A Portable Neurofeedback Training System for Attention Improvement Based on High-Performance Edge CNN Accelerator

Chunhua He, Wei Huang, Yi Huang, Haojie Wang, Heng Wu, Songqing Deng, and Maojin Liang

***Abstract*—Currently, many people around the world, especially children and youth, are facing the problem of attention deficit. Neurofeedback training (NFT) is proved to be an effective method for improving the attention level. However, current NFT typically requires the use of computers or other nonportable devices, which limits the application scenarios of this technique. Therefore, a portable NFT system based on high-performance edge AI accelerator is proposed in this article. More specifically, the real-time single-channel electroencephalogram (EEG) and electrocardiogram (ECG) signals of the trainees are first collected by the wearable flexible headband and patch. The wavelet packet decomposition algorithm is adopted to decompose and denoise the collected signals. The processed signals are classified by a convolutional neural network model, and an AI accelerator is designed to run this model for portability. The classification results are fed back to the trainees in real-time with a serious game to achieve the closed-loop regulation of their attentions. Finally, a single-blind controlled experiment is conducted to verify the effectiveness of the proposed system. The experimental results indicate that the attention levels of subjects trained with the proposed system are significantly improved ($p < 0.05$), and the attention-related EEG indicator theta/beta ratio decreased by an average of 21.85%.**

***Index Terms*—AI accelerator, attention improving, convolutional neural network, portable neurofeedback training (NFT) system.**

Received 4 May 2025; revised 22 May 2025 and 16 June 2025; accepted 29 June 2025. Date of publication 2 July 2025; date of current version 8 September 2025. This work was supported in part by the Key Research and Development Program of Guangdong under Grant 2023B0303010004; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515220127 and Grant 2024A1515011168; in part by the Guangzhou Key Research and Development Program under Grant 2024B03J1251; and in part by the Smart Medical Innovation Technology Center, GDUT under Grant ZYZX24-037. *(Chunhua He, Wei Huang, and Yi Huang contributed equally to this work.) (Corresponding authors: Heng Wu; Songqing Deng; Maojin Liang.)*

Chunhua He is with the School of Computer, Guangdong University of Technology, Guangzhou 510000, China, and also with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510000, China (e-mail: hechunhua@pku.edu.cn).

Wei Huang and Haojie Wang are with the School of Computer, Guangdong University of Technology, Guangzhou 510000, China (e-mail: 2122205212@mail2.gdut.edu.cn; whj_mail 2022@163.com).

Yi Huang is with the School of Integrated Circuits, Tsinghua University, Beijing 100000, China (e-mail: yihuang@ustc.edu).

Heng Wu is with the School of Automation, Guangdong University of Technology, Guangzhou 510000, China (e-mail: heng.wu@foxmail.com).

Songqing Deng is with the Department of Obstetrics and Gynecology and the Guangxi Hospital Division, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China (e-mail: dengsq@mail.sysu.edu.cn).

Maojin Liang is with the Department of Otolaryngology, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou 510120, China (e-mail: liangmj3@mail.sysu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2025.3585233

I. Introduction

CURRENTLY, many people around the world, especially children and youth, are facing the problem of attention deficit, which seriously affects their learning and work. And this issue is even worse in remote areas with scarce medical resources. In recent years, the rapid development of the Internet of Medical Things (IoMT) has promoted the popularization of portable healthcare systems, providing effective solutions to the above-mentioned issue.

In the past, methods such as mindfulness meditation [1], music intervention [2], aerobic exercise [3], and yoga [4] were adopted to improve attention. Although these methods are effective, they have the drawbacks of being time-consuming, slow to take effect, and prone to causing fatigue. Hence, some digital methods were developed to solve this problem, such as games used for cognitive training [5] and games based on virtual reality and augmented reality [6]. Although they are interesting and can effectively alleviate boredom, a pure game can easily lead to addiction. In addition, transcranial magnetic stimulation (TMS), as a painless and noninvasive medical technique, can be also applied to improve attention [7]. Although this technique has significant effects, it requires expensive medical equipment, making it difficult to popularize.

Biofeedback training (BFT) is a medical technique that can improve cognitive function, regulate mental responses, and treat various behavioral disorders. It feeds back imperceptible physiological signals to patients in the form of comprehensible numbers, sounds, or images via professional instruments, enabling patients to learn to regulate and control physiological processes. Common biofeedback methods include functional Nearinfrared Spectroscopy (fNIRS) [8], electroencephalogram (EEG) [9], electromyogram (EMG) [10], and electrocardiogram (ECG) [11]. Among them, EEG is widely adopted due to its advantages of low cost, noninvasiveness, and high temporal resolution. Neurofeedback training (NFT), as a subdomain of BFT, is a neural regulation technique that utilizes EEG to detect and regulate brain activity to improve cognitive abilities. Many studies have proven that EEG-based NFT can not only improve the attention of healthy individuals [12], [13], [14], but is also widely applied to the treatment of diseases such as attention deficit and hyperactivity disorder (ADHD) [15], autism [16], anxiety disorder [17], and depression [18]. Compared with traditional drug therapy, this noninvasive intervention method has fewer side effects and better patient compliance. Overall, EEG-based NFT is a

promising method for improving attention, but so far it still has some limitations.

Accurately detecting changes in the real-time attention level of trainees is the most crucial step in the NFT process, which directly affects the training effect. With the rapid development of AI, machine learning (ML) methods have been applied to improve the detection accuracy of attention levels in NFT processes [19], [20]. For the EEG signal, it can be divided into five waves in accordance with the frequency range: 1) delta (0.5∼4 Hz); 2) theta (4∼8 Hz); 3) alpha (8∼14 Hz); 4) beta (14∼30 Hz); and 5) gamma (30∼64 Hz). Related researchers have found that the fluctuations in attention levels can cause changes in the activity of these EEG waves [21], [22], [23]. Therefore, the power of these EEG waves is considered a feature that can characterize changes in attention levels. ML methods focus on manually extracting these attention-related EEG features from raw EEG signals, and then using them to train ML models. The trained ML model is capable of detecting attention levels based on input EEG features. However, manual feature extraction may miss key dynamic information in the EEG signals, leading to a decrease in accuracy [24]. Compared with ML models, deep learning (DL) models can automatically extract advanced robust features from raw data, providing superior solutions for improving detection accuracy [25], [68]. Hence, the NFT system proposed in this article is implemented based on DL.

Besides, the devices used for NFT also have limitations. Currently, in most NFT systems, the processing of EEG signals needs to be completed on computers or other nonportable devices [19], [26], [27], [28], [29], [30], which makes formal NFT only available in laboratories or hospitals. This restriction hinders the promotion of the NFT technique. Therefore, it is necessary to develop a real-time NFT system that can run on mobile devices. However, DL models with complex structures require a large amount of computing resources. In order to meet the requirements of the real-time NFT systems for low power consumption and low latency, lightweight DL models are a more suitable choice.

As for the location for collecting EEG signals, neuroscience studies have found that the prefrontal lobe is closely related to advanced cognitive functions such as attention, memory, and decision-making [31]. And EEG signals from the frontal lobe have been proven to contain rich attention-related information [32]. Besides, as the forehead is not covered by hair, it is easier to obtain clean signals. For these reasons, the forehead is a convenient location for collecting EEG signals. Recently, single-channel EEG signals have been adopted in some studies to detect attention and have achieved promising results [24], [33], [34], [35]. For convenience, we also use single-channel EEG signals of Fp1 channel to detect attention levels. However, information required by the single-channel EEG is not enough for high-accuracy attention recognition, since it lacks spatial information [35]. Therefore, to supplement other biofeedback signals is very significant. Considering that ECG signals are easily obtained, so ECG signals can be combined with EEG signals and applied for attention detection [36], [37], [69].

As for the selection of training tools, boring feedback methods may reduce the enthusiasm of trainees, thereby affecting the training effect. Serious games generally refer to games designed to impart knowledge and train skills. They can provide fun and motivation for trainees during repeated training processes, making them suitable as a training tool for NFT [38].

Overall, NFT is an effective method for improving attention and has the advantages of safety, noninvasiveness, and good patient compliance. However, existing NFT techniques still have the following limitations.

1) Existing NFT systems generally use ML methods to detect the real-time attention level of trainees, while ML methods rely on manual feature extraction. EEG is a complex physiological signal with high temporal resolution, and simply quantifying it into several feature values can easily miss important information contained within, leading to a decrease in the performance of NFT systems.
2) In existing NFT systems, the processing and classification of EEG signals are usually done on computers or other nonportable electronic devices, which reduces their convenience and limits their application scenarios. Especially for children and adolescents in areas with poor medical conditions, it is difficult to receive effective attention training.

To address the aforementioned limitations and challenges, we propose a portable NFT system based on high-performance edge CNN accelerator, aimed at expanding the clinical application scenarios of NFT technique, which is meaningful for the popularization of daily attention training. The main contributions of this research are as follows.

1) We use a multimodal physiological signal composed of five types of EEG waves and denoised ECG signals as samples to train a lightweight CNN model. The trained model can accurately recognize three types of attention states.
2) We design a high-performance edge CNN accelerator that can run the attention classification model in real-time on embedded devices, and the model inference speed is faster than that of computers. Furthermore, it also has the feature of reconfigurability, which can support CNN models of various structures.

Compared with current NFT systems, the proposed NFT system can run entirely on mobile devices without the need for nonportable devices such as computers, allowing users to train their attention at home or school. In addition, it also has the advantages of low latency and low power consumption. Therefore, the proposed NFT system can be applied to a wider range of real-world scenarios.

Table I shows the comparison results between the proposed NFT system and other neurofeedback systems used to improve attention in terms of input signal, neuromarker or model used, device used, and system accuracy.

II. Materials and Methods

A. Hardware System Design

1) Hardware Introduction: The novel portable NFT system for attention improvement based on high-performance edge CNN accelerator mainly consists of a flexible fabric headband

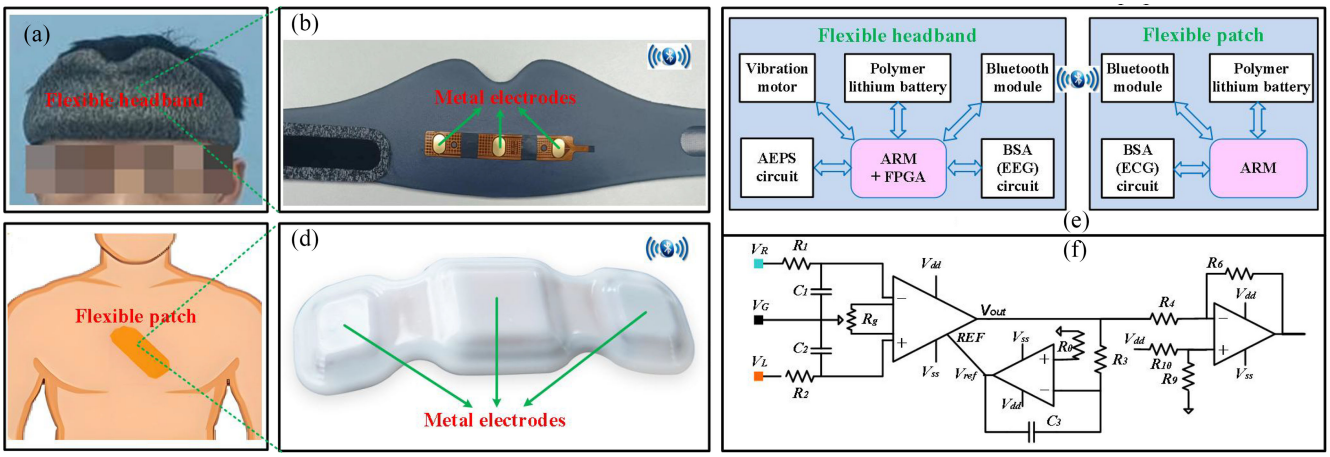

Fig. 1. (a) Flexible headband for collecting single-channel EEG signals. (b) Flexible fabric headband. (c) Flexible patch for collecting single-channel ECG signals. (d) Flexible silicon patch. (e) Structure of the hardware system. (f) Structure of the BSA circuit.

TABLE I
COMPARISON WITH OTHER NEUROFEEDBACK SYSTEMS USED TO IMPROVE ATTENTION

Ref	Input signal	Neuromarker or model	Device	Accuracy
[19]	EEG	IRF	Laptop	79.34%
[20]	EEG	SVM	Desktop	NA
[26]	EEG	Theta/Beta ratio	Laptop	NA
[64]	EEG	Alpha/Theta ratio	ProCamp 5	NA
[65]	EEG	P300	Desktop	NA
[66]	EEG	Sample entropy	Desktop	NA
[67]	EEG	SMR/Theta ratio	TV	NA
Our work	EEG + ECG	CNN	Edge CNN accelerator	95.81%

and a flexible silicon patch. For data acquisition of the single-channel EEG signals of the forehead, a flexible fabric headband is developed, as shown in Fig. 1(a) and (b). For data acquisition of the single-channel ECG signals of the chest, a flexible silicon patch is developed, as shown in Fig. 1(c) and (d). The system overview of the flexible headband and flexible patch is illustrated in Fig. 1(e). The headband system is composed of a field-programmable gate array (FPGA), a microprogrammed control unit (MCU), a bioelectrical signal acquisition (BSA) circuit, an acupoint electric pulse stimulation (AEPS) circuit, a vibration motor, a Bluetooth module, and a polymer lithium battery. The FPGA is used as an edge AI accelerator to fulfill the CNN algorithm. The model number of the FPGA is XCKU5P (manufactured by Xilinx Company Ltd., California, USA). The ARM MCU is used to acquire the EEG signals and conduct the data preprocessing. The model number of the MCU is CS32A010 (manufactured by Chipsea Company Ltd., Shenzhen, China). It has a 24-bit Σ-Δ-type analog-to-digital converter (ADC) with a data conversion rate of 8 ksps, enabling it to handle high-speed changing signals. The AEPS circuit and the vibration motor are applied to stimulate trainees once they are at an extremely low level of attention. The BSA circuit, as depicted in Fig. 1(f), is adopted to detect the EEG signals. The model numbers of the instrumentation amplifier (IA) and operational amplifier (OA) in the circuit are SGM620 and SGM8959-2 (both manufactured by SGMICRO Company Ltd., Shenzhen, China), respectively. The common mode rejection ratio (CMRR) of the circuit is greater than 100 dB, which meets clinical standards. The reliability of the metal electrode has been tested in our previous work [39]. The polymer lithium battery is used as the power supply to avoid the power frequency interference. The Bluetooth module of the headband is used to communicate with the patch and smartphone.

On the other hand, the patch system is composed of a MCU, a BSA circuit, a Bluetooth module, and a polymer lithium battery. Similarly, the ECG signals are detected by the BSA circuit, and then acquired by the ARM MCU. After data preprocessing, the raw ECG signals are sent to the headband by the Bluetooth module for attention recognition. It should be noted that the BSA circuit used for detecting the ECG signals is similar to that for detecting the EEG signals, and the difference is the amplification times. Since the ECG signals are stronger than the EEG signals, the total gain of the BSA circuit used for detecting the ECG signals is set to 1000, while that of the BSA circuit used for detecting the EEG signals is set to 9910. The sampling frequencies of the two ADCs of the two ARM MCUs are both set to 512 Hz. Thus, the EEG signals and ECG signals can be easily acquired by the two portable Internet of Things (IoT) devices. It should be emphasized that before wearing our signal acquisition devices, it is necessary to wipe the skin of the acquisition site with alcohol to minimize the impedance between the electrode and the skin as much as possible, so as to obtain higher quality signals.

2) Signal-to-Noise Ratio Measurement: signal-to-noise ratio (SNR) is an important performance evaluation indicator for signal acquisition equipment. A commercial bioelectric signal acquisition board (manufactured by OpenBCI Company Ltd., New York, USA) was selected for comparison. Specifically, we collected two 10-s eye blink signals with a frequency of 1 Hz from the same person using the developed signal acquisition equipment and OpenBCI board. Then, the SNR of the collected blink signals was calculated separately to compare the performance differences

between our equipment and the standard commercial equipment.

B. Bioelectric Signal Acquisition Experiment Method

1) Experimental Purpose: This experiment is designed to acquire raw single-channel EEG and ECG signals in three attention levels, which will be used to train the DL model. Many psychological studies have shown that tasks of different difficulty levels can stimulate different attention levels in subjects [40], [41]. Hence, the same method is adopted in this experiment. The definitions of the three attention levels are as follows.

1) low attention (LA) is defined as a resting state that typically occurs during closed-eye rest.
2) medium attention (MA) is defined as a relaxed state that typically occurs during the execution of simple tasks.
3) high attention (HA) is defined as a focused state that typically occurs during the execution of difficult tasks.

2) Experimental Materials: In order to stimulate different attention levels of the subjects, various tasks of varying difficulty were provided, including the Stroop color test, Schulte table, mental arithmetic, and silent reading, which have been adopted in similar studies [26], [42], [43]. In addition, the valence section of the self-assessment manikin (SAM) is a commonly used attention self-assessment scale [19], [44]. It was adopted in this experiment to evaluate the attention levels of subjects.

3) Participants: The main target audience for NFT is children and youth. Considering that children have difficulty completing difficult attention-inducing tasks, 29 young people aged 16 to 27 were recruited to participate in this experiment (male: 17, female: 12, age: 21.34 $\pm$ 3.02). All recruited subjects meet the following selection criteria.

1) The test score for the Wechsler adult intelligence scale (WAIS) is above 90 points.
2) The vision test results do not show color blindness or color weakness.
3) The test score for the Adult ADHD self-report scale (ASRS) is below 17 points.

Considering that the physiological data characteristics of participants with a history of mental illness may differ from those of healthy individuals, and that previous participation in similar experiments may cause pre-existing psychological effects, we excluded individuals with a history of mental illness or similar experimental experiences through interviews.

4) Experimental Procedure: The experiment was conducted in a quiet and constant temperature laboratory with electromagnetic shielding. The subjects sat on the chair and completed all tasks according to the instructions, as shown in Fig. 2.

The entire experiment was divided into three stages, such as rest, pretest, and main test. During the rest, the subjects were asked to rest with their eyes closed for 600 s to collect bioelectrical signals in LA. During the pretest, the subjects were asked to complete each task once to familiarize themselves with the rules and adapt to changes in attention in advance. During the main test, each task was presented on a laptop with a fixed presentation time of 15 s. The order of task presentations is random, and the frequency of each task presentation is equal. After completing one task, the subjects need to evaluate their attention levels during the task via the SAM. Subsequently, the bioelectrical signals collected within these 15 s will be saved as a sample on another laptop. If the evaluation time exceeds 5 s, it may mean that the subject is unable to make a clear choice, and the collected bioelectrical signals will be discarded as an ineffective sample. The process of completing one task and one evaluation is defined as one trial. Each subject was asked to complete 80 such trials, and for every 20 trials completed, they would have a 90-s break to adjust their attention levels. Besides, samples that were severely contaminated due to the excessive body movements during the task will be excluded. After careful screening, we finally obtained 3167 samples. The number and proportion of samples for each attention level are shown in Table II.

Fig. 2. Experimental platform for collecting the bioelectrical signals.

TABLE II
NUMBER AND PROPORTION OF SAMPLES FOR EACH ATTENTION LEVEL

Attention level	Number	Proportion
LA	1207	38.11%
MA	951	30.03%
HA	1009	31.86%

C. Data Preprocessing Method

1) Data Segmentation: In similar studies, the length of the window used to segment EEG signals is 2 or 3 s [24], [45]. A normal person's heart rate is about 1$\sim$2 Hz. To include as many heartbeats as possible in an ECG fragment, the obtained samples were further divided into 3-s fragments. Finally, 15 835 (i.e., 3167*5) bioelectrical signal fragments were obtained.

2) EEG Data Processing: The introduction section points out that the five EEG waves (i.e., delta, theta, alpha, beta, and gamma waves) are closely related to attention. Thus, it is necessary to separate and extract these key EEG waves from the raw EEG signals. Wavelet packet decomposition (WPD) is very suitable for processing nonstationary random signals. It projects the time series signal onto the orthogonal wavelet basis function space and then decomposes the signal into different frequency bands. In this article, WPD is adopted

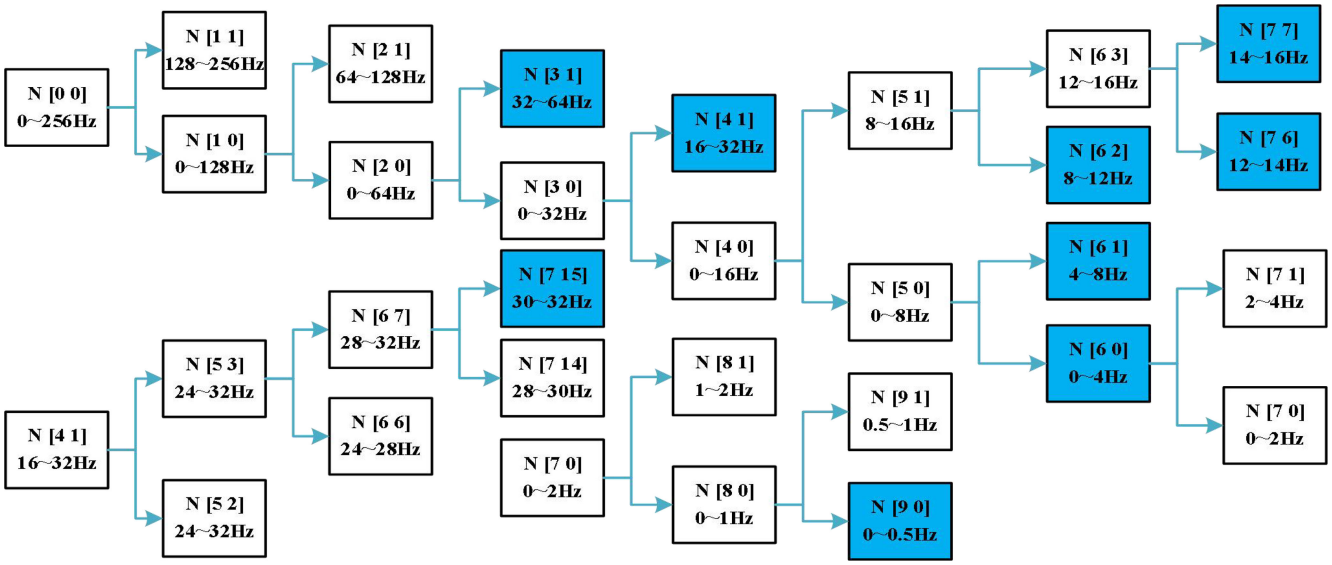

Fig. 3. Structure diagram of the decomposition tree of the nine-level wavelet packet transform.

to extract five EEG waves from the raw EEG signals, and the extraction process is as follows.

The structure diagram of the decomposition tree of the nine-level wavelet packet transform is shown in Fig. 3. In this figure, the frequency bands of each node are given, and the key nodes used for subsequent calculations are marked in blue. The five target EEG waves can be extracted by

$$\begin{cases} \mathrm{wpt}_{\mathrm{EEG}} = \mathrm{wpdec}(S_{\mathrm{EEG}}\ 9\ 'db6') \\ \mathrm{baseline1} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [9\ 0]) \\ S_{\mathrm{delta}} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [6\ 0]) - \mathrm{baseline1} \\ S_{\mathrm{theta}} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [6\ 1]) \\ S_{\mathrm{alpha}} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [6\ 2]) + \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [7\ 6]) \\ S_{\mathrm{beta}} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [4\ 1]) + \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [7\ 7]) \\ \qquad - \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [7\ 15]) \\ S_{\mathrm{gamma}} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [3\ 1]) + \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{EEG}}\ [7\ 15]) \end{cases} \tag{1}$$

where S_{EEG} is the raw EEG signal, *wpdec* is a 1-D WPD function used to decompose the input signal, wpt_{EEG} is a wavelet packet tree object corresponding to the 9th-level decomposition of the input signal, and w*procoef* is a wavelet packet reconstruction function used to reconstruct the decomposed signal. Here, [4 1] represents node 1 in the 4th layer, and similarly, [6 2] represents node 2 in the 6th layer. Because the sampling frequency is 512 Hz, the reconstructed signal of node 0 in the 9th layer represents the baseline signal with a frequency range of 0∼0.5 Hz. Here, $0.5{=}512/2^{(9+1)}$. The purpose of removing the baseline signal is to filter out any possible low-frequency noise, such as electrooculography (EOG) artifacts. S_{delta}, S_{theta}, S_{alpha}, S_{beta}, and S_{gamma} are the signals of delta, theta, alpha, beta, and gamma waves after WPD processing, respectively. Here, theta/beta ratio (TBR) (TBR) is an important EEG indicator related to the attention level, and the calculation equations are as shown in

$$\begin{cases} E_{\mathrm{theta}} = \int_4^8 \mathrm{AM}_{\mathrm{theta}}^2(f)\mathrm{df} \\ E_{\mathrm{beta}} = \int_{14}^{30} AM_{\mathrm{beta}}^2(f)\mathrm{df} \end{cases} \tag{2}$$

$$\mathrm{TBR} = E_{\mathrm{theta}} / E_{\mathrm{beta}} \tag{3}$$

where $AM_{\mathrm{theta}}(f)$ and $AM_{\mathrm{beta}}(f)$ are the amplitude functions of theta wave and beta wave, respectively.

3) ECG Data Processing: During the process of collecting ECG signals, the subjects sat quietly on the chair, so the influence induced by motion artifacts can be ignored. Similarly, the baseline drift in the ECG signal should be removed, and here WPD is also adopted. The processing equation is shown in

$$\begin{cases} \mathrm{wpt}_{\mathrm{ECG}} = \mathrm{wpdec}(S_{\mathrm{ECG}}\ 9\ 'db6') \\ \mathrm{baseline2} = \mathrm{wprcoef}(\mathrm{wpt}_{\mathrm{ECG}}\ [9\ 0]) \\ Y_{\mathrm{ECG}} = S_{\mathrm{ECG}} - \mathrm{baseline2} \end{cases} \tag{4}$$

where S_{ECG} is the raw ECG signal, *baseline2* is the baseline drift in the frequency range of 0∼0.5 Hz, and Y_{ECG} is the ECG signal without baseline drift.

D. Attention Level Recognition Model

1) Model Structure and Training: In this article, a lightweight CNN model is constructed to recognize the attention levels of trainees during NFT processes. The ECG signals without baseline drift and five EEG waves were used to train the CNN model. The structure of the CNN model is shown in Fig. 4. It contains three convolutional layers, three max-pooling layers, two fully connected layers, and a dropout layer. The convolutional layer is utilized to extract features from input signals. The size of the convolution kernels for all convolutional layers is 3*3, and ReLU is selected as the activation function. The function of the max-pooling layers is to reduce parameters while retaining the main features, and the size of all pooling layers is 2*2. The fully connected layer maps the features obtained from the convolutional and pooling layers to the sample labeling space, playing the role of a classifier. Dropout is a frequently-used regularization method in DL that can randomly select some nodes to participate in the model prediction. The dropout rate is set to 0.2.

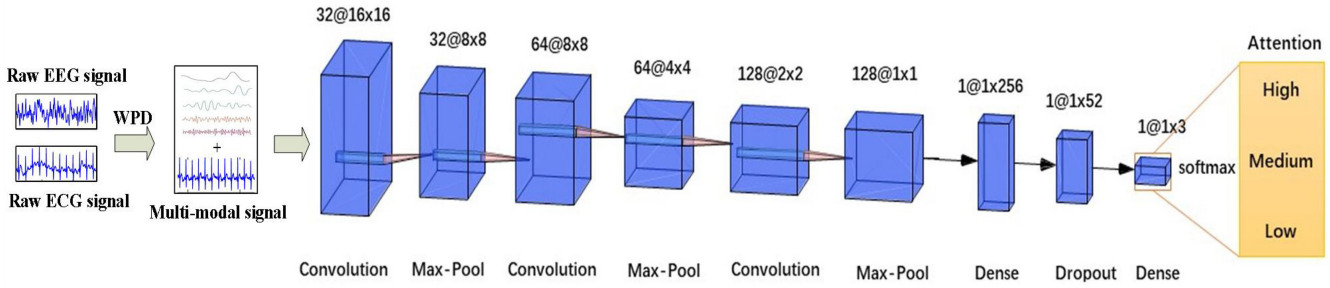

Fig. 4. Structure of the CNN model for recognizing the attention levels of trainees during NFT processes.

The output layer is a fully connected layer that outputs the probabilities of the three different types of attention levels, and the final prediction result is determined by the maximum probability. For model training, the backpropagation algorithm is utilized to learn and optimize network parameters. The Adam optimizer is adopted with a learning rate set to 1e-4. The loss function is cross-entropy loss. The batch size and epoch are set to 32 and 100, respectively.

2) Model Test: Precision (PR), Recall (RE), F1-score (F1-score), and Accuracy (ACC) are adopted as evaluation metrics for model performance, as defined in (5). Leave-one-out cross-validation (LOOCV) is adopted to evaluate the recognition ability of the model. For LOOCV, samples from each subject are used as the test set in turn, while samples from the remaining subjects are used as the training set

$$\begin{cases} \mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \\ \mathrm{RE} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \\ F1-\mathrm{score} = \frac{2\times\mathrm{PR}\times\mathrm{RE}}{\mathrm{PR+RE}} \\ \mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \end{cases} \tag{5}$$

where TP stands for the number of true positives, FP stands for the number of false positives, TN stands for the number of true negatives, and FN stands for the number of false negatives.

In addition, to compare the effects of the multimodal signal and the single-modal signal, the EEG and ECG signals are, respectively, used to train the model. And the same validation method is adopted to evaluate the performance of the model.

3) Hardware Acceleration: While using a laptop or desktop for CNN inference can handle real-time dataflow, it is not a portable and efficient solution for future applications of NFT. Conversely, relying on embedded cores often results in suboptimal performance, compromising real-time processing capabilities. To address these challenges, we propose a dedicated embedded CNN accelerator. Unlike the high-performance computing (HPC)-oriented CNN accelerators discussed in [57], [58], [59], [60], and [61], our design strikes an optimal balance between power, area, and performance. Fig. 5 illustrates the architecture of our hardware design. It features 8 matrix memory load units (LoadUs) and 8 store units. Our matrix memory LoadU is specifically engineered for efficient matrix and submatrix access. It directly fetches data from memory in a convolutional style (addressing the data in the form of $m*l_{matrix}+n+i*l_{submatrix}+j$, m = m+stride_m, n = n+stride_n, i = i+1, j = j+1). This approach allows for dynamic configuration of submatrix size ($l_{submatrix}$), the starting point

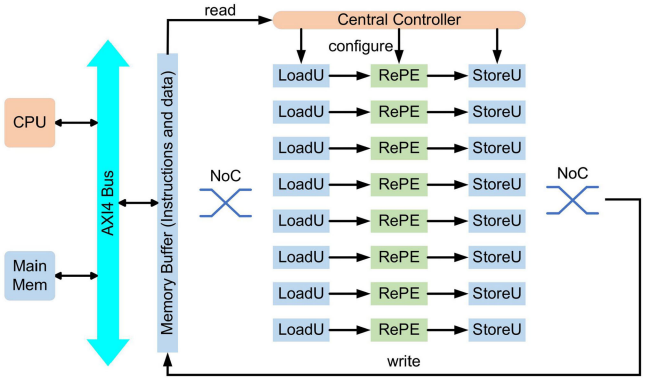

Fig. 5. Architecture of the proposed efficient CNN accelerator.

of the submatrix (m, n), and the strides. Therefore, there is no need to spatially unroll the convolutional layer from four nested loops into a matrix multiplication layout with only two nested loops to apply the affine access (addressing the data in the form of $i*l_{matrix}+j$, i = i+1, j = j+1). Such an unroll technique adopted in [57] and Google's TPU [58] replicates the data in the overlapped area of adjacent submatrix to form new nonoverlapped submatrix and reorganizes the submatrix into a single row of new matrix. By doing so, the convolution operation is unrolled into the matrix multiplication at the cost of increasing memory usage. In contrast, our matrix memory LoadUs efficiently handle the complex addressing pattern inherent in convolution operations without requiring data replication. This eliminates the need for additional memory storage, while maintaining performance comparable to unrolled architectures. As a result, our design significantly enhances efficiency by reducing memory overhead and optimizing data access.

On the computation side, we devise a fully-pipelined multiply-accumulate (MAC) unit capable of timely data consumption to meet real-time performance requirements. While the baseline MAC unit is primarily intended for multiplication and accumulation operations in convolutional and fully connected layers, we enhance its utility by reusing its adder to form a configurable subtractor/adder. This modification enables support for additional operations such as max pooling, ReLU, and basic element-wise additions, thereby improving computational resource utilization. We refer to this enhanced unit as the reconfigurable processing element (RePE), as illustrated in Fig. 5. The RePE not only supports arithmetic

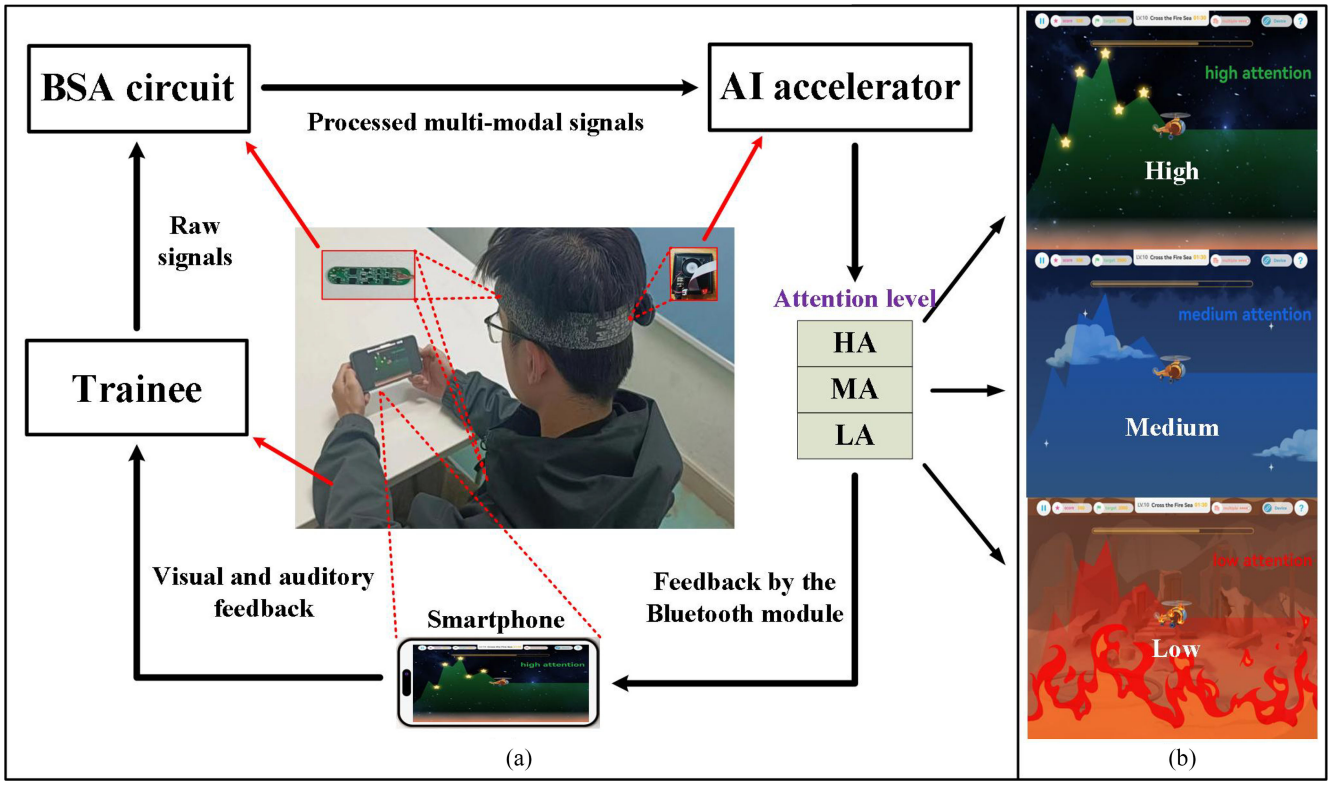

Fig. 6. (a) Closed-loop feedback system framework for attention training. (b) GUIs of different attention levels in the serious game.

operations in convolutional and fully connected layers but also accommodates fundamental functions in ReLU, max pooling, average pooling, and other commonly-used layers.

To improve flexibility, the accelerator incorporates an instruction memory and a centralized controller. Moreover, all compute units are fully reconfigurable, allowing the accelerator to dynamically adapt to changes in layer dimensions or to support the integration of new layers in the CNN model. With native floating-point support, as discussed in [62], the hardware can approximate exponential functions required by sigmoid or tanh activations through multiplication and accumulation. By integrating additional division units, the architecture can be further extended to support a broader set of nonlinear functions and complex model structures. Importantly, in CNN inference tasks, the final softmax layer is often omitted to expedite computation. This is because the softmax function preserves the partial order of class scores, and its exclusion does not affect the correctness of classification outcomes.

The accelerator undergoes multiple reconfigurations during the execution of a CNN model, as illustrated in Fig. 4. For example, it may first be configured for a convolutional layer, followed by a max pooling layer. In each stage, LoadUs are programmed to fetch specific submatrices by setting iteration patterns and submatrix lengths. Simultaneously, the RePEs are reconfigured to execute context-specific operations—for instance, as MAC units during convolution, or as comparators during max pooling. Upon completion of all layers, the final output is retrieved by the host CPU via the AXI4 bus from the accelerator's memory, completing the inference process.

E. Portable NFT System Framework

The number of patients with attention deficits is constantly increasing, but NFT is usually only available in hospitals or laboratories and needs to be completed under the guidance of professional doctors or trainers. Therefore, it is necessary to provide a treatment plan that is convenient enough and can be conducted at home. In order to meet this demand, a portable NFT system framework is described in this section. Its detailed design is as follows.

The closed-loop feedback system framework for attention training is described in Fig. 6(a), and a serious game developed on the Cocos platform that can be installed on smartphones is used as a training tool for NFT. The graphical user interfaces (GUIs) of different attention levels in the game is shown in Fig. 6(b). In this game, the movement of the virtual airplane is controlled by the trainee's attention level, and a real-time attention level is fed back to the trainee in both visual and auditory forms.

More specifically, the real-time bioelectrical signals of the trainee are first collected by the proposed wearable IoT devices and processed by WPD. The CNN model running on the AI accelerator is then used to classify the processed signals, and the classification result is sent to the smartphone via the Bluetooth module and displayed in real-time in the upper right corner of the screen. The flight altitude of the virtual aircraft is determined by the attention level, and the higher the level, the higher the flight altitude. When the trainee remains in the LA state, the virtual airplane will gradually fall, and a red warning will appear on the screen and fast-paced background music will be played to remind the trainee to stay focused.

Furthermore, the AEPS circuit and the vibration motor can be applied to stimulate trainees once they are at an extremely low level of attention for a long time. As the trainee's attention gradually increases, the background of the screen will turn blue (i.e., MA level) or green (i.e., HA level), accompanied by the slow-paced light music. In this way, the changes in different attention levels can be transformed in real-time into changes in the game scene. The trainee needs to constantly adjust attention level during the training process to ensure the virtual aircraft can reach the finish line smoothly.

F. Attention Training Experiment Method

1) Experimental Purpose and Grouping: This single-blind controlled experiment is designed to verify the effectiveness of the proposed portable NFT system. To eliminate the impact of the placebo effect, the following three groups were set up.

1) The NFT group used the proposed portable NFT system to train attention.
2) The serious game training (SGT) group used serious games without the neurofeedback mechanism to train attention. All serious games were selected from a brain-training app called Lumosity, which is widely applied to improve cognitive functions such as attention and memory [46], [47].
3) The blank control group did not conduct any training.

2) Participants: To ensure the rationality of the experimental samples, all candidate subjects were undergraduate students from the same university. The selection criteria for the subjects in this experiment are consistent with those of the signal acquisition experiment. After screening, 18 eligible college students aged 20 to 22 participated in this experiment (male: 11, female: 7, age: 21.33 $\pm$ 0.69). Before the experiment, all subjects were informed of the usage of the experimental devices and signed informed consent forms. They were randomly assigned to the NFT group (male: 5, female: 1, age: 21.50 $\pm$ 0.55), the SGT group (male: 4, female: 2, age: 21.17 $\pm$ 0.75), and the blank control group (male: 2, female: 4, age: 21.33 $\pm$ 0.82). And the subjects in each group were unaware of the training methods of other groups and the grouping situation.

3) Evaluation Method: integrated visual and auditory continuous performance test (IVA-CPT) is a neuropsychological evaluation method widely used to assess attention and inhibitory control [48], [49], [50]. Before the test, the subject sat 15 to 24 inches away from the computer screen and gripped the mouse in a comfortable position. During the test, the subject was instructed to click the mouse when they saw or heard a "1" (target) and not click the mouse when they saw or heard a "2" (error). The reaction time and number of incorrect responses of the subject will be recorded, and these raw data will be converted into standard quotients (the mean value is 100, with a standard deviation of 15) based on a normative database [51]. In this article, the full-scale attention quotient (FSAQ) is adopted to assess the attention level of the subject (its value is positively correlated with attention level).

In addition, NFT mainly improves attention by regulating brain activity. And the EEG indicator TBR has been proven to be negatively correlated with attention level [26], [52], [53]. Thus, TBR is used to evaluate changes in the attention-related brain activity of subjects during the experiment. To reduce the error caused by EEG signals fluctuations, at the end of training, the 5-min EEG signal of the subject was first collected, and then the signal was evenly divided into five segments and the TBR of each segment was calculated. Finally, the average of these TBRs was calculated as the current TBR of the subject, denoted as TBR_{cur}. The calculation method for TBR has been provided in Section II-C. Due to the significant differences in TBR among different subjects [26], the rate of change of TBR_{cur} relative to baseline was adopted to reflect the changes in TBRs of the subjects, as defined in

$$TBR_{cr} = (TBR_{cur} - TBR_{baseline}) / TBR_{baseline}. \quad (6)$$

To ensure the reliability of the experimental results, we measured the test-retest reliability of the two attention evaluation methods mentioned above. Specifically, the subjects were first subjected to an IVA-CPT and EEG assessment, and their FSAQs and TBRs were recorded. After a 7-day interval, the same evaluation methods were used to assess the same group of subjects, and their FSAQs and TBRs were recorded. Then, the Pearson correlation coefficient (PCC) between the two sets of evaluation results was calculated to assess their correlation. The closer the absolute value of PCC is to 1, the stronger the correlation. After calculation, the PCC between the two sets of FSAQs is 0.97 ($p < 0.05$), and the PCC between the two sets of TBRs is 0.92 ($p < 0.05$). Therefore, both attention evaluation methods have good reliability and stability.

4) Experimental Procedure: At the beginning of the experiment, all subjects underwent an IVA-CPT, and the FSAQs in the test results were recorded as the baseline. The TBRs of subjects in the NFT and SGT groups were recorded as $TBR_{baseline}$. During the experiment, the NFT and SGT groups were required to complete two training sessions per week, each lasting 30 min. After each training session, their FASQs and TBR_{cr} were recorded. The blank control group maintained a normal daily routine, and their FASQs need to be recorded twice a week. The experiment lasted for three weeks, and all subjects successfully completed the experiments.

5) Statistical Method: One-way ANOVAs were conducted on the FSAQs of each group of subjects after each training to evaluate whether there were significant differences in their FSAQs. If present ($p<0.05$), Bonferroni method will be further used for post hoc multiple comparisons to evaluate the significance of differences in FSAQs between groups.

III. Results and Discussion

A. SNR Comparison Results

The time-domain waveforms and spectra of eye blink signals collected using our signal acquisition equipment and OpenBCI board are presented in Fig. 7. In the figure, SNR(L) and SNR(H) represent the SNR of the low-frequency and high-frequency parts of the blink signal, respectively. The calculation results of SNR indicate that the signal collected using the OpenBCI board has significant low-frequency noise, with the amplitude of the noise close to that of the blink signal,

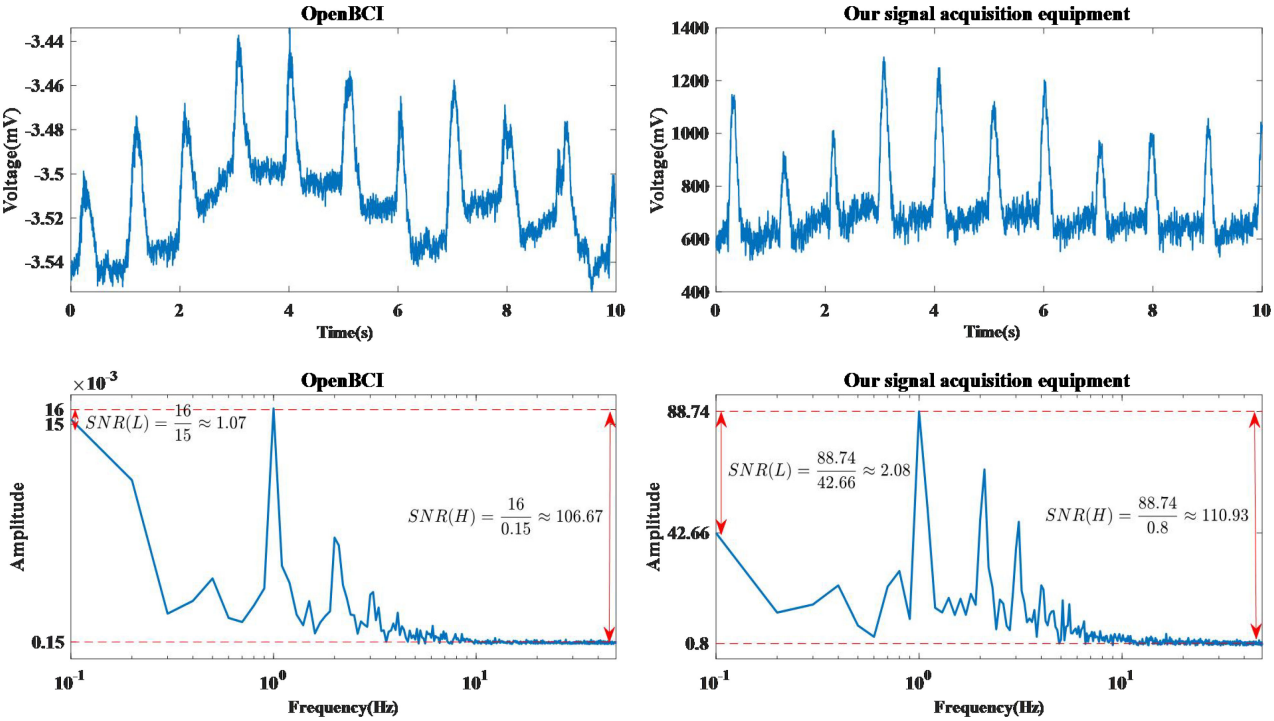

Fig. 7. Time-domain waveforms and spectra of eye blink signals collected using our signal acquisition equipment and OpenBCI board.

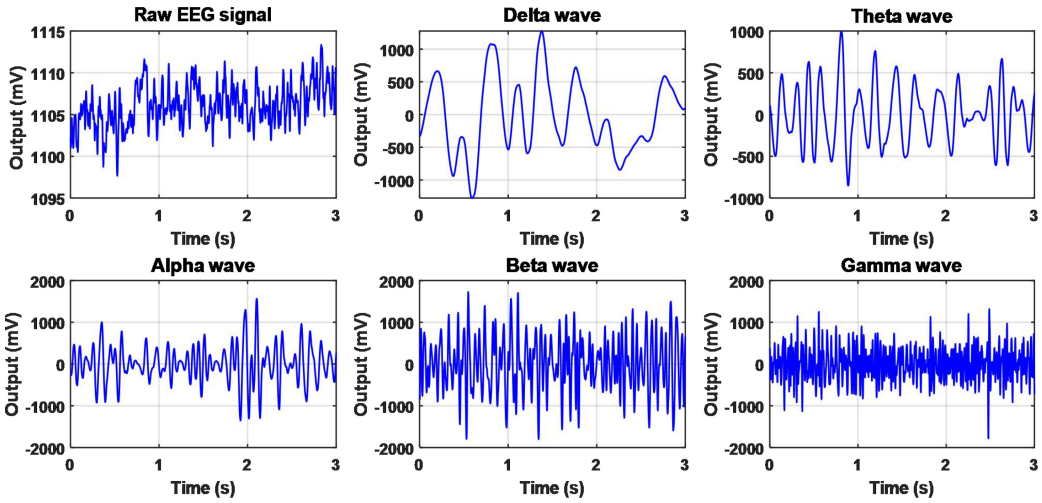

Fig. 8. Delta, theta, alpha, beta, and gamma waves extracted from the raw EEG signal by WPD processing.

while the low-frequency noise in the signal collected using our equipment is relatively smaller. Furthermore, both devices can effectively suppress high-frequency noise in the signal, and our equipment has a better suppression effect.

Overall, in comparison with the commercial bioelectric signal acquisition equipment, the signal collected using our equipment shows a higher SNR, further demonstrating the effectiveness of our signal acquisition equipment.

B. Data Preprocessing Results

Delta, theta, alpha, beta, and gamma waves can be extracted from the raw EEG signals by WPD processing, as shown in Fig. 8. Furthermore, a 10-s raw ECG signal is collected for testing, as shown in Fig. 9. It is clear that there is an obvious random drift in the raw ECG signal. After WPD processing, the baseline drift can be largely removed, as shown in Fig. 9. Therefore, the data preprocessing method is effective.

C. Model Classification Results

The accuracy and loss curves of model training and validation are shown in Fig. 10. The performance comparison of the multimodal signal and the single-modal signal is shown in Table III. These results show that the performance of the multimodal signal is superior to those of two types of single-mode signals, and the precision, recall, F1-score, and accuracy are all larger than 95.7%. It demonstrates that the use of the multimodal signal can effectively improve the recognition accuracy of attention level. In other words, it is feasible to

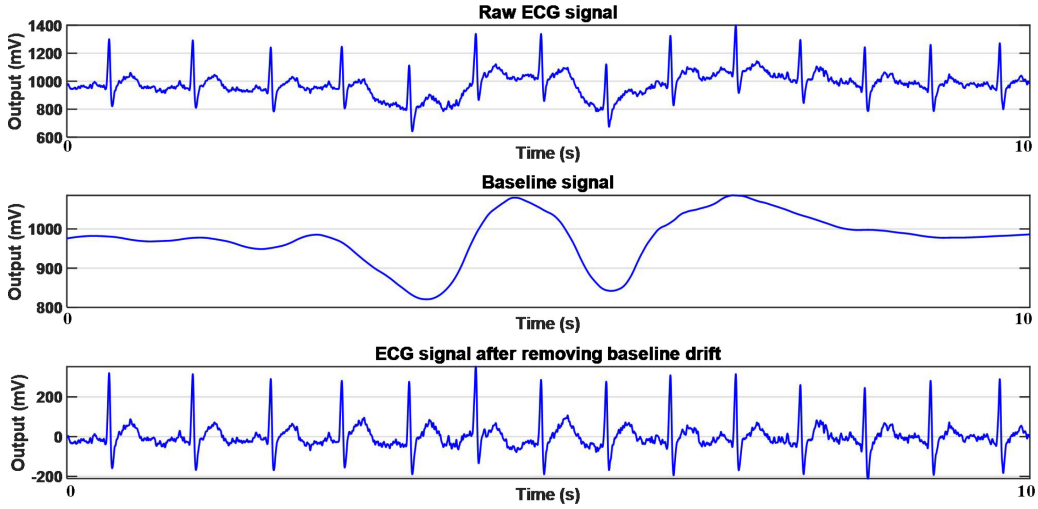

Fig. 9. Baseline drift in the raw ECG signal is removed by WPD processing.

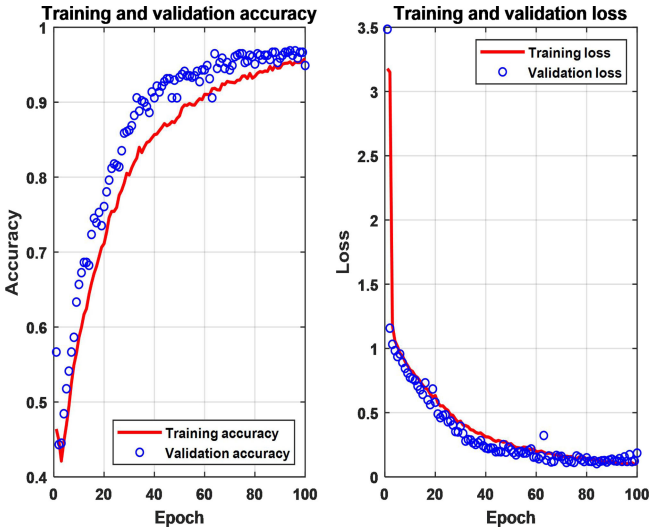

Fig. 10. Accuracy and loss curves of model training and validation.

TABLE III
PERFORMANCE COMPARISON OF THE MULTIMODAL SIGNAL AND THE SINGLE-MODAL SIGNAL

Signal	PR	RE	F1-score	ACC
EEG	89.11%	89.37%	89.24%	89.35%
ECG	75.37%	75.32%	75.34%	75.24%
EEG + ECG	95.77%	95.91%	95.84%	95.81%

use the ECG signal as a supplement to the single- channel EEG signal to improve the performance of the proposed NFT system.

D. Hardware Acceleration Results

The proposed high-performance CNN accelerator is fully implemented in Chisel3, and is evaluated as an System-on-Chip (SoC) design on Xilinx KU5P FPGA with a microblaze core. We fully synthesize, place and route it on Vivado 2021.1 and Vitis 2021.1, and the performance results are acquired by real runs on the FPGA. The FPGA device and layout for fulfilling the proposed CNN accelerator are shown in Fig. 11. After placement and routing, our design consists of 73630 LUTs, 32279 FFs, and 360 block RAMs, 160 DSP, operating at a frequency of 68 MHz with a power of 979 mW. The operating frequency is not pushed further yet to keep a balance between power and performance. On the real runs on the KU5P board, one inference only consumes 15.18 ms, while running the same workload on a 300 MHz MicroBlaze core takes 4.09 s, which clearly does not meet real-time requirements. Thus, the proposed FPGA prototype shows a significant 269× speedup compared to the embedded core. When compared to a high-performance AMD CPU (Ryzen 5 5600G core at 4 GHz), our accelerator still shows a 7.48× speedup. To make a fair performance comparison with related hardware designs, our accelerator is also synthesized with Design Compiler 2020 in the 28 nm technology node provided by TSMC.

Table IV shows the comparison results of our accelerator and state-of-the-art EEG-related accelerator ASIC designs. When the operating frequency of the proposed accelerator is pushed to 250 MHz in the ASIC design, the inference time decreases to 4.13 ms and the power increases to 157.36 mW. While our design achieves comparable frequency, area, and power efficiency, our inference latency is higher. This disparity arises because prior works target significantly smaller neural network models tailored for emotion recognition, resulting in orders-of-magnitude fewer operations and lower SRAM requirements. However, such compact models fall short in our application context, where high inference accuracy is critical, making their model architectures inadequate. Therefore, it is reasonable for our accelerator to have a slightly larger area and power consumption. Notably, our design supports 32-bit floating-point computation, enabling high numerical precision and broad applicability across complex DL workloads.

E. Attention Training Results

The variation curve of the average FSAQ of each group is shown in Fig. 12. The results of the one-way ANOVAs

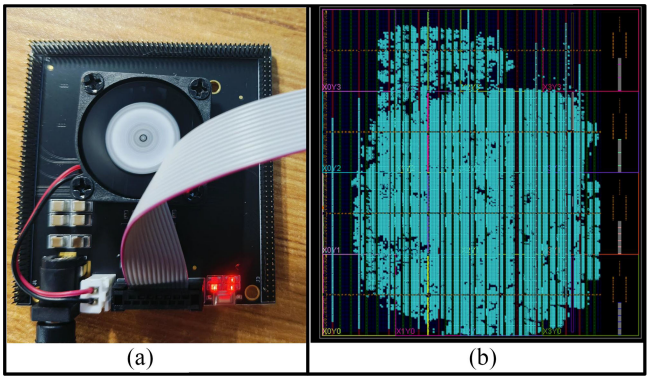

Fig. 11. (a) FPGA device for fulfilling the proposed CNN accelerator. (b) Layout for the proposed CNN accelerator.

TABLE IV
COMPARISON OF DIFFERENT HARDWARE ACCELERATOR DESIGNS

SPEC	[54]	[55]	[63]	[56]	Our work
Model	HDC	LRCN	CNN	HDC	CNN
Technology node	28 nm	16 nm	28 nm	16 nm	28 nm
Area (mm^2)	0.093	1.288	NA	0.596	1.388
Power (mW)	35.54	48.24	29.5	67.98	157.36
Frequency (MHz)	455	125	70	500	250
Inference time	1 us	1.9 us	0.284ms	628 ns	4.13 ms

and post hoc multiple comparisons are shown in Table V. The statistical results indicate that there is no significant difference in baseline FSAQs among the three groups ($p > 0.05$), proving that the baseline of the experiment is fair. After the experiment, the FSAQs of the NFT group are significantly higher than those of the other two groups ($p < 0.05$), and the FSAQs of the SGT group are significantly higher than those of the blank control group ($p < 0.05$). In addition, the average FSAQ of the NFT group increased by 14.50 compared to baseline FSAQ, while the average FSAQ of the SGT group only increased by 5.84 compared to baseline FSAQ. The above results indicate that both the proposed NFT system and serious games without the neurofeedback mechanism can effectively improve attention, and the former has a better effect than the latter. This further proves that the neurofeedback mechanism plays an important role in improving attention. It is worth noting that the FSAQs of the NFT group begin to be significantly higher than those of the blank control group after the second training, while the FSAQs of the SGT group begin to be significantly higher than those of the blank control group after the fifth training, which proves that the proposed NFT system takes effect faster than serious games without the neurofeedback mechanism.

As for the changes in attention-related brain activity of the subjects, the TBR_{cr} of the NFT group and SGT group during the training process is shown in Table VI. It can be observed that 5/6 of the subjects in the NFT group have a significant decrease in TBR compared to baseline, with an average decrease of 21.85%, while only half of the subjects in the SGT group have a significant decrease in TBR compared to baseline, with an average decrease of 17.14%. To sum up,

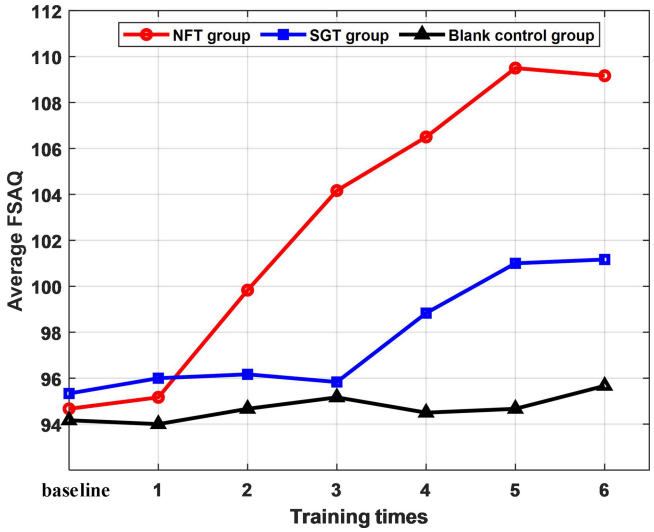

Fig. 12. Variation curve of the average FSAQ of each group in the attention training experiment.

the training effect of NFT group is the best, indicating that the proposed NFT system for attention improvement based on high-performance edge CNN accelerator is effective.

F. Discussion on Equipment Safety Compliance

Safety compliance is a necessary consideration for medical devices. The security compliance of the proposed NFT system can be discussed from the following two aspects.

In terms of hardware, first, the hardware devices used in the system are powered by built-in batteries with low operating voltage, so they are safe for the human body and there is no hidden danger of lightning surge. Second, electrostatic protection was applied to the circuit to eliminate the risk of electro-static discharge (ESD). Third, optocoupler isolation was applied to the equipment to suppress electromagnetic interference, which can improve the reliability of data transmission. Fourth, the PCB is placed inside a shielding shell, which can effectively reduce radiation. Furthermore, the material of the electrodes has good biocompatibility, and the devices used to fix the electrodes are made of flexible materials, so long-term wearing will not cause discomfort or skin damage. In terms of software, the virtual objects in the serious game we design are completely controlled by the user's attention level, which is different from traditional video games and therefore not easily addictive to users.

G. Discussion on Clinical Application and Implication

In general, the main audience for neurofeedback intervention is the healthy population and ADHD patients. As far as our study is concerned, the recruited subjects are all healthy young individuals. The results of the clinical controlled trial have preliminarily demonstrated that the proposed NFT system is effective in improving the attention of healthy adult populations. Therefore, in clinical practice, the proposed NFT system can be directly applied to train the attention of healthy young people. When facing patients with attention deficit problems, it is necessary to collect their EEG and

TABLE V
RESULTS OF THE ONE-WAY ANOVAS AND POST HOC MULTIPLE COMPARISONS

	Mean and standard deviation of FSAQ			One-way ANOVA		P-value of post hoc multiple comparison		
Times	NFT group (I)	SGT group (II)	Blank control group (III)	F	P-value	I vs II	I vs III	II vs III
baseline	94.67 ± 2.07	95.33 ± 3.50	94.16 ± 4.17	0.182	0.835	/	/	/
1	95.17 ± 2.48	96.00 ± 3.10	94.00 ± 3.79	0.602	0.560	/	/	/
2	99.83 ± 1.94	96.17 ± 2.48	94.67 ± 3.14	6.423	0.010	0.078	0.010	0.984
3	104.17 ± 1.83	95.83 ± 2.79	95.17 ± 4.67	13.759	< 0.001	0.002	0.001	1.000
4	106.50 ± 1.64	98.83 ± 3.76	94.50 ± 3.51	22.789	< 0.001	0.002	< 0.001	0.088
5	109.50 ± 3.27	101.00 ± 2.45	94.67 ± 4.18	29.185	< 0.001	0.002	< 0.001	0.016
6	109.17 ± 3.54	101.17 ± 2.14	95.67 ± 3.98	25.136	< 0.001	0.002	< 0.001	0.035

TABLE VI
TBR_{cr} OF THE NFT GROUP AND SGT GROUP DURING THE TRAINING PROCESS

	TBR_{cr} of the NFT group						TBR_{cr} of the SGT group					
Times	Subject1	Subject2	Subject3	Subject4	Subject5	Subject6	Subject7	Subject8	Subject9	Subject10	Subject11	Subject12
baseline	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
1	-3.46%	1.41%	-17.10%	-3.60%	33.27%	-14.01%	-20.02%	33.05%	0.04%	27.40%	-14.16%	33.06%
2	-6.89%	-21.27%	-21.10%	-5.12%	39.19%	-10.99%	-1.90%	13.11%	57.03%	-33.25%	-21.01%	30.31%
3	-11.13%	-33.26%	-23.46%	-7.33%	-78.21%	2.73%	-23.81%	11.17%	-1.46%	-13.29%	-38.61%	21.17%
4	-15.70%	-25.11%	-11.09%	-18.19%	2.47%	-12.80%	-41.09%	-0.80%	27.81%	-11.08%	-22.29%	-37.99%
5	-9.56%	-27.23%	-30.47%	-10.18%	83.16%	-16.19%	-3.46%	10.16%	-34.90%	-1.26%	1.96%	-3.40%
6	-10.73%	-34.15%	-31.89%	-10.78%	23.35%	-21.72%	-31.37%	12.31%	-7.69%	1.01%	5.79%	-12.35%

ECG data to retrain the classification model to ensure its performance, considering that the physiological data characteristics of patients may differ from those of healthy individuals. When facing ADHD patients with more severe conditions, neurofeedback intervention alone may have little effect and requires combined intervention with medication. As for the intervention plan, it varies for different target groups. For healthy population, typically 4–5 training sessions per week are sufficient. For ADHD patients, appropriate intervention plans should be developed based on the severity of symptoms, and typically 8-10 training sessions per week are necessary. It should be emphasized that the intervention plan needs to be adjusted accordingly based on the improvement of the patient's symptoms.

The most prominent potential benefits of the proposed NFT system are security and portability, as it has almost no potential side effects and is more convenient to use. Furthermore, as attention deficit problems are more common in children and adolescents, design based on the serious game is more in line with their interests and can improve their motivation for training. As for the limitations, the main problem with current neurofeedback intervention methods in clinical practice is the poor treatment effect on patients with severe ADHD symptoms. The main intervention method for such patients is medication, and neurofeedback intervention can generally only be used as an auxiliary means. Our method also failed to break through this limitation. In addition, ADHD patients (especially ADHD children) often exhibit excessive hyperactivity symptoms, which may cause signal acquisition devices to detach or electrodes to loosen. Therefore, interventions for such patients generally need to be completed with the accompaniment of a doctor or guardian.

IV. CONCLUSION

Attention training is meaningful and valuable for our daily learning and work. NFT is an effective and safe method for improving the attention level. To address some of the limitations of current NFT technique, a portable NFT system based on flexible IoT devices and an edge CNN accelerator is proposed in this article. More specifically, a multimodal bioelectrical signal containing single-channel EEG and ECG signals is used to train a lightweight CNN model, enabling it to accurately detect the subject's attention level during the NFT process. In order to achieve portability, a high-performance CNN accelerator is designed to run the CNN model. Then, a serious game is developed as a training tool. It can provide real-time attention feedback to subjects in visual and auditory forms to achieve closed-loop regulation of their attention levels. Finally, a single-blind controlled experiment is conducted to verify the effectiveness of the proposed system. The FSAQ of IVA-CPT and the attention-related EEG indicator TBR are adopted to evaluate the attention levels of the subjects. The experimental results show that the CNN

model can achieve an average recognition accuracy of 95.81%. And the average FSAQ of subjects trained with the NFT system increased by 14.50, and the average TBR decreased by 21.85%.

In summary, the proposed NFT system is very effective in improving attention and has the advantages of convenience and portability. Nevertheless, there are still some limitations to our work. First, the sample size for training and testing the CNN model is very limited. In the future, more samples containing a wider population will be used to train and test the model to further improve its accuracy and generalization ability. Second, the sample size for validating the proposed system is also limited. In the future, its effectiveness will be validated in a more diverse population, such as children and ADHD patients. Third, the attention of subjects receiving neurofeedback intervention has not been evaluated in the long term. For a clinical intervention system used for attention training, the lasting effect of training is a necessary consideration factor. However, attention is easily distracted. Factors such as environmental noise, excessive stress, fatigue, lack of sleep, and emotional problems can all cause attention fluctuations. These unpredictable and uncontrollable factors can affect the objectivity of assessing the lasting effects of NFT and make it difficult to maintain long-term improvements in attention. Therefore, we provide the following solution for follow-up studies reference.

Consider developing an app for monitoring attention, which includes a complete attention assessment methodology system, such as the IVA-CPT and TBR assessment used in this article. In this way, it is convenient to monitor users' attention levels for a long time. Once the user's attention level decreases or they find it difficult to concentrate during daily learning or work, timely NFT can be conducted.

Data Availability Statement:

The hardware data manuals, detailed code for signal preprocessing algorithms and CNN model, and engineering files for the designed CNN accelerator are publicly available at the link: https://github.com/hy509/A-Portable-NFT-System

Ethical Statement:

There are no potential risks associated with this research. All subjects have signed informed consent forms. The authors promise to protect any physiological data related to the personal privacy of the subjects generated in this research. This research has been approved by the Ethics Review Association of Guangdong University of Technology, with ethics approval number GDUTXS20250044.

REFERENCES

[1] M. G. M. Álvarez et al., "Effects of Web-based mindfulness training on psychological outcomes, attention, and neuroplasticity," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, Art. no. 22635.

[2] K.U. Yuka, Z. Shuo, and T. Motomi, "The effect of music intervention on attention in children: Experimental evidence," *Front. Neurosci.*, vol. 14, p. 757, Jul. 2020.

[3] A. Mehren et al., "Acute effects of aerobic exercise on executive function and attention in adult patients with ADHD," *Front. Psychiat.*, vol. 10, Mar. 2019, Art. no. 444174.

[4] A. M. Busch, C. A. Modica, and E. R. Sheridan, "The effect of yoga on anxiety, attention and social-emotional symptoms in preschool children: A pilot quasi-experimental study," *Child Psychiat. Human Develop.*, vol. 56, no. 2, pp. 570–579, 2025.

[5] G. Savulich, E. Thorp, T. Piercy, K. A. Peterson, J. D. Pickard, and B. J. Sahakian, "Improvements in attention following cognitive training with the novel 'decoder' game on an iPad," *Front. Behav. Neurosci.*, vol. 13, Jan. 2019, Art. no. 427567.

[6] S. Sarker, A. Md Linkon, F. H. Bappy, M. F. Rabbi, and M. Nahid, "Improving joint attention in children with autism: A VR-AR enabled game approach," *Int. J. Eng. Adv. Technol.*, vol. 10, no. 3, pp. 77–80, 2021.

[7] A. Jannati, L. M. Oberman, A. Rotenberg, and A. Pascual-Leone, "Assessing the mechanisms of brain plasticity by transcranial magnetic stimulation," *Neuropsychopharmacology*, vol. 48, no. 1, pp. 191–208, 2023.

[8] I. Lee, D. Kim, S. Kim, H. J. Kim, U. S. Chung, and J. J. Lee, "Cognitive training based on functional near-infrared spectroscopy neurofeedback for the elderly with mild cognitive impairment: A preliminary study," *Front. Aging Neurosci.*, vol. 15, Jul. 2023, Art. no. 1168815.

[9] A. Dehghani, H. Soltanian-Zadeh, and G.-A. Hossein-Zadeh, "Neural modulation enhancement using connectivity-based EEG neurofeedback with simultaneous fMRI for emotion regulation," *Neuroimage*, vol. 279, Oct. 2023, Art. no. 120320.

[10] S. L. Toepp, M. V. Mohrenschildt, and A. J. Nelson, "An EMG-based biofeedback system for tailored interventions involving distributed muscles," *IEEE Sensors J.*, vol. 23, no. 22, pp. 28095–28109, Nov. 2023.

[11] T. C. Ribeiro et al., "Assessing effectiveness of heart rate variability biofeedback to mitigate mental health symptoms: A pilot study," *Front. Physiol.*, vol. 14, 2023, Art. no. 1147260.

[12] J. H. Gruzelier, "EEG-neurofeedback for optimising performance. I: A review of cognitive and affective outcome in healthy participants," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 124–141, Jul. 2014.

[13] H.J. Engelbregt et al., "Short and long-term effects of sham-controlled prefrontal EEG-neurofeedback training in healthy subjects," *Clin. Neurophysiol.*, vol. 127, no. 4, pp. 1931–1937, 2016.

[14] S. Aliakbaryhosseinabadi, D. Farina, and N. Mrachacz-Kersting, "Real-time neurofeedback is effective in reducing diversion of attention from a motor task in healthy individuals and patients with amyotrophic lateral sclerosis," *J. Neural Eng.*, vol. 17, no. 3, 2020, Art. no. 36017.

[15] X. Luo et al., "A randomized controlled study of remote computerized cognitive, neurofeedback, and combined training in the treatment of children with attention-deficit/hyperactivity disorder," *Eur. Child Adolescent Psychiat.*, vol. 32, no. 8, pp. 1475–1486, 2023.

[16] L. Konicar et al., "Volitional modification of brain activity in adolescents with autism spectrum disorder: A Bayesian analysis of slow cortical potential neurofeedback," *NeuroImage Clin.*, vol. 29, Jan. 2021, Art. no. 102557.

[17] J. Micoulaud-Franchi, C. Jeunet, A. Pelissolo, and T. Ros, "EEG neurofeedback for anxiety disorders and post-traumatic stress disorders: A blueprint for a promising brain-based therapy," *Current Psychiat. Rep.*, vol. 23, pp. 1–14, Oct. 2021.

[18] S.Y. Wang et al., "The effects of alpha asymmetry and high-beta down-training neurofeedback for patients with the major depressive disorder and anxiety symptoms," *J. Affect. Disord.*, vol. 257, pp. 287–296, Oct. 2019.

[19] B. Wang, Z. Xu, T. Luo, and J. Pan, "EEG-based closed-loop neurofeedback for attention monitoring and training in young adults," *J. Healthcare Eng.*, vol. 2021, no. 1, 2021, Art. no. 5535810.

[20] M. Zhang, J. Zhang, and D. Zhang, "ATVR: An attention training system using multitasking and neurofeedback on virtual reality platform," in *Proc. IEEE Int. Conf. Artif. Intell. Virtual Reality (AIVR)*, 2019, pp. 159–1593.

[21] J. Kamiński, A. Brzezicka, M. Gola, and A. Wróbel, "Beta band oscillations engagement in human alertness process," *Int. J. Psychophysiol.*, vol. 85, no. 1, pp. 125–128, 2012.

[22] P. Kaushik, A. Moye, M. V. Vugt, and P. P. Roy, "Decoding the cognitive states of attention and distraction in a real-life setting using EEG," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 20649.

[23] I. P. Bodala, J. Li, N. V. Thakor, and H. Al-Nashash, "EEG and eye tracking demonstrate vigilance enhancement with challenge integration," *Front. Human Neurosci.*, vol. 10, p. 273, Jun. 2016.

[24] F. Fahimi, Z. Zhang, W. B. Goh, T. S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, 2019, Art. no. 26007.

[25] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, 2019, Art. no. 31001.

[26] H. Cai, Y. Zhang, H. Xiao, J. Zhang, B. Hu, and X. Hu, "An adaptive neurofeedback method for attention regulation based on the Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15829–15838, Nov. 2021.

[27] D. Mahmood, H. Nisar, R. Nawaz, V. V. Yap, and C. Y. Tsai, "Attention-related power and functional connectivity modulation associated with long-term alpha neurofeedback training," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, Art. no. 105431.

[28] Z. Hao, C. He, Y. Ziqian, L. Haotian, and L. Xiaoli, "Neurofeedback training for children with ADHD using individual beta rhythm," *Cogn. Neurodyn.*, vol. 16, no. 6, pp. 1323–1333, 2022.

[29] D. van Son, W. van der Does, G. P. Band, and P. Putman, "EEG theta/beta ratio neurofeedback training in healthy females," *Appl. Psychophysiol. Biofeedback*, vol. 45, no. 3, pp. 195–210, 2020.

[30] M. Blume, R. Schmidt, J. Schmidt, A. Martin, and A. Hilbert, "EEG neurofeedback in the treatment of adults with binge-eating disorder: A randomized controlled pilot study," *Neurotherapeutics*, vol. 19, no. 1, pp. 352–365, 2023.

[31] M. J. Kane and R. W. Engle, "The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective," *Psychonomic Bull. Rev.*, vol. 9, no. 4, pp. 637–671, 2002.

[32] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, "Classification of EEG signals to identify variations in attention during motor task execution," *J. Neurosci. Methods*, vol. 284, pp. 27–34, Jun. 2017.

[33] Z. Liang, X. Wang, J. Zhao, and X. Li, "Comparative study of attention-related features on attention monitoring systems with a single EEG channel," *J. Neurosci. Methods*, vol. 382, Dec. 2022, Art. no. 109711.

[34] G. Shahaf, U. Nitzan, G. Erez, S. Mendelovic, and Y. Bloch, "Monitoring attention in ADHD with an easy-to-use electrophysiological index," *Front. Human Neurosci.*, vol. 12, p. 32, Feb. 2018.

[35] L. W. Ko, O. Komarov, W. K. Lai, W. G. Liang, and T. P. Jung, "Eyeblink recognition improves fatigue prediction from single-channel forehead EEG in a realistic sustained attention task," *J. Neural Eng.*, vol. 17, no. 3, 2020, Art. no. 36015s.

[36] C. Carreiras, A. Lourenço, H. Aidos, H. P. da Silva, and A. L. Fred, "Unsupervised analysis of morphological ECG features for attention detection," in *Proc. Int. Joint Conf. Comput. Intell.*, 2015, pp. 437–453.

[37] A. Belle, R. H. Hargraves, and K. Najarian, "An automated optimal engagement and attention detection system using electrocardiogram," *Comput. Mathe. Methods Med.*, vol. 2012, no. 1, 2012, Art. no. 528781.

[38] H. Huang, M.-C. Shin, J. Lee, and S.-H. Yoon, "Effective neurofeedback training of large electroencephalogram signals using serious video games," *IEEE Access*, vol. 11, pp. 112175–112188, 2023.

[39] X. Wang, W. Huang, C. He, H. Wu, J. Lin, and L. Cheng, "A flexible EEG acquisition headband with high reliability and high signal-to-noise ratio," *IEEE Sensors J.*, vol. 24, no. 9, pp. 14370–14379, May 2024.

[40] Y. Chen, S. Martinez-Conde, S. L. Macknik, Y. Bereshpolova, H. A. Swadlow, and J.-M. Alonso, "Task difficulty modulates the activity of specific neuronal populations in primary visual cortex," *Nat. Neurosci.*, vol. 11, no. 8, pp. 974–982, 2008.

[41] D. A. Washburn and R. T. Putney, "Attention and task difficulty: When is performance facilitated?" *Learn. Motivation*, vol. 32, no. 1, pp. 36–47, 2001.

[42] P. Bouny, L. M. Arsac, Y. Pratviel, A. Boffet, E. Touré Cuq, and V. Deschodt-Arsac, "A single session of SMR-neurofeedback training improves selective attention emerging from a dynamic structuring of brain–heart interplay," *Brain Sci.*, vol. 12, no. 6, p. 794, 2022.

[43] M. Mohammadpour and S. Mozaffari, "Classification of EEG-based attention for brain computer interface," in *Proc. 3rd Iran. Conf. Intell. Syst. Signal Process. (ICSPIS)*, 2017, pp. 34–37.

[44] Y. Li, X. Li, M. Ratcliffe, L. Liu, Y. Qi, and Q. Liu, "A real-time EEG-based BCI system for attention recognition in ubiquitous environment," in *Proc. Int. Workshop Ubiquitous Affective Awareness Intell. Interact.*, 2011, pp. 33–40.

[45] W. Wan, X. Cui, Z. Gao, and Z. Gu, "Frontal EEG-based multi-level attention states recognition using dynamical complexity and extreme gradient boosting," *Front. Human Neurosci.*, vol. 15, Jun. 2021, Art. no. 673955.

[46] E. Ruiz-Marquez, A. Prieto, J. Mayas, P. Toril, J. M. Reales, and S. Ballesteros, "Effects of nonaction videogames on attention and memory in young adults," *Games Health J.*, vol. 8, no. 6, pp. 414–422, 2019.

[47] H. Y. Ho, M. D. Chen, C. C. Tsai, and H. M. Chen, "Effects of computerized cognitive training on cognitive function, activity, and participation in individuals with stroke: A randomized controlled trial," *NeuroRehabilitation*, vol. 51, no. 1, pp. 79–89, 2022.

[48] H. Chen, X. Hu, J. Gao, H. Han, X. Wang, and C. Xue, "Early effects of repetitive transcranial magnetic stimulation combined with sertraline in adolescents with first-episode major depressive disorder," *Front. Psychiat.*, vol. 13, Jul. 2022, Art. no. 853961.

[49] Y.-C. Wang et al., "A randomized, sham-controlled trial of high-definition transcranial direct current stimulation on the right orbital frontal cortex in children and adolescents with attention-deficit hyperactivity disorder," *Front. Psychiat.*, vol. 14, Feb. 2023, Art. no. 987093.

[50] I. Moreno-García, S. Meneres-Sancho, C. Camacho-Vara de Rey, and M. Servera, "A randomized controlled trial to examine the posttreatment efficacy of neurofeedback, behavior therapy, and pharmacology on ADHD measures," *J. Attention Disord.*, vol. 23, no. 4, pp. 374–383, 2019.

[51] T. P. Tinius, "The integrated visual and auditory continuous performance test as a neuropsychological measure," *Arch. Clin. Neuropsychol.*, vol. 18, no. 5, pp. 439–454, 2003.

[52] D. van Son, F. M. De Blasio, J. S. Fogarty, A. Angelidis, R. J. Barry, and P. Putman, "Frontal EEG theta/beta ratio during mind wandering episodes," *Biol. Psychol.*, vol. 140, pp. 19–27, Jan. 2019.

[53] A. Angelidis, W. van der Does, L. Schakel, and P. Putman, "Frontal EEG theta/beta ratio as an electrophysiological marker for attentional control and its test-retest reliability," *Biol. Psychol.*, vol. 121, pp. 49–52, Dec. 2016.

[54] A. Menon et al., "A highly energy-efficient hyperdimensional computing processor for biosignal classification," *IEEE Trans. Biomed. Circuits Syst.*, vol. 16, no. 4, pp. 524–534, Aug. 2022.

[55] C. J. Yang, W. C. Li, M. T. Wan, and W. C. Fang, "Real-time EEG-based affective computing using on-chip learning long-term recurrent convolutional network," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2021, pp. 1–5.

[56] J. Y. Li and W. C. Fang, "An edge AI accelerator design based on HDC model for real-time EEG-based emotion recognition system with RISC-V FPGA platform," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Singapore, 2024, pp. 1–5.

[57] T. Yuan, W. Liu, J. Han, and F. Lombardi, "High performance CNN accelerators based on hardware and algorithm co-optimization," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 1, pp. 250–263, Jan. 2021.

[58] N. P. Jouppi et al., "In-datacenter performance analysis of a tensor processing unit," in *Proc. ACM/IEEE 44th Annu. Int. Symp. Comput. Archit. (ISCA)*, Toronto, ON, Canada, 2017, pp. 1–12.

[59] S. Kala, B. R. Jose, J. Mathew, and S. Nalesh, "High-performance CNN accelerator on FPGA using unified Winograd-GEMM architecture," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 12, pp. 2816–2828, Dec. 2019.

[60] L. Bai, Y. Lyu, and X. Huang, "RoadNet-RT: High throughput CNN architecture and SoC design for real-time road segmentation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 2, pp. 704–714, Feb. 2021.

[61] S. Ran, B. Wu, K. Chen, and W. Liu, "VLCP: A high-performance FPGA-based CNN accelerator with vector-level cluster pruning," in *Proc. 18th ACM Int. Symp. Nanoscale Archit. (NANOARCH)*, New York, NY, USA, 2024, pp. 1–6.

[62] N. N. Schraudolph, "A fast, compact approximation of the exponential function," *Neural Comput.*, vol. 11, no. 4, pp. 853–862, May 1999.

[63] Y. D. Huang, K. Y. Wang, Y. L. Ho, C. Y. He, and W. C. Fang, "An edge AI system-on-chip design with customized convolutional-neural-network architecture for real-time EEG-based affective computing system," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Nara, Japan, 2019, pp. 1–4.

[64] R. Nooripour, M. G. Viki, N. Ghanbari, F. Farmani, and F. Emadi, "Alpha/theta neurofeedback rehabilitation for improving attention and working memory in female students with learning disabilities," *OBM Neurobiol.*, vol. 8, no. 3, pp. 1–20, 2024.

[65] S. Noble, E. Woods, T. Ward, and J. Ringwood, "Accelerating P300-based neurofeedback training for attention enhancement using iterative learning control: A randomised controlled trial," *J. Neural Eng.*, vol. 21, no. 2, 2024, Art. no. 26006.

[66] T. Suhail and A. Vinod, "An online neurofeedback game using consumer-grade EEG for improving attention and memory skills," in *Proc. 9th Int. Conf. Intell. Inf. Technol.*, 2024, pp. 487–492.

[67] M. Gacek et al., "Effects of school-based neurofeedback training on attention in students with autism and intellectual disabilities," *J. Autism Develop. Disord.*, vol. 2024, pp. 1–12, May 2024.

[68] H. Khan et al., "Visionary vigilance: Optimized YOLOV8 for fallen person detection with large-scale benchmark dataset," *Image Vis. Comput.*, vol. 149, Sep. 2024, Art. no. 105195.
[69] M. J. Ahmed et al., "CardioGuard: AI-driven ECG authentication hybrid neural network for predictive health monitoring in telehealth systems," *SLAS Technol.*, vol. 29, no. 5, Oct. 2024, Art. no. 100193.

Chunhua He received the B.S. degree in microelectronics from Sun Yat-sen University, Guangzhou, China, in 2010, and the M.S. and Ph.D. degrees in microelectronics and solid-state electronics from Peking University, Beijing, China, in 2013 and 2018, respectively.

From 2013 to 2017, he was an Engineer with the No.5 Electronics Research Institute, Ministry of Industry and Information Technology, Guangzhou, China. From 2017 to 2019, he was an Engineer with Midea Group, Foshan, China. From 2019 to 2021, he was a Senior Engineer with Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou, China. He joined the School of Computer, Guangdong University of Technology in 2021, Guangzhou, where he is currently an Associate Professor. From 2024 to 2025, he is also a Visiting Scholar with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. His current research interests include the design and application of MEMS sensors and artificial intelligence algorithms for biomedical engineering.

Wei Huang received the B.S. degree in software engineering from Beijing Union University, Beijing, China, in 2021. He is currently pursuing the M.S. degree in software engineering with the School of Computing, Guangdong University of Technology, Guangzhou, China.

His current research interests include artificial intelligence, brain–computer interfaces, and ADHD treatment.

Yi Huang received the B.E. degree in electrical science and Technology from the University of Science and Technology of China, Hefei, China, in 2021, and the master's degree in IC engineering from Tsinghua University, Beijing, China, in 2024.

He is currently a Research Assistant of IC Engineering with Tsinghua University. His current research interests include reconfigurable architecture design and compilation optimization for reconfigurable computing.

Haojie Wang received the B.S. degree in computer network engineering from Guangdong University of Technology, Guangzhou, China, in 2024.

His current research interests include neural networks and intelligent detection systems.

Heng Wu received the B.S. degree in electronic and information engineering from Wuhan University of Science and Technology, Wuhan, China, in 2009, and the M.S. and Ph.D. degrees in optics and mechanical manufacture and automation from South China University of Technology, Guangzhou, China, in 2012 and 2017, respectively.

He is currently an Associate Professor with the School of Automation, Guangdong University of Technology, Guangzhou. His research interests are in the fields of optical imaging, optical measurement, machine vision, and image processing.

Songqing Deng received the M.D. degree (8-year program) from Sichuan University, Chengdu, China, in 2013.

Since 2013, she has been with the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, where she is currently an Attending Doctor. She is currently specializing in perinatal medicine and high-risk pregnancy management. Her clinical and research interests include gestational diabetes mellitus, perinatal depression, and the application of AI in health management and medicine.

Maojin Liang received the master's degree (7-year program) and the M.D. degree from Sun Yat-sen University, Guangzhou, China, in 2009 and 2015, respectively.

Since 2009, he has been with the Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou, where he is currently a Deputy Chief Physician, and a Master Supervisor. He is currently specialized in otology and neurotology. His clinical and research interest includes development and influencing factors of auditory centers, middle ear related diseases, eustachian tube related diseases, and deaf children. He is also interested in ADHD treatment and the application of AI in health management.