SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

June 2026, Vol. 69, Iss. 6, 162204:1–162204:16
https://doi.org/10.1007/s11432-025-4830-4

CFAN: convolutional frequency-attention network for ECG-based emotion recognition

Ziman YE[1], Hao ZHENG[1], Geng HAN[1] & Fang DENG[1,2*]

[1]*School of Automation, Beijing Institute of Technology, Beijing 100081, China*
[2]*Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China*

Received 10 February 2025/Revised 6 July 2025/Accepted 23 August 2025/Published online 13 May 2026

Abstract Emotion recognition using electrocardiogram (ECG) signals is a promising research direction with broad applications ranging from healthcare to human-computer interaction. However, mainstream convolutional neural network (CNN) and transformer-based models rarely exploit frequency-domain characteristics of ECG signals explicitly, limiting their cross-subject generalization capability. To address these limitations, we propose a convolutional frequency-attention network (CFAN) that integrates a frequency-aware attention mechanism with dynamic convolution to effectively extract and combine frequency and temporal features. CFAN comprises three key components: a frequency-aware attention module (FAAM), an attention-guided convolution neural network (AG-CNN) block, and a multilayer perceptron (MLP) classifier, all working synergistically to enhance the efficacy of emotion recognition. We evaluate CFAN using leave-one-subject-out cross-validation by employing the WESAD dataset and further fine-tune the framework using data for individual subjects to reduce the inter-subject variability. CFAN outperforms state-of-the-art methods, achieving an accuracy of 76.06% and an F1-score of 0.75, providing an accurate and efficient solution for ECG-based emotion recognition.

Keywords emotion recognition, electrocardiogram, convolutional neural network, attention mechanism, signal processing

Citation Ye Z M, Zheng H, Han G, et al. CFAN: convolutional frequency-attention network for ECG-based emotion recognition. Sci China Inf Sci, 2026, 69(6): 162204, https://doi.org/10.1007/s11432-025-4830-4

1 Introduction

The expression of emotions plays a pivotal role in human communication, serving as a foundation for social interaction and understanding. Affective computing, a multidisciplinary field, seeks to endow machines with human-like capabilities to observe, interpret, and generate affective features, thereby enabling more natural and intuitive human-machine interactions [1]. Emotion recognition (ER), a cornerstone of this endeavor, leverages advancements in artificial intelligence, psychophysiology, cognitive science, and computer science [2, 3]. The applications of ER span various domains, demonstrating transformative potential in areas such as entertainment [4], healthcare [5, 6], education [7], and conversational agents [8, 9].

Emotion characteristics can be extracted from various multimodal signals, including facial expressions [10, 11], speech [12,13], conversation [14,15], body gestures [16,17], and physiological signals [18,19]. However, the reliability of modalities other than physiological signals (e.g., facial expressions, speech, text, and body gestures) is often limited by social masking, where individuals may consciously or unconsciously conceal their true emotions [20]. In contrast, physiological signals are less prone to voluntary control, providing more accurate and objective measures of emotional states [21]. Furthermore, ER based on physiological signals is well-suited for integration into human-computer interaction systems such as somatosensory and wearable health monitoring devices, offering enhanced responsiveness and improved user engagement [22, 23].

ER based on physiological signals leverages diverse data sources, including electrocardiography (ECG) [24, 25], electrodermal activity (EDA) [26,27], photoplethysmography (PPG) [28,29], respiration (RESP) [30], electromyography (EMG) [31,32], eye movements [33,34], and electroencephalography (EEG) [35–37]. Among these physiological signals, ECG has earned distinction as a promising modality for ER because of its high sensitivity to emotional changes, non-invasive nature, and cost-effective measurement capabilities [23, 38]. ECG measures changes in potential using electrodes placed on the body's surface, capturing the electrical activity of the heart [39]. Agrafioti

* Corresponding author (email: dengfang@bit.edu.cn)

et al. [40] demonstrated a strong correlation between ECG signals and emotional states, further highlighting the potential of ECG for advancing emotion recognition research.

Significant research efforts have been devoted to ECG-based ER. Existing approaches can be broadly categorized into two main streams: (1) methods that rely on feature engineering, such as the extraction of frequency-domain features [41–43], time-domain features [42, 44, 45], and nonlinear characteristics [41, 42], which are then fed into machine learning models [43,45] or neural networks for classification; and (2) methods that directly utilize raw ECG signals as the input, employing advanced model architectures such as convolutional neural networks (CNNs) [46–48] and Transformers [46, 49]. These architectures have demonstrated effectiveness in general spatial and temporal feature extraction but are often inefficient for processing ECG signals because of their high temporal resolution and quasi-periodic nature. In contrast to existing reports, this study proposes a novel approach that explicitly incorporates frequency information to address these limitations. The main contributions of this study are as follows.

(1) Existing studies rarely utilize the quasi-periodicity of the ECG and are thus inefficient. This study proposes a novel model for ECG-based ER named convolutional frequency-attention network (CFAN), which captures frequency information and then uses it to enhance the 1D convolution of temporal ECG signals.

(2) CFAN first employs the fast-Fourier transform (FFT) and an attention mechanism to identify and extract a set of frequency factors specific to each ECG sample. Using these identified frequency factors, CFAN then performs dynamic convolutions to obtain frequency-aware temporal features.

(3) We evaluate the proposed method by employing leave-one-subject-out (LOSO) cross-validation settings using the public WESAD dataset. We also explore the performance by fine-tuning the settings, demonstrating that CFAN can effectively capture ECG features and consistently improve the classification performance.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive review of previous studies on ECG-based ER. Section 3 details the architecture and components of the proposed model. Section 4 describes the experimental setup in detail. Section 5 presents the experimental results and analysis. Finally, Section 6 presents the conclusion of the study and discusses potential future directions.

2 Related work

2.1 ER and ECG

ECG signals are quasi-periodic in nature and exhibit characteristic waveforms within each cardiac cycle, including the P-wave, QRS complex, and T-wave [50]. ECG signals carry rich physiological information and have thus emerged as one of the most widely studied biosignals for ER [51]. Agrafioti et al. [40] demonstrated a strong correlation between ECG patterns and the emotional states of humans, highlighting the potential of ECG for emotion recognition tasks.

Wang et al. [41] exploited the time-frequency domain features, waveform characteristics, and nonlinear properties of ECG signals for classifying the emotional states of drivers by employing a back-propagation neural network. Similarly, Hsu et al. [42] extracted 34 features spanning the time domain, frequency domain, and nonlinear dynamics, using a least squares support vector machine (LS-SVM) as the classifier. Sepúlveda et al. [43] introduced wavelet scattering for extracting multi-scale features from ECG signals, enabling simultaneous analysis in the time and frequency domains. Khan et al. [52] leveraged the continuous wavelet transform for generating input features and designed a CNN-LSTM architecture for classifying emotional states. Chen et al. [44] focused on the statistical and rhythmic features extracted from ECG signals, utilizing a Bi-LSTM model to improve the classification accuracy. Kumar et al. [53] utilized the discrete wavelet transform to extract discriminative features from ECG signals for ER.

These studies collectively highlight the significance of frequency-domain features, among others, in ECG-based ER. The findings demonstrate the criticality of developing models that effectively capture frequency information to fully leverage the potential of ECG signals in decoding human emotions.

2.2 Deep learning methods in ECG-based ER

Sarkar et al. [54] proposed a two-stage framework, where the first stage is learning robust ECG representations through self-supervised tasks involving six signal transformations and the second is partially fine-tuning the model for ER. Behinaein et al. [46] proposed a deep learning model that combines convolutional layers with a transformer mechanism for stress detection using ECG signals. Ye et al. [55] developed a novel, online cross-subject emotion recognition method using hypergraph-based online transfer learning, which focuses on learning high-order correlations among data to enhance the adaptability of the model to new subjects. This approach highlights the potential

of transfer learning techniques for addressing inter-subject variability in ECG-based ER, paving the way for more versatile and robust models. Similarly, Vazquez-Rodriguez et al. [49] presented a transformer-based, self-supervised learning approach, demonstrating the effectiveness of attention mechanisms in building contextualized signal representations. These studies collectively underscore the potential of transformer architectures in capturing long-range dependencies and relevant features in ECG signals for ER.

Fan et al. [47] proposed a deep CNN incorporating an improved convolutional block attention module, which leverages channel and spatial attention mechanisms to enhance feature representation from ECG signals. In a related study, Fang et al. [56] utilized random convolutional kernels to process ECG signals for extracting multi-scale features, achieving improved classification of various emotional states. These approaches demonstrate that CNNs can effectively handle the complexity and variability inherent in ECG data for ER.

As demonstrated by these studies, CNN and transformer architectures are the most widely adopted structures for ECG-based ER. However, these models are often inefficient in capturing the frequency-domain features of ECG signals because of the high temporal resolution and quasi-periodic nature of the signals.

3 Methodology

3.1 Problem formulation

ECG-based ER can be viewed as a time series classification problem. Given an input ECG signal $\boldsymbol{X}$, the objective is to classify the emotional state into emotion categories $\boldsymbol{C}$. Mathematically, the task can be described as follows.

Let $\boldsymbol{X}_i = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ be a set of ECG signal samples, where each $\boldsymbol{x}_j \in \mathbb{R}^L$ is a one-dimensional sequence of length L representing an ECG signal segment for subject i. The goal is to learn a function $f : \mathbb{R}^L \to \boldsymbol{C}$, which maps the input signal sample $\boldsymbol{x}_i$ to one of the emotion categories.

3.2 Overview of CFAN

From a medical and physiological perspective, ECG signals exhibit typical waveform characteristics that are concentrated in a few specific frequency bands [57]. These signals are inherently quasi-periodic and consist of repetitive waveforms such as the P-wave, QRS complex, and T-wave, each corresponding to distinct physiological events in the cardiac cycle [58]. This property makes ECG signals particularly suitable for modeling approaches that can explicitly leverage frequency information. Motivated by this prior knowledge of ECG signals, we design CFAN for generating sample-specific frequency features from the input and dynamically adapt the CNN kernels for ER using ECG signals.

CFAN integrates a frequency-aware attention mechanism and dynamic convolution. The overall architecture consists of three main components, as illustrated in Figure 1. (i) Frequency-aware attention module (FAAM). This module extracts frequency domain features from the input ECG signals by applying FFT to generate the frequency spectrum. An attention mechanism then emphasizes important frequency components, producing refined frequency features. These features are further processed through a multi-layer perceptron (MLP) to generate frequency factors. (ii) Attention-guided convolution neural network block (AG-CNN). Operating in the time domain, this block incorporates the frequency factors to generate frequency-aware temporal features. The module effectively bridges the frequency and temporal domains, producing enhanced frequency-aware temporal features. (iii) MLP classifier. The frequency-aware temporal features are first processed through an adaptive average pooling layer to reduce dimensionality, followed by a fully connected (FC) layer.

3.3 Balanced sliding window sampling

Most existing studies segment ECG signals using fixed-step windowing, which may limit the sample diversity given the quasi-periodic nature of ECG waveforms. Additionally, class imbalance remains a common issue in ECG datasets. Because valid segments can theoretically start at any point within a cardiac cycle, we propose balanced sliding window sampling, as shown in Figure 2, in which a fixed number of windows per sample is randomly selected to enhance the diversity while promoting class balance during training.

Balanced sliding window sampling improves the model performance by addressing two key issues: it enforces class balance by extracting a fixed number of samples per recording and enhances phase diversity by randomly positioning the windows. This exposes the model to a wider range of temporal patterns while preventing the over-representation of longer signals, leading to more robust and generalizable learning.

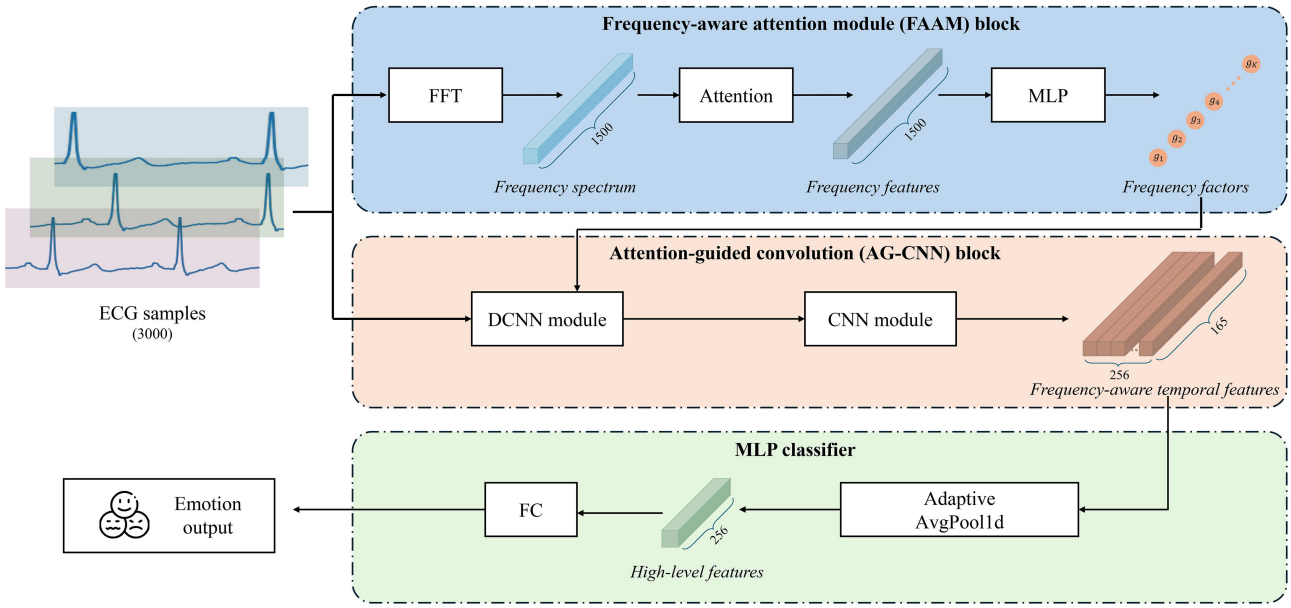

Figure 1 (Color online) Overall architecture of CFAN. The FAAM block applies attention to the frequency-domain features to emphasize the most informative parts of the signal. AG-CNN block combines dynamic convolution and normal convolution, adjusting its kernel weights based on the input frequency factors to capture sample-relevant temporal features. MLP classifier processes the learned features and outputs the final classification result corresponding to the predicted emotion.

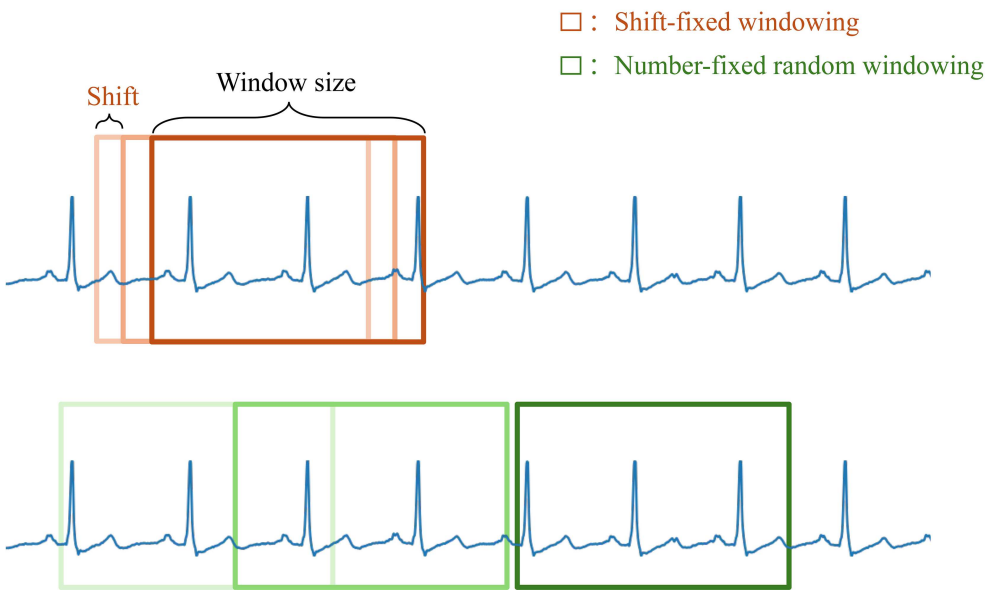

Figure 2 (Color online) Windowing methods. The orange regions show shift-fixed windowing with uniform spacing, and the green segments represent balanced sliding window sampling, which randomly selects a fixed number of windows to ensure class balance and phase diversity.

3.4 FAAM block

Emotional states, particularly those involving arousal or stress, are known to modulate the autonomic nervous system, which in turn alters the heart rate variability [59]. These changes are typically observed in the power spectrum of the ECG signal, especially within two key frequency bands: the low-frequency (LF) band (0.04–0.15 Hz) and the high-frequency (HF) band (0.15–0.4 Hz) [60]. A high LF/HF ratio is often associated with sympathetic dominance, such as during stress, whereas a lower ratio suggests parasympathetic activation, such as during relaxation [61].

As shown in Figure 3, the FAAM block processes the raw ECG signal by first converting it into the frequency domain using a real-valued FFT. Given an input ECG signal $\boldsymbol{x}_i$, the FFT computes the real and imaginary components of the signal's frequency domain representation.

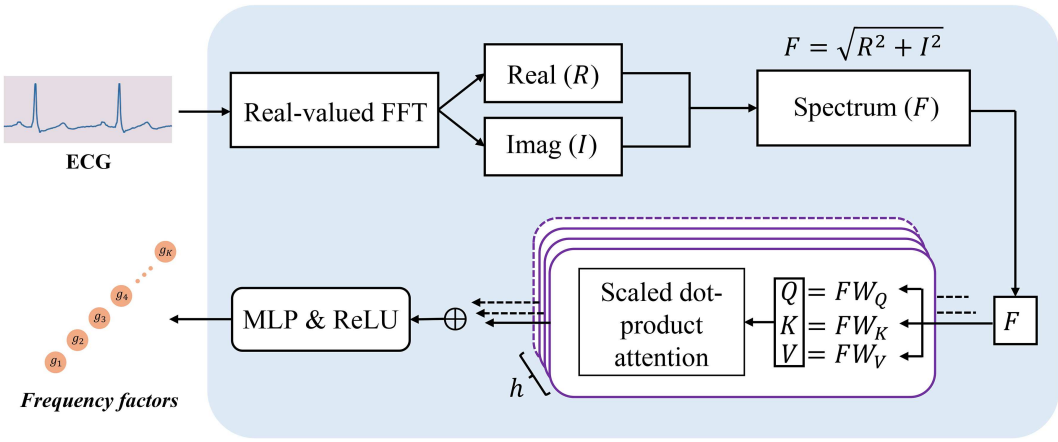

Figure 3 (Color online) FAAM block.

Specifically, the real part $\boldsymbol{R}_{\boldsymbol{x}_i}$ and imaginary part $\boldsymbol{I}_{\boldsymbol{x}_i}$ are calculated as follows:

$$\boldsymbol{R}_{\boldsymbol{x}_i} = \mathrm{Re}\{\mathrm{FFT}(\boldsymbol{x}_i)\}, \quad \boldsymbol{I}_{\boldsymbol{x}_i} = \mathrm{Im}\{\mathrm{FFT}(\boldsymbol{x}_i)\}, \tag{1}$$

where Re and Im denote the real and imaginary parts, respectively. The frequency domain representation of the signal is then obtained by concatenating the real and imaginary components: the magnitude of the frequency spectrum $\boldsymbol{F}(f)$ is determined by combining these components

$$\boldsymbol{F}_{\boldsymbol{x}_i} = \sqrt{\boldsymbol{R}_{\boldsymbol{x}_i}^2 + \boldsymbol{I}_{\boldsymbol{x}_i}^2}. \tag{2}$$

This frequency spectrum serves as the foundation for the ensuing attention mechanism. In the scaled dot-product attention, we use the spectrum $\boldsymbol{F}(\boldsymbol{x})$ to generate query, key, and value matrices, denoted as $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$, respectively. These matrices are derived through learned linear transformations of the frequency spectrum

$$\boldsymbol{Q}_{\boldsymbol{x}_i} = \boldsymbol{W}_{\boldsymbol{Q}}\boldsymbol{F}_{\boldsymbol{x}_i}, \quad \boldsymbol{K}_{\boldsymbol{x}_i} = \boldsymbol{W}_{\boldsymbol{K}}\boldsymbol{F}_{\boldsymbol{x}_i}, \quad \boldsymbol{V}_{\boldsymbol{x}_i} = \boldsymbol{W}_{\boldsymbol{V}}\boldsymbol{F}_{\boldsymbol{x}_i}, \tag{3}$$

where $\boldsymbol{W}_{\boldsymbol{Q}}$, $\boldsymbol{W}_{\boldsymbol{K}}$, and $\boldsymbol{W}_{\boldsymbol{V}}$ are weight matrices that project the frequency spectrum into the query, key, and value spaces, respectively.

The attention scores are computed by taking the dot product of the query and key matrices, followed by scaling and the application of a softmax function to obtain the attention weights

$$\mathrm{Attention}(\boldsymbol{Q}_{\boldsymbol{x}_i}, \boldsymbol{K}_{\boldsymbol{x}_i}, \boldsymbol{V}_{\boldsymbol{x}_i}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}_{\boldsymbol{x}_i}\boldsymbol{K}_{\boldsymbol{x}_i}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}_{\boldsymbol{x}_i}. \tag{4}$$

Here, d_k is the dimensionality of the key. The softmax function ensures that the attention scores sum to unity, effectively allowing the model to focus on the most relevant frequency components.

The attention output is then passed through an MLP, which consists of a linear transformation followed by a ReLU activation

$$\boldsymbol{F}\boldsymbol{F}_{\boldsymbol{x}_i} = \mathrm{ReLU}(\boldsymbol{W}_h \cdot \mathrm{Attention}(\boldsymbol{Q}_{\boldsymbol{x}_i}, \boldsymbol{K}_{\boldsymbol{x}_i}, \boldsymbol{V}_{\boldsymbol{x}_i})), \tag{5}$$

where $\boldsymbol{W}_h$ is a learnable weight matrix; the resulting feature representation $\boldsymbol{F}\boldsymbol{F}_{\boldsymbol{x}_i} \in \mathbb{R}^{K_f}$ is the frequency factor, which captures the key frequency information from the ECG signal and serves as the input for the next block in the model, the AG-CNN block.

By incorporating this attention mechanism, the FAAM block effectively highlights important frequency patterns in the ECG signal that are highly correlated with emotion. This allows the model to focus on subtle changes in the signal that are crucial for ER.

3.5 AG-CNN block

As shown in Figure 4, the AG-CNN block is designed to dynamically adapt the convolutional filters based on the frequency factors from the FAAM block, ensuring that the most relevant parts of the signal are emphasized in the feature extraction. The AG-CNN block operates via a dynamic convolution mechanism that adjusts the

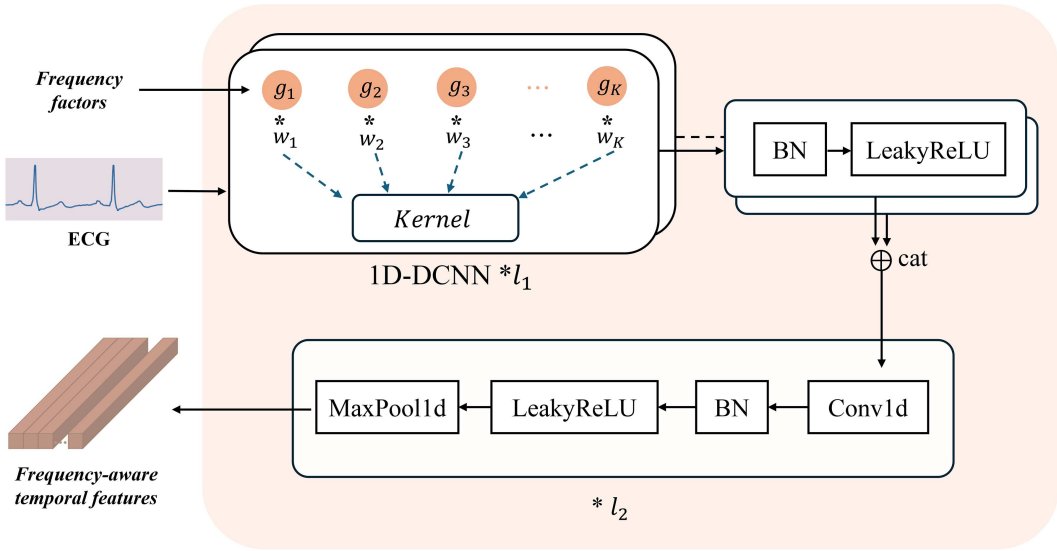

Figure 4 (Color online) AG-CNN block.

convolutional kernels for each input based on the learned attention weights. The dynamic convolutional process in this block involves multiple kernel sizes for capturing features on different temporal scales.

The learned frequency factors $\boldsymbol{FF}_{\boldsymbol{x}_i}$ modulate the convolutional weights $\boldsymbol{W}_k$. K_f is the number of dynamic filters. The aggregated weight for convolution is thus given by

$$\boldsymbol{W}_{agg} = \sum_{k=1}^{K} \boldsymbol{FF}_{\boldsymbol{x}_i} \boldsymbol{W}_k. \tag{6}$$

The dynamic convolution operation is then applied to the input ECG signal sample

$$\boldsymbol{y} = \text{Conv1d}(\boldsymbol{x}_i, \boldsymbol{W}_{agg}, \text{stride}, \text{padding}), \tag{7}$$

where stride and padding are hyperparameters controlling the step size of the convolution and the zero-padding, respectively. The dynamic convolution process in the AG-CNN block is performed twice with different kernel sizes (e.g., k_1 and k_2) to capture both short-term and long-term dependencies in the signal. The outputs of the two convolution operations are concatenated along the channel dimension

$$\boldsymbol{y}_{concat} = \text{ReLU}(\text{BN}(\boldsymbol{y}_1)) \oplus \text{ReLU}(\text{BN}(\boldsymbol{y}_2)), \tag{8}$$

where $\oplus$ denotes concatenation and BN represents batch normalization, which is applied to the outputs to improve the training stability and achieve faster convergence.

After the dynamic convolution, two standard 1D convolutional modules are applied sequentially in the AG-CNN block to further refine the extracted features. Each convolutional module consists of a 1D convolution operation with a fixed kernel, BN for stabilizing the distribution of features, and a leaky ReLU activation function that introduces non-linearity while allowing small gradients for negative values. This helps retain subtle characteristics of the ECG signal that might be important for ER. Max pooling is used to reduce the temporal dimensions, retaining essential features while providing some degree of translation invariance.

Formally, let $\boldsymbol{z}^{(1)}$ and $\boldsymbol{z}^{(2)}$ be the outputs of the first and second convolutional layers, respectively. Each convolution operation Conv1d$(\cdot)$ in this part is defined as

$$\boldsymbol{z}^{(i)} = \text{MaxPool1d}(\text{LeakyReLU}(\text{BN}(\text{Conv1d}(\boldsymbol{z}^{(i-1)})))), \tag{9}$$

where $\boldsymbol{z}^{(0)} = \boldsymbol{y}_{concat} \in \mathbb{R}^{B \times C_{out} \times L'}$ is the concatenated output from the dynamic convolution stage. The application of max pooling progressively reduces the length of the feature maps, focusing on high-level abstractions in the ECG signal.

3.6 MLP classifier

The MLP classifier in the CFAN model serves as the final stage, where the learned features from the AG-CNN block are processed to output the emotion class. This block uses MLP to map the high-level feature representations to the target labels.

After the high-level feature extraction process in the AG-CNN block, the feature map $\boldsymbol{z}^{(2)}$ is passed through an adaptive average pooling layer, which reduces the dimensionality of the feature map by averaging over the temporal dimension

$$\boldsymbol{f}_{avg} = \text{AdaptiveAvgPool1d}(\boldsymbol{z}^{(2)}). \tag{10}$$

This results in a fixed-size feature vector $\boldsymbol{f}_{avg}$. The pooled features are then passed through an FC layer that maps the features to the final output space, which corresponds to the number of emotion classes $n_{classes}$. The output of the FC layer is computed as

$$\boldsymbol{o} = \boldsymbol{W}_{fc}\boldsymbol{f}_{avg} + \boldsymbol{b}_{fc}, \tag{11}$$

where $\boldsymbol{o} \in \mathbb{R}^{B\times n_{classes}}$ is the output prediction for each class, $\boldsymbol{W}_{fc}$ is the weight matrix of the FC layer, and $\boldsymbol{b}_{fc}$ is the bias term.

Finally, a softmax activation function is applied to the output to convert the logits into probability scores for each class

$$\hat{\boldsymbol{y}} = \text{softmax}(\boldsymbol{o}), \tag{12}$$

where $\hat{\boldsymbol{y}} \in \mathbb{R}^{B\times n_{classes}}$ represents the predicted probabilities for each emotion class. The class with the highest probability is selected as the predicted emotion for each input ECG signal.

3.7 Loss function

The model is trained to minimize the classification error by optimizing a cross-entropy loss function, which is defined as

$$\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\sum_{i=1}^{n}\sum_{j=1}^{n_{classes}} y_{ij}\log(\hat{y}_{ij}), \tag{13}$$

where $\boldsymbol{y} \in \{0,1\}^{n\times n_{classes}}$ is the true label matrix, $y_{ij} = 1$ if sample i belongs to class $\boldsymbol{C}_j$, and $\hat{\boldsymbol{y}} \in [0,1]^{n\times n_{classes}}$ is the predicted probability for each class obtained from the model.

The classification problem can be solved using supervised learning techniques, where the model f is trained on labeled ECG data $\boldsymbol{X}$ and the corresponding emotion labels $\boldsymbol{y}$, where the objective is to find the optimal parameters that minimize $\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})$.

To address potential overfitting arising from the small size of the dataset (15 subjects) and the complex network architecture, we implemented dropout with a rate of 0.3 after each attention and FC layer.

4 Experiments

4.1 Datasets

The WESAD dataset [62] is a publicly available dataset for wearable stress and affect detection. Table 1 presents a detailed overview of the dataset. It contains data from 15 subjects, each of whom participated in a laboratory study involving the completion of a series of tasks designed to induce stress and affect. The Baseline modality is induced using carefully selected reading materials to define the neutral condition. Amusement is elicited through engaging video clips designed to evoke positive emotional responses [63], while Stress is induced using the Trier social stress test (TSST) [64], a widely recognized and validated method for eliciting stress. These stimuli were chosen because of their established effectiveness in reliably triggering the intended emotional states, ensuring the robustness and relevance of the dataset. The data were collected using a chest strap (RespiBAN Professional[1)]) and wrist-worn sensors (empatica E4[2)]). In this study, we use the ECG data from the chest-worn sensors, which provide a one-channel signal at a sampling rate of 700 Hz. The WESAD dataset is available for download from the UCI machine learning repository[3)].

1) https://www.pluxbiosignals.com/.
2) https://www.empatica.com/e4-wristband.
3) https://archive.ics.uci.edu/ml/datasets/WESAD.

Table 1 Summary of components and stimuli characteristics of the WESAD dataset.

Factor	WESAD	
Subjects	15 (12 male, 3 female)	
Label	Baseline/Amusement/Stress	
Stimuli	Baseline	Reading material
	Amusement	Video clips [63]
	Stress	TSST [64]
Duration	Baseline	About 20 min
	Amusement	About 6.5 min
	Stress	About 10 min
Sampling rate	700 Hz	

4.2 Evaluation metrics

We evaluated the performance of the algorithm model using the average accuracy and F1-score, which were calculated as follows. TP_i indicates true positive, TN_i indicates true negative, FP_i indicates false positive, and FN_i indicates false negative for subject i, and C is the number of subjects.

The average accuracy across all subjects is calculated as follows:

$$\text{Average Accuracy} = \frac{1}{C}\sum_{i=1}^{C}\frac{\mathrm{TP}_i + \mathrm{TN}_i}{\mathrm{TP}_i + \mathrm{TN}_i + \mathrm{FP}_i + \mathrm{FN}_i}. \tag{14}$$

The average F1-score across all subjects is thus

$$\text{Average F1-score} = \frac{1}{C}\sum_{i=1}^{C}\frac{2\times\mathrm{TP}_i}{2\times\mathrm{TP}_i + \mathrm{FP}_i + \mathrm{FN}_i}. \tag{15}$$

4.3 Preprocessing

For the WESAD dataset [62], we first applied a band-pass finite impulse response filter with a frequency range of 0.05–150 Hz to remove low- and high-frequency noise and artifacts. The data were then downsampled from 700 to 300 Hz to reduce the computational complexity while preserving essential information. Finally, we standardized the data across subjects by applying user-specific z-score normalization [54].

Deep learning models require large datasets, but WESAD includes only 15 subjects. To compensate for this shortcoming, we segmented the ECG signals into 10 s windows, increasing the sample count and standardizing the input size for more effective training. To address any class imbalance and improve the generalization, we applied balanced sliding window sampling with overlapping windows and randomized offsets (Figure 2). For each class and subject, 500 segments were randomly extracted, ensuring balanced and diverse training samples.

4.4 Experimental settings

We employed two types of experimental settings: (i) LOSO cross-validation and (ii) LOSO pre-training with subject-dependent fine-tuning, as shown in Figure 5. Each method is explained in the following paragraphs.

LOSO setting, as shown in Figure 5(b), is a form of cross-validation in which the data are divided such that each subject in the dataset is used once as a test set while the remaining subjects form the training set. This method is particularly advantageous in scenarios with limited data availability and high inter-subject variability because it ensures that the model is tested on unseen subjects in each iteration. The WESAD dataset comprises 15 available subjects. In each iteration, one subject is designated as the test set, while the remaining subjects are used as the training set. This process was repeated 15 times, ensuring that each subject served as the test set exactly once. The final performance metric is computed as the average across all iterations. This approach prevents data leakage and provides a more accurate evaluation of the model's generalization performance. The primary role of the validation set is to guide model selection under the assumption that the validation and test sets follow similar distributions. However, in the LOSO setting, this assumption does not hold because each test subject differs from the training subjects. Consequently, using a separate validation set is unnecessary in this context.

As shown in Figure 5(c), the fine-tuning setting follows the same data partitioning strategy as LOSO. However, model training is conducted in two stages: pretraining and fine-tuning. In the pretraining stage, the model is trained on the entire training subset (i.e., all subjects except the test subject). This step allows the model to

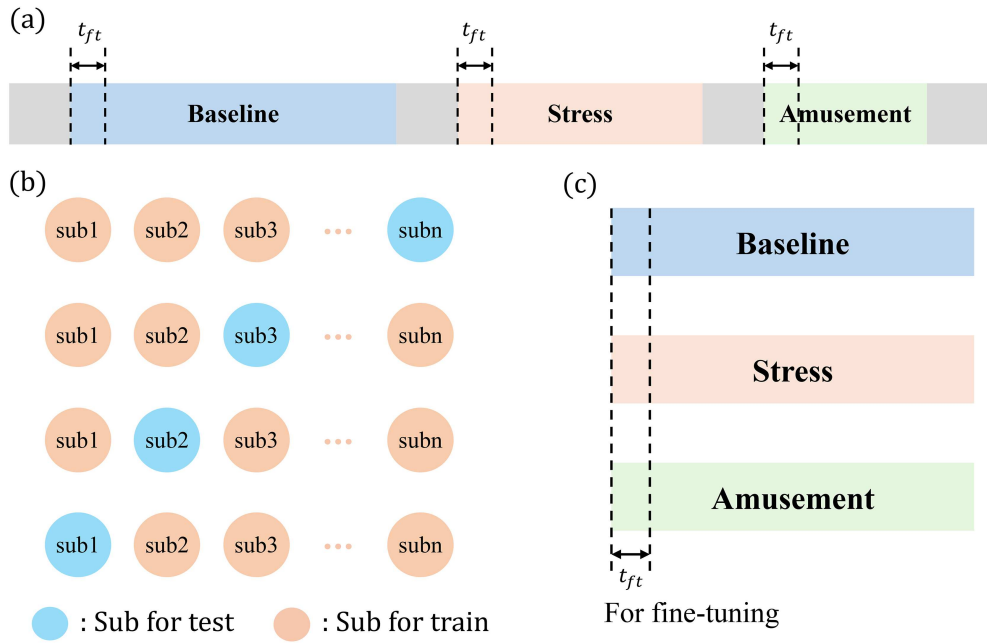

Figure 5 (Color online) Illustration of the experimental settings. (a) Example of data collection protocol for a single subject, including the Baseline, Stress, and Amusement phases. (b) LOSO cross-validation setting. In each fold, one subject is reserved as the test set while the others are used as the training set. (c) Fine-tuning setting. The beginning of t_{ft} for each phase is used for fine-tuning.

learn generalizable feature representations from a broader dataset, providing strong initialization for subsequent fine-tuning. In the fine-tuning stage, we use an initial segment of the test subject's data to adapt the model to the subject-specific characteristics, while the remaining portion serves as the final test set. Specifically, instead of selecting data randomly, the earliest available segment of the test subject's data is used for fine-tuning. This design choice reflects real-world application scenarios, where early data from a new subject can be used for rapid model adaptation before further predictions are made. This approach ensures that the model learns subject-specific patterns while minimizing the risk of data leakage. The process is repeated for each subject in the dataset, ensuring that each subject is used as the test set once. The final model performance is reported as the average across all iterations.

4.5 Implementation details

For the implementation developed herein, the embedding dimension of the FAAM block is set to 128, with two attention heads. For the frequency factor, the dimensionality K is set to 18, and the number of dynamic filters is 32. The dynamic convolution operations utilize kernel sizes of 35 and 17. Appropriate padding is applied to maintain the input-output dimensional consistency. The subsequent CNN layers adopt ratio coefficients of 15 and 8, with output channel dimensions of 64 and 256, respectively. These hyperparameters, including the kernel sizes and channel configurations, were determined based on preliminary experiments aimed at balancing the complexity and performance of the model. The training protocol remains consistent across all subjects. The model is trained using the Adam optimizer, with an initial learning rate of 1×10^{-4}. The cross-entropy loss is selected as the objective function to guide training. The training is conducted for a maximum of 100 epochs with a batch size of 1028.

5 Results and analysis

In this section, we first present and statistically analyze the accuracy and F1-score of the developed method in comparison with state-of-the-art approaches. We then conduct ablation studies to examine the contribution of each component in the CFAN model. Furthermore, we examine the impact of key hyperparameters, including the kernel size in the AG-CNN layer and the dimensionality of K, on the model performance.

Table 2 Comparison of accuracy results across models in both LOSO and fine-tuning settings. Superscripts denote statistical significance. a: $p < 0.001$, b: $p < 0.01$, c: $p < 0.05$. ACC values are expressed as percentages.

	Metric	CFAN	KNN	AB	RF	CNN	GRU	TimesNet	Transformer	CapsuleNet
LOSO	ACC	57.62±14.28	48.15±10.62	50.24±15.18	52.43±15.80	55.55±10.28	47.00±12.72^{c}	58.02±16.22	43.11±8.66^{b}	43.08±13.21^{b}
	F1-score	0.4990±0.1714	0.4675±0.1007	0.4663±0.1606	0.4873±0.1625	0.4906±0.1048	0.4376±0.1394	0.4305±0.1673	0.3368±0.0935^{b}	0.3991±0.1327
	AUC	0.7472±0.1094	0.6744±0.1106	0.7089±0.1567	0.7289±0.1671	0.7465±0.1106	0.7020±0.1450	0.6933±0.1682	0.6246±0.1094^{b}	0.6297±0.1340^{c}
15 s	ACC	70.13±17.50	44.45±11.96	51.24±17.63	53.63±15.33	58.33±11.38^{c}	53.43±11.01^{b}	61.75±16.88	40.94±7.82^{a}	44.43±12.83^{a}
	F1-score	0.6791±0.1912	0.4158±0.1160	0.4809±0.1879	0.5155±0.1570	0.5720±0.1187	0.5118±0.1098^{b}	0.4962±0.1753^{c}	0.3181±0.0921^{a}	0.4148±0.1268^{a}
	AUC	0.8563±0.1086	0.6039±0.1098	0.6343±0.1322	0.7186±0.1166	0.7928±0.0918	0.7272±0.1236^{b}	0.7168±0.1492^{c}	0.6194±0.0929^{a}	0.6494±0.1228^{a}
30 s	ACC	73.64±16.92	48.98±11.07	58.58±14.81	60.17±15.80^{c}	62.11±12.07^{c}	54.37±11.06^{b}	68.03±16.34^{b}	43.98±6.61^{a}	45.00±13.84^{a}
	F1-score	0.7309±0.1712	0.4790±0.1178	0.5693±0.1554	0.5986±0.1593^{b}	0.6138±0.1263^{c}	0.5205±0.1088^{a}	0.5016±0.1676	0.3481±0.0926^{a}	0.4215±0.1344^{a}
	AUC	0.8720±0.0977	0.6592±0.0937	0.7018±0.1030	0.7945±0.1130^{b}	0.8093±0.0905	0.7353±0.1220^{b}	0.7302±0.1523^{c}	0.6389±0.0921^{a}	0.6524±0.1276^{a}
40 s	ACC	76.06±14.87	49.24±13.24	64.38±11.77^{a}	63.69±13.54^{b}	65.35±11.92^{c}	55.43±10.13^{a}	68.04±16.43^{a}	44.40±7.13^{a}	45.09±14.06^{a}
	F1-score	0.7515±0.1560	0.4846±0.1351	0.6306±0.1224^{a}	0.6315±0.1390^{a}	0.6446±0.1244	0.5313±0.0980^{a}	0.5394±0.1649^{a}	0.3522±0.0926^{a}	0.4193±0.1373^{a}
	AUC	0.8867±0.0907	0.6771±0.1233	0.7473±0.0830^{c}	0.8250±0.1047^{a}	0.8274±0.0854	0.7571±0.1195^{b}	0.7570±0.1505^{b}	0.6429±0.0966^{a}	0.6575±0.1309^{a}

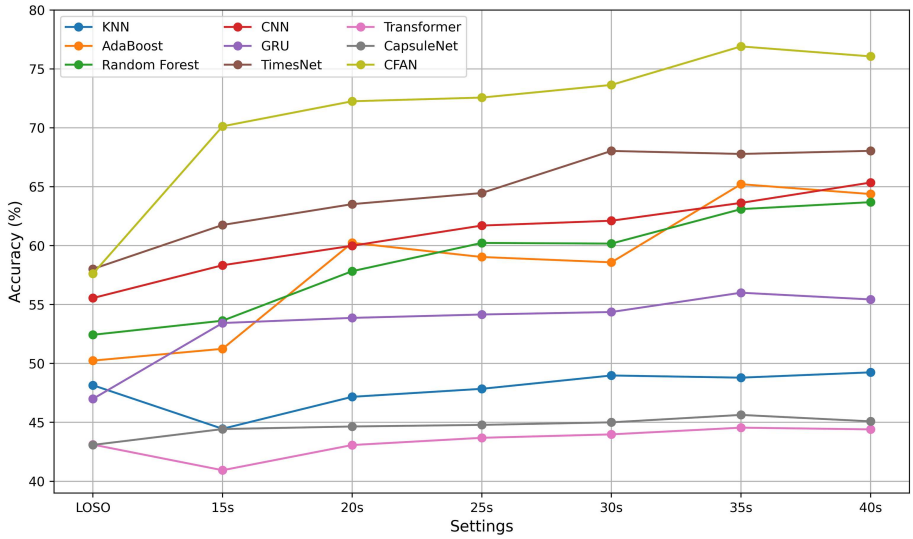

Figure 6 (Color online) Comparison of accuracy results across models in both LOSO and fine-tuning settings.

5.1 Performance comparison

To evaluate the effectiveness of the proposed CFAN model, the metrics were quantitatively compared against those of multiple baseline approaches, including manual feature selection methods (k-nearest neighbor (KNN), adaptive boosting (AB), random forest (RF)) and automatic feature learning methods (CNN, GRU, Transformer, TimesNet [65], and CapsuleNet [66]).

Table 2 provides a comprehensive comparison of the ACC, F1-score, and AUC for the various models employing both LOSO cross-validation and fine-tuning settings, and Figure 6 illustrates the corresponding ACC trends. Additionally, the fine-tuning setting is evaluated with different input signal durations. In the LOSO setting, CFAN achieves the best overall performance, with an F1-score of 0.4990 and an AUC of 0.7472. Although TimesNet yields a marginally higher ACC of 58.02% compared to CFAN's 57.62%, its F1-score of 0.4305 and AUC of 0.6933 are substantially lower, suggesting that its accuracy advantage is achieved at the cost of reduced balanced performance across classes. Among the traditional models, RF performs best, reaching 52.43% accuracy, despite ranking below CFAN and CNN. Deep learning models such as CNN, at 55.55%, demonstrate competitive performance, whereas GRU provides slightly lower but comparable results. Transformer and CapsuleNet underperform, with AUC values not exceeding 0.63 in this setting. In the fine-tuning setting, CFAN consistently outperforms all models as the signal duration increases, reaching its peak performance at 40 s with 76.06% accuracy, an F1-score of 0.7515, and an AUC of 0.8867. This trend suggests that longer signals capture richer temporal information, enabling better fine-tuning for emotion classification. Among the traditional models, RF and AB both show considerable improvement with longer signals, reaching 63.69% and 64.38% accuracy at 40 s, respectively. In contrast, KNN exhibits limited improvement, indicating weaker adaptability to increasing input complexity. Across all fine-tuning durations, CFAN consistently surpasses the other deep learning models. CNN, which achieves 65.35% accuracy at 40 s, still lags behind CFAN. Notably, the performance gap widens as the signal duration increases, as indicated by the statistically significant differences (superscripts a, b, and c denote $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively). Overall,

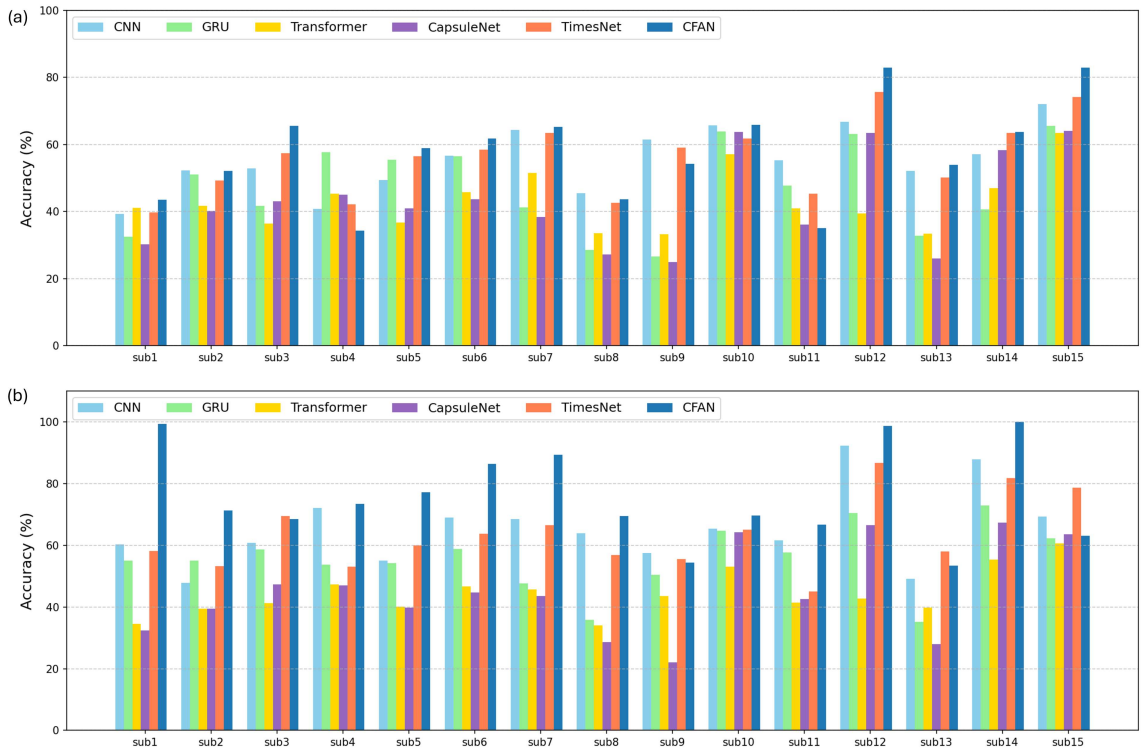

Figure 7 (Color online) Accuracy for each subject using CFAN with LOSO cross-validation (a) and 40-s fine-tuning settings (b).

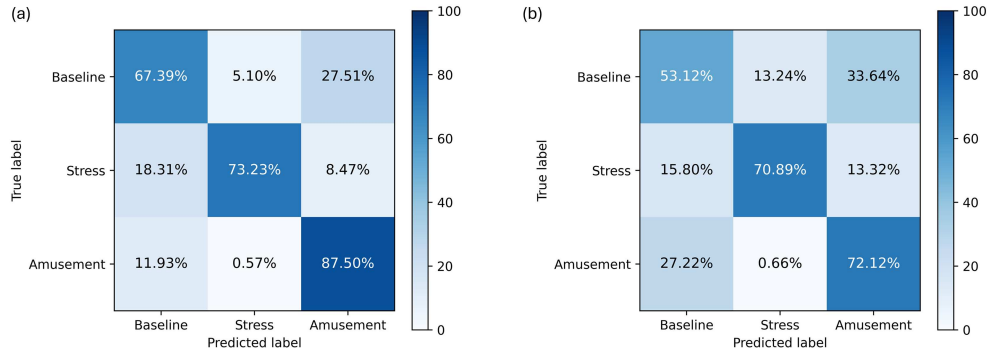

Figure 8 (Color online) Results of confusion matrices for CFAN (a) and CNN (b) employing 40-s fine-tuning settings.

these results highlight the robustness of CFAN, particularly in the fine-tuning setting, reinforcing its potential as a state-of-the-art model for ER in both LOSO and fine-tuning scenarios.

Figure 7 presents the classification accuracy for all subjects using the CFAN model and the comparative models with both LOSO cross-validation and fine-tuning settings. The performance varies notably across subjects, with some models achieving relatively high accuracy, suggesting that the task is significantly influenced by inter-subject variability. In the LOSO setting (Figure 7(a)), CFAN consistently outperforms the other models across most subjects, with particularly strong results for sub3 and sub12. Despite some degree of inter-subject variability, CFAN demonstrates greater robustness and stability compared to the other models, underscoring its superior generalization capability across unseen subjects. In the 40-s fine-tuning setting (Figure 7(b)), CFAN achieves substantial accuracy improvements across all subjects, with particularly high performance for sub1, sub6, sub7, sub12, and sub14, where the accuracy exceeds 80%. In contrast, the other baseline models exhibit inconsistent improvements, where CNN and TimesNet occasionally perform on par with CFAN, but fail to maintain competitive accuracy across most subjects. These results underscore the adaptability and reliability of CFAN, reinforcing its effectiveness in emotion recognition tasks.

Figure 8 presents the confusion matrices for CFAN (a) and CNN (b) under the 40-s fine-tuning condition. The CFAN model demonstrates superior classification performance, as indicated by the more pronounced diagonal in

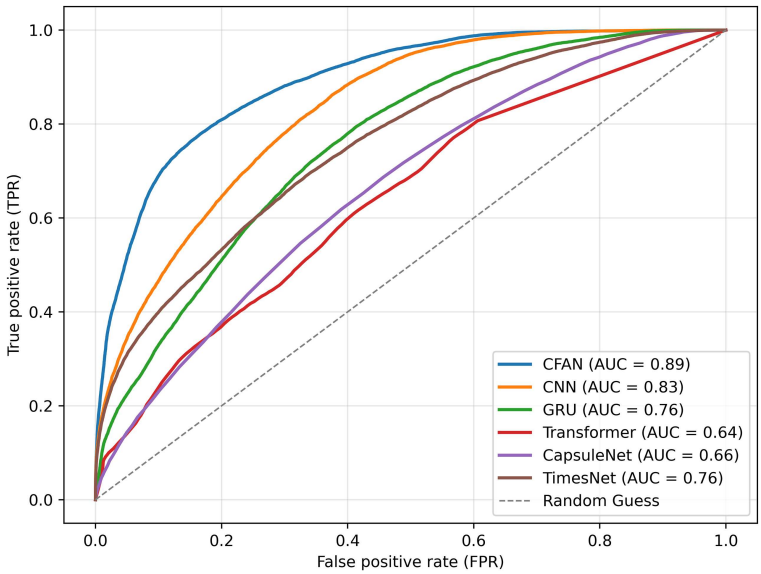

Figure 9 (Color online) ROC curves with 40-s fine-tuning setting.

Table 3 Results of ablation study for CFAN.

Method	ACC (%)	F1-score
Normal CNN	55.46	0.4694
AG-CNN (CFAN w/o FAAM)	57.74	0.4995
CFAN (AG-CNN + FAAM)	58.57	0.5191

its confusion matrix, reflecting higher accuracy across all emotional states. Specifically, CFAN achieves 87.50% accuracy in classifying the "Amusement" state, with minimal misclassification among other categories. In contrast, CNN exhibits a higher rate of off-diagonal errors, particularly in the "Baseline" class, where CNN frequently confuses other emotional states. Across all emotional states, CFAN consistently outperforms CNN, demonstrating notably higher precision for both the "Amusement" and "Stress" classes. These results underscore the potential of CFAN for ER.

Figure 9 presents the receiver operating characteristic (ROC) curves for different models using the 40-s fine-tuning setting. The AUC values are reported to quantify the overall classification performance of each model. The CFAN model achieves the highest AUC of 0.89, demonstrating its superior ability to distinguish between classes while maintaining an optimal balance between the true positive rate and false positive rate. CNN and TimesNet follow closely, with competitive AUCs of 0.83 and 0.82, respectively, indicating robust but slightly inferior performance compared to that of CFAN. In contrast, GRU achieves a moderate AUC value of 0.76. Transformer and CapsuleNet exhibit the weakest performance, with AUC values of 0.64 and 0.66, respectively. This suggests that these architectures do not effectively capture the underlying emotional patterns within the 40 s signals, possibly because of limitations in modeling temporal dependencies or sensitivity to fine-tuning adjustments. The dashed diagonal line represents the random guess performance, serving as a baseline reference. Overall, the ROC curves confirm the superior discriminative capability of CFAN, highlighting its significant performance advantage compared to the baseline models.

5.2 Ablation study

Table 3 presents the results of the ablation study, which evaluates the contribution of different components in the proposed CFAN model. The comparison includes the standard CNN model, AG-CNN model, meaning CFAN model without FAAM, and the full CFAN model.

The results in Table 3 demonstrate that the normal CNN achieves an accuracy of 55.46% and an F1-score of 0.4694. By incorporating dynamic parameter generation, AG-CNN demonstrates a performance improvement, where the accuracy and F1-score increased to 57.74% and 0.4995, respectively. This suggests that dynamically adjusting the CNN parameters during training allows the model to better adapt to variations in the ECG signals. Finally, the full CFAN model achieves the highest accuracy of 58.57% and F1-score of 0.5191, demonstrating the

Table 4 Performance comparison for different kernel sizes.

	Kernel size								
	17	21	25	27	29	31	35	41	55
ACC (%)	56.63	55.59	54.48	56.77	54.40	57.56	58.57	56.04	54.56
F1-score	0.4993	0.5155	0.5034	0.4962	0.4813	0.5078	0.5191	0.5008	0.4589

Table 5 Comparison of performance with variation of K_f.

	K_f									
	8	12	16	18	20	24	28	32	36	60
ACC (%)	54.22	54.30	58.12	58.57	55.50	55.94	54.53	54.59	54.29	56.60
F1-score	0.4919	0.4668	0.5120	0.5191	0.5016	0.4640	0.4487	0.4600	0.4641	0.4722

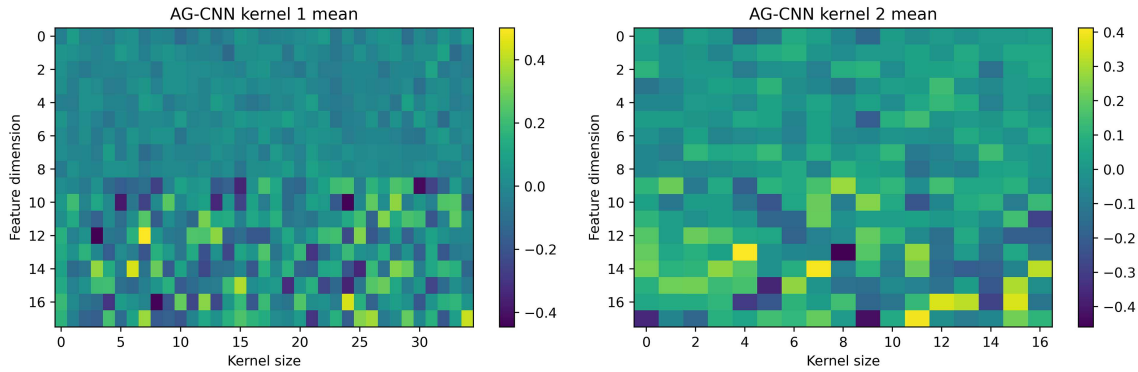

Figure 10 (Color online) Visualizations of the learned convolutional kernels using AG-CNN.

effectiveness of its frequency-aware design in adapting to the quasi-periodic nature of ECG signals and providing improved performance compared to standard CNN and AG-CNN architectures.

5.3 Effect of kernel size and frequency factor K_f in AG-CNN layer

Table 4 summarizes the effect of different kernel sizes in the first layer of AG-CNN. The kernel size significantly influences early-stage feature extraction by controlling the receptive field. We evaluated kernel sizes ranging from 17 to 55. The best performance was achieved with a kernel size of 35, yielding 58.57% accuracy and an F1-score of 0.5191. Smaller kernels (17, 21) and larger kernels (41, 55) caused the accuracy and F1-score to decrease.

These results indicate that a kernel size of 35 offers an effective trade-off, being small enough to capture local details yet large enough to maintain contextual awareness. In contrast, overly small kernels may miss broader temporal patterns, and overly large kernels may oversmooth features, weakening the model's discriminative power. Thus, selecting an appropriate kernel size is critical for optimizing the performance of AG-CNN.

Table 5 presents the impact of varying K_f, which determines both the dimension of the frequency factor and the number of dynamic filters in AG-CNN. We evaluated K_f values ranging from 8 to 60, using the accuracy and F1-score as performance metrics.

The model performs best at $K_f = 18$, achieving 58.57% accuracy and an F1-score of 0.5191. Lower K_f values (e.g., 8 and 12) result in reduced performance (accuracy below 55%, F1-scores under 0.50), possibly arising from limited frequency representation. Increasing K_f beyond 18 also leads to a performance decline (e.g., 54.59% accuracy at $K_f = 32$), indicating that overly large values may introduce redundant or noisy features.

These findings suggest that proper tuning of K_f is crucial. An overly small value limits the model's capacity to encode spectral information, whereas excessively large values may hinder generalization. A moderate setting, such as $K_f = 18$, provides a favorable trade-off between the model complexity and representational power.

5.4 Visualization of AG-CNN kernels

To further demystify the "black-box" nature, visualizations of the learned convolutional kernels by AG-CNN were averaged across the output dimensions, as shown in Figure 10.

After training, different kernels learn to focus on specific input dimensions and frequency-dependent characteristics of the ECG signal. For instance, some kernels become specialized in capturing HF transients typically found

Table 6 Comparison of performance of model in 5-class Valence and Arousal classification using DREAMER dataset [67]. Metrics reported as Accuracy/F1-score (mean ± std). CFAN achieves the best performance across both dimensions.

Model	Valence	Arousal
	Accuracy/F1-score	Accuracy/F1-score
CNN	0.6940 ± 0.038/0.6900 ± 0.031	0.6201 ± 0.044/0.6066 ± 0.036
GRU	0.7230 ± 0.036/0.7070 ± 0.030	0.6408 ± 0.041/0.6237 ± 0.034
Transformer	0.6810 ± 0.042/0.6860 ± 0.035	0.6228 ± 0.047/0.6075 ± 0.038
TimesNet	0.7250 ± 0.034/0.7100 ± 0.028	0.6435 ± 0.040/0.6282 ± 0.033
CapsuleNet	0.7510 ± 0.031/0.7390 ± 0.027	0.6660 ± 0.038/0.6525 ± 0.032
CFAN (ours)	0.7860 ± 0.030/0.7720 ± 0.025	0.6939 ± 0.035/0.6822 ± 0.030

in the QRS complex, while others respond more strongly to LF components associated with broader waves such as the P-wave or T-wave.

5.5 Additional evaluation of CFAN using DREAMER dataset

To further evaluate the generalizability of the proposed CFAN model, we conducted additional experiments using the DREAMER dataset, which differs from WESAD in terms of the subject composition and emotional elicitation paradigms. Table 6 summarizes the performance of CFAN compared to several representative baselines, including deep learning models (Transformer, CNN, GRU, TimesNet, CapsuleNet), for 5-class classification tasks for the Valence and Arousal dimensions. The mean accuracy and F1-score, used as evaluation metrics, are presented with the standard deviations, providing a more comprehensive view of the performance stability.

As shown in Table 6, CFAN achieves the best overall results for the Valence (0.7860/0.7720) and Arousal (0.6939/0.6822) classification, consistently outperforming all baselines in terms of both the predictive performance and reliability. Although models such as CapsuleNet and TimesNet also provide competitive results, CFAN demonstrates superior robustness across varying emotional conditions. These findings underscore the strong generalization capability and practical applicability of CFAN in diverse emotion recognition scenarios based on physiological signals.

6 Conclusion

The CFAN model was developed for ER using ECG signals, integrating an FAAM block and an AG-CNN block to effectively capture both frequency-domain and temporal patterns. Evaluation using the WESAD dataset with both LOSO and subject-dependent settings employing pretraining and fine-tuning demonstrated that CFAN significantly outperforms state-of-the-art models in terms of recognition accuracy and F1-score. Ablation studies confirm the impact of key architectural components and hyperparameter choices, especially the frequency-factor dimension and kernel size, on the performance of the model. These results demonstrate the robustness and effectiveness of CFAN for emotion recognition tasks, highlighting its strong potential for ECG-based emotion detection applications.

Although CFAN demonstrates high accuracy in controlled environments, real-world deployment poses challenges because of inter-subject variability in ECG signals. Addressing this variability through personalization techniques, such as transfer learning or subject-specific fine-tuning, could further enhance the robustness of the model. Additionally, improvements in data augmentation, unsupervised learning, and model optimization for low-latency processing would support scalable, real-time applications on wearable devices. In summary, CFAN holds considerable potential, but further development is essential for ensuring adaptability and scalability across diverse, real-world conditions.

Acknowledgements The authors would like to thank https://ubi29.informatik.uni-siegen.de/usi/data_wesad.html for providing the open dataset of ECG signals.

References

1 Wang Y, Song W, Tao W, et al. A systematic review on affective computing: emotion models, databases, and recent advances. Inf Fusion, 2022, 83: 19–52
2 Guo Y, Tang C, Wu H, et al. GNN-based multi-source domain prototype representation for cross-subject EEG emotion recognition. Neurocomputing, 2024, 609: 128445
3 Al-Saadawi H F T, Das B, Das R. A systematic review of trimodal affective computing approaches: text, audio, and visual integration in emotion recognition and sentiment analysis. Expert Syst Appl, 2024, 255: 124852
4 Izountar Y, Benbelkacem S, Otmane S, et al. VR-PEER: a personalized exer-game platform based on emotion recognition. Electronics, 2022, 11: 455
5 Huang C W, Wu B C Y, Nguyen P A, et al. Emotion recognition in doctor-patient interactions from real-world clinical video database: initial development of artificial empathy. Comput Meth Prog Bio, 2023, 233: 107480
6 Fang A C, Zhong P, Pan F, et al. Impact of emotional states on tinnitus sound therapy efficacy based on ECG signals and emotion recognition model. J Neurosci Meth, 2024, 409: 110213

7 El Hammoumi O, Benmarrakchi F, Ouherrou N, et al. Emotion recognition in e-learning systems. In: Proceedings of the 6th International Conference on Multimedia Computing And Systems (ICMCS), 2018. 1–6
8 Jiang C L, Zhang C H, Ji Y, et al. An affective chatbot with controlled specific emotion expression. Sci China Inf Sci, 2022, 65: 202102
9 Koptyra B, Ngo A, Radliński Ł, et al. Clarin-emo: training emotion recognition models using human annotation and chatgpt. In: Proceedings of International Conference on Computational Science, 2023. 365–379
10 Cui R X, Chen W Z, Li M Y. Emotion recognition using cross-modal attention from EEG and facial expression. Knowl-Based Syst, 2024, 304: 112587
11 Zhu J C, Ding Y, Liu H W, et al. Emotion knowledge-based fine-grained facial expression recognition. Neurocomputing, 2024, 610: 128536
12 Kang X L. Speech emotion recognition algorithm of intelligent robot based on ACO-SVM. Int J Cogn Computing Eng, 2025, 6: 131–142
13 Fan Y H, Huang H M, Han H. Hierarchical convolutional neural networks with post-attention for speech emotion recognition. Neurocomputing, 2025, 615: 128879
14 Liu Y J, Li J, Wang X P, et al. EmotionIC: emotional inertia and contagion-driven dependency modeling for emotion recognition in conversation. Sci China Inf Sci, 2024, 67: 182103
15 Poria S, Majumder N, Mihalcea R, et al. Emotion recognition in conversation: research challenges, datasets, and recent advances. IEEE Access, 2019, 7: 100943
16 Leong S C, Tang Y M, Lai C H, et al. Facial expression and body gesture emotion recognition: a systematic review on the use of visual data in affective computing. Comput Sci Rev, 2023, 48: 100545
17 Wu J T, Zhang Y J, Sun S Y, et al. Generalized zero-shot emotion recognition from body gestures. Appl Intell, 2022, 52: 8616–8634
18 Wan C T, Xu C P, Chen D Y, et al. Emotion recognition based on a limited number of multimodal physiological signals channels. Measurement, 2025, 242: 115940
19 Alarcao S M, Fonseca M J. Emotions recognition using EEG signals: a survey. IEEE Trans Affect Comput, 2017, 10: 374–393
20 Houssein E H, Hammad A, Ali A A. Human emotion recognition from EEG-based brain-computer interface using machine learning: a comprehensive review. Neural Comput Applic, 2022, 34: 12527–12557
21 Farabbi A, Polo E M, Barbieri R, et al. Comparison of different emotion stimulation modalities: an eeg signal analysis. In: Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022. 3710–3713
22 Deng F, Ding N, Ye Z M, et al. Wearable ubiquitous energy system. Sci China Inf Sci, 2021, 64: 1–3
23 Li Y, Gu Y Z, Qian S, et al. A stretchable, ionic conductive, and adhesive patch electrode with ultra-low on-skin impedance for electrophysiological signal recording. Sci China Inf Sci, 2025, 68: 129402
24 Sayed Ismail S N M, Ab. Aziz N A, Ibrahim S Z. A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG). J King Saud Univ-Com, 2022, 34: 3539–3558
25 Nita S, Bitam S, Heidet M, et al. A new data augmentation convolutional neural network for human emotion recognition based on ECG signals. BioMed Signal Proces, 2022, 75: 103580
26 Veeranki Y R, Posada-Quintero H F, Swaminathan R. Transition network-based analysis of electrodermal activity signals for emotion recognition. IRBM, 2024, 45: 100849
27 Joudeh I O, Cretu A M, Guimond S, et al. Prediction of emotional measures via electrodermal activity (EDA) and electrocardiogram (ECG). In: Proceedings of the 29th Signal Processing and Communications Applications Conference (SIU), 2021. 1–4
28 Goshvarpour A, Goshvarpour A. Asymmetric measures of polar Chebyshev chaotic map for discrete/dimensional emotion recognition using PPG. BioMed Signal Proces, 2025, 100: 107089
29 Mellouk W, Handouzi W. CNN-LSTM for automatic emotion recognition using contactless photoplythesmographic signals. BioMed Signal Proces, 2023, 85: 104907
30 Zhang Q, Chen X X, Zhan Q Y, et al. Respiration-based emotion recognition with deep learning. Comput Ind, 2017, 92-93: 84–90
31 Xu M H, Cheng J, Li C, et al. Spatio-temporal deep forest for emotion recognition based on facial electromyography signals. Comput Biol Med, 2023, 156: 106689
32 Han E G, Kang T K, Lim M T. Physiological signal-based real-time emotion recognition based on exploiting mutual information with physiologically common features. Electronics, 2023, 12: 2933
33 Zhu M, Wu Q Z, Bai Z L, et al. EEG-eye movement based subject dependence, cross-subject, and cross-session emotion recognition with multidimensional homogeneous encoding space alignment. Expert Syst Appl, 2024, 251: 124001
34 Rodger H, Sokhn N, Lao J, et al. Developmental eye movement strategies for decoding facial expressions of emotion. J Exp Child Psychol, 2023, 229: 105622
35 Li C, Hou Y M, Song R C, et al. Multi-channel EEG-based emotion recognition in the presence of noisy labels. Sci China Inf Sci, 2022, 65: 140405
36 Luo G, Han Y T, Xie W C, et al. GCD-JFSE: graph-based class-domain knowledge joint feature selection and ensemble learning for EEG-based emotion recognition. Knowl-Based Syst, 2025, 309: 112770
37 He R J, Jie Y W, Tong W, et al. A parallel neural networks for emotion recognition based on EEG signals. Neurocomputing, 2024, 610: 128624
38 Xu X Z, Suo Y X, Zhao Y, et al. A dry-electrode enabled ECG-on-chip with arrhythmia-aware data transmission. Sci China Inf Sci, 2025, 68: 122405
39 Liu X W, Wang H, Li Z J, et al. Deep learning in ECG diagnosis: a review. Knowl-Based Syst, 2021, 227: 107187
40 Agrafioti F, Hatzinakos D, Anderson A K. ECG pattern analysis for emotion detection. IEEE Trans Affect Comput, 2011, 3: 102–115
41 Wang X Y, Guo Y Q, Ban J, et al. Driver emotion recognition of multiple-ECG feature fusion based on BP network and D-S evidence. IET Intel Trans Sys, 2020, 14: 815–824
42 Hsu Y L, Wang J S, Chiang W C, et al. Automatic ECG-Based emotion recognition in music listening. IEEE Trans Affect Comput, 2020, 11: 85–99
43 Sepúlveda A, Castillo F, Palma C, et al. Emotion recognition from ECG signals using wavelet scattering and machine learning. Appl Sci, 2021, 11: 4945
44 Chen T, Yin H F, Yuan X H, et al. Emotion recognition based on fusion of long short-term memory networks and SVMs. Digit Signal Process, 2021, 117: 103153
45 Alam A, Urooj S, Ansari A Q. Design and development of a non-contact ECG-based human emotion recognition system using SVM and RF classifiers. Diagnostics, 2023, 13: 2097
46 Behinaein B, Bhatti A, Rodenburg D, et al. A transformer architecture for stress detection from ecg. In: Proceedings of the 2021 ACM International Symposium on Wearable Computers, 2021. 132–134
47 Fan T Q, Qiu S, Wang Z L, et al. A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition. Comput Biol Med, 2023, 159: 106938
48 Khare S K, Blanes-Vidal V, Nadimi E S, et al. Emotion recognition and artificial intelligence: a systematic review (2014-2023) and research recommendations. Inf Fusion, 2024, 102: 102019
49 Vazquez-Rodriguez J, Lefebvre G, Cumin J, et al. Transformer-based self-supervised learning for emotion recognition. In: Proceedings of the 26th International Conference on Pattern Recognition (ICPR), 2022. 2605–2612
50 Wulan N, Wang W, Sun P Z, et al. Generating electrocardiogram signals by deep learning. Neurocomputing, 2020, 404: 122–136
51 Nikolova D, Petkova P, Manolova A, et al. ECG-based emotion recognition: overview of methods and applications. In: Proceedings of the Advances in Neural Networks and Applications (ANNA'18), 2018. 1–5

52 Khan A N, Ihalage A A, Ma Y, et al. Deep learning framework for subject-independent emotion detection using wireless signals. PLoS ONE, 2021, 16: e0242946
53 Kumar A, Kumar A. Human emotion recognition using machine learning techniques based on the physiological signal. BioMed Signal Proces, 2025, 100: 107039
54 Sarkar P, Etemad A. Self-supervised ECG representation learning for emotion recognition. IEEE Trans Affect Comput, 2020, 13: 1541–1554
55 Ye Y L, Pan T J, Meng Q H, et al. Online ECG emotion recognition for unknown subjects via hypergraph-based transfer learning. In: Proceedings of IJCAI, 2022. 3666–3672
56 Fang A C, Pan F, Yu W C, et al. ECG-based emotion recognition using random convolutional kernel method. BioMed Signal Proces, 2024, 91: 105907
57 Tereshchenko L G, Josephson M E. Frequency content and characteristics of ventricular conduction. J ElectroCardiol, 2015, 48: 933–937
58 Clifford G D, Azuaje F, Mcsharry P, et al. ECG statistics, noise, artifacts, and missing data. In: Advanced Methods and Tools for ECG Data Analysis. Boston: Artech House, 2006. 55–85
59 Zhu J, Ji L, Liu C. Heart rate variability monitoring for emotion and disorders of emotion. Physiol Meas, 2019, 40: 064004
60 Shi H Y, Yang L C, Zhao L L, et al. Differences of heart rate variability between happiness and sadness emotion states: a pilot study. J Med Biol Eng, 2017, 37: 527–539
61 Geisler F C M, Vennewald N, Kubiak T, et al. The impact of heart rate variability on subjective well-being is mediated by emotion regulation. Pers Indiv Differ, 2010, 49: 723–728
62 Schmidt P, Reiss A, Duerichen R, et al. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction, 2018. 400–408
63 Samson A C, Kreibig S D, Soderstrom B, et al. Eliciting positive, negative and mixed emotional states: a film library for affective scientists. Cognition Emotion, 2016, 30: 827–856
64 Kirschbaum C, Pirke K M, Hellhammer D H. The 'trier social stress test' — a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology, 2008, 28: 76–81
65 Wu H X, Hu T G, Liu Y, et al. Timesnet: temporal 2d-variation modeling for general time series analysis. ArXiv: 2210.02186
66 Jiao Y, Qi H M, Wu J. Capsule network assisted electrocardiogram classification model for smart healthcare. Biocybern BioMed Eng, 2022, 42: 543–555
67 Katsigiannis S, Ramzan N. DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE J Biomed Health, 2017, 22: 98–107