 3

[27] Justin Storbeck and Raeya Maswood. Happiness increases verbal and spatial working memory capacity where sadness does not: Emotion, working memory and executive control. *Cogn Emot*, 30(5):925–938, 2016. 3

[28] Laurie Kelly McCorry. Physiology of the autonomic nervous system. *American journal of pharmaceutical education*, 71(4), 2007. 3

[29] Sepideh Sadaghiani and Andreas Kleinschmidt. Brain networks and α-oscillations: structural and functional foundations of cognitive control. *Trends in cognitive sciences*, 20(11):805–817, 2016. 3

[30] Stephen H. Fairclough. Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145, 2009. 3

[31] Ronglong Xiong, Fanmeng Kong, Xuehong Yang, Guangyuan Liu, and Wanhui Wen. Pattern recognition of cognitive load using EEG and ECG signals. *Sensors*, 20(18):5122, 2020. 3

[32] Keyvan Mahjoory, Elena Cesnaite, Friederike U Hohlefeld, Arno Villringer, and Vadim V Nikulin. Power and temporal dynamics of alpha oscillations at rest differentiate cognitive performance involving sustained and phasic cognitive control. *Neuroimage*, 188:135–144, 2019. 3

[33] Stephen H Fairclough, Louise Venables, and Andrew Tattersall. The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56(2):171–184, 2005. 3

[34] Frederic M Stoll, Charles RE Wilson, Maïlys CM Faraut, Julien Vezoli, Kenneth Knoblauch, and Emmanuel Procyk. The effects of cognitive control and time on frontal beta oscillations. *Cerebral cortex*, 26(4):1715–1732, 2016. 3

[35] Grace M Clements, Daniel C Bowie, Mate Gyurkovics, Kathy A Low, Monica Fabiani, and Gabriele Gratton. Spontaneous alpha and theta oscillations are related to complementary aspects of cognitive control in younger and older adults. *Frontiers in human neuroscience*, 15:621620, 2021. 3

[36] Thomas Hinault, Kevin Larcher, Natalja Zazubovits, Jean Gotman, and Alain Dagher. Spatio–temporal patterns of cognitive control revealed with simultaneous electroencephalography and functional magnetic resonance imaging. *Human Brain Mapping*, 40(1):80–97, 2019. 3

[37] 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, 2012. 3

[38] Pavlo Antonenko, Fred Paas, Roland Grabner, and Tamara Van Gog. Using electroencephalography to measure cognitive load. *Educational psychology review*, 22:425–438, 2010. 3

[39] Ming Li, Junjun Pan, Yang Gao, Yang Shen, Fang Luo, Ju Dai, Aimin Hao, and Hong Qin. Neurophysiological and subjective analysis of VR emotion induction paradigm. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3832–3842, 2022. 4

[40] 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. 4

[41] Corentin Gonthier, Brooke N. Macnamara, Michael Chow, Andrew R. A. Conway, and Todd S. Braver. Inducing proactive and reactive control shifts in the AX-CPT. *Frontiers in Psychology*, 7:1822, 2018. 4

[42] Panagiotis Kourtesis, Danai Korre, Simona Collina, Leonidas AA Doumas, and Sarah E MacPherson. Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of virtual reality everyday assessment lab (VR-EAL), a neuropsychological test battery in immersive virtual reality. *Frontiers in Computer Science*, 1:12, 2020. 4

[43] Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083, 2008. 5

[44] S P Fitzgibbon, D MW Powers, K J Pope, and C R Clark. Removal of eeg noise and artifact using blind source separation. *Journal of Clinical Neurophysiology*, 24(3):232–243, 2007. 5

[45] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004. 5

[46] M. Chaumon, Dvm Bishop, and N. A. Busch. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J Neurosci Methods*, 250:47–63, 2015. 5

[47] Fred Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5:258, 2017. 5

[48] Mathias Benedek and Christian Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *psychophysiology*, 47(4):647–658, 2010. 5

[49] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018. 7

[50] Peter Alfeld. A trivariate clough—tocher scheme for tetrahedral data. *Computer Aided Geometric Design*, 1(2):169–181, 1984. 7

[51] Grace M Clements, Mate Gyurkovics, Kathy A Low, Arthur F Kramer, Diane M Beck, Monica Fabiani, and Gabriele Gratton. Dynamics of alpha suppression index both modality specific and general attention processes. *NeuroImage*, 270:119956, 2023. 9

[52] Lisa-Marie Vortmann and Felix Putze. Exploration of person-independent bcis for internal and external attention-detection in augmented reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–27, 2021. 9

[53] Francesco Chiossi, Changkun Ou, and Sven Mayer. Exploring physiological correlates of visual complexity adaptation: Insights from eda, ecg, and eeg data for adaptation evaluation in VR adaptive systems. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2023. 9

[54] Sawon Pratiher, Karuna P Sahoo, Mrinal Acharya, Ananth Radhakrishnan, Sazedul Alam, Scott E Kerick, Nilanjan Banerjee, Nirmalya Ghosh, and Amit Patra. Affective physiological state analysis and interpretable predictive modeling of marksmanship in Go/NoGo VR shootingdifficulty task. *UMBC Student Collection*, 2022. 9

[55] Francesco Chiossi, Johannes Zagermann, Jakob Karolus, Nils Rodrigues, Priscilla Balestrucci, Daniel Weiskopf, Benedikt Ehinger, Tiare Feuchtner, Harald Reiterer, Lewis L Chuang, et al. Adapting visualizations and interfaces to the user. *it-Information Technology*, 64(4-5):133–143, 2022. 9

[56] John E Muñoz, Luis Quintero, Chad L Stephens, and Alan Pope. Taxonomy of physiologically adaptive systems and design framework. In *International Conference on Human-Computer Interaction*, pages 559–576, 2021. 9