

Stress Severity Detection in College Students Using Emotional Pulse Signals and Deep Learning

Mi Li , Junzhe Li , Yanbo Chen , and Bin Hu , *Fellow, IEEE*

Abstract—College students face increasing stress from difficulties with studies, employment, and social interactions, which, if left unaddressed, may lead to depression and physical illnesses. Currently, the detection of stress severity relies on self-assessment scales, while machine learning or deep learning-based approaches primarily focus on classification. This study proposes an approach using pulse signals containing emotional cues and deep learning to automatically detect the severity of stress in college students. First, pulse signals of 177 college students were collected using photoplethysmography (PPG) during they watched five virtual reality (VR) emotional scenes, including calm, sadness, happiness, fear, and tension. Pulse rate variability (PRV) and discrete PPG (dPPG) were extracted from these signals as input for detecting stress severity. Then, the proposed stress detection framework, 1DCNN-BiLSTM + Cross-Attention + XGBoost, was employed to detect stress severity, incorporating an emotional Cross-Attention mechanism. The impact of induced emotions on stress severity detection performance was examined. The results indicated that stress severity detection in emotional scenes outperformed in calm. Furthermore, the detection performance that integrates multiple emotions surpassed single emotions. The fusion of PRV and dPPG signals yielded the best detection performance. This study provides an end-to-end automated approach for detecting stress severity in college students.

Index Terms—Stress severity (SS), emotion, pulse rate variability (PRV), discrete PPG (dPPG), cross-attention (CA).

Received 25 April 2024; revised 16 January 2025; accepted 24 February 2025. Date of publication 4 March 2025; date of current version 15 September 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFA0706200, in part by the National Natural Science Foundation of China under Grant 61602017, Grant 61632014, and Grant 61627808, and in part by the National Basic Research Programme of China under Grant 2014CB744600. Recommended for acceptance by T. Zhang. (*Corresponding author: Bin Hu*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Ethics Committee at Beijing Anding Hospital, Capital Medical University, Beijing, China under Application No. 2019054.

Mi Li is with the School of Information Science and Technology, Beijing University of Technology, Beijing 100021, China, also with the Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing 100124, China, also with the Engineering Research Center of Intelligent Perception and Autonomous Control, Ministry of Education, Beijing 100124, China, and also with the Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China (e-mail: limi@bjut.edu.cn).

Junzhe Li and Yanbo Chen are with the School of Information Science and Technology, Beijing University of Technology, Beijing 100021, China (e-mail: lijunzhe@emails.bjut.edu.cn; yanbochen@emails.bjut.edu.cn).

Bin Hu is with the Institute of Engineering Medicine, Beijing Institute of Technology, Beijing 100811, China, and also with the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China (e-mail: bh@lzu.edu.cn).

Digital Object Identifier 10.1109/TAFFC.2025.3547753

I. INTRODUCTION

THE mental health of college students is primarily characterized by high stress, and poor psychological endurance and adaptability [1], and they are more prone to mental health issues [2], [3], [4]. Additionally, they may also lead to insomnia, a weakened immune system, and other health issues [5].

Currently, stress detection, including among college students, primarily relies on self-reporting through health questionnaires or self-rating scales [6], [7]. Although it serves as a useful tool, it is undeniable that it has certain limitations [8]. Accordingly, there has been increasing interest in research on stress detection methods grounded in deep learning techniques in an effort to enhance detection accuracy in recent years. However, these studies have primarily focused on stress classification methods [9], [10], with a lack of research on methods for detecting the severity of stress.

The detection of stress severity is crucial because severe stress can evoke intense fear even in situations that are not threatening, potentially leading to psychological conditions like post-traumatic stress disorder [11]. Stress can also impair reward-related brain regions, including the nucleus accumbens, leading to decreased reward preference and social avoidance, resulting in emotional disorders such as decreased happiness and, in severe cases, depression [12]. Therefore, identifying and intervening early with individuals who have high levels of stress severity, including college students, can facilitate timely psychological counseling and intervention, effectively preventing further development into serious mental disorders. However, to date, there has been a scarcity of machine learning-based methods capable of automatically detecting stress severity, especially among college students.

Previous research has shown that chronic stress can result in symptoms such as excessive stress, mood swings, and other related issues [13], [14], [15], whereas severe stress can precipitate serious emotional disorders, including clinical depression [11], [12]. This suggests a close association between stress and emotions. However, current research on stress detection methods has not considered the influence of emotions. It is clear that incorporating emotional signals could significantly enhance stress detection performance.

In stress detection research, heart rate variability (HRV), extracted from electrocardiogram (ECG) data, and pulse rate variability (PRV), derived from photoplethysmography (PPG) signals, are the most frequently employed physiological signals [16], [17], [18]. PPG signals are particularly sensitive to

changes in stress levels [19], [20]. Nevertheless, the PRV signal extracted from PPG only reflects the interval between adjacent two peaks, losing the peak signal. Therefore, this study also extracts the peak signal from PPG, referred to as the discrete PPG (dPPG) signal, to enhance the recognition of stress severity.

The main work and contributions of this study include:

- 1) To improve stress detection performance, this study proposes a hybrid framework for stress level detection that combines deep learning and ensemble learning: 1DCNN-BiLSTM + Cross-Attention + XGBoost. Multiple deep learning methods are utilized for automatic feature extraction, while ensemble learning (XGBoost) is used for the inference and evaluation of stress levels.
- 2) This study, for the first time, examines the impact of induced emotional states on stress severity detection, and finds that stress severity detection in emotional scenarios outperforms that in calm conditions, with performance being especially better in negative emotional scenes. This indicates that stress can lead to negative emotions. Furthermore, the study has found that fusing multiple emotional states can effectively improve detection performance, achieving higher accuracy than any single emotion.
- 3) A pulse signal dataset with multiple emotional cues (ePulse) has been constructed, filling the gap in current datasets for stress assessment research lacking emotional cues.

II. RELATED WORK

At present, stress detection methods primarily include self-rating scale questionnaires [21] and biochemical marker detection [22], [23]. Self-rating scales are the most common method, although they have been proven to be a valuable tool for stress detection through extensive experimental research and clinical practice [6], [7], [8], it is clear that they have certain limitations. Biochemical marker detection, which involves measuring cortisol levels in saliva, blood, and urine [22], [23], necessitates specialized equipment and conditions.

This section primarily emphasizes the investigation of machine learning (ML), deep learning (DL), and stress detection methods for college students.

A. Stress Detection Methods Based on Machine Learning

In the context of machine learning-based stress detection methods, signals such as electrocardiography (ECG), electroencephalography (EEG), galvanic skin response (GSR), and photoplethysmography (PPG) are commonly collected, and machine learning algorithms, including k-nearest neighbors (KNN), linear discriminant analysis (LDA), and support vector machine (SVM), are subsequently used for stress classification.

For example, early studies such as those conducted by Melillo et al. [24] involved nonlinear analysis of HRV using features such as approximate entropy and correlation dimension, with LDA being employed for stress detection. Bong et al. [25] employed ECG signals to extract features such as heart rate, average R-peak amplitude, and average interval, for stress identification

using KNN and SVM. Cinaz et al. [26] extracted HRV signals from ECG and applied LDA, KNN, and SVM for three-class stress classification (low, medium, high).

In recent years, for instance, Xia et al. [27] proposed an early stress detection method, which involved employing the montreal imaging stress task (MIST) to induce participants, simultaneously collecting EEG and ECG signals, manually extracting features, and utilizing SVM for stress classification detection. Arsalan et al. [28] proposed a stress recognition approach in public speaking contexts, where they collected PPG, EEG, and GSR from 40 participants during rest and public speaking activities, manually extracting features and employing SVM for binary stress classification. Jiao et al. [29] utilized features extracted from PPG signals, such as PRV in the time and frequency domains, for binary stress classification using SVM.

Apart from the traditional machine learning methods mentioned above, alternative architectures such as Broad Learning System (BLS), as an emerging machine learning method [30], [31], can also be applied to stress detection. The core idea of BLS is to achieve more efficient model training and prediction via incremental learning and feature extension, which combines existing features with new learning capabilities. Gong et al. [31] applied BLS to process EEG and eye movement signals, ultimately improving the model's accuracy in emotion recognition. This suggests that alternative architectures such as BLS have potential application value in stress detection research.

B. Stress Detection Methods Based on Deep Learning

Deep learning has demonstrated state-of-the-art capabilities in areas like computer vision and natural language processing [32], [33], [34]. Recently, it has also attracted significant interest in the realm of stress detection. He et al. [35] utilized HRV signals extracted from ECG for stress binary classification, and the results indicated that compared to traditional ML methods such as LDA and SVM, CNN exhibited higher accuracy in stress classification. Li et al. [36] collected ECG and GSR signals using both chest-worn and wrist-worn sensors, and they employed one-dimensional convolutional neural networks (1DCNN) and multilayer perceptron (MLP) to achieve higher accuracy in the binary classification of psychological stress levels compared to traditional ML methods.

However, CNNs can only extract local spatial features from signals and are unable to capture temporal information between features. Long-Short-Term Memory (LSTM) networks can address this limitation by capturing the temporal relationships between these spatial features. Therefore, some researchers use LSTM to recognize stress. For example, Xia et al. [37] utilized the sliding window technology to divide the raw EEG signal into multiple time slices and then extracted intra-slice and inter-slice features for input into a multibranch LSTM for stress recognition (low, medium, and high). Zhang et al. [38] constructed a BiLSTM network with self-attention and cross-attention mechanisms to recognize stress and non-stress states using blood volume pulse (BVP) and electrodermal activity (EDA) signals. Moreover, some researchers have proposed the

CNN-LSTM model to enhance stress detection performance. Rastgo et al. [9] introduced a multimodal fusion approach employing a CNN-LSTM network to identify the presence or absence of stress in drivers using physiological signals including ECG, EDA, and GSR. Mou et al. [39] used a CNN-LSTM model, which assigns different weights to the features from different modalities (eye, vehicle, and environmental data) for feature fusion and conducts three-class stress classification (low, medium, and high). Tanwar et al. [40] also applied an attention mechanism to assign different weights to features of different modalities (ECG and EDA), and constructed a CNN-LSTM model to detect stress versus non-stress states.

In addition, Graph Convolutional Networks (GCNs) have been widely applied in fields such as social network analysis [41], recommendation systems [42], image recognition [43], natural language processing [34], and speech recognition [44]. Moreover, GCNs related to stress detection have also been extensively studied in the field of affective computing [45]. A study has also used GCNs and HRV signals for stress detection, achieving promising results in psychological stress binary classification tasks [46].

The MuSe Stress 2022 challenge has significantly contributed to the advancement of stress-detection research using deep learning. It mainly focuses on stress detection through the Valence and Arousal dimensions [47], [48]. Christ et al. [48], the organizer of this challenge, used facial expression features and an LSTM-RNN model for stress detection, thus establishing the baseline for the competition. Park et al. [49] employed Transformer and LSTM models with the audio, video, and text features provided by the challenge and achieved remarkable results. Li et al. [50] took audio, video, physiological, and text data as inputs to a Gated Recurrent Unit with self-attention, exceeding the baseline results provided by the MuSe organizers. Similarly, Liu et al. [51] proposed an LSTM-CNN model with self-attention for stress detection, which was based on audio, video, and text features. Also in 2022, Park et al. [52] developed a Transformer model with only an encoder and a linear layer to detect stress using audio, video, and text features. This model performed better than both the LSTM and the original Transformer models. Finally, He et al. [53] introduced an attention mechanism that could capture both single-modal and multimodal features and integrated it into a TCN-Transformer model. By using multimodal inputs (including audio, video, physiological signals, and text), this model achieved excellent performance in stress detection.

C. Stress Detection Methods for College Students

There are numerous statistics-based stress surveys and analyses conducted for college students. Barbayannis et al. [54] conducted a study to examine the correlation between academic stress and psychological well-being among college students during the COVID-19 pandemic using self-assessment questionnaires. These findings revealed a statistically significant association between perceived academic stress and mental well-being, showing a negative correlation between higher stress levels and lower levels of happiness. The study highlighted the increased

susceptibility of college students to stress compared to other demographic groups. Kenchannavar et al. [55] collected data from 486 engineering students by using self-assessment questionnaires to assess stress levels. Using statistical one-way analysis of variance (ANOVA), they found that 7.81% reported low stress levels, 63.58% experienced moderate stress, and 24.27% faced significant stress, emphasizing the varying degrees of stress experienced within this cohort. Asif et al. [56] conducted a research to evaluate the occurrence of stress, anxiety, and depression among 500 students at Sialkot University in Pakistan, using the DASS-21 questionnaire. The statistical analysis revealed prevalence rates of 75%, 88.4%, and 84.4% for stress, anxiety, and depression, respectively, among college students. Zhan et al. [57] surveyed 1,586 Chinese college students to evaluate their mental health status during the COVID-19 pandemic. Analyzing the responses from 1,586 participants through online distribution of self-assessment questionnaires measuring stress, depression, and anxiety, their statistical analysis indicated that 67.50% of students experienced moderate to severe stress, with depression detection rates reaching 43.77%, and 20.60% reporting symptoms of anxiety. This underscores the widespread high levels of stress among college students during the pandemic. These findings highlight the pressing requirement for tailored interventions and support networks to tackle the mental health issues confronting college students, particularly amidst global crises such as the COVID-19 pandemic.

ML and DL methods have also been utilized for stress detection among college students. Zhang et al. [58] collected ECG data from 16 college students to detect stress using SVM. Liu et al. [59] assessed stress levels in college students using EEG data collected from 90 students, and used radial basis function neural networks and improved extreme learning machines for stress classification. Mane et al. [60] and Hafeez et al. [61] analyzed alpha, beta, and theta signals in EEG data and used CNN and LSTM for stress detection. Tian et al. [62] investigated psychological stress recognition using deep learning techniques with ECG data collected from 60 college students.

From the above stress detection based on ML and DL, the main focus is on the classification, while there is a lack of research on regression methods for the severity of pressure. Additionally, stress is also an emotional disorder [11], [12], however, so far, no research has been conducted on the impact of emotions on stress detection based on ML and DL. Moreover, the PRV signal reflects the interval signal between two adjacent peaks of the PPG signal, which has been proven effective for stress detection [25]. However, the peak signal of the PPG signal has been overlooked.

This study proposes the use of the dPPG signal, which represents the peak signals of the Pulse signal for stress detection. Furthermore, research has established a strong link between stress and emotions [63], [64], where stress can result in emotional disturbances and even developing into depression [11], [12]. Therefore, the objective of this research is to examine the impact of various induced emotional signals on the detection of stress severity and to integrate multiple induced emotional signals to improve detection performance.

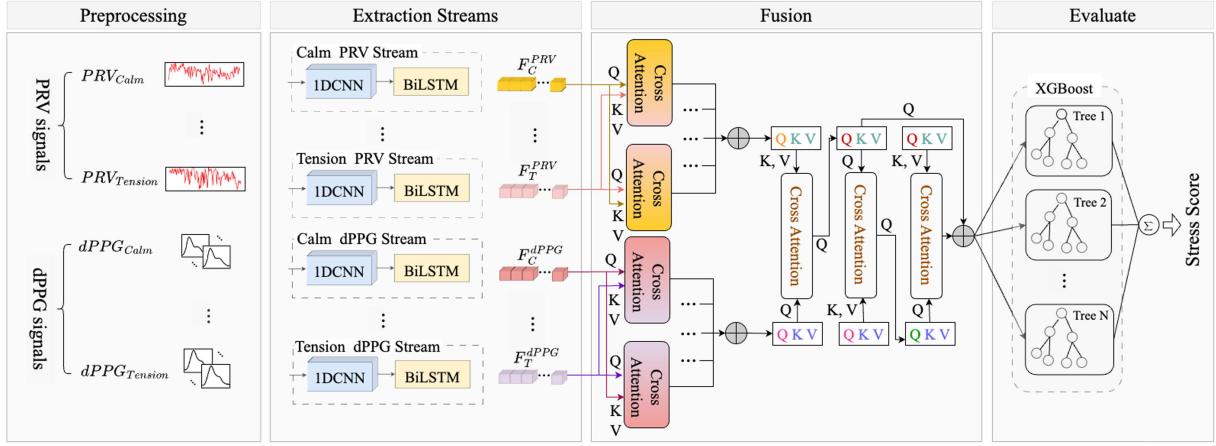


Fig. 1. The overall framework for detecting the severity of stress among college students. In this framework, PRV and dPPG signals are processed as inputs to generate stress level scores for college students. Five induced emotions from PRV and dPPG (calm, sadness, happiness, fear, and tension) are fed into their respective spatio-temporal feature streams (1DCNN + BiLSTM) to extract spatio-temporal features. Each emotional feature is paired with another, and their importance is encoded through cross-attention. Then, the features of PRV and dPPG are fused using multimodal cross-attention. Finally, XGBoost is employed for stress level detection.

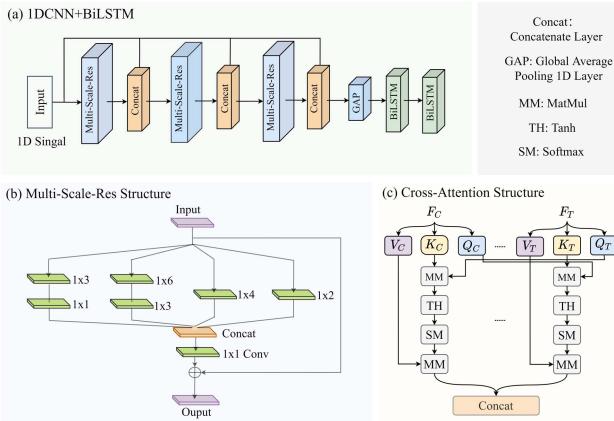


Fig. 2. Components of the frame structure in Fig. 1. (a) Structure of the 1DCNN+BiLSTM stream (DB); (b) Multi-scale-res structure of the 1DCNN module; (c) Cross-Attention module structure (CA).

III. PROPOSED DETECTION METHODS

In this section, we put forward a comprehensive framework for evaluating the stress severity among college students: 1DCNN-BiLSTM + Cross-Attention + XGBoost, which takes PRV and dPPG as input signals, as depicted in Fig. 1.

A. 1DCNN-BiLSTM

As illustrated in Fig. 1, five independent streams are employed to extract emotional features, each composed of 1DCNN and BiLSTM. Their structure is presented in Fig. 2(a). Here, 1DCNN encodes both local and global modes and extracts multi-level and multi-scale spatial feature representations from the input signal. This module consists of multi-scale convolution and residual connection (Multi-Scale-Res), as shown in Fig. 2(b). The BiLSTM module encodes the temporal dependencies and context within the features.

B. Cross-Attention

In this article, the cross-attention comprises two parts. The first part is emotional cross-attention, which is utilized to calculate the significance of different emotional features; The second part is multimodal cross-attention, used to fuse PRV and dPPG modal features.

1) Emotional cross-attention: The emotional cross-attention module assigns importance to features from various emotional streams, and calculates attention weights to assess the significance of each emotion within the features, as shown in Fig. 2(c).

For example, we have two emotional features F_C and F_T , and form their query and key-value pairs.

Then, the similarity score between Q_C and K_T is calculated using matrix multiplication.

$$S_{C\&T} = Q \cdot K^T = Q_C \cdot K_T^T \quad (1)$$

Next, the hyperbolic tangent function is used for nonlinear activation, and softmax is applied for normalization.

$$W_{C\&T} = SM(TH(S_{C\&T})) \quad (2)$$

Finally, the value vector V_T is weighted and summed using the emotion attention weight $W_{C\&T}$ to obtain the final output vector.

$$O_{C\&T} = W_{C\&T} \cdot V_T \quad (3)$$

2) Multimodal cross-attention This study uses multimodal cross-attention to fuse the features of PRV and dPPG. First, a bidirectional interaction is established between different modal features, ensuring that information can flow not only from one modal to another but also in reverse. Then, cross-attention is applied again to ensure that the final feature representation contains sufficient multimodal information.

TABLE I
STRESS SCALE SCORE

Stress Score	Severity
<= 14	Not have
15 - 18	Mild
19 - 25	Moderate
26 - 33	Moderately
>= 34	Severe

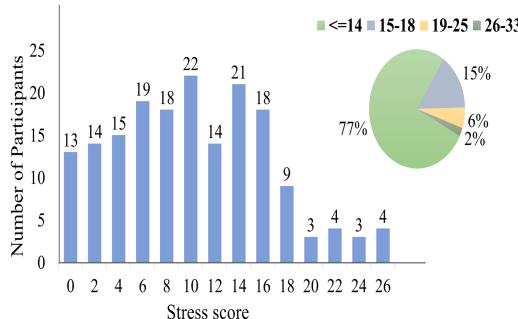


Fig. 3. Distribution of stress scores among 177 participants.

C. XGBoost

XGBoost is an ensemble learning algorithm that achieves accurate predictions by integrating the prediction results of multiple weak decision tree models. In this study, the ensemble learning regression model XGBoost is used to decode emotional features and infer the severity of stress.

IV. DATA

A. Participants and Self-Rating Scales

Our study recruited 177 undergraduate volunteers from the Beijing University of Technology. Participants' stress scores were obtained using the DASS-21 self-rating scale, which has the advantage of high reliability and validity, allowing for a rapid assessment of stress levels [65], [66], [67]. The scale comprises three self-reported components: Stress, Anxiety, and Depression, each subscale comprising 7 items. The correlation between stress scores assessed with the DASS-21 and stress levels is illustrated in Table I. Fig. 3 illustrates the distribution of stress scores among the 177 college students. The gender ratio of participants is 89:88 (89 males and 88 females), with an average age of 20.37 ± 2.97 years and an average education level of 14.50 ± 1.49 years.

B. Ways for Inducing Emotions

Emotion induction ways play a crucial role in emotion research [68], [69]. In this study, we employed virtual reality (VR) technology to construct five VR scenes designed to elicit emotions of calm, sadness, happiness, fear, and tension. Each scene lasted for 90 seconds, and simultaneously inducing physiological and psychological responses in individuals through visual and auditory stimuli. In comparison to emotional pictures and two-dimensional videos, VR scenes offer a greater

TABLE II
EVALUATION OF EMOTIONAL SCENE AROUSAL (30 PARTICIPANTS) DURING DESIGN PHASE

	Very High	High	Moderate	Low	None
Calm	9	21			
Sadness	19	8	3		
Happiness	22	6	2		
Fear	18	6	6		
Tension	12	9	7	2	

sense of immersion, thereby enhancing individuals' emotional experiences [70], [71].

Sadness and happiness represent negative and positive emotional experiences, respectively, and are closely related to mental health [68], [69], [72], [73], [74]. Fear is considered as a manifestation of physiological and psychological stress [75], while tension is common in stressful situations, especially when facing challenges or threats [76], [77]. Calm, as a low-intensity emotional state, is generally regarded as an important indicator of mental health [78], and we use it as the emotional baseline.

Sadness Scene: This scene shows a girl crying in front of a dark cemetery, with warm memories engraved on the tombstone. The surrounding environment is gloomy and spacious, creating a heavy and sorrowful atmosphere. In this scene, participants are guided to feel emotions of loss and grief.

Happiness Scene: This scene is set in a joyful park, where participants can see colorful fireworks blooming in the night sky and children playing and laughing around. The park is shaded by green trees, with flowers in bloom and a pleasant atmosphere pervading the air, aiming to arouse participants' happiness and excitement.

Fear Scene: This scenario simulates an urgent threat, in which participants face a seemingly attacking character. The environment is dim, with compact music and intimidating sounds in the background, enhancing the sense of fear. In this situation, participants experience strong feelings of anxiety and fear.

Tension Scene: The scene is set on the edge of a cliff, with participants standing high and overlooking the abyss. Ahead, there is a person walking on a tightrope between two mountains, swaying and shaking, accompanied by wind and tense background music in the scene. This situation makes participants feel highly tense.

Calm Scene: This scene is set in a quiet classroom and hallway, with neatly-arranged tables and chairs in the classroom and a quiet environment. In this scene, participants feel peaceful and tranquil, being able to relax both physically and mentally.

In the design of VR emotional scenes, a 5-point scale was adopted to evaluate the level of emotional arousal. The arousal levels were rated as follows: 5-Very High, 4-High, 3-Moderate, 2-Low, and 1-None. A total of 30 participants took part in this evaluation. The results are shown in Table II. This evaluation helped us verify the rationality of emotional scene design, ensuring that the scenes could arouse the expected emotional responses.

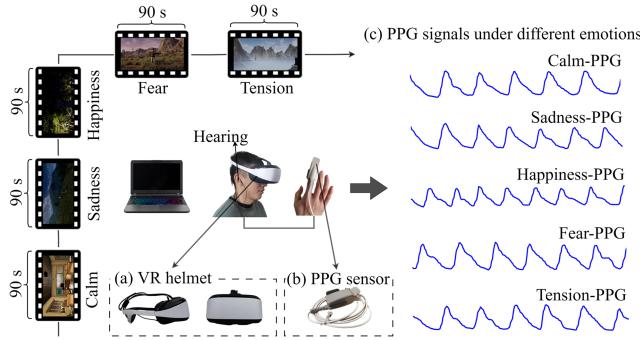


Fig. 4. Data collection process. Participants wore VR helmets and headphones, with the PPG sensor affixed to their right index finger. They were required to view five VR emotional scenes, including scenes of calm, sadness, happiness, fear, and tension, without engaging in any additional interactive tasks. Each VR scene lasted for 90 seconds, leading to a total data collection time of 7 minutes and 30 seconds. The PPG sensor captured participants' fingertip PPG signals in real-time at a sampling frequency of 45 Hz. The PPG signals presented in the figure are from Participant 001. They are simply shown to demonstrate the PPG signals under different emotional states, without any specific selection criterion. These signals were extracted within a period (approximately 30 ms) shortly after the signal collection started.

C. Data Collection

We developed an automated system for acquiring PPG signals, incorporating a high-performance laptop, virtual reality (VR) helmet, PPG sensor, and specialized data acquisition software, as depicted in Fig. 4.

First, participants completed the DASS-21 self-rating questionnaire to obtain stress scores through self-reporting.

Next, participants wore VR helmets and headphones, with the PPG sensor affixed to their right index finger. During the data collection phase, participants needed only to view five VR emotional scenes, including scenes of calm, sadness, happiness, fear, and tension, without engaging in any additional interactive tasks. The study designed nine pseudo-random scene sequences in total. Each sequence ensures that there is no fixed order of arrangement among different emotional scenes, such as CSHFT, SHFTC, HFTCS, FTCSH, THSFC, SCFTH, FCTHS, HTCSF, and TSCHF. Here, C stands for Calm, S for Sadness, H for Happiness, F for Fear, and T for Tension. During data collection, to control potential order effects, participants were randomly assigned one of these scene sequences. Participants viewed each VR scene for 90 seconds, leading to a total data collection time of 7 minutes and 30 seconds. The PPG sensor captured participants' fingertip PPG signals in real-time, sampling at frequency of 45 Hz. Since no rest periods were incorporated between VR scenes, we ignored the emotional response data in the first 5 seconds after each scene began to address emotional overflow.

Finally, the pulse dataset containing emotional cues (ePulse) was constructed, consisting of 5 types of raw induced emotional PPG signals (calm, sadness, happiness, fear, and tension) from 177 college students. Each individual entry included stress scale scores, depression scale scores, and anxiety scale scores as labels, along with basic information such as ID, gender, and age.

TABLE III
EVALUATION OF EMOTIONAL SCENE AROUSAL (177 PARTICIPANTS) AFTER DATA COLLECTION

	Very High	High	Moderate	Low	None
Calm	86	91			
Sadness	87	72	13	5	
Happiness	98	76	3		
Fear	122	44	8	3	
Tension	62	70	18	27	

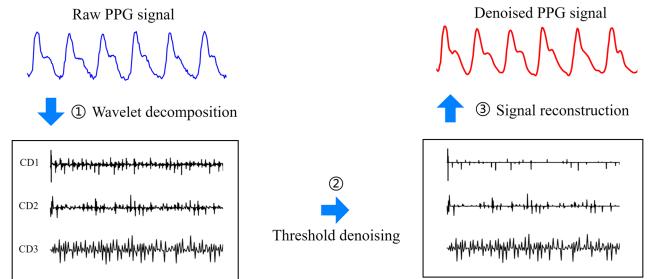


Fig. 5. Wavelet denoising for PPG signal. The raw PPG signal (in blue) is decomposed into three high-frequency coefficients, CD1, CD2, and CD3, through wavelet analysis. Using soft threshold functions to process CD1, CD2, and CD3, achieving wavelet denoising. After denoising, reconstruct the three-layer waveform to obtain the denoised PPG signal (in red).

After collecting the data, we conducted a post-survey among 177 participants, using the same 5-point scale to evaluate their levels of emotional arousal. The results are shown in Table III.

The survey results show that although there are individual differences in emotional experiences, the emotional scenes were generally highly consistent with the emotional labels.

V. DATA PREPROCESSING

A. Denoising of PPG Signals

Utilizing the wavelet threshold denoising technique [79], noise and artifacts are removed from PPG signals, a process that involves wavelet decomposition, threshold denoising, and subsequent signal reconstruction. Fig. 5 illustrates the contrasting results of PPG signal denoising before and after the application of the technique.

B. Extraction of PRV Signals

PRV refers to the periodic variations between consecutive pulses [80], [81]. The extraction process, as shown in Fig. 6(b), utilizes a threshold peak detection algorithm [20] from the PPG signal, as depicted in Fig. 6(a).

In general, the length of a single pulse wave cycle is approximately 0.75 s [82], [83]. Initially, a window containing $N = 0.75 \times fs$ points is set, encompassing the temporal sequence S of N points, represented as $S = [S_1, S_2, \dots, S_N]$. Subsequently, a binary search is applied, starting from $S_{N/2}$ as the initial point of the algorithm to locate the maximum point in the sequence. The window moves from the start to the end of the PPG signal at an interval of one pulse cycle, and the maximum is searched within it. If no maximum is found, the

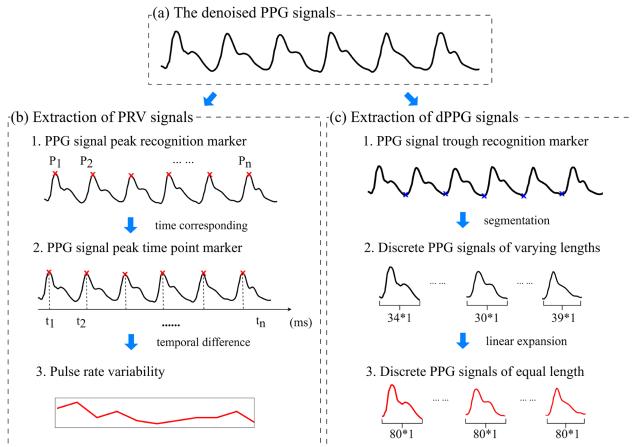


Fig. 6. The PRV and dPPG signals were extracted utilizing both the threshold peak detection algorithm [20] and the trough valley detection algorithm [84]. (a) Denoised PPG signals. (b) Extraction of PRV signals. Peaks are marked with red X, and the corresponding time points are determined based on these peaks, followed by differentiation to obtain the PRV signals. (c) Extraction of dPPG signals. Troughs are identified with blue X, and the PPG signal is accordingly segmented, followed by linear interpolation to obtain dPPG signals of equal length.

window shifts to the right by one point to form a new window $S = [S_2, S_3, \dots, S_{N+1}]$, with $S_{N/2+1}$ as the new starting point. The search is repeated until all peak points of the entire PPG signal are located.

All main wave peaks are marked (see Fig. 6(b-1)), $[P_1, P_2, \dots, P_n]$, where $n \in N$. The time points corresponding to these peak points are obtained by calculation (see Fig. 6(b-2)), $[t_1, t_2, \dots, t_n]$, where $n \in N$. The time points are differentially processed to obtain the time differences between peaks $t_n - t_{n-1}$, where $n \in N$, thereby obtaining the PRV signal, as illustrated in Fig. 6(b-3).

C. Extraction of dPPG Signals

The process for extracting the dPPG signal is shown in Fig. 6(c). We use a valley detection algorithm [84] to locate the positions of the valleys in the PPG signal, which are situated at the lowest points of each PPG (see Fig. 6(c-1)). Each pair of valley points corresponds to one pulse wave cycle. The PPG signal is segmented into several dPPG signals based on the positions of the valley points (see Fig. 6(c-2)). The detailed process is as follows:

- 1) Locate the valleys, $F(x)$, in the one-dimensional PPG signal according to (4), and record their positions, x , of each valley, storing them in the set D .

$$\begin{aligned} F(x) &< F(x-1) \text{ and } F(x) \geq F(x+1) \\ D &= \{x_i\}, i \in N \end{aligned} \quad (4)$$

- 2) Traverse the set D , and denote the discrete pulse signal as f_i . Then, the individual discrete pulse wave can be obtained by (5).

$$f_i = F(D[i]:D[i+1]) \quad (5)$$

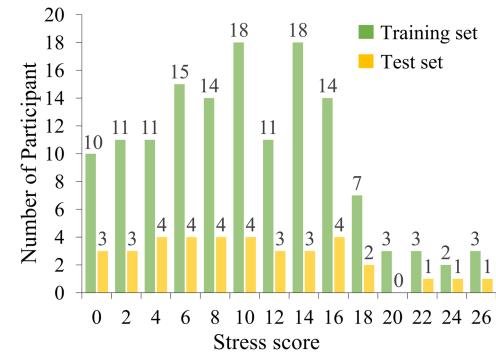


Fig. 7. Distribution of stress scores among 140 participants in the training set and 37 in the test set.

- 3) Since deep convolutional networks require consistent data input sizes, but the number of sample points in dPPG signals varies widely, resulting in varying lengths of several dPPGs, as shown in Fig. 6(c-2), further steps involve using linear interpolation to extend any discrepant-length dPPG signals to the same length. This process results in obtaining the dPPG signal, as shown in Fig. 6(c-3).

VI. EXPERIMENT

The study involved 177 college students, with each participant providing samples for five induced emotions: calm, sadness, happiness, fear, and tension. To ensure the independence of the training and test sets, the data were randomly split into a training set of 140 individuals and a test set of 37 individuals. To ensure consistent sample proportions for each category in the training and test sets, we adopted the method of group random sampling (stress score > 18 and stress score ≤ 18). The distribution of stress labels in the training and test sets is shown in Fig. 7. Although we try our best to make the data distribution in the training and test sets as consistent as possible, the partitioning of the validation set during model training is random, which can't guarantee the consistency of data distribution among each validation set, nor between the validation set and test sets. To address this issue, we utilized the XGBoost regression model, which has advantages in handling missing values and imbalanced samples, to minimize these effects and enhance the model's generalization performance and robustness. Model training employed 5-fold cross-validation. The test set was then used to assess the models from the 5-fold cross-validation, and the results were reported as mean values with standard deviations (mean \pm sd).

The experiments were carried out using Python 3.7 and the deep learning framework Keras to construct the regression detection model. These training and test processes were performed on a system running Windows 11, which was equipped with a GTX 3060 14 GB graphics card to facilitate efficient computation.

The mean squared error (MSE) was employed as the loss function during the model training process. Evaluation metrics for the test results included the mean absolute error (MAE) and the root mean square error (RMSE), as defined in (6) and (7),

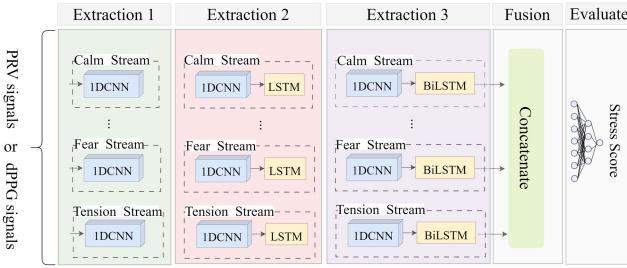


Fig. 8. The ablation experiment for the 1DCNN + BiLSTM model, including three conditions: 1DCNN, 1DCNN + LSTM, and 1DCNN + BiLSTM.

TABLE IV
COMPARISON OF ABLATION RESULTS FOR STRESS DETECTION BASED ON DIFFERENT FEATURE EXTRACTION MODULES (MEAN \pm SD)

Feature extraction module	PRV		dPPG	
	MAE	RMSE	MAE	RMSE
1DCNN	5.13 \pm 0.08	6.16 \pm 0.09	5.49 \pm 0.10	6.61 \pm 0.18
1DCNN + LSTM	5.09 \pm 0.16	6.02 \pm 0.15	5.44 \pm 0.07	6.46 \pm 0.09
1DCNN + BiLSTM	4.96 \pm 0.09	5.92 \pm 0.07	5.29 \pm 0.04	6.32 \pm 0.10

respectively.

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (7)$$

where, y_i is the true label, \hat{y}_i is the predicted value.

VII. RESULTS AND DISCUSSION

According to the stress detection framework in Fig. 1, the experiments in this section include four parts: 1) An ablation experiment for feature encoding module ‘1DCNN-BiLSTM’; 2) An emotional ablation experiment; 3) A cross-attention mechanism for stress detection experiment; 4) An ablation experiment for XGBoost algorithm.

A. Ablation Experiment for 1DCNN-BiLSTM

As shown in Fig. 8, the ablation experiments of the 1DCNN-BiLSTM model were conducted on three modules: 1DCNN (Extraction 1), 1DCNN + LSTM (Extraction 2), and 1DCNN + BiLSTM (Extraction 3). The corresponding model training curves are displayed in Fig. 9.

The experimental results (see Table IV) indicate that while the multi-scale 1DCNN is capable of extracting both local and global spatial pattern features from the signal, the generation of emotion is inherently time-dependent. Therefore, the detection performance is enhanced by using either LSTM or BiLSTM, as both models can encode the temporal dependencies within the time series signal. Notably, BiLSTM, which can further encode temporal emotional context dependencies, yields the best performance.

Although the error of BiLSTM on the validation set is higher than that on the test set, statistical tests show that this difference

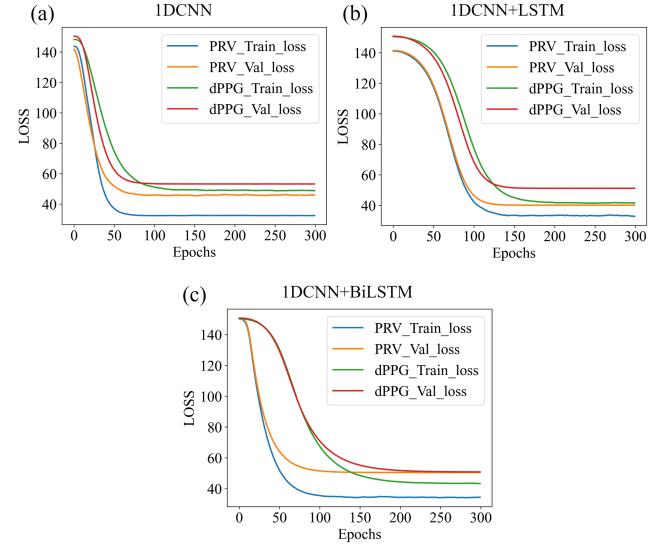


Fig. 9. Training loss curves for the fusion of five induced emotions based on PRV or dPPG signals. (a) 1DCNN. (b) 1DCNN + LSTM. (c) 1DCNN + BiLSTM.

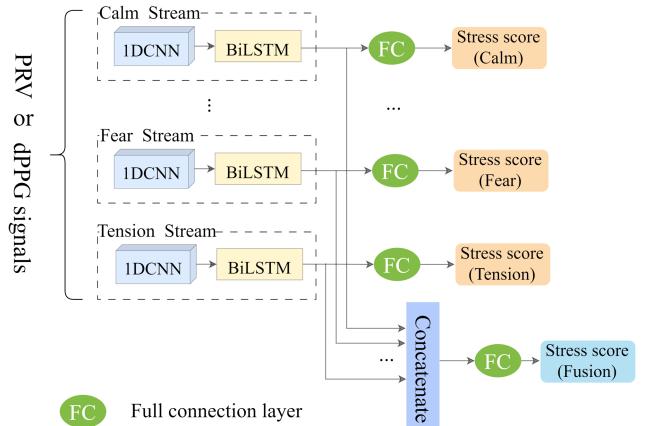


Fig. 10. Emotion ablation experiments for PRV and dPPG, and fusion experiments of PRV and dPPG for various induced emotions.

is not statistically significant ($P > 0.05$). Furthermore, the statistical test results of Table IV are presented in Table V. For RMSE, whether it is the PRV signal or the PPG signal, the errors of the 1DCNN + BiLSTM and 1DCNN + LSTM models are smaller than those of 1DCNN. Additionally, the errors of the 1DCNN + BiLSTM model are smaller than those of 1DCNN + LSTM, both of which are statistically significant ($P < 0.05$). Taking all factors into account, we have selected the BiLSTM model. The advantage of BiLSTM is that it can capture the correlation information in time series from two directions (forward and backward), which is helpful for improving the learning of sequence features and knowledge representation.

B. Ablation Experiment for Emotion

The experimental methods are shown in Fig. 10, and this section presents four experiments focusing on stress detection:

TABLE V
STATISTICAL T-TEST IN TABLE IV FOR MAE/RMSE (P -VALUE)

	Feature extraction module	1DCNN	1DCNN + LSTM	1DCNN + BiLSTM
PRV	1DCNN	-	0.079 ⁿ /0.046 ^a	0.048 ^a /0.026 ^a
	1DCNN + LSTM	-	-	0.058 ^c /0.049 ^a
	1DCNN + BiLSTM	-	-	-
dPPG	1DCNN	-	0.158 ⁿ /0.037 ^a	0.008 ^b /0.018 ^a
	1DCNN + LSTM	-	-	0.039 ^a /0.046 ^a
	1DCNN + BiLSTM	-	-	-

where, “a” and “b” indicate statistical significance (a, $P < 0.05$; b, $P < 0.01$), “c” indicates marginal significance (c, $0.05 < P < 0.059$), and “n” indicates no statistical significance ($P > 0.05$).

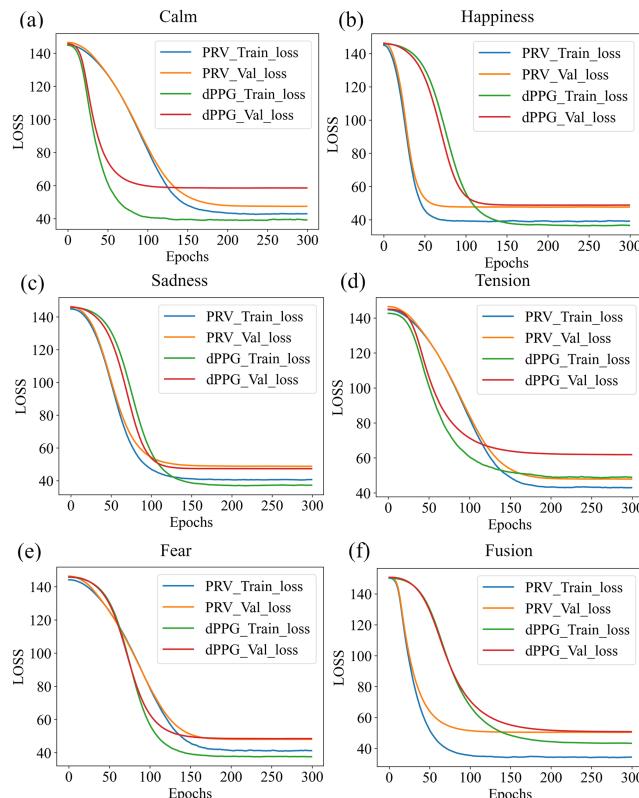


Fig. 11. Training loss curves for induced emotions including calm, happiness, sadness, tension, and fear, as well as emotion fusion based on PRV or dPPG signals. (a) Calm. (b) Happiness. (c) Sadness. (d) Tension. (e) Fear. (f) Fusion of 5 emotions for PRV and dPPG signals.

- 1) The impact of different induced emotional signals from PRV;
- 2) The impact of different induced emotional signals from dPPG;
- 3) The influence of fusing multiple induced emotional signals from PRV; and 4) The influence of fusing multiple induced emotional signals from dPPG. Fig. 11 shows the training loss curves, and Table VI provides the corresponding results for the test set. The statistical test results for Table VI are shown in Table VII.

It can be seen that, 1) whether it is the PRV signal or the PPG signal, the detection error of all emotions (MAE or RMSE, or both) is smaller than that of calm ($P < 0.05$), suggesting that stress is a form of emotional disorder; 2) the detection error of all negative emotions (RMSE) is smaller than that for happiness (P

TABLE VI
COMPARISON OF STRESS DETECTION OF EMOTIONAL ABLATION AND FUSION (MEAN \pm SD)

Emotion	PRV		dPPG	
	MAE	RMSE	MAE	RMSE
Calm	5.29 \pm 0.21	6.27 \pm 0.23	5.66 \pm 0.20	6.78 \pm 0.24
Happiness	5.27 \pm 0.11	6.36 \pm 0.15	5.57 \pm 0.12	7.07 \pm 0.09
Sadness	5.16 \pm 0.21	6.24 \pm 0.26	5.53 \pm 0.27	6.67 \pm 0.35
Tension	5.09 \pm 0.13	6.26 \pm 0.17	5.45 \pm 0.28	6.86 \pm 0.30
Fear	4.99 \pm 0.05	6.17 \pm 0.14	5.32 \pm 0.15	6.38 \pm 0.12
Fusion	4.96 \pm 0.09	5.92 \pm 0.07	5.29 \pm 0.04	6.32 \pm 0.10

< 0.05), implying that measurement under negative emotions is more accurate; 3) the fusion error of multiple emotions (RMSE) is smaller than that of a single emotion ($P < 0.05$), except for the borderline significance between the PRV signal and the dPPG signal ($P < 0.059$), indicating that stress impacts a range of emotions.

C. Stress Severity Detection Based on Cross-Attention

Fig. 12 illustrates the experimental method, which includes three experiments: 1) Cross-attention fusion of PRV signals for various emotions; 2) Cross-attention fusion of dPPG signals for various emotions; 3) Fusion of PRV and dPPG signals for various emotions.

Fig. 13 displays the model training loss curves. The experimental results, as shown in Table VIII, indicate that the cross-attention fusion of both PRV and dPPG signals achieves the lowest evaluation error and the best performance. Since PRV reflects autonomic nervous system activity and is associated with autonomic dysregulation due to stress [85], while dPPG provides insights into arterial blood pressure, indicating the direct impact of stress on the cardiovascular system [86], [87], the integration of PRV and dPPG data offers complementary perspectives on stress-induced responses in these systems, thereby improving the performance of stress severity detection.

D. Ablation Experiment for XGBoost

The fully connected layer in the PRV and dPPG cross-attention fusion in Fig. 12 is replaced with ML methods KNN, SVR, GBDT, and XGBoost for stress level detection. The relevant hyperparameter settings are detailed in Table IX, and the

TABLE VII
STATISTICAL T-TEST IN TABLE VI FOR MAE/RMSE (*P*-VALUE)

	Emotion	Calm	Happiness	Sadness	Tension	Fear	Fusion
PRV	Calm	-	.181 ⁿ /.047 ^a	.049 ^a /.057 ^c	.043 ^a /.067 ⁿ	.047 ^a /.038 ^a	.024 ^a /.014 ^a
	Happiness		-	.078 ⁿ /.048 ^a	.034 ^a /.048 ^a	.021 ^a /.041 ^a	.046 ^a /.009 ^b
	Sadness			-	.287 ⁿ /.341 ⁿ	.174 ⁿ /.329 ⁿ	.001 ^b /.049 ^a
	Tension				-	.658 ⁿ /.589 ⁿ	.000 ^b /.048 ^a
	Fear					-	.061 ⁿ /.017 ^a
	Fusion						-
dPPG	Calm	-	.219 ⁿ /.017 ^a	.044 ^a /.049 ^a	.034 ^a /.047 ^a	.046 ^a /.048 ^a	.019 ^a /.018 ^a
	Happiness		-	.053 ^c /.048 ^a	.024 ^a /.048 ^a	.029 ^a /.045 ^a	.015 ^a /.000 ^b
	Sadness			-	.098 ⁿ /.084 ⁿ	.107 ⁿ /.005 ^b	.033 ^a /.007 ^b
	Tension				-	.319 ⁿ /.051 ^c	.036 ^a /.034 ^a
	Fear					-	.057 ^c /.059 ^c
	Fusion						-

where, “a” and “b” indicate statistical significance (a, $P < 0.05$; b, $P < 0.01$), “c” indicates marginal significance (c, $0.05 < P < 0.059$), and “n” indicates no statistical significance ($P > 0.05$).

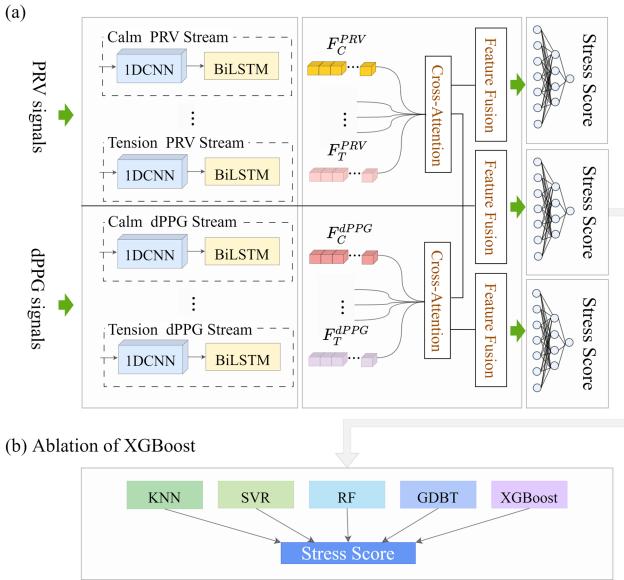


Fig. 12. Stress severity detection based on Cross-Attention and XGBoost. (a) Stress severity detection based on CA of calm, happiness, sadness, fear, and tension emotions using PRV or dPPG signals. Here, extracted features of various emotions from PRV or dPPG are fused with Cross-Attention and then processed through their respective dual-layer fully connected layers for stress detection. The fused features from PRV’s CA fusion and dPPG’s CA fusion (concatenation) are combined through a fully connected layer (labeled as ‘middle’) for stress detection. (b) Describes the XGBoost ablation experiment method, where various machine learning algorithms will replace the middle fully connected layer.

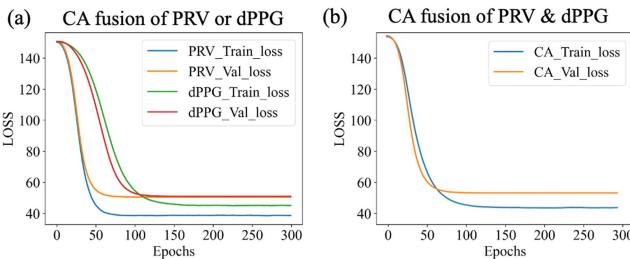


Fig. 13. Training loss curves. (a) CA fusion training using PRV or dPPG signals. (b) CA fusion training using both PRV and dPPG signals.

TABLE VIII
COMPARISON OF EMOTIONAL CROSS-ATTENTION (CA) FUSION RESULTS BASED ON PRV AND dPPG SIGNALS (MEAN \pm SD)

Feature Fusion (CA)	MAE	RMSE
dPPG	$5.16 \pm 0.06^{**}$	$6.16 \pm 0.02^{**}$
PRV	$4.89 \pm 0.16^*$	$5.84 \pm 0.11^*$
PRV & dPPG	4.79 ± 0.12	5.62 ± 0.14

*: $P < 0.05$, **: $P < 0.01$, statistical significance.

TABLE IX
HYPERPARAMETER SETTINGS BASED ON MACHINE LEARNING MODELS

Models	Hyperparameters	Optimum value
KNN	N Neighbors = (1, 10)	N Neighbors = 9
	Leaf Size = (1, 10)	Leaf Size = 3
SVR	C = (0.75, 0.9, 1.0, 1.25)	C = 1.0
RF	N Estimators = (10, 300)	N Estimators = 26
	Max Depth = (1, 15)	Max Depth = 7
	Min Samples Leaf = (1, 15)	Min Samples Leaf = 4
GBDT	Learning Rate = (0.001, 0.1)	Learning Rate = 0.01
	N Estimators = (10, 300)	N Estimators = 210
	Max Depth = (1, 15)	Max Depth = 5
	Min Samples Leaf = (1, 15)	Min Samples Leaf = 10
XGBoost	Learning Rate = (0.001, 0.1)	Learning Rate = 0.01
	N Estimators = (10, 300)	N Estimators = 225
	Max depth = (1, 15)	Max depth = 8
	Colsample Bytree = (0.5, 1.0)	Colsample Bytree = 0.8
	Gamma = (0.001, 1.0)	Gamma = 0.17
	Scale Pos Weight = (1, 15)	Scale Pos Weight = 12
	Subsample = (0.5, 1.0)	Subsample = 0.7

experimental results are presented in Table X. Our findings indicate that XGBoost achieves the highest detection performance among other ML methods, as shown in Table X and Fig. 14, with statistical significance. XGBoost is an ensemble learning approach based on decision trees, which trains multiple models iteratively. It enhances the model’s generalization by techniques like weighted random sampling to prevent overfitting, leading to superior detection performance.

TABLE X
COMPARISON OF ABLATION RESULTS IN XGBOOST (MEAN \pm SD)

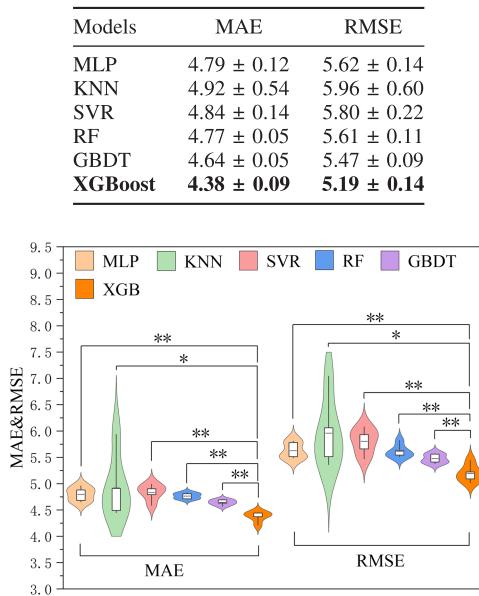


Fig. 14. The detection error of XGBoost is significantly lower than other methods. *: $P < 0.05$, **: $P < 0.01$, statistical significance.

VIII. CONCLUSION

College students' mental health is a societal concern. Early detection and intervention can prevent the risk of developing further psychological disorders. To detect stress levels among college students, we have constructed a pulse dataset (ePulse), which includes data from 177 college students and five containing emotional cues signals: calm, sadness, happiness, fear, and tension. These pulse signals are labeled with stress, anxiety, and depression scores. To the best of our knowledge, this is the first dataset that incorporates multiple emotional information and multiple labels in pulse signals, making it valuable for studying stress in college students as well as for investigating the influence of emotions on other psychological states.

This study, for the first time, investigates the impact of induced emotional states on stress severity detection. It is found that stress severity detection in emotional scenarios is better than that in calm conditions, especially in negative emotional scenarios where the detection performance is superior. This suggests that stress can lead to negative emotions. Additionally, the study has shown that integrating multiple emotional states can effectively enhance detection performance, achieving higher accuracy than any single emotion. Meanwhile, the fusion of PRV and dPPG attained higher evaluation performance, indicating that stress is related to various physiological features.

This study utilized the ensemble learning method (XGBoost) to infer and evaluate stress levels. XGBoost combines multiple weak learners (decision trees) to form a strong learner. It calculates feature weights based on the contribution of features in the decision tree building process using the gradient boosting algorithm. Thus, when constructing decision tree models, XGBoost assigns higher weights to features with high contribution, and

the fusion of multiple emotional features at different contribution levels enhances stress detection performance.

While this study enriches the methods of evaluating mental stress and can serve as a tool for detecting the severity of stress among college students, the exploration of psychological stress still remains a broad field. Although our study achieved good results in detecting the stress levels of college students, there are still some limitations. The selection of research subjects is limited to college students, and there is a lack of studies on a broader population, which results in an insufficient understanding of the impact of factors such as gender and age. To further improve the performance of stress detection, we plan to delve into the following aspects in future research: 1) Enlarging the scope of participants to include more diverse populations, such as adolescents and working professionals; 2) Investigating the impact of other induced emotions on stress detection to gain a more comprehensive understanding of the relationship between emotions and stress; 3) Collecting additional modalities of information, such as facial videos and audio signals, and adopting multimodal fusion methods to improve the accuracy of stress level detection.

DATA AVAILABILITY AND ETHICS STATEMENT

The data supporting the findings of this study can be available on reasonable request. Every participant in the study gave their consent by signing an informed consent form. Additionally, this research received approval from the Ethics Committee at Beijing Anding Hospital, Capital Medical University, Beijing, China.

REFERENCES

- [1] H. Zhang, "A novel deep learning model for analyzing psychological stress in college students," *J. Elect. Comput. Eng.*, vol. 2022, pp. 1–11, 2022.
- [2] E. Ramón-Arribés, V. Gea-Caballero, J. M. Granada-López, R. Juárez-Vela, B. Pellicer-García, and I. Antón-Solanas, "The prevalence of depression, anxiety and stress and their associated factors in college students," *Int. J. Environ. Res. Public Health*, vol. 17, no. 19, 2020, Art. no. 7001.
- [3] N. Kent, F. Alhowaymel, K. Kalmakis, L. Troy, and L. M. Chiodo, "Development of the college student acute stress scale (CSASS)," *Perspectives Psychiatr. Care*, vol. 58, no. 4, pp. 2998–3008, 2022.
- [4] I. J. Ratul, M. M. Nishat, F. Faisal, S. Sultana, A. Ahmed, and M. A. Al Mamun, "Analyzing perceived psychological and social stress of university students: A machine learning approach," *Heliyon*, vol. 9, no. 6, 2023, Art. no. e17307.
- [5] A. Seiler, C. P. Fagundes, and L. M. Christian, "The impact of everyday stressors on the immune system and health," in *Stress Challenges and Immunity in Space: From Mechanisms to Monitoring and Preventive Strategies*. Berlin, Germany: Springer, 2020, pp. 71–92.
- [6] F. Frisoni, F. Sicari, S. Settimeri, and E. M. Merlo, "Clinical psychological assessment of stress: A narrative review of the last 5 years," *Clin. Neuropsychiatry*, vol. 18, no. 2, pp. 91–100, 2021.
- [7] A. D. Crosswell and K. G. Lockwood, "Best practices for stress measurement: How to measure psychological stress in health research," *Health Psychol. Open*, vol. 7, no. 2, 2020, Art. no. 2055102920933072.
- [8] E. Andreou et al., "Perceived stress scale: Reliability and validity study in Greece," *Int. J. Environ. Res. Public Health*, vol. 8, no. 8, pp. 3287–3298, 2011.
- [9] M. N. Rastgoor, B. Nakisa, F. Maire, A. Rakotonirainy, and V. Chandran, "Automatic driver stress level classification using multimodal deep learning," *Expert Syst. Appl.*, vol. 138, 2019, Art. no. 112793.
- [10] A. Sundaresan, B. Penchina, S. Cheong, V. Grace, A. Valero-Cabré, and A. Martel, "Evaluating deep learning EEG-based mental stress classification in adolescents with autism for breathing entrainment BCI," *Brain Inform.*, vol. 8, no. 1, 2021, Art. no. 13.

- [11] H. Li et al., "Generalized fear after acute stress is caused by change in neuronal cotransmitter identity," *Science*, vol. 383, no. 6688, pp. 1252–1259, 2024.
- [12] F. Cathomas et al., "Circulating myeloid-derived MMP8 in stress susceptibility and depression," *Nature*, vol. 626, no. 8000, pp. 1108–1115, 2024.
- [13] D. L. Haakonsen et al., "Stress response silencing by an E3 ligase mutated in neurodegeneration," *Nature*, vol. 626, no. 8000, pp. 874–880, 2024.
- [14] Y. Bouter et al., "Chronic psychosocial stress causes increased anxiety-like behavior and alters endocannabinoid levels in the brain of C57Bl/6J Mice," *Cannabis Cannabinoid Res.*, vol. 5, no. 1, pp. 51–61, 2020.
- [15] R. P. Asplund, A. Jäderlind, I. H. Björk, B. Ljótsson, P. Carlbring, and G. Andersson, "Experiences of internet-delivered and work-focused cognitive behavioral therapy for stress: A qualitative study," *Internet Interv.*, vol. 18, 2019, Art. no. 100282.
- [16] J. Schoffl, I. Pozzato, D. Rodrigues, M. Arora, and A. Craig, "Pulse rate variability: An alternative to heart rate variability in adults with spinal cord injury," *Psychophysiology*, vol. 60, no. 11, 2023, Art. no. e14356.
- [17] Y. S. Chen, Y. Y. Lin, C. C. Shih, and C. D. Kuo, "Relationship between heart rate variability and pulse rate variability measures in patients after coronary artery bypass graft surgery," *Front. Cardiovasc. Med.*, vol. 8, 2021, Art. no. 749297.
- [18] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affective Comput.*, vol. 13, no. 1, pp. 440–460, First Quarter 2022.
- [19] P. H. Charlton, P. Celka, B. Farukh, P. Chowienczyk, and J. Alastruey, "Assessing mental stress from the photoplethysmogram: A numerical study," *Physiol. Meas.*, vol. 39, no. 5, 2018, Art. no. 054001.
- [20] X. Zhang et al., "Photoplethysmogram-based cognitive load assessment using multi-feature fusion model," *ACM Trans. Appl. Percep.*, vol. 16, no. 4, pp. 1–17, 2019.
- [21] O. V. Crowley et al., "The interactive effect of change in perceived stress and trait anxiety on vagal recovery from cognitive challenge," *Int. J. Psychophysiol.*, vol. 82, no. 3, pp. 225–232, 2011.
- [22] S. H. Seo, J. T. Lee, and M. Crisan, "Stress and EEG," *Convergence Hybrid Inf. Technol.*, vol. 27, pp. 413–424, 2010.
- [23] D. H. Hellhammer, S. Wüst, and B. M. Kudielka, "Salivary cortisol as a biomarker in stress research," *Psychoneuroendocrinology*, vol. 34, no. 2, pp. 163–171, 2009.
- [24] P. Melillo, M. Bracale, and L. Peccia, "Nonlinear heart rate variability features for real-life stress detection. Case study: Students under stress due to university examination," *Biomed. Eng. Online*, vol. 10, no. 1, 2011, Art. no. 96.
- [25] S. Z. Bong, M. Murugappan, and S. Yaacob, "Analysis of electrocardiogram (ECG) signals for human emotional stress classification," in *Proc. 1st Int. Conf. Trends Intell. Robot. Automat. Manuf.*, Springer, Berlin, Germany, 2012, pp. 198–205.
- [26] B. Cinaz, B. Arnrich, R. La Marca, and G. Tröster, "Monitoring of mental workload levels during an everyday life office-work scenario," *Pers. Ubiquitous Comput.*, vol. 17, pp. 229–239, 2013.
- [27] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," *Biomed. Signal Process. Control*, vol. 46, pp. 18–32, 2018.
- [28] A. Arsalan and M. Majid, "Human stress classification during public speaking using physiological signals," *Comput. Biol. Med.*, vol. 133, 2021, Art. no. 104377.
- [29] Y. Jiao et al., "Feasibility study for detection of mental stress and depression using pulse rate variability metrics via various durations," *Biomed. Signal Process. Control*, vol. 79, 2023, Art. no. 104145.
- [30] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
- [31] X. Gong, C. L. P. Chen, B. Hu, and T. Zhang, "CiABL: Completeness-induced adaptative broad learning for cross-subject emotion recognition with EEG and eye movement signals," *IEEE Trans. Affective Comput.*, vol. 15, no. 4, pp. 1970–1984, Fourth Quarter 2024.
- [32] H. Chen, J. Ru, H. Long, J. He, T. Chen, and W. Deng, "Semi-supervised adaptive pseudo-label feature learning for hyperspectral image classification in Internet of Things," *IEEE Internet of Things J.*, vol. 11, no. 19, pp. 30754–30768, Oct. 2024.
- [33] H. Zhao, Y. Gao, and W. Deng, "Defect detection using shuffle Net-CASSD lightweight network for turbine blades in IoT," *IEEE Internet of Things J.*, vol. 11, no. 20, pp. 32804–32812, Oct. 2024.
- [34] H. Ren et al., "Graph convolutional networks in language and vision: A survey," *Knowl. Based Syst.*, vol. 251, 2022, Art. no. 109250.
- [35] J. He, K. Li, X. Liao, P. Zhang, and N. Jiang, "Real-time detection of acute cognitive stress using a convolutional neural network from electrocardiographic signal," *IEEE Access*, vol. 7, pp. 42710–42717, 2019.
- [36] R. Li and Z. Liu, "Stress detection using deep neural networks," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 11, 2020, Art. no. 285.
- [37] L. Xia et al., "MuLHiTA: A novel multiclass classification framework with multibranch LSTM and hierarchical temporal attention for early detection of mental stress," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 9657–9670, Dec. 2023.
- [38] X. Zhang et al., "Dynamic alignment and fusion of multimodal physiological patterns for stress recognition," *IEEE Trans. Affective Comput.*, vol. 15, no. 2, pp. 685–696, Second Quarter 2024.
- [39] L. Mou et al., "Driver stress detection via multimodal fusion using attention-based CNN-LSTM," *Expert Syst. Appl.*, vol. 173, 2021, Art. no. 114693.
- [40] R. Tanwar, O. C. Phukan, G. Singh, P. K. Pal, and S. Tiwari, "Attention based hybrid deep learning model for wearable based stress recognition," *Eng. Appl. Artif. Intell.*, vol. 127, 2024, Art. no. 107391.
- [41] F. Yu, Y. Fang, Z. Zhao, J. Bei, T. Ren, and G. Wu, "CAGNet: A context-aware graph neural network for detecting social relationships in videos," *Vis. Intell.*, vol. 2, no. 22, pp. 1–14, 2024.
- [42] Y. Zheng, C. Gao, L. Chen, D. Jin, and Y. Li, "DGCN: Diversified recommendation with graph convolutional networks," in *Proc. Web Conf.*, 2021, pp. 401–412.
- [43] S. Jia, S. Jiang, S. Zhang, M. Xu, and X. Jia, "Graph-in-graph convolutional network for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 1157–1171, Jan. 2024.
- [44] D. Chandola, E. Altarawneh, M. Jenkin, and M. Papagelis, "SERC-GCN: Speech emotion recognition in conversation using graph convolutional networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2024, pp. 76–80.
- [45] B. Chen, C. L. P. Chen, and T. Zhang, "GDDN: Graph domain disentanglement network for generalizable EEG emotion recognition," *IEEE Trans. Affective Comput.*, vol. 15, no. 3, pp. 1739–1753, Third Quarter 2024.
- [46] V. Adarsh and G. R. Gangadharan, "Mental stress detection from ultra-short heart rate variability using explainable graph convolutional network with network pruning and quantisation," *Mach. Learn.*, vol. 113, pp. 5467–5494, 2024.
- [47] S. Amiriparian et al., "MuSe 2022 challenge: Multimodal humour, emotional reactions, and stress," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 7389–7391.
- [48] L. Christ et al., "The MuSe 2022 multimodal sentiment analysis challenge: Humor, emotional reactions, and stress," in *Proc. 3rd Int. Multimodal Sentiment Anal. Workshop Challenge*, 2022, pp. 5–14.
- [49] H. Park, G. Kim, J. Oh, A. V. Messem, and W. D. Neve, "Interpreting stress detection models using SHAP and attention for MuSe-Stress 2022," *IEEE Trans. Affective Comput.*, to be published, doi: 10.1109/TAFFC.2024.3488112.
- [50] J. Li et al., "Hybrid multimodal feature extraction, mining and fusion for sentiment analysis," in *Proc. 3rd Int. Multimodal Sentiment Anal. Workshop Challenge*, New York, NY, USA, 2022, pp. 81–88.
- [51] Y. Liu, W. Sun, X. Zhang, and Y. Qin, "Improving dimensional emotion recognition via feature-wise fusion," in *Proc. 3rd Int. Multimodal Sentiment Anal. Workshop Challenge*, New York, NY, USA, 2022, pp. 55–60.
- [52] H. Park et al., "Towards multimodal prediction of time-continuous emotion using pose feature engineering and a transformer encoder," in *Proc. 3rd Int. Multimodal Sentiment Anal. Workshop Challenge*, New York, NY, USA, 2022, pp. 47–54.
- [53] Y. He et al., "Multimodal temporal attention in sentiment analysis," in *Proc. 3rd Int. Multimodal Sentiment Anal. Workshop Challenge*, New York, NY, USA, 2022, pp. 61–66.
- [54] G. Barbayannis, M. Bandari, X. Zheng, H. Baquerizo, K. W. Pecor, and X. Ming, "Academic stress and mental well-being in college students: Correlations, affected groups, and COVID-19," *Front. Psychol.*, vol. 13, 2022, Art. no. 886344.
- [55] H. H. Kenchannavar, S. D. Perur, U. P. Kulkarni, and R. Hegde, "Statistical analysis of stress levels in students pursuing professional courses," in *Proc. Int. Conf. Inf. Commun. Technol. Syst.*, Springer, Singapore, 2020, pp. 469–477.
- [56] S. Asif, A. Mudassar, T. Z. Shahzad, M. Raouf, and T. Pervaiz, "Frequency of depression, anxiety and stress among university students," *Pakistan J. Med. Sci.*, vol. 36, no. 5, pp. 971–976, 2020.
- [57] H. Zhan, C. Zheng, X. Zhang, and M. Yang, "Chinese college students' stress and anxiety levels under COVID-19," *Front. Psychiatry*, vol. 12, 2021, Art. no. 615390.

- [58] J. Zhang, W. Wen, F. Huang, and G. Liu, "Recognition of real-scene stress in examination with heart rate features," in *Proc. 9th Int. Conf. Intell. Hum. Mach. Syst. Cybern.*, 2017, pp. 26–29.
- [59] L. Liu, Y. Ji, Y. Gao, T. Li, and W. Xu, "A novel stress state assessment method for college students based on EEG," *Comput. Intell. Neurosci.*, vol. 2022, no. 1, 2022, Art. no. 4565968.
- [60] S. A. M. Mane and A. Shinde, "StressNet: Hybrid model of LSTM and CNN for stress detection from electroencephalogram signal (EEG)," *Results Control Optim.*, vol. 11, 2023, Art. no. 100231.
- [61] M. A. Hafeez and S. Shakil, "EEG-based stress identification and classification using deep learning," *Multimedia Tools Appl.*, vol. 83, no. 14, pp. 42703–42719, 2023.
- [62] Y. Tian, "Identification and modeling of college students' psychological stress indicators for deep learning," *Sci. Program.*, vol. 2022, no. 1, 2022, Art. no. 6048088.
- [63] Y. Wu, R. Gu, Q. Yang, and Y. J. Luo, "How do amusement, anger and fear influence heart rate and heart rate variability?," *Front. Neurosci.*, vol. 13, 2019, Art. no. 1131.
- [64] O. Carr, F. Andreotti, K. E. A. Saunders, A. C. Bilderbeck, G. M. Goodwin, and M. De Vos, "Linking changes in heart rate variability to mood changes in daily life," in *Proc. Comput. Cardiol.*, 2017, pp. 1–4.
- [65] A. M. Ali, A. A. Alkhamees, H. Hori, Y. Kim, and H. Kunugi, "The depression anxiety stress scale 21: Development and validation of the depression anxiety stress scale 8-item in psychiatric patients and the general public for easier mental health measurement in a post COVID-19 world," *Int. J. Environ. Res. Public Health*, vol. 18, no. 19, 2021, Art. no. 10142.
- [66] O. N. Medvedev, "Depression anxiety stress scales (DASS-21) in international contexts," in *International Handbook of Behavioral Health Assessment*, Berlin, Germany: Springer, 2023, pp. 1–15.
- [67] C. H. Cao, C. Y. Dang, X. Zheng, W. G. Chen, I. H. Chen, and J. H. Gamble, "The psychometric properties of the DASS-21 and its association with problematic internet use among Chinese college freshmen," *Healthcare*, vol. 11, no. 5, 2023, Art. no. 700.
- [68] M. Li, W. Zhang, B. Hu, J. Kang, Y. Wang, and S. Lu, "Automatic assessment of depression and anxiety through encoding pupil-wave from HCl in VR scenes," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 20, no. 2, pp. 1–22, 2024.
- [69] M. Li, J. Zhang, J. Song, Z. Li, and S. Lu, "A clinical-oriented non-severe depression diagnosis method based on cognitive behavior of emotional conflict," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 1, pp. 131–141, Feb. 2023.
- [70] R. Somarathna, T. Bednarz, and G. Mohammadi, "Virtual reality for emotion elicitation—A review," *IEEE Trans. Affective Comput.*, vol. 14, no. 4, pp. 2626–2645, Fourth Quarter 2023.
- [71] S. Vlahovic, I. Slivar, M. Silic, L. Skorin-kapov, and M. Suznjevic, "Exploring the facets of the multiplayer VR gaming experience," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 20, no. 9, pp. 1–24, 2024.
- [72] B. L. Fredrickson, "The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions," *Amer. Psychol.*, vol. 56, no. 3, 2001, Art. no. 218.
- [73] M. Li, Y. Wang, C. Yang, Z. Lu, and J. Chen, "Automatic diagnosis of depression based on facial expression information and deep convolutional neural network," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 5, pp. 5728–5739, Oct. 2024.
- [74] M. Li, Z. Lu, Q. Cao, J. Gao, and B. Hu, "Automatic assessment method and device for depression symptom severity based on emotional facial expression and pupil-wave," *IEEE Trans. Instrum. Meas.*, vol. 73, Oct. 2024, Art. no. 2531215.
- [75] K. N. Ochsner and J. J. Gross, "The cognitive control of emotion," *Trends Cogn. Sci.*, vol. 9, no. 5, pp. 242–249, 2005.
- [76] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos, "Stress detection using wearable physiological sensors," in *Proc. Int. Work Conf. Interplay Between Natural Artif. Computation*, Springer, Cham, 2015, pp. 526–532.
- [77] J. W. Mason, "A historical view of the stress field," *J. Hum. Stress*, vol. 1, no. 2, pp. 22–36, 1975.
- [78] J. J. Gross and O. P. John, "Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being," *J. Pers. Social Psychol.*, vol. 85, no. 2, 2003, Art. no. 348.
- [79] H. Wang, "Noise reduction in pulse signal using the wavelet packet transform and median filtering," in *Proc. Int. Workshop Educ. Technol. Training 2008 Int. Workshop Geosci. Remote Sens.*, 2008, pp. 675–679.
- [80] E. Mejia-Mejia and P. A. Kyriacou, "Duration of photoplethysmographic signals for the extraction of pulse rate variability indices," *Biomed. Signal Process. Control*, vol. 80, 2023, Art. no. 104214.
- [81] E. Yuda, K. Yamamoto, Y. Yoshida, and J. Hayano, "Differences in pulse rate variability with measurement site," *J. Physiol. Anthropol.*, vol. 39, pp. 1–6, 2020.
- [82] E. Merdjanovska and A. Rashkovska, "A framework for comparative study of databases and computational methods for arrhythmia detection from single-lead ECG," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 11682.
- [83] O. Desmedt, O. Corneille, O. Luminet, J. Murphy, G. Bird, and P. Maurage, "Contribution of time estimation and knowledge to heartbeat counting task performance under original and adapted instructions," *Biol. Psychol.*, vol. 154, 2020, Art. no. 107904.
- [84] P. H. Charlton et al., "Detecting beats in the photoplethysmogram: Benchmarking open-source algorithms," *Physiol. Meas.*, vol. 43, no. 8, 2022, Art. no. 085007.
- [85] D. S. Goldstein, "Stress and the 'extended' autonomic system," *Autonomic Neurosci.*, vol. 236, 2021, Art. no. 102889.
- [86] P. D. Lambiase, S. N. Garfinkel, and P. Taggart, "Psychological stress, the central nervous system and arrhythmias," *QJM: Int. J. Med.*, vol. 116, no. 12, pp. 977–982, 2023.
- [87] F. Vancheri, G. Longo, E. Vancheri, and M. Y. Henein, "Mental stress and cardiovascular health—Part I," *J. Clin. Med.*, vol. 11, no. 12, 2022, Art. no. 3353.



Mi Li received the PhD degree from the Beijing University of Technology, in 2012. She is a professor with the School of Information Science and Technology, Beijing University of Technology. Her research interests mainly include affective computing, machine learning, computational psychophysiology, and cognitive neuroscience.



Junzhe Li received the master's degree from the School of Information Science and Technology, Beijing University of Technology. His research interests include artificial intelligence, affective computing, and machine learning.



Yanbo Chen is currently working toward the doctor's degree with the School of Information Science and Technology, Beijing University of Technology. Her research interests include artificial intelligence, affective computing, and machine learning.



Bin Hu (Fellow, IEEE) received the PhD degree in computer science from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China, in 1998. He is a professor with the Institute of Engineering Medicine, Beijing Institute of Technology, Beijing; and Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China. His research interests include pervasive computing, computational psychophysiology, and data modeling.