IoT-29974-2023

A Novel Emotion Recognition Method Based on the Feature Fusion of Single-lead EEG and ECG Signals

Xiaoman Wang, Jianwen Zhang, Chunhua He*, Heng Wu*, Lianglun Cheng

***Abstract*—Emotions are complex, and people vary greatly in their accuracy in recognizing their own emotions and those of others. With advances in computer science and neuroscience, there is a desire to use automated techniques to help people identify emotions. Bio-electrical signals have been proven effective for emotion detection, but the acquisition of conventional ECG and EEG requires medical-specific equipment, which is very expensive, uncomfortable, and inconvenient due to the large number of electrodes and the hair-covered scalp. In this paper, a novel emotion recognition method based on the feature fusion of single-lead EEG and ECG signals is proposed, using the LSTM-MLP-based model and the CNN-based model for feature fusion and classification, respectively, with five-fold cross-validation for validation. The ECG and EEG signals of 15 participants were collected in five states: happy, relaxed, calm, sad, and afraid, each of which was stimulated using the participants' own proposed music. Various time-domain features, frequency-domain features, and nonlinear features were extracted from the ECG and EEG signals. Experimental results demonstrate that the accuracy of emotion recognition and classification of signals captured by the proposed device can reach 92.08% using the CNN model. While using the LSTM-MLP feature fusion model, the accuracy figure can be improved to 95.07%. The results of the ablation experiment indicate that the feature fusion approach does improve the accuracy of recognition. It is demonstrated that the proposed device and emotional recognition approach are effective and feasible.**

***Index Terms*—EEG, ECG, HRV, wearable convenience devices, emotion recognition, feature fusion networks**

I. Introduction

Emotions play a key role in people's daily perception, cognition, and decision-making processes [1], which is critical in human communication and interaction. However, there is no scientific consensus among scientists and psychologists on the definition of emotions, and people vary greatly in their accuracy in identifying their own and others' emotions [2]. As the fields of computer science and artificial intelligence evolve, there is a desire to use these techniques to help people recognize emotions. This is a relatively new area of research. In the field of healthcare, if the emotions of patients can be accessed in real time, doctors can adjust their treatment measures at the right time, which is beneficial to the patient's treatment process. Emotion recognition systems also have a wide range of applications in many fields, such as education and entertainment [3].

When people are in different emotional states, their various physiological indicators respond differently, which is why many scientists record biological signals and images for emotion recognition. Common biological signals are ECG, EMG, EEG, respiratory rate, blood pressure, etc. Among these measurements, bioelectric signals represent neural signals, which provide a direct measure of human emotional responses and have attracted increasing interest in the field in recent years [4].

The voltage fluctuations generated by ionic currents in neurons in the brain are recorded by the EEG signal. Placing electrodes at the scalp allows measurement of the synchronized sum of postsynaptic potentials during excitation of numerous pyramidal cells, which is a non-invasive method of electrophysiological monitoring. The EEG spectrum ranges from 0.5 to 45 Hz and is divided into five different frequency bands [5]. There have been many studies [6]–[8] that have used EEG for emotion recognition and classification, and they have used various linear and nonlinear features of EEG, such as the energy of the 5 bands, Fractal Dimension (FD) [9]–[11], Detrended Fluctuation Analysis (DFA) [12], Hurst Exponent (HE) [13], and Approximate Entropy (ApEn) [14].

Numerous studies have already found that emotion recognition based on multichannel EEG and ECG signals is feasible with high accuracy. However, these devices are usually set up in hospitals or laboratories for medical examinations and scientific studies. With their poor comfort level and complex operation, it often requires a long experiment preparation phase. With regard to the convenience of use, it is difficult to industrialize these devices as they are usually not convenient for daily life, entertainment, or real-

Manuscript received May 17, 2023; revised July 19, 2023, September 1, 2023; accepted September 25, 2023. This work was partially supported by the National Natural Science Foundation of China (Grant Nos. U22A2012, 62104047 and 62173098), Natural Science Foundation of Guangdong Province (Grant No. 2023A1515010291), and Basic and Applied Basic Research Project of Guangzhou Basic Research Program (Grant No. 2023A04J1707). (Corresponding author: Chunhua He, Heng Wu).

Xiaoman Wang, Jianwen Zhang, Chunhua He and Lianglun Cheng are with the School of Computer, Heng Wu is with the School of Automation, Guangdong University of Technology, Guangzhou 510000, China. (e-mail: 2112205067@mail2.gdut.edu.cn, 3121004845@mail2.gdut.edu.cn, hechunhua@pku.edu.cn, llcheng@gdut.edu.cn, heng.wu@foxmail.com).

time feedback for disease treatment. As for the cost, these devices are very expensive.

There are also many publicly available datasets available to researchers conducting research on emotion recognition, such as DEAP, SEED, etc. [15]. But one of the common features of these databases is the use of dedicated, non-portable, and expensive equipment to capture the contained biosignals. Although the use of such devices allows the study of affect-related EEG and physiological signals, it limits the application of the proposed algorithms in the limited and controlled space that can accommodate the use of such devices [16].

In order to apply EEG-based emotion recognition technology in daily life, convenient wearable devices are inevitable, and electrode distribution needs to be reduced from the whole brain to the prefrontal lobe [17]. The prefrontal cortex is located in the anterior part of the frontal lobe of the brain and is involved in processes related to higher-level cognitive activity, personality performance, and decision-making. By simply applying the electrodes to the forehead, the EEG signal can be measured without hair interference, making the measurement convenient. It has been shown [18] that emotion accumulation is positively correlated with the left middle frontal lobe (the dorsolateral part of the prefrontal cortex), and emotion recognition based on prefrontal EEG may have great promise for real-life applications. [16] captured EEG and ECG signals in different emotional states using portable, wearable, wireless, low-cost, and off-the-shelf devices, demonstrating through his research the promise of using low-cost devices for emotion recognition applications. Sachin Taran's proposed method [19] utilized single bipolar EEG channels to classify emotions and greatly reduced the complexity of emotion recognition-based BCI systems. [20] achieved binary emotion recognition using single-channel EEG, demonstrating that single-channel EEG is sufficient for emotion recognition.

The reduction in the number of electrodes collected brings about a reduction in the amount of EEG data, which inevitably reduces the accuracy of the recognition. To compensate for the reduced accuracy, other physiological signals are introduced, and multimodal features are adopted by scientists. Facial microexpressions [21], electrocardiograms (ECG) [22], and speech signals [23] were introduced for emotion recognition tasks. Acquisition of facial micro-expressions usually requires a camera to capture video recordings, and speech signals are captured from the subject's voice during a conversation; in comparison, features of heart rate are more convenient to acquire.

The time interval between heartbeats is not fixed; the vibrations of a healthy heart are complex and constantly changing [24]. Heart rate variability, which is the fluctuation of the time interval between adjacent heartbeats, can be used to analyze the ongoing interplay of sympathetic and parasympathetic influences on heart rate, generating information about autonomic flexibility and thus representing the ability to regulate emotional responses [25]. The R wave is the wave with the highest amplitude in each heartbeat waveform and can be used to identify each heartbeat, usually to compute HRV by calculating the RR interval. There have been many studies that have used heart rate variability for affective classification [3], using features such as time-domain features, frequency-domain features, and nonlinear features [26]–[28]. With the 12 leads used in conventional medical ECG to determine the status of each part of the heart [29], HRV only needs to locate the R-peak, so only the I-lead in the limb leads is needed to measure the waveform of the ECG and locate the R-peaks.

Therefore, to address the above-mentioned advantages and disadvantages, we would like to develop a set of convenient wearable devices that can capture bioelectric signals in daily life and provide real-time feedback on the user's emotional state, including a brain ring for collecting EEG signals (see **Fig. 1**) and a bracelet for collecting ECG signals. For the convenience of the device, the electrode acquisition part of the brain ring will touch the part of the forehead that is not covered by hair. To implement such a device, we have the following problems to solve:

(1) design a circuit that can capture bioelectric signals;

(2) design a data processing method to acquire features with physiological and physical significance.

(3) design models for feature fusion and emotion recognition, and the algorithm for recognition has to be lightweight enough to be embedded in our device to implement edge computing.

In this paper, an EEG acquisition device based on prefrontal EEG and an ECG acquisition device based on limb I leads are proposed. The EEG and ECG signals of participants in different emotional states are acquired using our self-developed devices, and the wavelet decomposition algorithm is introduced to calculate the HRV features of the ECG and various time-domain features, frequency-domain features, and nonlinear features of the EEG. Finally, the LSTM-MLP feature fusion model and the CNN model are used for emotion recognition and classification of multimodal features based on EEG features and HRV features, respectively. A five-fold cross-validation method is demonstrated to illustrate the results of identification. Ablation experiments were performed to investigate the contribution of EEG and HRV features.

Fig.1. Conceptual diagram of a brain loop for capturing EEG signals

II. Methods

A. Devices and Circuits

For the acquisition of ECG and EEG signals, separate circuits were designed, both with a sampling rate of 1 kHz.

The ECG signal is a highly sensitive bioelectrical signal. The most common frequencies in the ECG signal are 0.67-5 Hz; QRS wave complexes can be detected in the range of 10-50 Hz; and T waves can be detected in the range of 1-7 Hz. The main sources of ECG signal noise are motion artifacts (5-50 Hz), baseline drift, and industrial frequency interference (50/60 Hz). To attenuate various interferences, a band-pass filter of 0.16-34 Hz is used in our ECG acquisition circuit equipment with only one channel. The EEG signal is generated under the skull and other tissues and is already very weak when transmitted to the scalp. EEG has a spectral range of 0.5-45 Hz and is often subject to cardiac, myoelectric, and industrial frequency interference. In order to attenuate various interferences, a band-pass filter of 0.16–80 Hz is used in our EEG acquisition circuit equipment. An 18650 lithium battery is used to power the device in order to eliminate industrial frequency interference. An experimental platform for bioelectricity acquisition was established, as shown in **Fig. 2**(a). A series of experiments can be performed on this test platform.

The circuit used to acquire the bioelectric signal is shown in **Fig. 2**(b). Vdd and Vss are connected to the positive and negative power supplies, respectively. Among them, the Instrumentation Amplifier (IA) is used to amplify the differential input signal by approximately 100–1000 times. To avoid saturation distortion of the amplified signal, two first-order High Pass Filters (HPF) are used to filter the DC voltage. Besides, two first-order Low Pass Filters (LPF) are used to suppress high-frequency noise and obtain a clear, objective signal. To add a DC offset to the AC signal, an adder is used to make the output signal between 0 and Vdd, and then the output is obtained through the Analog to Digital Converter (ADC) of the Microprocessor Control Unit (MCU).

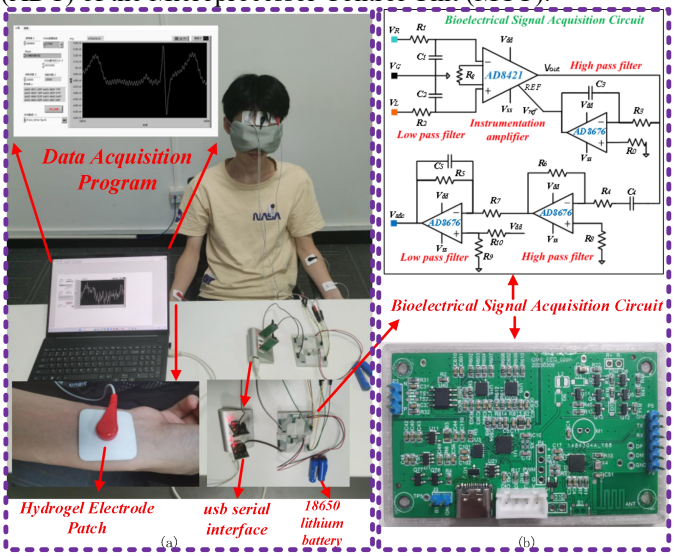

Fig.2. (a). The experimental platform of bioelectrical acquisition experiment. (b). The Bioelectrical Signal Acquisition (BSA) circuit

To reduce noise, the AD8421 was selected as a high-precision instrumentation amplifier as well as the AD8676 as a noise analyzer, both of which were designed and manufactured by Analog Devices, Inc. In addition, the CS32A010 (manufactured by Chippsee Co., Shenzhen, China) was selected as the 32-bit MCU (ARM Cortex-M0 core).

Circuits for ECG and EEG acquisition are similar in structure, but the main difference lies in the parameter settings of the resistor and capacitor. Since the amplitude of the raw ECG signal is almost 10 times higher than that of the raw EEG signal, the total gain required for the latter is almost 10 times higher than that of the former. Rg is used to adjust the gain of IA, which is set to 100Ω for ECG acquisition and to 10Ω for EEG acquisition. The DC offset is equal to half of Vdd. In this way, the total gain of the ECG signal acquisition circuit is 1000, while the total gain of the ECG signal acquisition circuit is 9910. Hence, tiny bioelectric signals can be filtered and amplified.

During the experiment, the subject's skin was lightly wiped with alcohol, and then a hydrogel snap-on electrode sheet was applied to the subject's skin for acquiring bioelectric signals. A USB device was used to connect the input of the circuit to the laptop computer, and a program written in LabVIEW was used to control the start and end of the acquired signals. The acquired signals were stored on the hard disk of the laptop computer during the signal acquisition process. During the experiment, the three acquisition electrodes of the ECG board were attached to the left and right upper limbs, and the three acquisition electrodes of the EEG board were attached to the forehead.

B. Participants

Fifteen participants took part in this experiment (male: 13, female: 2, age: 23.93 ± 3.53). All participants had normal hearing, no history of neurological, psychiatric, or other serious illnesses that might affect the results of the experiment, and were not involved in professional music training. Prior to the experiment, each participant was informed of the details of the procedure and signed an informed consent form.

C. Experiment

Music is often used to stimulate emotions, but each person's preference for music may affect the intensity and activation of emotions [30]–[32]. Thus, each participant was asked to select four songs from their familiar songs for the experiment, which belonged to the four emotions of happy, relaxed, sad, and scared, corresponding to high valence-high arousal (HVHA), high valence-low arousal (HVLA), low valence-high arousal (LVHA), and low valence-low arousal (LVLA) in Russell's two-dimensional model of emotion [33]. The experimental flow is shown in **Fig. 3**. During the experiment, resting-state ECG and EEG were collected for 5 min, and then participants were asked to listen to the songs of their choice for 5 min each, with a break of 1 min. Single-channel prefrontal EEG and ECG were collected using the equipment described in II.A. Hence, the dataset used in this paper was obtained, and the differences with other publicly available datasets are shown in Table I.

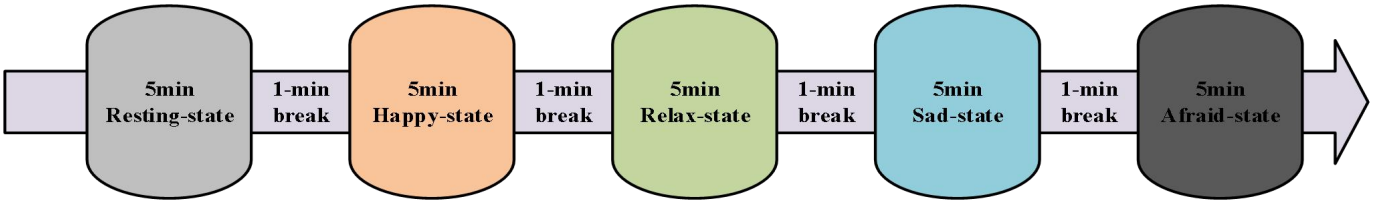

Fig. 3. Experimental procedure

TABLE I
COMPARISON BETWEEN DIFFERENT DATASETS

Dataset	Particip ants	stimulus material	data	emotions
SEED-V [34]	16	video	EEG, eye movement	5: Happy, Sad, Disgust, Neutral, Fear
DREAM ER [16]	23	audio-visual	EEG, ECG	9:amusement, excitement, happiness, calmness, anger, disgust, fear, sadness and surprise.
Our dataset	15	music	EEG, ECG	5:calm, happy, relax, sad, afraid

D. Data Preprocessing and Feature Extraction

1) Data expansion

Due to the small amount of sample data, it is necessary to expand the data. Kim et al. [21] used a 1s sliding window to segment the EEG data into small segments, sliding 1% at a time with 99% overlap. Sepúlveda et al. [22] split the ECG signal into 20-s segments for analysis. Through research, psychologists have found that a video stimulus lasting at least 1 minute can induce a single emotion [16], [35]. Hence, to align the ECG and EEG data segments, the details of the modality used are as follows: a sliding window of 1 minute in length was used to cut the ECG and EEG data, sliding 10% at a time. The EEG data and ECG data from the same time period were aligned as one sample.

2) Feature extraction and processing

Wavelet packet decomposition (WPD), also known as optimal subband tree structure, is able to perform modern time-frequency analysis and processing of various types of non-smooth random signals effectively. The wavelet packet transform can decompose the collected signal into multiple two-dimensional parameters (time, position, and frequency) to realize the feature decomposition of the signal in different frequency bands and different moments. WPD can be achieved by the Mallat algorithm or the lifting algorithm. In this paper, the sampling frequency of the devices is set to 1000 Hz, so the reconstructed signal of the 10th level node 0 represents the low-frequency signal in the 0-0.5 Hz band, where $0.5 \approx 1000/2^{10+1}$. The structure diagram of the decomposition tree of the ten-level tower wavelet transform is shown in **Fig. 4**.

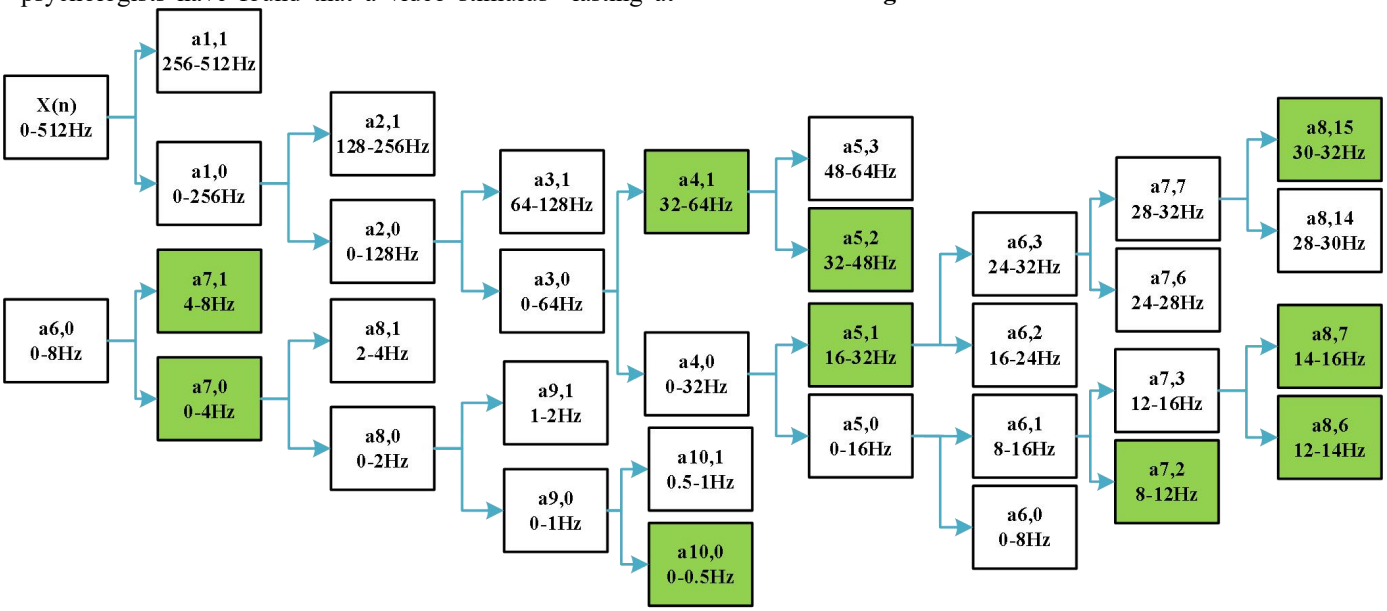

Fig. 4. WPD process for a decomposition level of 10

For EEG data, the 1 minute EEG data was divided into frames with a frame length of 4 s and a frame repetition rate of 50%, and the 1 minute EEG data was divided into 29 frames, and then the data of each frame was subjected to feature extraction. Firstly, the signal was processed using wavelet decomposition reconstruction to remove the baseline data below 0.5 Hz, and then 5 features were extracted, including δ-wave energy (0.5-4 Hz), θ-wave energy (4-8 Hz), α-wave energy (8-14 Hz), β-wave energy (14-30 Hz), γ-wave energy (>30 Hz). The relevant

expressions are as follows:

$$\begin{cases} wpt_{EEG} = wpdec(X_{\text{EEG}},10,'db5') \\ baseline_{EEG} = wprcoef(wpt_{EEG},[10\ \ 0]) \\ delta_{EEG} = wprcoef(wpt_{EEG},[7\ \ 0]) - baseline_{EEG} \\ theta_{EEG} = wprcoef(wpt_{EEG},[7\ \ 1]) \\ alpha_{EEG} = wprcoef(wpt_{EEG},[7\ \ 2]) \\ \qquad + wprcoef(wpt_{EEG},[8\ \ 6]) \\ beta_{EEG} = wprcoef(wpt_{EEG},[8\ \ 7]) \\ \qquad + wprcoef(wpt_{EEG},[5\ \ 1]) \\ \qquad - wprcoef(wpt_{EEG},[8\ \ 15]) \\ gamma_{EEG} = wprcoef(wpt_{EEG},[8\ \ 15]) \\ \qquad + wprcoef(wpt_{EEG},[5\ \ 2]) \end{cases} \tag{1}$$

where, *wpdec* is a one-dimensional wavelet packet decomposition function that receives three parameters, namely the raw signal, the decomposition level, and the wavelet function, and returns a wavelet packet tree object. *Wprcoef* is a wavelet packet reconstruction function that is used to reconstruct the signal and can receive a one-dimensional or two-dimensional input signal. In this paper, the selected wavelet packet decomposition has a level of 10, and the selected wavelet function is: *db5*. wpt_{EEG} is a wavelet packet tree object corresponding to the 10th-level decomposition of the input signal X_{EEG}. [10 0] denotes the first node of the 10th level decomposition. $delta_{ECG}$, $theta_{EEG}$, $alpha_{EEG}$, $beta_{EEG}$, and $gamma_{EEG}$ are the signals of δ, θ, α, β, and γ waves after WPD processing. Fast Fourier Transform (FFT) is used to convert the time domain signal to the frequency domain. The energy integrals of the corresponding frequency bands are used to calculate the energy of each band. The relevant expressions are as follows:

$$\begin{cases} DeltaEnergy = \int_{0.5}^{4} DelAm^2(f)df \\ ThetaEnergy = \int_{4}^{8} TheAm^2(f)df \\ AlphaEnergy = \int_{8}^{14} AlpAm^2(f)df \\ BetaEnergy = \int_{14}^{30} BetAm^2(f)df \\ GammaEnergy = \int_{30}^{45} GamAm^2(f)df \end{cases} \tag{2}$$

where *DelAm*, *TheAm*, *AlpAm*, *BetAm*, and *GamAm* are the amplitudes of δ, θ, α, β, and γ-wave spectra after WPD analysis and FFT, respectively.

Then, the other 10 features of the EEG signal are calculated using the PyEEG [36] package, including Petrosian Fractal Dimension (PFD), Higuchi Fractal Dimension (HFD), Hjorth Mobility and Complexity, Spectral Entropy, Singular Value Decomposition Entropy (SVD Entropy), Fisher Information, and Approximate Entropy (ApEn), Detrended Fluctuation Analysis (DFA), and Hurst Exponent (Hurst). These 10 features, together with the 5 previously introduced features, constitute the features of one frame of the EEG signal. A 29*15 feature matrix was obtained from 1 minute of EEG data.

For ECG data, the standards for HRV measurements were described by a joint European and US working group in 1996 [37]. In this paper, the signal was processed using a wavelet packet decomposition and reconstruction-based method for each minute of ECG data to remove the baseline drift. The relevant expressions are as follows:

$$\begin{cases} wpt_{ECG} = wpdec(X_{\text{ECG}},10,'db5') \\ baseline_{ECG} = wprcoef(wpt_{ECG},[10\ \ 0]) \\ Y_{ECG} = X_{ECG} - baseline_{ECG} \end{cases} \tag{3}$$

where, $baseline_{ECG}$ is the baseline drift within 0-0.5 Hz and Y_{ECG} is the ECG signal after debaselining.

Then the location of the R peaks was located to obtain the RR interval data. The following conditions were used to determine whether a peak was an R-peak or not:

1. The R-peak is the highest peak in the ECG cluster, so a queue value of MinPeakHeight is set, and peaks above this queue value enter the alternative set.

2. The number of heartbeats in a minute for a normal human is within the range of 60-120, so the spacing of R peaks is between 0.5 s and 1 s. Then the ECG data is segmented, and the length of each segment is 0.5 s. If there are multiple peaks within a segment, the highest one enters the alternative set.

3. The peak prominence of the R peak will be more prominent than the peaks of other waves, so a peak prominence queue is set, MinPeakProminence, and the peak prominence exceeds this queue into the alternative set.

After locating the position of the R-peaks, the mean heart rate (*HRMEAN*) can be calculated over the interval. The spacing matrix RR between adjacent R peaks can also be calculated, and some key time-domain parameters about HRV can be obtained based on the RR array, including the mean value of RR intervals (*RRMEAN*), the overall standard deviation (*SDNN*), the root mean square of successive differences (*RMSSD*), the standard deviation of successive differences (*SDSD*), and the number of adjacent RR intervals with a difference greater than 50 ms (*PNN50*).

The frequency-domain analysis of HRV is also very important. When analyzing HRV from the frequency domain perspective, the power spectrum estimation is performed first. In this paper, Welch's spectral density estimation function is adopted for power spectrum estimation. Hence, the frequency domain analysis parameters for HRV can be obtained based on the power spectrum estimation, including the power in the very low frequency band (*VLF*), the power in the low frequency band (*LF*), the power in the high frequency band (*HF*), and the total signal power (*TP*).

Due to the inevitable limitations of using only linear analysis methods, nonlinear parameters are also introduced into the HRV analysis. In this paper, two non-linear parameters, *SD1* and *SD2*, were introduced.

In summary, a 1*12 feature matrix can be obtained from 1 minute of ECG data.

E. Model

Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, K Nearest Neighbor, Gaussian Plain Bayes, and Multi-Layer Perceptron Classifier were adopted as baseline models. Besides, two models were designed and applied for emotion recognition and classification.

Long Short Term Memory (LSTM) is an artificial neural network suitable for classification and processing based on time series data, as well as correcting for gradient disappearance [38]. Multiple LSTM layers are used for the pre-processing of the EEG feature matrix. A multilayer perceptron is a forward-structured artificial neural network that maps a set of input vectors to a set of output vectors. The HRV features are pre-processed using a multilayer perceptron. A feature fusion model was designed to linearly fuse the ECG feature processing results with the EEG feature matrix processing results, as shown in **Fig. 5(a)**, and then a multilayer perceptron was used for classification [39].

The ECG features first pass through 2 linear implicit layers with 128 and 64 nodes, respectively, and output a tensor of size 1*64, which is named *B1*. The EEG features first pass through two bidirectional LSTM layers with 64 nodes each, and then pass through a LSTM layers with outputs of 32 and another LSTM layers with 16 nodes. The last column vector of the final output is taken and named *B2*. The two outputs *B1* and *B2* are linearly connected into *B*, as shown in (4), and then passed through two linear implicit layers, and finally, five nodes representing five different states are output.

$$B = \begin{bmatrix} B1 & B2 \end{bmatrix} \tag{4}$$

Another model used is based on convolutional neural networks. Convolutional layers have been widely adopted for a variety of tasks, including emotion recognition [40]. ECG and EEG features are first expanded and linearly spliced, and then fed into a deep network for learning. The model contains two convolutional layers and two pooling layers with a convolutional kernel size of 3 * 3, as shown in **Fig. 5(b)**. The padding size is set to 1, and the pool size is set to 2 × 2. The output of the deep network is then flattened and passed through two linear implicit layers, with the final output of five nodes representing five different states.

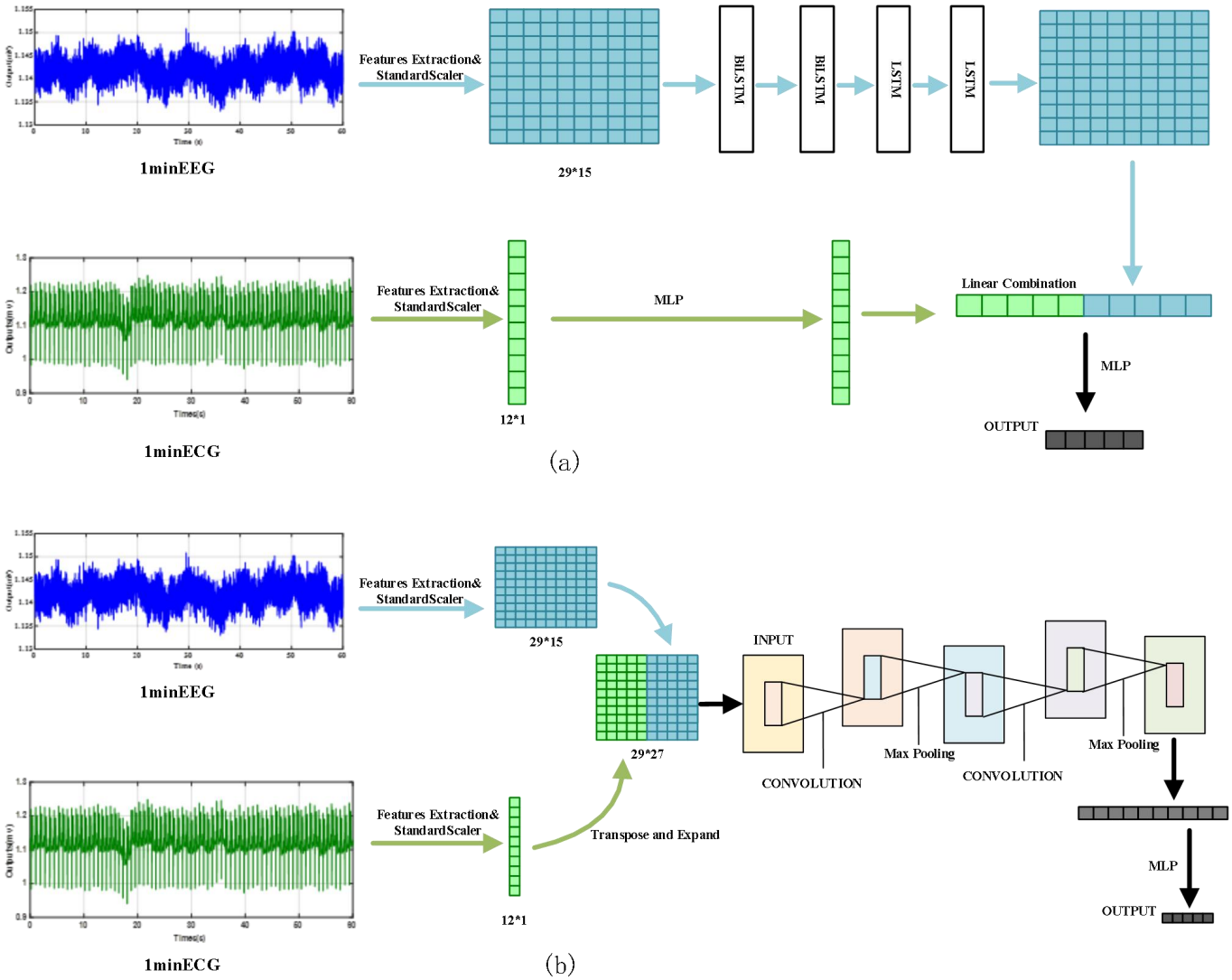

Fig. 5. (a) LSTM-MLP Feature Fusion Model Structure Diagram; (b) CNN Model Structure Diagram

The backpropagation algorithm is a gradient descent algorithm for training a multilayer artificial neural network. The network is trained using the backpropagation algorithm, and the loss is calculated for each training epoch using the cross-entropy loss as the loss function. In this paper, the learning rate of the model is set to 0.001, and the model is trained for 200 epochs to find the least lossy result.

Five-fold cross-validation was adopted for model training and testing, and each feature is standardized. In this paper, the normalization process is as follows: each column of data in the training set is normalized with the training set normalized parameters, and then the test set data is also normalized with the training set normalized parameters.

III. Results

A. EEG Analysis Results

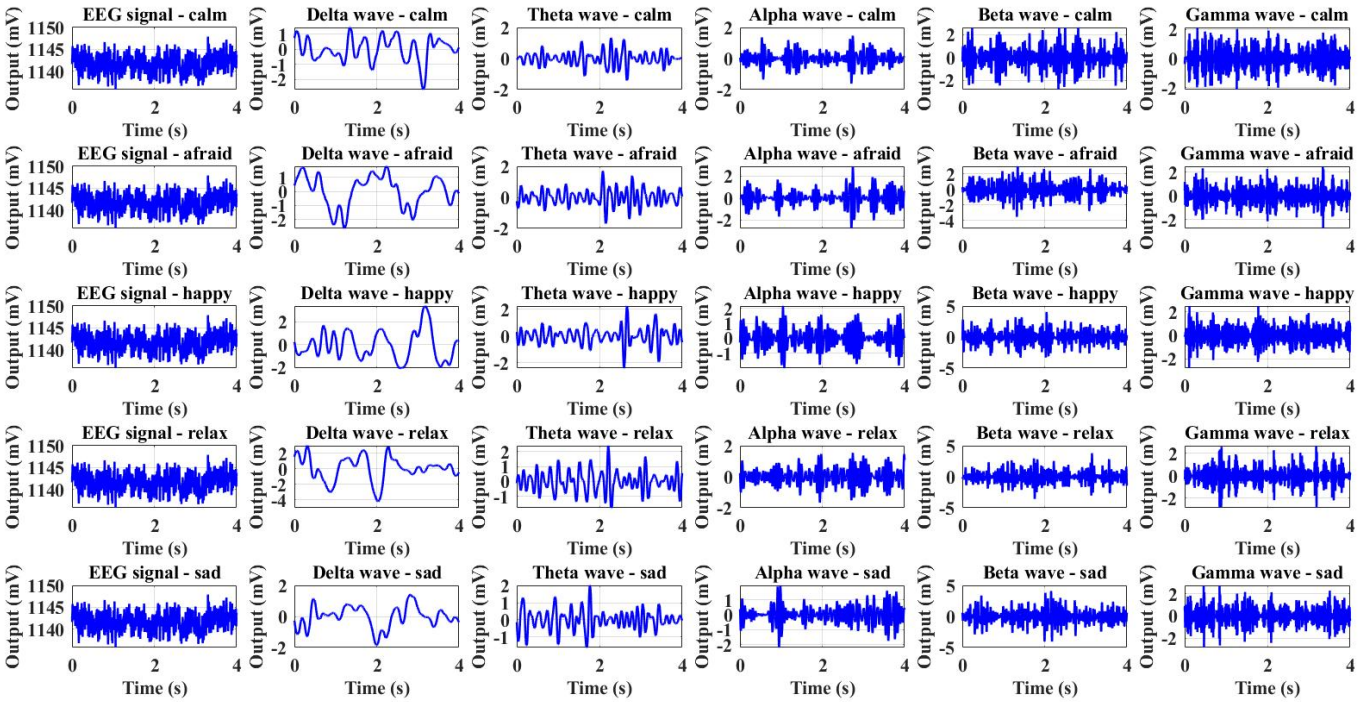

Fig. 6. Delta, theta, alpha, beta and gamma waves are separated from the raw EEG signal of the five emotional states with WPD analysis.

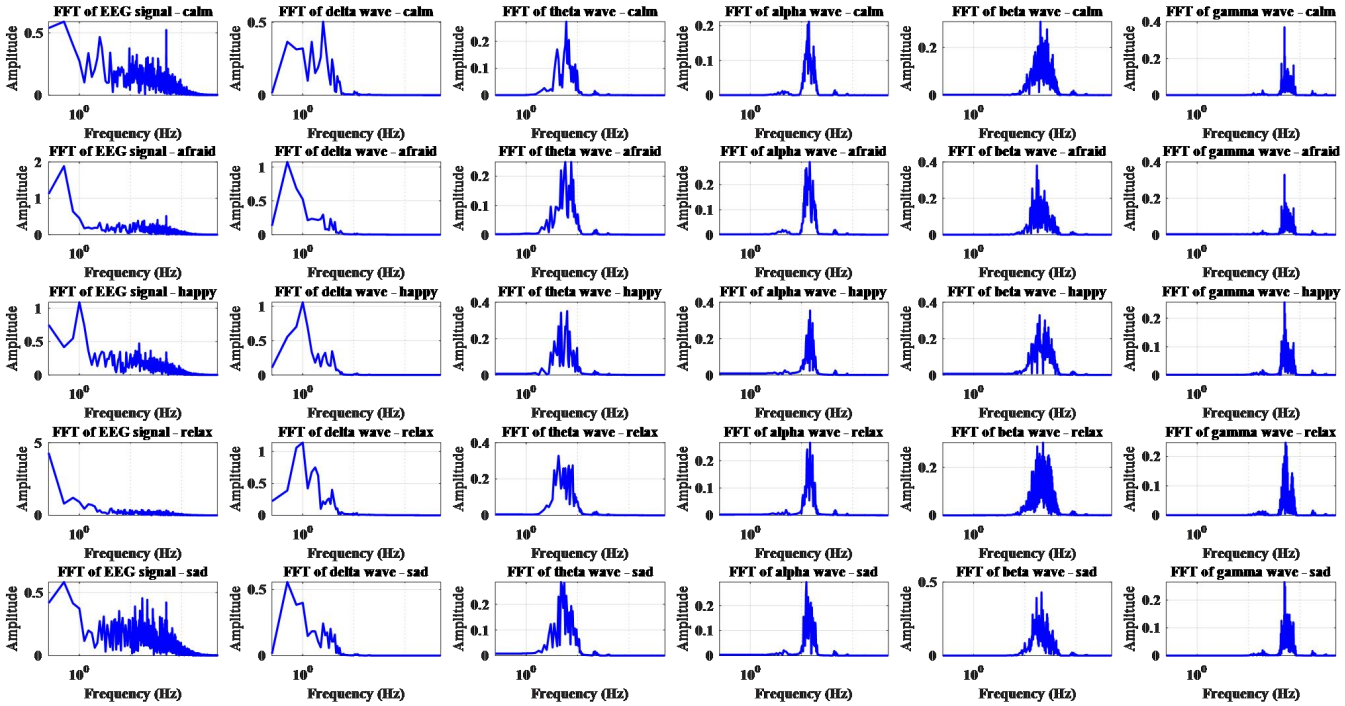

Fig. 7. FFT analysis is conducted to delta, theta, alpha, beta and gamma waves of the five emotional states, and the spectrums of the five waves according with the definitions.

The wavelet packet decomposition and reconstruction algorithm was performed on the raw EEG signal obtained from the BSA circuit, from which the delta, theta, alpha, beta, and gamma waves were separated, as shown in **Fig. 6**. The five waves were subjected to FFT processing to convert the time-domain signal into a frequency-domain signal, as shown in **Fig. 7**. It can be seen from the figure that the frequency spectra of the five waves meet the definition, thus the proposed signal acquisition and separation method is effective. According to the equation proposed in the previous section, the EEG features can be calculated.

B. HRV Analysis Results

The wavelet packet decomposition and reconstruction algorithm is performed on the raw ECG signal obtained from the BSA circuit, from which the baseline drift signal is separated. After removing the baseline drift from the raw signal, R-peak localization analysis was performed. In the ECG waveform in the figure, the R-peak can be accurately located, which is consistent with the results of manual R-peak labeling. Therefore, the BSA circuit and the peak detection method are effective. Hence, the RR interval array and HRV characteristics can be obtained. The raw and baseline ECG signals and the R-peak localization after de-baselining for the five moods are shown in **Fig. 8.**

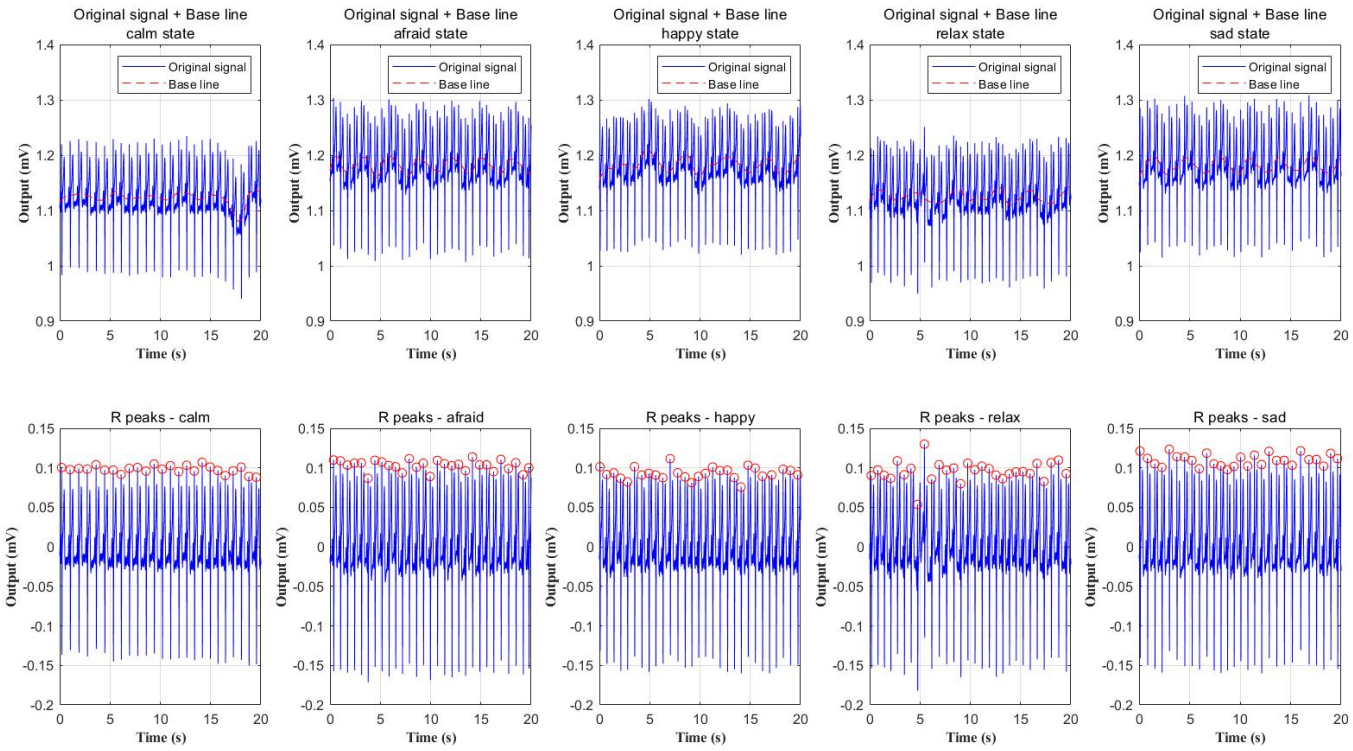

Fig. 8. The raw and baseline signal and.results of removing the baseline and R-peak localization of ECG of five states.

C. Emotion Recognition and Classification Results

The precision (*PR*), recall (*RE*), *F1-score*, and accuracy (*ACC*) were defined to evaluate the performance of the classification model.

$$\begin{cases} PR = \dfrac{TP}{TP + FP} \\ RE = \dfrac{TP}{TP + FN} \\ F1 - Score = \dfrac{2 \times precision \times recall}{precision + recall} \\ ACC = \dfrac{TP + TN}{TP + TN + FP + FN} \end{cases} \tag{5}$$

The recognition abilities of the two models and the baseline models are shown in Table II, which shows that the precision, recall, f1 score, and accuracy of the two models for emotions are higher than those of the baseline models. The recognition abilities of the two models for the five emotional states are shown in Table III, with the values representing the mean of the five-fold cross-validation. As for the CNN model, the recognition precision for afraid, calm, happy, and relax exceeds 90%. While using the LSTM-MLP feature fusion model, the recognition precision exceeds 95% for five emotional states and exceeds 96% for calm emotions. The accuracy of the CNN model for emotion recognition and classification reaches 92.08%. While using the LSTM-MLP feature fusion model, the accuracy can be improved by 2.93%, achieving 95.07%. By comparing, it can be seen that the LSTM-MLP feature fusion model outperforms the CNN in every recognition ability. Furthermore, the experimental results of both models are much greater than the chance probability of 20%, which indicates that it is effective and feasible to capture bioelectrical signals for emotion recognition by the proposed device.

TABLE II
ACCURACY METRIC FOR EVALUATING CLASSIFICATION MODELS

model	*PR*	*RE*	*F1-Score*	*ACC*
LogisticRegression	0.32	0.31	0.31	0.36
SVC	0.44	0.39	0.37	0.44
DecisionTreeClassifier	0.57	0.57	0.57	0.59
RandomForestClassifier	0.69	0.66	0.67	0.70
KNeighborsClassifier	0.42	0.42	0.41	0.46
GaussianNB	0.33	0.29	0.27	0.35
MLPClassifier	0.46	0.45	0.45	0.48
LSTM-MLP	0.9479	0.9469	0.9470	0.9507
CNN	0.9149	0.9135	0.9135	0.9208

TABLE III
ACCURACY METRIC FOR FIVE STATES

	LSTM-MLP			CNN		
	PR	*RE*	*F1-Score*	*PR*	*RE*	*F1-Score*
afraid	0.9404	0.9611	0.9505	0.9112	0.9079	0.9088
calm	0.9649	0.9723	0.9685	0.9584	0.9696	0.9639
happy	0.9422	0.9440	0.9428	0.9058	0.9045	0.9047
relax	0.9409	0.9289	0.9342	0.9086	0.8965	0.9018
sad	0.9513	0.9281	0.9391	0.8904	0.8891	0.8884
mean	0.9479	0.9469	0.9470	0.9149	0.9135	0.9135

D. Ablation Experiment Results

Through model training, we found that the LSTM-MLP feature fusion model outperforms the CNN model in terms of recognition ability, but the contribution of EEG features and HRV features to the recognition ability of the model is unknown. Hence, we designed and conducted ablation experiments based on the LSTM-MLP feature fusion model to investigate the recognition ability of single EEG features and single HRV features on emotions.

TABLE IV
THE RESULTS OF THE ABLATION EXPERIMENT ON LSTM-MLP FEATURE FUSION MODEL

features	*PR*	*RE*	*F1-Score*	*ACC*
EEG	0.9285	0.9258	0.9267	0.9316
HRV	0.7620	0.7564	0.7567	0.773
EEG-HRV	0.9479	0.9469	0.9470	0.9507

Experimental results show that when only HRV features are applied for emotion recognition and classification, the recognition accuracy is 77.3%. When only EEG features are adopted, the accuracy improves to 93.16%. However, the accuracy achieves the best of 95.07% based on the features fusion of the EEG and ECG signals. It demonstrates that the feature fusion approach is effective and feasible. Both EEG and HRV features are significant in the emotion recognition task and can complement each other.

IV. CONCLUSION

In this paper, a circuit and device were first proposed for acquiring bio-electric signals, and then ECG and EEG signals were collected from 15 participants in different emotional states stimulated by music, and emotion recognition and classification were performed. In the data pre-processing process, a wavelet packet transform-based raw signal processing algorithm was presented. For the ECG signals, R-peak localization is performed, and various HRV feature values are calculated. For EEG signals, energy features of various bands and various nonlinear features are extracted. Finally, the LSTM-MLP feature fusion model and the CNN model were designed and used for emotion recognition and classification, respectively. The experimental results show that the LSTM-MLP feature fusion model outperforms the CNN model in every dimension for the recognition of signals captured by the device. The accuracy of the CNN model for emotion recognition and classification reaches 92.08%. While using the LSTM-MLP feature fusion model, the accuracy can be improved by 2.99%, achieving 95.07%. The results of the ablation experiment demonstrate that the recognition accuracy using only EEG features (accuracy: 93.16%) or only HRV features (accuracy: 77.3%) is lower than that of EEG-HRV feature fusion (accuracy: 95.07%). Therefore, the suggested method of emotion recognition and classification based on the proposed circuits is effective and feasible. This low-cost, non-invasive, and convenient method can be applied to wearable devices for entertainment, education, and the therapeutic process of mood disorder diseases.

Considering that human emotions are complex in real life and each person has individual differences, more experiments are needed to verify the stability and applicability of the method, and the next step is to integrate the method into a convenient wearable device. We will integrate our circuits and models into the EEG headband and heart rate band, transfer data using Bluetooth and WiFi, and complete it as an IoT product so that it can be widely used in daily life.

REFERENCES

[1] J. Panksepp, *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press, 2004.

[2] B. Garcia-Martinez, A. Martinez-Rodrigo, R. Alcaraz, and A. Fernandez-Caballero, "A Review on Nonlinear Methods Using Electroencephalographic Recordings for Emotion Recognition," *IEEE Trans. Affective Comput.*, vol. 12, no. 3, pp. 801–820, Jul. 2021, doi: 10.1109/TAFFC.2018.2890636.

[3] M. A. Hasnul, N. A. A. Aziz, S. Alelyani, M. Mohana, and A. A. Aziz, "Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare—A Review," *Sensors*, vol. 21, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/s21155015.

[4] Y. Ding, X. Hu, Z. Xia, Y.-J. Liu, and D. Zhang, "Inter-Brain EEG Feature Extraction and Analysis for Continuous Implicit Emotion Tagging During Video Watching," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 92–102, Jan. 2021, doi: 10.1109/TAFFC.2018.2849758.

[5] S. Sanei, *Adaptive processing of brain signals*. John Wiley & Sons, 2013.

[6] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "EEG-Based BCI Emotion Recognition: A Survey," *Sensors*, vol. 20, no. 18, p. 5083, Sep. 2020, doi: 10.3390/s20185083.

[7] N. S. Suhaimi, J. Mountstephens, and J. Teo, "EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–19, Sep. 2020, doi: 10.1155/2020/8875426.

[8] R. Kumawat and M. Jain, "EEG based Emotion Recognition and Classification: a Review," *IRJASH*, vol. 3, no. Special Issue ICITCA-2021 5S, pp. 1–10, May 2021, doi: 10.47392/irjash.2021.131.

[9] S. Hatamikia and A. M. Nasrabadi, "Recognition of emotional states induced by music videos based on nonlinear feature extraction and som classification," in *2014 21th Iranian Conference on Biomedical Engineering (ICBME)*, IEEE, 2014, pp. 333–337. doi: 10.1109/ICBME.2014.7043946.

[10] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer*, vol. 32, pp. 347–358, 2016, doi: 10.1007/s00371-015-1183-y.

[11] E. Ruiz-Padial and A. J. Ibáñez-Molina, "Fractal dimension of EEG signals and heart dynamics in discrete emotional states," *Biological psychology*, vol. 137, pp. 42–48, 2018, doi: 10.1016/j.biopsycho.2018.06.008.

[12] A. Banerjee *et al.*, "Study on brain dynamics by non linear analysis of music induced EEG signals," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 110–120, 2016, doi: 10.1016/j.physa.2015.10.030.

[13] S. Z. Bong, K. Wan, M. Murugappan, N. M. Ibrahim, Y. Rajamanickam, and K. Mohamad, "Implementation of wavelet packet transform and non linear analysis for emotion classification in stroke patient using brain signals," *Biomedical signal processing and control*, vol. 36, pp. 102–112, 2017, doi: 10.1016/j.bspc.2017.03.016.

[14] W.-L. Chu, M.-W. Huang, B.-L. Jian, and K.-S. Cheng, "Analysis of EEG entropy during visual evocation of emotion in schizophrenia," *Annals of general psychiatry*, vol. 16, pp. 1–9, 2017, doi: 10.1186/s12991-017-0157-z.

[15] M. Zangeneh Soroush, K. Maghooli, S. K. Setarehdan, and A. Motie Nasrabadi, "A Review on EEG Signals Based Emotion Recognition," *Int Clin Neurosci J*, vol. 4, no. 4, pp. 118–129, Oct. 2017, doi: 10.15171/icnj.2017.01.

[16] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan. 2018, doi: 10.1109/JBHI.2017.2688239.

[17] Z. Gao, X. Cui, W. Wan, W. Zheng, and Z. Gu, "Long-range correlation analysis of high frequency prefrontal electroencephalogram oscillations for dynamic emotion recognition," *Biomedical Signal Processing and Control*, vol. 72, p. 103291, Feb. 2022, doi: 10.1016/j.bspc.2021.103291.

[18] M. Résibois *et al.*, "The neural basis of emotions varies over time: different regions go with onset- and offset-bound processes underlying emotion intensity," *Social Cognitive and Affective Neuroscience*, vol. 12, no. 8, pp. 1261–1271, Aug. 2017, doi: 10.1093/scan/nsx051.

[19] S. Taran and V. Bajaj, "Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method," *Computer Methods and Programs in Biomedicine*, vol. 173, pp. 157–165, May 2019, doi: 10.1016/j.cmpb.2019.03.015.

[20] S. Toraman and Ö. Dursun, "GameEmo-CapsNet: Emotion Recognition from Single-Channel EEG Signals Using the 1D Capsule Networks," *Traitement du Signal*, vol. 38, pp. 1689–1698, Dec. 2021, doi: 10.18280/ts.380612.

[21] H. Kim, D. Zhang, L. Kim, and C.-H. Im, "Classification of Individual's discrete emotions reflected in facial microexpressions using electroencephalogram and facial electromyogram," *Expert Systems with Applications*, vol. 188, p. 116101, Feb. 2022, doi: 10.1016/j.eswa.2021.116101.

[22] A. Sepúlveda, F. Castillo, C. Palma, and M. Rodriguez-Fernandez, "Emotion Recognition from ECG Signals Using Wavelet Scattering and Machine Learning," *Applied Sciences*, vol. 11, no. 11, p. 4945, May 2021, doi: 10.3390/app11114945.

[23] Y. Shen, H. Yang, and L. Lin, "Automatic Depression Detection: An Emotional Audio-Textual Corpus and a GRU/BiLSTM-based Model." arXiv, Feb. 14, 2022. doi: 10.48550/arXiv.2202.08210.

[24] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health*, vol. 5, 2017, doi: 10.3389/fpubh.2017.00258.

[25] B. M. Appelhans and L. J. Luecken, "Heart Rate Variability as an Index of Regulated Emotional Responding," *Review of General Psychology*, vol. 10, no. 3, pp. 229–240, Sep. 2006, doi: 10.1037/1089-2680.10.3.229.

[26] C.-J. Yang, N. Fahier, W.-C. Li, and W.-C. Fang, "A Convolution Neural Network Based Emotion Recognition System using Multimodal Physiological Signals," in *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, Sep. 2020, pp. 1–2. doi: 10.1109/ICCE-Taiwan49838.2020.9258341.

[27] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay, and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2019, doi: 10.1109/ACCESS.2018.2883213.

[28] A. S. Anusha, J. Jose, S. P. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomed. Phys. Eng. Express*, vol. 4, no. 6, p. 065001, Sep. 2018, doi: 10.1088/2057-1976/aadbd4.

[29] C. and Sensors, "12 Lead ECG Placement Guide | Cables & Sensors," *Cables and Sensors*. https://www.cablesandsensors.com/pages/12-lead-ecg-placement-guide-with-illustrations (accessed May 05, 2023).

[30] K. Corrigall, "Music: The language of emotion," 2013, pp. 299–326.

[31] M. Reybrouck, P. Vuust, and E. Brattico, "Brain Connectivity Networks and the Aesthetic Experience of Music," *Brain Sciences*, vol. 8, no. 6, Art. no. 6, Jun. 2018, doi: 10.3390/brainsci8060107.

[32] C. L. Krumhansl, "Music: A Link Between Cognition and Emotion," *Curr Dir Psychol Sci*, vol. 11, no. 2, pp. 45–50, Apr. 2002, doi: 10.1111/1467-8721.00165.

[33] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[34] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 715–729, Jun. 2022, doi: 10.1109/TCDS.2021.3071170.

[35] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition and Emotion*, vol. 24, no. 7, pp. 1153–1172, Nov. 2010, doi: 10.1080/02699930903274322.

[36] F. S. Bao, X. Liu, and C. Zhang, "PyEEG: An Open Source Python Module for EEG/MEG Feature Extraction," *Computational Intelligence and Neuroscience*, vol. 2011, p. e406391, Mar. 2011, doi: 10.1155/2011/406391.

[37] "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.

[38] A. Garg, A. Kapoor, A. K. Bedi, and R. K. Sunkaria, "Merged LSTM Model for emotion classification using EEG signals," in *2019 International Conference on Data Science and Engineering (ICDSE)*, Sep. 2019, pp. 139–143. doi: 10.1109/ICDSE47409.2019.8971484.

[39] Y.-P. Lin, C.-H. Wang, T.-L. Wu, S.-K. Jeng, and J.-H. Chen, "Multilayer perceptron for EEG signal classification during listening to emotional music," in *TENCON 2007 - 2007 IEEE Region 10 Conference*, Oct. 2007, pp. 1–3. doi: 10.1109/TENCON.2007.4428831.

[40] G. Du, W. Zhou, C. Li, D. Li, and P. X. Liu, "An Emotion Recognition Method for Game Evaluation Based on Electroencephalogram," *IEEE Trans. Affective Comput.*, vol. 14, no. 1, pp. 591–602, Jan. 2023, doi: 10.1109/TAFFC.2020.3023966.

XIAOMAN WANG received the B.S. degree in Network Engineering from South China Normal University, Guangzhou, China, in 2020. She is now pursuing the M.S. degree in Guangdong University of Technology, Guangzhou, China. Her current research interests include biomedical engineering and artificial intelligence.

JIANWEN ZHANG is currently pursuing the B.S. degrees in Computer Science and Technology at the School of Computer, Guangdong University of Technology. His research interests include EEG and emotion recognition.

CHUNHUA HE received the B.S. degree in microelectronics from Sun Yat-sen University, Guangzhou, China, in 2010, and the M.S. degree and Ph.D. degree in microelectronics 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 Co., Ltd., Guangzhou, China. He joined the School of Computer, Guangdong University of Technology in 2021, Guangzhou, China, where he is currently an Associate Professor. His current research interests include the design and application of MEMS sensors and artificial intelligence algorithms for biomedical engineering.

HENG WU received the B.S. degree in electronic and information engineering from the Wuhan University of Science and Technology, Wuhan, China, in 2009, and received the M.S. and Ph.D. degrees in optics and mechanical manufacture and automation from the South China University of Technology, Guangzhou, China, in 2012 and 2017, respectively. He joined the School of Automation, Guangdong University of Technology in 2017, Guangzhou, China, where he is currently an Associate Professor. His research interests are in the fields of optical imaging, optical measurement, machine vision, and image processing.

LIANGLUN CHENG received the B.S. degree and M.S. degree in automatic control engineering from Huazhong University of Science and Technology, Wuhan, China, in 1988 and 1992, respectively. He received the Ph.D. degree in automatic control engineering from Institute of Automation of Chinese Academy of Sciences, Beijing, China, in 1999.

He joined the School of Computer, Guangdong University of Technology in 1992, Guangzhou, China, where he is currently a Professor and Dean. His current research interests include internet of things and information physical integration system, interconnection and fusion of heterogeneous network, big data of industrial process, and high performance computing.

Prof. Cheng is currently an expert with special subsidy from the government of the State Council, director of the national and local joint Engineering Research Center for integrated technology of Intelligent Manufacturing Information Physics fusion system, director of the Key Laboratory of Guangdong Information Physics fusion system, and deputy director of Qian Xuesen Innovation Committee of CASS, Director of China automation society, member of China computer society, vice director of Guangdong automation society, executive vice president of Guangdong Robotics Society.