Journal of Integrative Neuroscience

J. Integr. Neurosci. **2025**; 24(8): 41547
https://doi.org/10.31083/JIN41547

Original Research

EEG-ERnet: Emotion Recognition based on Rhythmic EEG Convolutional Neural Network Model

Shuang Zhang[1,2,†], Chen Ling[3,†], Jingru Wu[3], Jiawen Li[1,3,4,*], Jiujiang Wang[1,2], Yuanyu Yu[1,2], Xin Liu[5], Jujian Lv[3,*], Mang I Vai[4,6], Rongjun Chen[3,7]

[1]Key Laboratory of Numerical Simulation of Sichuan Provincial Universities, School of Mathematics and Information Sciences, Neijiang Normal University, 641000 Neijiang, Sichuan, China
[2]School of Artificial Intelligence, Neijiang Normal University, 641004 Neijiang, Sichuan, China
[3]School of Computer Science, Guangdong Polytechnic Normal University, 510665 Guangzhou, Guangdong, China
[4]ZUMRI-LYG Joint Lab, Zhuhai UM Science and Technology Research Institute, 519031 Zhuhai, Guangdong, China
[5]School of Mathematics and Computer Science, Northwest Minzu University, 730030 Lanzhou, Gansu, China
[6]Department of Electrical and Computer Engineering, University of Macau, 999078 Macau, China
[7]Guangdong Provincial Key Laboratory of Intellectual Property and Big Data, Guangdong Polytechnic Normal University, 510665 Guangzhou, Guangdong, China
*Correspondence: lijiawen@gpnu.edu.cn (Jiawen Li); jujianlv@gpnu.edu.cn (Jujian Lv)
[†]These authors contributed equally.
Academic Editor: Bettina Platt
Submitted: 18 May 2025 Revised: 9 July 2025 Accepted: 25 July 2025 Published: 28 August 2025

Abstract

Background: Emotion recognition from electroencephalography (EEG) can play a pivotal role in the advancement of brain-computer interfaces (BCIs). Recent developments in deep learning, particularly convolutional neural networks (CNNs) and hybrid models, have significantly enhanced interest in this field. However, standard convolutional layers often conflate characteristics across various brain rhythms, complicating the identification of distinctive features vital for emotion recognition. Furthermore, emotions are inherently dynamic, and neglecting their temporal variability can lead to redundant or noisy data, thus reducing recognition performance. Complicating matters further, individuals may exhibit varied emotional responses to identical stimuli due to differences in experience, culture, and background, emphasizing the necessity for subject-independent classification models. **Methods**: To address these challenges, we propose a novel network model based on depthwise parallel CNNs. Power spectral densities (PSDs) from various rhythms are extracted and projected as 2D images to comprehensively encode channel, rhythm, and temporal properties. These rhythmic image representations are then processed by a newly designed network, EEG-ERnet (Emotion Recognition Network), developed to process the rhythmic images for emotion recognition. **Results**: Experiments conducted on the dataset for emotion analysis using physiological signals (DEAP) using 10-fold cross-validation demonstrate that emotion-specific rhythms within 5-second time intervals can effectively support emotion classification. The model achieves average classification accuracies of 93.27 $\pm$ 3.05%, 92.16 $\pm$ 2.73%, 90.56 $\pm$ 4.44%, and 86.68 $\pm$ 5.66% for valence, arousal, dominance, and liking, respectively. **Conclusions**: These findings provide valuable insights into the rhythmic characteristics of emotional EEG signals. Furthermore, the EEG-ERnet model offers a promising pathway for the development of efficient, subject-independent, and portable emotion-aware systems for real-world applications.

Keywords: electroencephalography; emotions; deep learning; convolutional neural networks; brain waves; cross-validation studies

1. Introduction

Emotions are key aspects of human behavior, communication, and decision-making, making accurate recognition vital for enhancing user experiences and advancing the development of more effective brain-computer interfaces (BCIs) [1]. Recently, detecting emotional states from brain waves on the scalp has provided a noninvasive method to understand human emotions. Thus, emotion recognition through electroencephalography (EEG) has emerged as a key component of affective computing [2–4]. EEG signals, generated by the brain's electrical activity, reflect the affective processes underlying different emotional states. EEG emotion recognition can detect subtle changes in brain activity that may not be discernible through other methods, such as facial expressions, body gestures, text, or speech [5–8], since these methods could fail to represent emotional nuances, especially in ambiguous external expressions [9]. EEG, in contrast, directly measures brain activity, providing a fundamental understanding of emotional processes and serving as a valuable tool for investigating the neural basis of emotions.

However, EEG emotion recognition presents several challenges due to the inherent complexity and noise in the system. First, EEG signals are highly susceptible to various noise and artifacts, including muscle artifacts, eye blinks, and external interference, all of which can degrade recognition accuracy [10]. Second, the high dimensionality and variability of EEG signals make extracting valu-

able features that can effectively distinguish between different emotions a challenging task. Therefore, feature extraction and classification models are necessary. Based on biomedical signal processing techniques, feature extraction involves identifying and selecting relevant features from EEG signals indicative of specific emotions [11–13]. This process is key for reducing data dimensionality and enhancing the ability to discern subtle emotional cues. Classification models, such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, cross-modal learning hybrid frameworks, and attention modules [14–17], have demonstrated considerable potential in overcoming the limitations of traditional approaches. Particularly, CNNs are renowned for their ability to learn hierarchical features, making them well-suited to capture spatial and temporal properties in EEG signals, which beneficially improve accuracy in emotion recognition [18].

Although CNN shows advances, three main limitations require refinement. The five brain rhythms, delta (δ, 0–4 Hz), theta (θ, 4–8 Hz), alpha (α, 8–13 Hz), beta (β, 13–30 Hz), and gamma (γ, >30 Hz) [19], play distinct roles in emotion recognition by reflecting various aspects of brain activity during neural processing. For instance, the δ rhythm is associated with deep relaxation or intense emotional states such as sadness [20]. The θ rhythm is linked to engagement, memory, and introspection [21]. In the context of valence classification, lower α activity tends to correspond with negative valence, while higher α power is often associated with positive emotional states or a relaxed baseline. This relationship implies α rhythm analysis to serve as an indicator for distinguishing between positive and negative emotional valence [22]. The β rhythm reveals active thinking and heightened states such as anxiety and stress [23]. The γ rhythm usually indicates the complex integration of emotions and intense experiences such as joy or excitement [24]. When extracted as features from EEG signals, these rhythms provide insights into emotion recognition. Nevertheless, conventional convolutional layers often obscure spectral characteristics, making it challenging to identify key rhythmic features vital for recognizing specific emotions. Moreover, EEG signals are time-series data representing the electrical activity over time. From a neuroscience perspective, emotions are dynamic and constantly evolving, usually shifting within seconds in response to stimuli [25]. Failing to account for temporal effects typically results in redundant data, which can amplify noise and disrupt the extraction of valuable features. Finally, the same stimulus can elicit different responses among individuals due to personal experiences, cultural influences, and backgrounds, which shape their psychological and cognitive mechanisms. In this regard, traditional subject-dependent models often struggle to generalize across individuals. Therefore, subject-independent models are preferred in real-world deployments to enhance their applicability.

To this end, this work proposes a subject-independent EEG emotion recognition model based on the depthwise parallel CNN. Its first step involves extracting power spectral densities (PSDs) from various rhythms to project the 2D images, which preserves a comprehensive representation of channel, rhythm, and temporal properties. Subsequently, a depthwise parallel CNN architecture, denoted as EEG-ERnet, is employed to train and test the rhythmic image features, enhancing its ability to distinguish diverse emotions. Experiments have been conducted on the database for emotion analysis using physiological signals (DEAP) dataset, a widely used benchmark for EEG emotion recognition, evaluating the effectiveness of the proposed model for arousal, valence, dominance, and liking tasks in a subject-independent approach. Hence, with the help of neural networks and rhythmic image features, the EEG-ERnet provides an innovative solution in this field.

The rest of this work is organized as follows: Section 2 reviews related work. Section 3 introduces the DEAP dataset and describes the EEG-ERnet, including the feature extraction for rhythmic-based 2D images and the network model design. Section 4 presents the experimental results, comparative study, and discussion. Finally, Section 5 shows the conclusion.

2. Related Work

Early approaches to EEG emotion recognition primarily rely on traditional machine learning techniques, which typically involve extracting features from EEG signals and employing classifiers such as support vector machine (SVM), k-nearest neighbors (k-NN), random forest (RF), linear discriminant analysis (LDA), decision tree, and rotation forest ensemble (RFE). For example, Subasi *et al.* [26] designed a framework that includes signal denoising using multi-scale principal component analysis (MSPCA), feature extraction through tunable Q wavelet transform (TQWT), dimension reduction via statistical methods, and RFE and SVM classifiers. Experiments on the DEAP dataset demonstrated that their framework achieved about 93% classification accuracy. Tuncer *et al.* [27] employed a fractal pattern feature generation function, termed the fractal Firat pattern (FFP), for emotion recognition. Their method contained decomposing EEG signals using TQWT and extracting fractal geometry features from decomposed signals through FFP. After that, an iterative chi-square selector (IChi2) was utilized for feature selection, followed by SVM, k-NN, and LDA classifiers. Experiments on the games-based emotion recognition system (GAMEEMO) dataset displayed a maximum accuracy of 99.82% with SVM. Salankar *et al.* [28] used an approach based on empirical mode decomposition (EMD) and second-order difference plot (SODP). Their method involved decomposing EEG signals into intrinsic mode functions (IMFs) using the EMD, followed by feature extraction from the SODP of these IMFs. The classifiers employed were SVM and a

two-hidden-layer multilayer perceptron (MLP). The experimental results from the DEAP dataset indicated 93.8% accuracy in the classifications of arousal and valence. Sarma and Barma [29] selected appropriate EEG segments based on random matrix theory (RMT) to achieve emotion recognition, utilizing PSDs as features. Their experiments were conducted on the Shanghai Jiao Tong University emotion EEG dataset (SEED) and DEAP datasets, yielding classification accuracies of 82.21% for valence and 86.03% for arousal, with the k-NN classifier offering the best performance. Although the above methods accomplished good results, they struggle to investigate the complex and dynamic characteristics of emotions across subjects. Hence, deep learning techniques have been considered.

Deep learning techniques have demonstrated remarkable capabilities in learning hierarchical features, making them well-suited for handling complex and high-dimensional EEG signals. Particularly, CNNs are beneficial for processing spatial and temporal patterns related to emotion recognition [30]. Regarding CNN, its convolutional layers extract local spatial features from multiple channels, and the pooling operations reduce dimensionality and maintain translation invariance. Moreover, the hierarchical architecture facilitates learning increasingly from low-level signal characteristics to high-level features. Recent studies have further enhanced CNN performance by incorporating attention modules [31–33], which allow it to concentrate on the most relevant spatial-temporal characteristics, aiming at cross-subject emotion recognition.

For instance, Yang *et al*. [34] designed a multi-column CNN for EEG emotion recognition, consisting of multiple modules that process temporal snapshots of EEG signals from the DEAP dataset. The final decision is generated through a weighted voting strategy, which combines the outputs of individual modules to enhance robustness and accuracy. They achieved accuracies of 90.01% and 90.65% for valence and arousal, respectively, demonstrating that the multi-column structure helps mitigate the impact of EEG variations. Hwang *et al*. [35] employed a CNN with topology-preserving differential entropy (DE) features to represent spatial information and enhance the resolution of EEG for emotion classification (positive, neutral, negative). Experiments on the SEED dataset yielded an accuracy of 90.41%, outperforming SVM with the radial basis function (RBF) kernel. Cui *et al*. [36] developed a DE-CNN-BiLSTM model, integrating DE, CNN, and bidirectional LSTM (BiLSTM) to process EEG signals. The DE features were extracted from different frequency bands and time slices, mapped into 4D tensors to represent brain spatial structure, and fed into the CNN for spatial feature learning. The BiLSTM was subsequently employed to capture the past and future temporal information. This model achieved accuracies of 94.86% for arousal and 94.02% for valence on the DEAP dataset, as well as 94.82% for the SEED dataset. Wang *et al*. [37] enhanced the resource efficiency of the CNN-based model by constructing six tasks through signal transformations to generate labels for the unlabeled EEG. A multi-task CNN was then trained to recognize these transformations. Next, the convolutional layers were frozen, and the fully connected layers were reconstructed for emotion recognition. Experiments on the SEED and DEAP datasets revealed that self-supervised learning can improve classification accuracy. On the SEED dataset, the average accuracy was 84.54% for the preprocessed data and 98.65% for the data with extracted DE features. For the DEAP dataset, the network acquired high F1-scores, with valence and arousal metrics attaining approximately 96% when trained on 20% of the data. Yao *et al*. [38] integrated a transformer and a CNN to extract spatial-temporal features for emotion recognition, which employed position encoding and multi-head attention mechanisms to represent channel positions and timing information from EEG. Two parallel transformer encoders extracted spatial and temporal features, which were then aggregated by a CNN and classified using softmax. Experiments conducted on the SEED and DEAP datasets showed that the model achieved accuracies of 96.67% on the SEED dataset and 95.73%, 96.95%, and 96.34% for the arousal-valence, arousal, and valence tasks on the DEAP dataset, respectively. Lu *et al*. [39] developed a convolution-multilayer perceptron network (CMLP-Net), where its architecture contained a temporal-stream shared convolution to extract shared features across consecutive time steps and reduce redundancy, a time-refinement temporal-spatial convolution to extract compelling temporal-spatial features, and a spatial interaction MLP to enhance the global spatial dependency of the features. Hence, CMLP-Net transformed 1D EEG signals into a 2D representation to better express spatial information. Experiments from the DEAP dataset revealed average accuracies of 98.65%, 98.70%, and 98.63% for valence, arousal, and dominance tasks. Qiao *et al*. [40] incorporated a temporal convolutional attention network to represent both local and global features of EEG signals. DE features were extracted and processed through a CNN to obtain local features. Subsequently, the self-attention mechanism was applied to enhance global feature extraction, followed by a BiLSTM network to investigate temporal dependencies. The experiments were performed on a self-collected dataset and the DEAP dataset, achieving average accuracies of 93.45% and 96.36% for valence and arousal, respectively. In addition to software-based deep learning models, recent research has explored hardware-efficient approaches for emotion recognition. For example, Ezilarasan and Leung [41] proposed an field programmable gate array (FPGA)-based architecture that extracts EEG features and classifies emotions using a lightweight approach, aiming to provide low-latency, energy-efficient processing suitable for embedded systems and illustrating the potential of real-time emotion-aware applications.

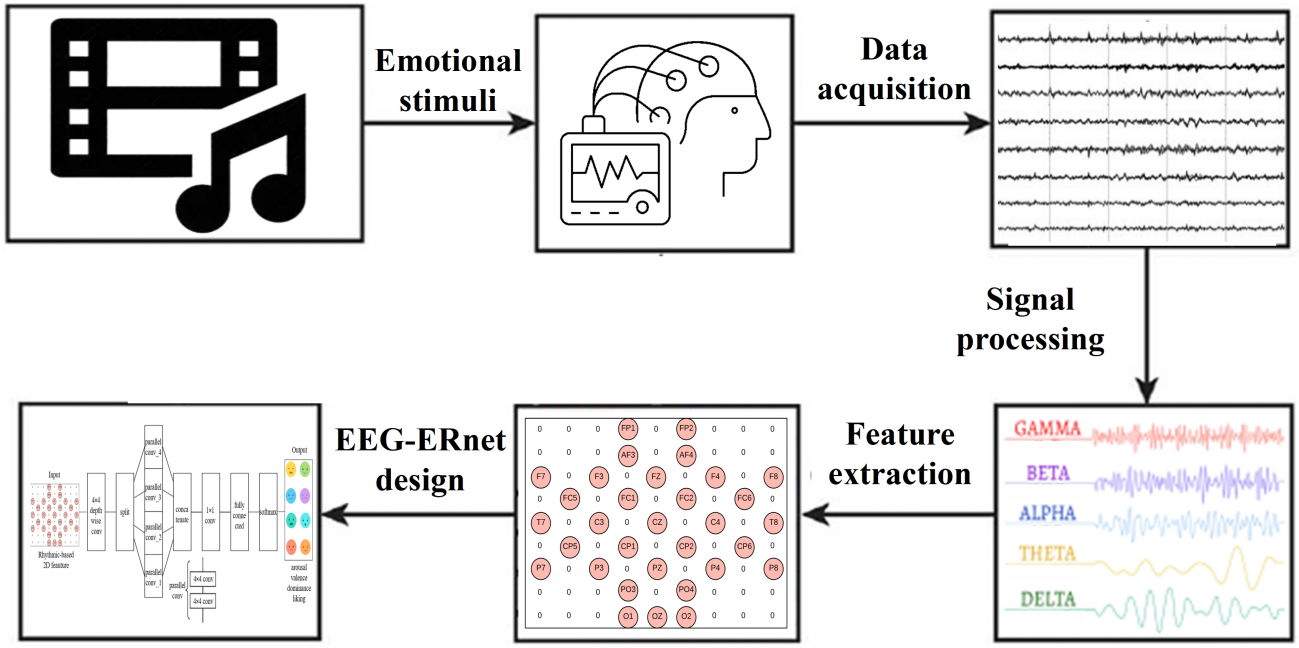

Fig. 1. The overall workflow of the proposed method. EEG, electroencephalography.

According to the related works discussed above, deep learning models, such as CNNs, offer potential for robust EEG emotion recognition. Nonetheless, they are black-box models that cannot provide insightful properties related to neuroscience knowledge. Therefore, understanding how and why rhythmic features influence emotion recognition is a challenging task. Additionally, reducing input temporal data is beneficial for deploying the model on resource-limited embedded devices. Meanwhile, a model that does not rely on individual data but is available for all subjects is more desired. Hence, it is meaningful to develop a subject-independent CNN model employing brain rhythms, which is the motivation of this work.

3. Proposed Method

3.1 Overall Workflow

For clarity, the overall workflow of the proposed method is illustrated in Fig. 1. First, the EEG signals are acquired from the DEAP dataset, which adopts music videos as stimuli. The multi-channel recordings are collected using a 32-channel system. Next, the short-time Fourier transform (STFT) is applied to convert each EEG from the time domain into the frequency domain, and PSDs are extracted based on various brain rhythms. Subsequently, rhythmic-based 2D images are projected by those extracted PSDs with spatial information, providing a comprehensive representation of channel, rhythm, and temporal properties. After that, the EEG-ERnet is designed using the depth-wise parallel CNN architecture, which is then employed for training and testing the rhythmic image features through 10-fold cross-validation. Finally, the most distinguishable brain rhythms associated with specific time intervals are investigated in detail for diverse emotion recognition tasks, offering subject-independent solutions in this field.

3.2 Data Acquisition

Data acquisition is the first step in EEG studies. As the primary objective is to design a subject-independent deep learning model, a publicly available dataset is a suitable choice. Therefore, the DEAP dataset, developed by Koelstra *et al.* [42], was evaluated. This dataset was designed to analyze emotional states, incorporating EEG recordings and comprehensive subjective assessments. Such information makes it valuable for cross-subject evaluation.

In detail, DEAP contains EEG recordings from 32 subjects (17 males and 15 females, aged 27.19 $\pm$ 4.45 years), each of whom watched 40 one-minute music videos prepared to evoke various emotions. The physiological data include 32-channel EEG signals sampled at 512 Hz, as well as peripheral signals such as electrooculogram (EOG), electromyogram (EMG), skin temperature, and respiration. Additionally, this dataset provides a preprocessed version, including downsampled signals at 128 Hz and the removal of artifacts.

Concerning the emotional scenarios, the DEAP dataset contains ratings for three fundamental dimensions: valence, arousal, and dominance, based on a 9-point scale provided by the subjects, i.e., self-assessment manikin (SAM), as illustrated in Fig. 2. Valence denotes the degree of pleasantness of an emotional state, from negative (sadness, fear) to positive (happiness, excitement). Arousal reflects the physiological activation level, from calm (relaxation, boredom) to excitement (stress, enthusiasm). Dominance refers to the sense of control over emotional experi-

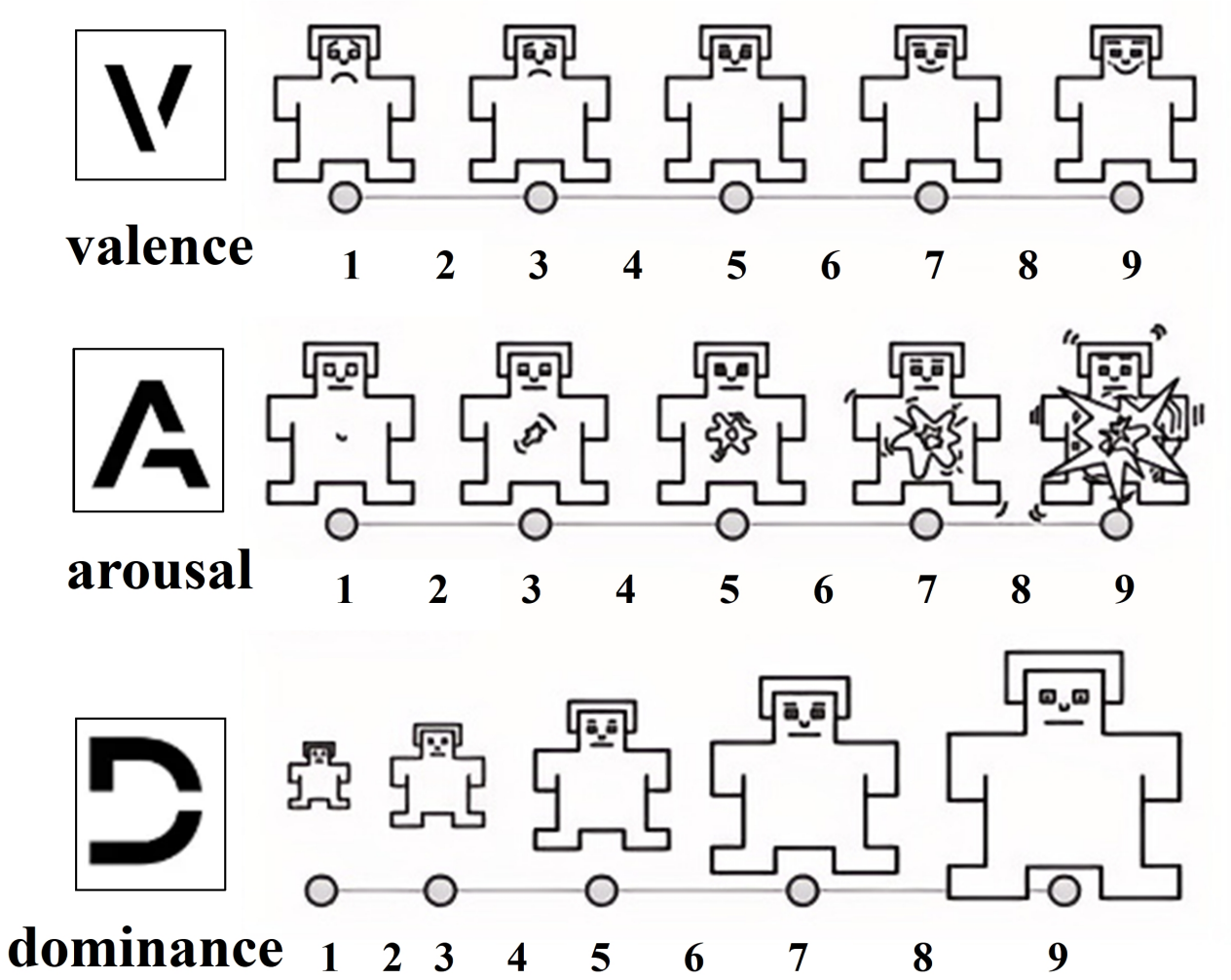

Fig. 2. Self-assessment manikin (SAM) in the DEAP dataset. DEAP, dataset for emotion analysis using physiological signals.

ences, ranging from submissive (fear, helplessness) to dominant (confidence, empowerment). These dimensions form a framework for describing emotional states, where specific emotions can be mapped to different regions within this space. For instance, high valence and high arousal correspond to emotions such as joy or excitement, while low valence and high arousal denote emotions like anger or fear. Meanwhile, the liking ratings are provided, offering an additional aspect to assess subjective preferences in emotion. Therefore, this work focuses on valence, arousal, dominance, and liking tasks, where the binary classification is based on a threshold of 5, following the common practice [43] with the DEAP dataset. After binarization, the class distributions for valence, arousal, dominance, and liking are found to be reasonably balanced, with high/low class ratios of approximately 54:46, 52:48, 51:49, and 53:47, respectively. While these ratios do not represent a perfect balance, they are sufficiently close to ensure fair model training without the need for class-weighting or data balancing techniques.

3.3 Feature Extraction

The initial step in feature extraction involves obtaining the PSDs, a prevalent process in EEG emotion recognition using CNN, as incorporating PSDs as features into a CNN helps the model discern valuable emotion-associated patterns [18,32]. To this end, the STFT is applied. The EEG $x(t)$ is divided into overlapping segments utilizing a sliding window function to reduce spectral leakage. In this work, the window function employs the Hamming window of 128 samples with 50% overlap, ensuring a balance between time and frequency resolution. The STFT is then acquired by:

$$\mathrm{STFT}_x(t,f) = \sum_{n=-\infty}^{\infty} x(n)w(n-t)e^{-j2\pi fn} \quad (1)$$

where $w(n\text{-}t)$ denotes the Hamming window function centered at time t, $e^{-j2\pi fn}$ represents the complex exponential term of the Fourier basis functions, which performs the transformation from the time domain to the frequency f.

The squared magnitude of the STFT provides the spectrogram from which the PSD is derived. To link with the

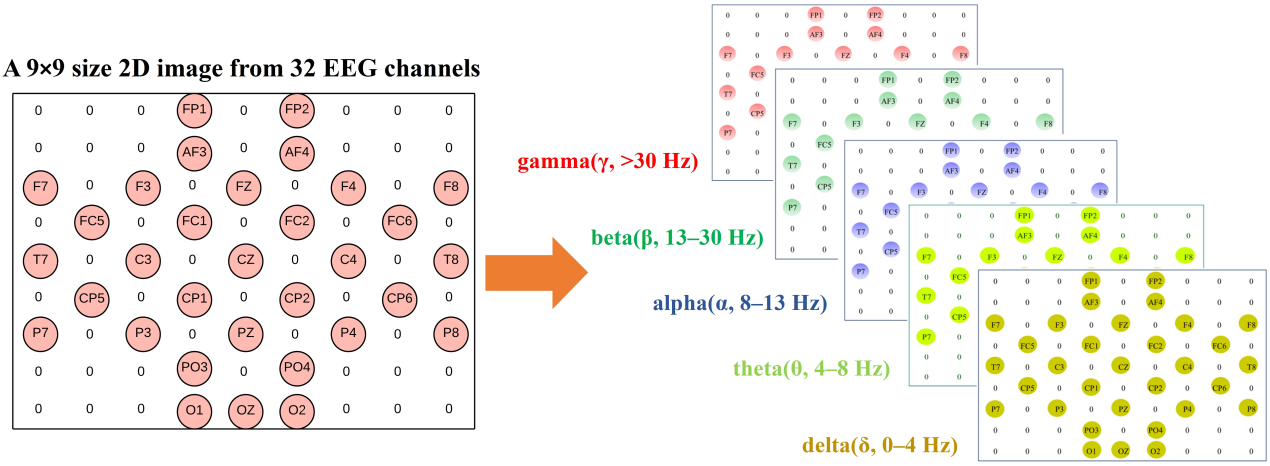

Fig. 3. The 9 × 9 rhythmic-based 2D image derived from the spectral features with 32 EEG channels.

brain rhythms, the frequency bins corresponding to each range are summed. Thus, the PSD for each brain rhythm is calculated by:

$$\mathrm{PSD}(t, f) = |\mathrm{STFT}_x(t, f)| \tag{2}$$

where the PSD is integrated over the predefined brain rhythms $B_K \in \{\delta, \theta, \alpha, \beta, \gamma\}$ to generate the feature:

$$P_k = \int_{f \in B_k} \mathrm{PSD}(t, f) df \tag{3}$$

Once the PSDs are extracted, they are organized into spectral features for each channel. To further enhance the ability to present spatial information, these spectral features are projected onto a 2D image that preserves the spatial arrangement of high-dimensional EEG channels, i.e., these features are arranged into 2D matrices $I_k \in \mathbb{R}^{H \times W}$. This 2D projection way, inspired by geographical mapping, maintains the geometric relationships between adjacent EEG channels [44]. Here, the 2D image is divided into a mesh of pixels, each representing a specific channel. Based on this, the spectral features acquired from the previous step are mapped onto the corresponding pixels, generating a spatially coherent brain rhythmic representation. Therefore, the 32-channel data is projected as a 9 × 9 size 2D image, and each rhythm generates its corresponding 2D image, as illustrated in Fig. 3.

Additionally, the temporal dynamics indicate the changes in emotional responses. As a result, the inputs are generated by smaller segments, each representing a specific 5-second interval, which helps analyze the EEG signals over time and removes redundant data. For each 5-second EEG data, the PSDs of various brain rhythms are extracted and projected onto a 9 × 9 image. After that, normalization is adopted to mitigate the impact of scaling differences and enhance the convergence speed during training. In this work, the min-max normalization is adopted, which rescales the rhythmic image features to a range of 0 to 1. A sample (DEAP, subject S12, γ rhythm, 0–5 s interval) is displayed in Fig. 4. Based on the above steps, the feature extraction process involves 12 × 5 × 40 (time intervals × brain rhythms × music videos) per subject, which serves as the input to the EEG-ERnet.

3.4 EEG-ERnet Design

Concerning EEG-ERnet, a depthwise parallel CNN architecture is applied to address the computational efficiency and accuracy requirements for EEG emotion recognition. Its kernel design phase involves a pre-test to determine parameters for the convolutional layers. The initialization method of the convolutional kernels has been evaluated through comparative analysis, where Gaussian distribution initialization demonstrates an improvement in accuracy over zero initialization, achieving a 45% enhancement (94.58% vs. 49.88%) in a preliminary investigation. Subsequently, the kernel size has been pre-tested with 2 × 2, 3 × 3, and 4 × 4 kernels, yielding accuracies of 91.01%, 92.69%, and 94.58%, respectively. Hence, the 4 × 4 kernel size is employed.

The traditional CNN structure consists of three continuous convolutional layers without max pooling layers. The convolution filter numbers are set at 64, 128, and 256, respectively, with a 1 × 1 convolution layer employed to reduce the number of output feature maps, alleviating computational pressure in the flatten layer. In the end, a fully connected layer with 1024 neurons and a softmax classifier is utilized. Nevertheless, this traditional structure is less effective in analyzing the 2D rhythmic image features, so the depthwise convolution layers are applied, which replace conventional convolution layers with depthwise ones, maintaining the same number of filters and sizes. The

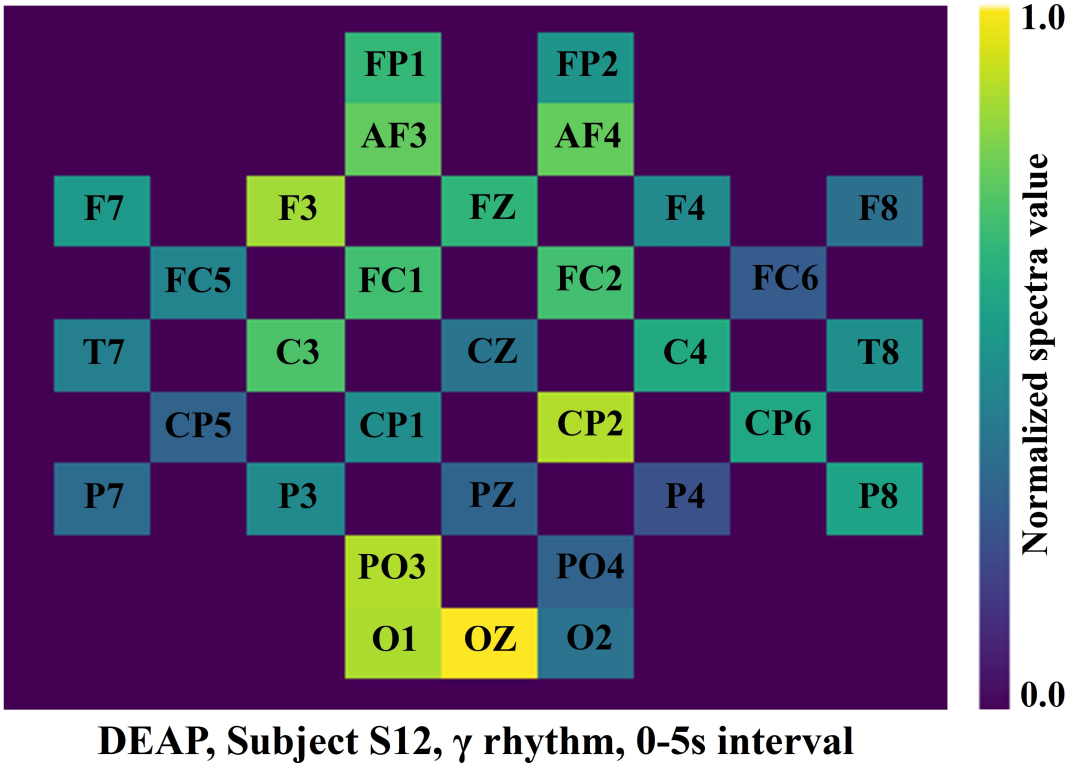

Fig. 4. A normalization sample for rhythmic image (DEAP, subject S12, γ rhythm, 0–5 s interval). The color bar corresponds to the strength of PSDs. PSDs, power spectral densities.

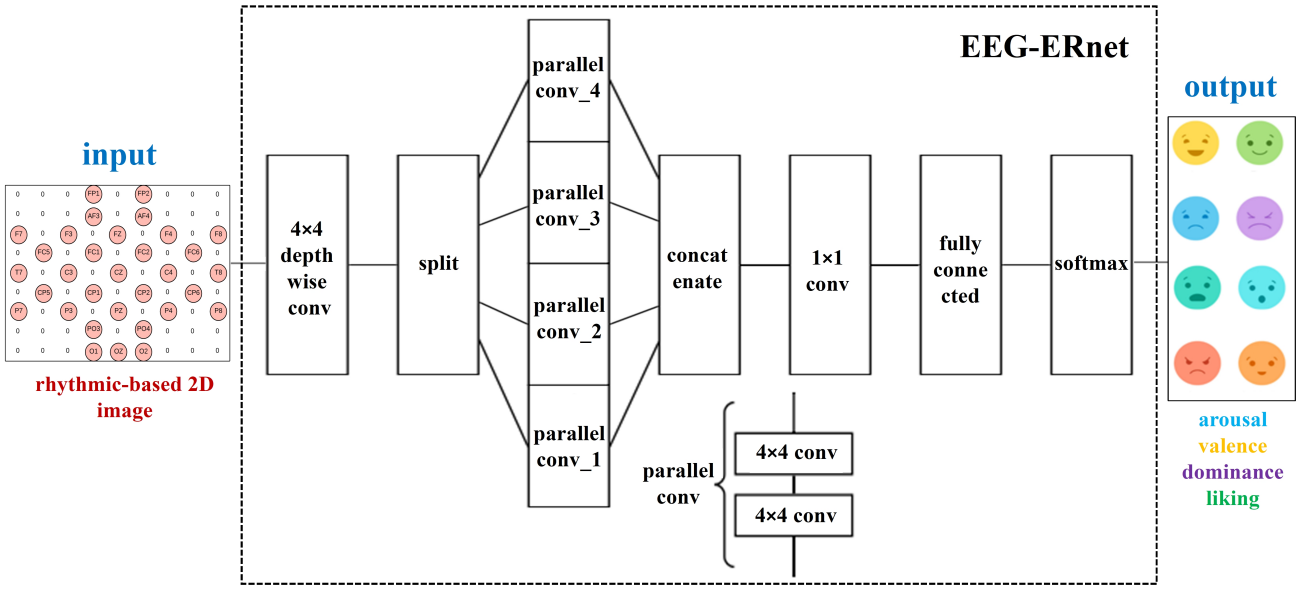

Fig. 5. The EEG-ERnet model based on depthwise parallel CNN. CNN, convolutional neural network.

depthwise convolution effectively reduces computational cost by processing each input data separately, which is more suitable for 2D features than the traditional structure.

Now, the first layer utilizes a depthwise convolution layer with 16 filters, generating 64 separate feature maps. A split layer then divides these feature maps into four sub-batches. Four parallel continuous CNNs are designed to process each sub-batch independently. Each continuous CNN consists of two convolution layers without pooling layers, with filter numbers set at 32 and 64, respectively. A concatenate layer is subsequently employed to stack the feature maps, followed by a 1 × 1 convolution layer with 64

filters to reduce the number of output feature maps. Finally, a fully connected layer with 1024 neurons and a softmax function is implemented. As this work aims at binary classification tasks, the output is a high (1) or low (0) state for arousal, valence, dominance, and liking tasks. The EEG-ERnet is depicted in Fig. 5, where its parameterized details are listed in Table 1, and a pseudocode-style algorithm (Algorithm 1) is described as follows:

Algorithm 1: Parallel EEG-ERnet Algorithm

Input: Four image views $\{I_1, I_2, I_3, I_4\}$

Output: Predicted label $y \in \{0, 1\}$

1. In parallel, compute each branch output.

2. Concatenate the outputs: $F = \text{Concat}(F_1, F_2, F_3, F_4)$.

3. Apply dropout, flattening, and a fully connected layer.

4. Compute the prediction y.

Beyond kernel size and initialization, several hyperparameters are optimized through empirical evaluation using a grid search on the training folds. The learning rate is set to 1×10^{-3}, and the batch size is chosen to be 32, balancing stability and efficiency. Training is conducted for up to 100 epochs, with early stopping if no improvement is observed for 10 consecutive epochs. A dropout rate of 0.5 is applied before the fully connected layer to reduce overfitting. The Adam optimizer is selected for its adaptive learning rate capability. ReLU is utilized as the activation function after the convolutional layers. Additionally, an L2 weight decay of 1×10^{-5} is employed on the convolutional layers to facilitate generalization. Such hyperparameter choices contribute to the stable performance of EEG-ERnet across subjects and tasks.

Table 1. The parameterized details of the EEG-ERnet model.

Compoment	Filter shape	Input size
cov_1	4 × 4 × 64	9 × 9 × 4
cov_2	4 × 4 × 128	9 × 9 × 64
cov_3	4 × 4 × 256	9 × 9 × 128
cov_4	1 × 1 × 64	9 × 9 × 256
fully connected	5184 × 1024	9 × 9 × 64
softmax	/	1 × 1024

Finally, to ensure subject independence, all preprocessing steps, including STFT, PSD computation, and min-max normalization, are applied independently to each trial. STFT and PSD are computed on a per-trial basis, utilizing predefined frequency ranges without any data outside the trial. The min-max normalization is performed on each rhythmic-based 2D image, with the minimum and maximum values calculated from that image alone. No statistics are shared across training and testing folds, providing strict separation and eliminating any potential information leakage in the subject-independent evaluation protocol. Then, a subject-wise 10-fold cross-validation is adopted. Specifically, the DEAP dataset, including 32 subjects, is partitioned such that each fold contains data from approximately 3–4 individual subjects, which are entirely withheld for testing, while the model is trained on data from the remaining subjects. This process is repeated ten times, so that each subject serves as the test set once. The strict separation of subjects between training and testing sets prevents subject-specific feature leakage, supporting the development of generalizable models suitable for cross-subject emotion recognition scenarios. The performance is evaluated based on the average accuracy across all ten folds, which can avoid overfitting the training data. Please note that, since brain rhythms and time intervals are key aspects assessed in this work, the training and testing are based on specific rhythmic image features extracted at the same 5-second interval across 32 subjects. To clarify the sequence of operations in EEG-ERnet, a pseudocode-style algorithm (Algorithm 2) is provided that summarizes the input, processing stages, and output. It begins with 2D input images and proceeds through the multi-branch CNNs with depthwise separable convolution layers, eventually producing the final classification output.

Algorithm 2: EEG-ERnet Classification Procedure

Input: rhythmic-based 2D image $I \in \mathbb{R}^{H \times W \times C}$

Output: Predicted label $y \in \{0, 1\}$

1. Apply depthwise separable convolutions in 4 parallel branches.

2. Perform ReLU activation, batch normalization, and max pooling on each branch.

3. Concatenate the outputs of all branches.

4. Apply dropout, flattening, and a fully connected layer.

5. Use a softmax function to obtain the final prediction.

Here, the computational complexity per branch is approximately $O(K^2 \cdot H \cdot W \cdot C + H \cdot W \cdot C \cdot D)$, where K is the kernel size, H and W are the image dimensions, C refers to the input channel size, and D denotes the number of filters. The final classification layer adds $O(N \cdot M)$ operations. This architecture is more efficient than traditional CNNs due to the use of depthwise separable convolutions, making EEG-ERnet suitable for real-time or embedded emotion recognition applications.

4. Results and Discussion

4.1 Experimental Results

In this work, MATLAB R2023b (The MathWorks Inc., Natick, MA, USA) was used for programming, and the random seed was set to 42. No learning rate scheduler was applied. Training was performed on an NVIDIA ray tracing eXtreme (RTX) 3090 graphics processing unit (GPU) using compute unified device architecture (CUDA) 11.6 (NVIDIA Corp., Santa Clara, CA, USA) on Ubuntu 20.04 (Canonical Ltd., London, UK), with early stopping applied after 10 epochs of no improvement. The same configu-

Table 2. The valence classification accuracies (mean ± standard deviation%) using the proposed method.

Time interval	δ rhyhtm	θ rhyhtm	α rhyhtm	β rhyhtm	γ rhyhtm
0–5 s	86.31 ± 3.74	88.54 ± 3.86	90.60 ± 2.01	86.65 ± 3.41	82.04 ± 5.21
5–10 s	89.88 ± 4.81	87.59 ± 4.40	90.63 ± 2.94	84.11 ± 6.64	82.76 ± 7.60
10–15 s	89.15 ± 3.45	87.47 ± 3.13	88.50 ± 3.13	80.65 ± 9.34	88.79 ± 2.08
15–20 s	89.18 ± 4.64	89.99 ± 3.57	85.15 ± 5.41	82.23 ± 7.45	86.18 ± 3.75
20–25 s	87.63 ± 3.29	83.82 ± 6.70	88.81 ± 4.57	80.36 ± 9.62	88.61 ± 4.71
25–30 s	88.96 ± 4.60	82.26 ± 8.34	87.76 ± 4.40	84.85 ± 8.63	90.60 ± 2.88
30–35 s	85.38 ± 6.90	83.56 ± 7.09	91.07 ± 2.95	82.01 ± 7.13	88.11 ± 3.56
35–40 s	83.60 ± 7.06	82.83 ± 8.05	87.44 ± 4.91	85.05 ± 7.10	89.68 ± 2.56
40–45 s	83.02 ± 6.76	87.46 ± 3.29	88.67 ± 3.33	87.66 ± 6.89	90.63 ± 4.96
45–50 s	88.93 ± 4.82	90.16 ± 3.52	89.90 ± 4.01	84.54 ± 7.30	88.22 ± 4.53
50–55 s	89.82 ± 2.43	90.97 ± 3.33	90.94 ± 3.08	88.89 ± 6.68	87.72 ± 6.87
55–60 s	86.20 ± 3.07	90.83 ± 3.61	93.27 ± 3.05	91.99 ± 5.24	87.46 ± 6.52

Table 3. The arousal classification accuracies (mean ± standard deviation%) using the proposed method.

Time interval	δ rhyhtm	θ rhyhtm	α rhyhtm	β rhyhtm	γ rhyhtm
0–5 s	87.23 ± 3.45	84.76 ± 4.26	86.43 ± 5.04	87.14 ± 2.51	87.79 ± 3.68
5–10 s	84.56 ± 6.12	85.44 ± 5.72	87.21 ± 4.78	89.16 ± 3.52	85.18 ± 5.03
10–15 s	88.91 ± 2.89	88.57 ± 3.67	84.73 ± 5.39	90.51 ± 3.26	85.60 ± 6.41
15–20 s	85.34 ± 5.78	87.05 ± 4.45	88.93 ± 2.98	90.75 ± 2.21	86.13 ± 3.39
20–25 s	85.67 ± 4.34	81.63 ± 7.12	82.35 ± 5.72	86.75 ± 2.57	88.58 ± 3.51
25–30 s	83.78 ± 6.98	86.03 ± 4.94	82.85 ± 5.11	86.17 ± 5.90	84.54 ± 5.64
30–35 s	86.45 ± 4.56	83.69 ± 5.53	86.28 ± 3.97	88.76 ± 2.37	85.74 ± 6.32
35–40 s	88.12 ± 3.98	88.33 ± 2.68	87.91 ± 3.42	84.05 ± 5.66	89.60 ± 2.09
40–45 s	84.89 ± 5.23	90.18 ± 4.71	88.21 ± 4.30	86.14 ± 5.79	88.90 ± 3.87
45–50 s	88.34 ± 3.12	87.86 ± 4.94	86.21 ± 4.69	90.87 ± 2.94	89.34 ± 4.01
50–55 s	87.56 ± 3.78	88.52 ± 4.30	86.62 ± 4.36	92.16 ± 2.73	88.69 ± 4.87
55–60 s	89.23 ± 2.45	90.41 ± 2.75	89.72 ± 2.69	91.40 ± 2.56	90.77 ± 2.35

ration was employed across all cross-validation folds and rhythm-interval evaluations. Extensive experiments were conducted based on four tasks: valence, arousal, dominance, and liking. Consequently, it is meaningful to identify distinguishable rhythms and appropriate 5-second intervals to recognize different emotions, offering insights into the characteristics of emotion recognition. To this end, all classification results were evaluated using the mean and standard deviation of accuracy across the 10-fold cross-validation for the valence, arousal, dominance, and liking tasks, as detailed in Tables 2,3,4,5. No inferential statistical tests were applied, as the primary goal of this work is to assess network model performance across rhythm-specific image inputs rather than test specific hypotheses. Also, please note that no demographic covariates were included due to the limited metadata available in the dataset.

Table 2 presents classification accuracies for valence, showing the 55–60 s interval where the α rhythm achieves the highest accuracy of 93.27 ± 3.05%. The α rhythm is typically associated with a relaxed yet attentive mental state, characterized by self-referential thought and introspection [45]. During the 55–60 s interval of the music video, it is plausible that most subjects have settled into a state of reflective calmness, which may enhance the neural correlates associated with valence [46]. This state could facilitate more precise differentiation between positive and negative stimuli. That means the heightened activity potentially enhances the neural representation of valence, resulting in improved classification performance.

Table 3 focuses on the arousal task and reveals that the β rhythm during the 50–55 s interval achieves the highest accuracy of 92.16 ± 2.73%. The β rhythm is generally associated with cognitive functions and increased alertness. According to the results, heightened arousal during the 50–55 s interval of the music video is associated with improved mental performance and emotional engagement, as reflected by the β rhythm. This finding aligns with the understanding that arousal can modulate neural activity [47]. Additionally, the β rhythm at 50–55 seconds could indicate a response to the music video stimulus, requiring a longer duration for cognitive processing, similar to the valence.

Table 4 indicates that the θ rhythm provides the highest dominance classification accuracy of 90.56 ± 4.44% during the 55–60 s interval. The θ rhythm is known to be involved in memory, learning, and decision-making processes [48]. Its association with the perception of dominance may stem from its role in integrating sensory infor-

Table 4. The dominance classification accuracies (mean ± standard deviation%) using the proposed method.

Time interval	δ rhyhtm	θ rhyhtm	α rhyhtm	β rhyhtm	γ rhyhtm
0–5 s	82.34 ± 6.78	82.04 ± 3.83	84.97 ± 5.77	82.67 ± 6.67	85.45 ± 4.89
5–10 s	81.12 ± 9.45	82.78 ± 5.34	82.27 ± 6.13	86.45 ± 6.34	88.23 ± 4.56
10–15 s	84.56 ± 5.89	88.06 ± 4.86	82.80 ± 8.59	84.89 ± 5.78	82.89 ± 6.23
15–20 s	82.23 ± 6.67	86.90 ± 5.54	82.92 ± 7.95	85.67 ± 5.45	82.67 ± 7.89
20–25 s	85.78 ± 5.12	88.32 ± 3.72	83.73 ± 6.94	86.34 ± 5.01	81.67 ± 9.23
25–30 s	84.12 ± 7.45	89.55 ± 4.01	83.26 ± 6.15	87.91 ± 4.78	81.23 ± 7.89
30–35 s	80.12 ± 9.56	87.21 ± 5.81	86.55 ± 4.07	87.78 ± 4.45	85.89 ± 2.56
35–40 s	80.45 ± 8.89	89.48 ± 6.24	87.17 ± 6.55	88.45 ± 5.12	86.23 ± 3.56
40–45 s	82.34 ± 6.78	88.32 ± 5.58	85.69 ± 4.29	90.01 ± 5.67	85.89 ± 5.34
45–50 s	82.45 ± 6.89	89.88 ± 3.75	87.17 ± 5.16	86.45 ± 7.34	84.23 ± 5.67
50–55 s	80.78 ± 6.12	89.62 ± 4.83	87.98 ± 6.26	89.67 ± 6.78	84.78 ± 6.23
55–60 s	85.89 ± 6.21	90.56 ± 4.44	88.81 ± 5.92	90.09 ± 4.00	84.89 ± 6.56

Table 5. The liking classification accuracies (mean ± standard deviation%) using the proposed method.

Time interval	δ rhyhtm	θ rhyhtm	α rhyhtm	β rhyhtm	γ rhyhtm
0–5 s	82.45 ± 6.89	84.78 ± 5.78	80.67 ± 8.67	85.34 ± 6.08	86.08 ± 5.78
5–10 s	80.34 ± 7.56	83.56 ± 7.45	83.45 ± 6.34	85.82 ± 6.45	86.68 ± 5.66
10–15 s	80.67 ± 7.89	84.12 ± 5.89	82.89 ± 7.78	85.92 ± 5.89	84.45 ± 5.45
15–20 s	81.34 ± 6.56	83.78 ± 7.56	80.67 ± 7.45	85.23 ± 5.67	85.78 ± 5.78
20–25 s	81.89 ± 7.12	82.12 ± 8.01	81.34 ± 6.01	84.78 ± 5.12	84.12 ± 6.12
25–30 s	82.56 ± 5.89	80.78 ± 8.78	77.01 ± 9.78	84.45 ± 5.89	84.45 ± 7.45
30–35 s	80.23 ± 6.56	80.45 ± 8.45	79.78 ± 5.45	83.12 ± 5.56	82.78 ± 6.78
35–40 s	80.89 ± 5.23	79.12 ± 8.12	78.45 ± 7.12	84.89 ± 6.23	82.12 ± 7.01
40–45 s	80.56 ± 7.89	77.78 ± 8.78	79.01 ± 6.67	81.34 ± 5.78	81.18 ± 7.56
45–50 s	80.23 ± 5.56	79.45 ± 8.45	80.45 ± 6.34	80.00 ± 7.46	82.12 ± 6.89
50–55 s	80.89 ± 6.23	81.12 ± 6.12	83.89 ± 4.91	77.78 ± 9.12	83.56 ± 6.45
55–60 s	81.56 ± 6.89	79.78 ± 9.78	83.67 ± 5.78	78.45 ± 8.89	82.78 ± 6.78

mation and processing complex emotions [49]. The enhanced classification accuracy observed in the 55–60 s interval could be due to the θ rhythm's involvement in binding emotional information with contextual cues, which are vital for recognizing dominance-related signals, such as music video stimuli.

Table 5 shows that the γ rhythm offers the highest classification accuracy of 86.68 ± 5.66% in the 5–10 s interval for the liking task. The γ rhythm is associated with higher cognitive functions, where its prominence in the early interval suggests a rapid neural processing mechanism for music video stimuli, mainly due to the brain's quick evaluation of sensory input, leading to an immediate emotional reaction [50]. This response is helpful for social interactions, where swift identification of positive stimuli is beneficial. That may be why most subjects determine their levels of liking at the beginning of emotional stimuli.

Finally, the varying accuracies across classification tasks may be attributed to inherent emotional complexity. Emotions, such as valence and arousal, are two fundamental dimensions that are widely recognized for understanding human emotions. Dominance and liking may involve more subjective and context-dependent interpretations. In this regard, the proposed EEG-ERnet demonstrates its ability to maintain impressive performance across individuals and tasks, since its architecture incorporates depthwise parallel CNNs that can effectively analyze both local and global features of rhythmic-based 2D images. The explainability of EEG-ERnet further contributes to providing insights into the brain rhythms and time intervals that are most beneficial in emotion recognition, improving its decision-making process in various cases across subjects. It is vital for real-world applications where emotional recognition systems should be adaptable to different users and contexts.

4.2 Comparative Study

A comprehensive comparative study was conducted to evaluate the proposed EEG-ERnet. First, in the ablation experiment, an initial baseline model consisting of three depthwise convolution layers, a 1 × 1 convolution layer, and a fully connected softmax function was compared. This initial version serves as the basis for the EEG-ERnet. The results are listed in Table 6.

Table 6 shows that the baseline model, which utilizes cascading depthwise convolution layers, suffers from a limitation in its architecture. Depthwise convolution, while computationally efficient, lacks feature integration. Hence,

Table 6. The accuracy comparison (mean ± standard deviation%) between the baseline and proposed EEG-ERnet.

Model	Valence	Arousal	Dominance	Liking
Baseline	64.68 ± 7.08	69.34 ± 6.96	64.06 ± 4.07	59.36 ± 9.25
EEG-ERnet	93.27 ± 3.05	92.16 ± 2.73	90.56 ± 4.44	86.68 ± 5.66

it cannot fully exploit the multi-dimensional nature of the data to make accurate classifications. In contrast, EEG-ERnet employs a parallel architecture that enhances feature integration, resulting in improved performance across all dimensions. Such an architectural advantage makes the EEG-ERnet more appropriate for cross-subject emotion recognition tasks.

A comprehensive overview of various EEG emotion recognition methods utilizing the DEAP dataset is offered by the comparative study presented in Table 7 (Ref. [14,18,29,34,51–55]). Different approaches have been assessed, including traditional machine learning algorithms such as k-NN, SVM, and RF, as well as more advanced deep learning techniques like CNN and its variants. Mahmoud *et al*. [18] employed a 2D-CNN with PSD features, achieving the highest valence and arousal recognition accuracy of 94.23% and 93.78%, respectively. However, their method did not cover all dimensions. The proposed EEG-ERnet has demonstrated impressive performance across all dimensions, achieving accuracies of 93.27% for valence, 92.16% for arousal, 90.56% for dominance, and 86.68% for liking. It is advantageous over previous works, particularly in terms of stable performance across subjects and emotional factors, which are key concerns.

Compared to methods that rely on temporal snapshots or spectrogram images, the EEG-ERnet enables the analysis of spatial and temporal information from rhythmic image features. It exhibits several advantages in EEG emotion recognition. First, by utilizing depthwise convolution layers, the model analyzes the spatial information of EEG signals more effectively than traditional CNNs. This architecture reduces computational costs and enhances the ability to identify spatial patterns associated with specific emotions. Second, the parallel processing of sub-batches offers more comprehensive feature integration, allowing the model to handle high-dimensional and complex data. As demonstrated in the experiments, such design choices collectively contribute to superior performance across four dimensions. Additionally, the results provide valuable insights into the characteristics of EEG emotion recognition, as they identify specific brain rhythms and time intervals associated with recognizing emotions. Such findings align with the known roles of these rhythms in emotional processing. Thus, the model identifies the key factors relevant to particular dimensions. Overall, the accuracies acquired in four dimensions demonstrate that it is well-suited to recognize the complex patterns in EEG signals associated with different emotions. It can be said that the EEG-ERnet provides an impressive solution that outperforms existing methods by addressing the limitations of incomplete dimensions in a subject-independent manner with only 5-second data sources.

4.3 Discussion

First, this work focuses on the five brain rhythms due to their strong neurophysiological grounding in the context of emotion recognition. While emerging techniques such as EMD can extract non-standard or adaptive frequency components, the proposed method prioritizes choice based on explainability and computational feasibility. Nevertheless, it is recognized that the potential of such methods, including high-frequency bands beyond 60 Hz, lies in uncovering additional emotion-relevant information. Future work will explore the integration of EMD-derived IMFs into the 2D image framework to enhance the discriminative ability for multiple-level classification tasks.

Second, the choice of a 5-second interval in this work is due to the balance between temporal resolution and spectral stability. This duration has also been used in emotion recognition and aligns with previous works [3,4] on the DEAP dataset. A shorter interval, like 2 seconds, may offer finer temporal granularity. Still, they could suffer from reduced frequency resolution, particularly for low-frequency bands. In contrast, a longer interval, such as 10 seconds, may average out dynamic changes and reduce the number of training samples. Although this work adopts a fixed non-overlapping 5-second interval, future work will investigate the effects of different segmentations, including overlapping and multi-scale windows, to improve real-time responsiveness and model performance.

Third, min-max normalization is applied to each rhythmic image to rescale PSD values to the [0, 1] range, which provides a consistent input scale across samples and reduces the impact of extreme values. While this method preserves the relative topographic and spectral structure of each input, it does not standardize inter-subject statistics. Therefore, to mitigate inter-subject variability, a subject-independent cross-validation involves training and testing on entirely different sets of individuals. Although no explicit domain adaptation is adopted, it promotes cross-subject generalization. Future work will incorporate advanced inter-subject normalization or domain-adaptive learning techniques, such as Riemannian alignment, statistical matching, or adversarial adaptation, to further enhance model robustness in highly diverse populations.

Next, the DEAP dataset assessed in this work includes EEG recordings from 32 subjects, spanning a reasonable

Table 7. A comparative study of EEG emotion recognition works that utilize the DEAP dataset.

Work	Classifier	Feature	Classification accuracy (%)			
			Valence	Arousal	Dominance	Liking
Wang *et al*. [14]	2D-CNN-LSTM	DEFM	91.92	92.31	/	/
Mahmoud *et al*. [18]	2D-CNN	PSD	94.23	93.78	89.54	/
Sarma and Barma [29]	k-NN, SVM, RF	PSD, CWT	82.21	86.03	/	/
Yang *et al*. [34]	Multi-column CNN with weighted sum	Temporal snapshots of EEG signals	90.01	90.65	/	/
Farokhah *et al*. [51]	Simplified 2D-CNN	Spectrogram images generated from ten selected EEG channels	89.31	91.28	/	/
Lin *et al*. [52]	Channel selection graph neural network	DE, PLI	90.74	91.00	/	/
Al-Asadi *et al*. [53]	Semi-supervised EEG-based emotion classifier by appropriate regularization terms	Raw EEG signals with two types of augmentations	88.44	91.77	/	/
Yilmaz *et al*. [54]	k-NN, SVM	AAG, SIFT	90.94	92.44	/	/
Wan *et al*. [55]	Light gradient boosting machine	PSD, DE, SASI, wavelet energy, entropy	84.03	84.37	/	/
This work	EEG-ERnet based on depthwise parallel CNN	Rhythmic-based 2D image	93.27	92.16	90.56	86.68

CNN, convolutional neural network; LSTM, long short-term memory; DEFM, differential entropy feature matrix; PSD, power spectral density; k-NN, k-nearest neighbors; SVM, support vector machine; RF, random forest; CWT, continuous wavelet transform; PLI, phase lag index; AAG, angle amplitude graphs; SIFT, scale-invariant feature transform; SASI, spectral asymmetry index; DE, differential entropy.

demographic range. But it does not represent specific clinical populations. Also, it is relatively small compared to datasets in other EEG-based studies, which may constrain the model's generalizability to broader or clinical populations. Meanwhile, in the DEAP dataset, the emotional responses may be influenced by cultural or psychological factors not accounted for, and no inferential statistical tests or covariate analyses are performed based on the results obtained, which limits the investigation of group differences, a limitation regarding personalized affective modeling. Future work will involve larger, more diverse studies and the introduction of covariate-aware modeling to enhance personalization.

Furthermore, EEG signals in the DEAP dataset may reflect overlapping cognitive phenomena beyond core emotional responses, including subjective perceptions like luck, expectation, or decision uncertainty. These non-emotional components can confound emotion recognition. To address this issue, the proposed method mitigates it by using band-specific rhythmic representations and parallel CNN branches, which help isolate emotion-relevant frequency dynamics. Meanwhile, the use of subject-independent training emphasizes generalizable emotion-related features while suppressing subject-specific cognitive noise. Future work will incorporate explicit component separation techniques such as adversarial learning to disentangle emotional signals from co-occurring cognitive influences.

Finally, this work identifies the brain rhythm and temporal interval combinations that are most informative for each emotional dimension. To this end, the experiments have thoroughly evaluated the model's performance across 60 rhythm-interval configurations in a controlled and consistent cross-validation setting. All evaluations are performed using subject-independent cross-validation, guaranteeing that performance is not inflated due to subject-specific overfitting. Therefore, the results add a layer of generalization to mitigate risk from testing multiple configurations. Such an approach is suitable for portable emotion-aware devices, as it utilizes fewer data sources with only 5-second data. In the future, a multi-configuration approach will be considered to enhance performance across various cases. Trivial baselines and nested cross-subjects will also be incorporated into framework-related applications, such as those for depression detection, to improve the model's advantage.

5. Conclusions

This work proposes the EEG-ERnet model, which employs a depthwise parallel CNN to classify the spatial and temporal features of emotional EEG signals. Using rhythmic-based 2D images extracted from multi-channel EEG recordings of specific 5-second intervals, the model can identify particular brain rhythms and time intervals most appropriate for recognizing different emotions. The experimental results from the DEAP dataset demonstrated that the α rhythm during the 55–60 s interval achieves the highest accuracy of 93.27 $\pm$ 3.05% for valence, the β rhythm during the 50–55 s interval shows the best perfor-

mance of 92.16 ± 2.73% for arousal, the θ rhythm provides the highest dominance classification accuracy of 90.56 ± 4.44% during the 55–60 s interval, and γ rhythm offers the highest classification accuracy of 86.68 ± 5.66% in the 5–10 s interval for the liking task. These findings demonstrate the model's ability to identify neural correlates and temporal dynamics in emotion recognition. Such an impressive performance indicates that the EEG-ERnet is suitable for recognizing complex emotional patterns in EEG signals and offering characteristics regarding brain rhythms and time intervals independently of the subject, which benefits the development of practical emotion-aware devices in daily applications.

Availability of Data and Materials

The EEG data analyzed during the current study are from a public dataset DEAP (http://www.eecs.qmul.ac.uk/mmv/datasets/deap). Other code will be made available from the corresponding author on reasonable request.

Author Contributions

SZ, CL, JRW, and JLi designed the research. SZ, JLi, and JLv performed the research. CL, JJW, YY, XL, and JLv analyzed the data. SZ, JLi, MIV, and RC interpreted the results. SZ, CL, and JLi revised the paper. JRW, XL, and JLv conceptualized the method. SZ, JLi, JLv, MIV, and RC administrated the project and made some contributions to the figures. SZ, CL, JLi, and RC investigated the dataset and supported funding acquisition. JJW, YY, XL, and MIV provided computing resource and supervised the research. SZ, CL, JLi, and JLv wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

The authors would like to appreciate the special contributions from Key Laboratory of Numerical Simulation of Sichuan Provincial Universities, ZUMRI-LYG Joint Lab, and Digital Content Processing and Security Technology of Guangzhou Key Laboratory.

Funding

This work was supported in part by the Guangzhou Science and Technology Plan Project under Grants 2024B03J1361 and 2023B03J1327, in part by the Research Fund of Key Laboratory of Numerical Simulation of Sichuan Provincial Universities under Grant 2024SZFZ007, in part by the Sichuan Science and Technology Program under Grant 2025ZNSFSC0780, in part by the Foundation of the 2023 Higher Education Science Research Plan of the China Association of Higher Education under Grant 23XXK0402, in part by the Foundation of the Sichuan Research Center of Applied Psychology (Chengdu Medical College) under Grant CSXL-25102, in part by the Neijiang Philosophy and Social Science Planning Project under Grant NJ2024ZD014, in part by the Guangdong Province Ordinary Colleges and Universities Young Innovative Talents Project under Grant 2023KQNCX036, in part by the Scientific Research Capacity Improvement Project of the Doctoral Program Construction Unit of Guangdong Polytechnic Normal University under Grant 22GPNUZDJS17, in part by the Graduate Education Demonstration Base Project of Guangdong Polytechnic Normal University under Grant 2023YJSY04002, in part by the Open Research Fund of State Key Laboratory of Digital Medical Engineering under Grant 2025-M10, and in part by the Research Fund of Guangdong Polytechnic Normal University under Grant 2022SDKYA015.

Conflict of Interest

The authors declare no conflict of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

The manuscript was written entirely by the authors. AI-based tools (Grammarly, ChatGPT-4.0) were used only for minor English language correction and grammar checking. All intellectual content, experiments, analysis, and interpretations were conceived, designed, and executed solely by the authors.

References

[1] Samal P, Hashmi M. Role of machine learning and deep learning techniques in EEG-based BCI emotion recognition system: A review. Artificial Intelligence Review. 2024; 57: 50. https://doi.org/10.1007/s10462-023-10690-2.

[2] Li JW, Barma S, Mak PU, Chen F, Li C, Li MT, *et al*. Single-Channel Selection for EEG-Based Emotion Recognition Using Brain Rhythm Sequencing. IEEE Journal of Biomedical and Health Informatics. 2022; 26: 2493–2503. https://doi.org/10.1109/JBHI.2022.3148109.

[3] Li J, Feng G, Ling C, Ren X, Liu X, Zhang S, *et al*. A Resource-Efficient Multi-Entropy Fusion Method and Its Application for EEG-Based Emotion Recognition. Entropy (Basel, Switzerland). 2025; 27: 96. https://doi.org/10.3390/e27010096.

[4] Li JW, Lin D, Che Y, Lv JJ, Chen RJ, Wang LJ, *et al*. An innovative EEG-based emotion recognition using a single channel-specific feature from the brain rhythm code method. Frontiers in Neuroscience. 2023; 17: 1221512. https://doi.org/10.3389/fnins.2023.1221512.

[5] Wei J, Hu G, Yang X, Luu AT, Dong, Y. Learning facial expression and body gesture visual information for video emotion recognition. Expert Systems with Applications. 2024; 237: 121419. https://doi.org/10.1016/j.eswa.2023.121419.

[6] Li J, Huang Y, Lu Y, Wang L, Ren Y, Chen R. Sentiment analysis using e-commerce review keyword-generated image with a

hybrid machine learning-based model. CMC-Computers, Materials & Continua. 2024; 80: 1581–1599.

[7] Khan M, Gueaieb W, El Saddik A, Kwon S. MSER: Multimodal speech emotion recognition using cross-attention with deep fusion. Expert Systems with Applications. 2024; 245: 122946. https://doi.org/10.1016/j.eswa.2023.122946.

[8] Gong L, Chen W, Li M, Zhang T. Emotion recognition from multiple physiological signals using intra-and inter-modality attention fusion network. Digital Signal Processing. 2024; 144: 104278. https://doi.org/10.1016/j.dsp.2023.104278.

[9] Liu H, Zhang Y, Li Y, Kong X. Review on Emotion Recognition Based on Electroencephalography. Frontiers in Computational Neuroscience. 2021; 15: 758212. https://doi.org/10.3389/fncom.2021.758212.

[10] Mumtaz W, Rasheed S, Irfan A. Review of challenges associated with the EEG artifact removal methods. Biomedical Signal Processing and Control. 2021; 68: 102741. https://doi.org/10.1016/j.bspc.2021.102741.

[11] Li J, Feng G, Lv J, Chen Y, Chen R, Chen F, *et al*. A Lightweight Multi-Mental Disorders Detection Method Using Entropy-Based Matrix from Single-Channel EEG Signals. Brain Sciences. 2024; 14: 987. https://doi.org/10.3390/brainsci14100987.

[12] Tolie HF, Ren J, Chen R, Zhao H, Elyan E. Blind sonar image quality assessment via machine learning: Leveraging micro-and macro-scale texture and contour features in the wavelet domain. Engineering Applications of Artificial Intelligence. 2025; 141: 109730. https://doi.org/10.1016/j.engappai.2024.109730.

[13] Yan Y, Ren J, Sun H, Williams R. Nondestructive quantitative measurement for precision quality control in additive manufacturing using hyperspectral imagery and machine learning. IEEE Transactions on Industrial Informatics. 2024; 20: 9963–9975.

[14] Wang T, Huang X, Xiao Z, Cai W, Tai Y. EEG emotion recognition based on differential entropy feature matrix through 2D-CNN-LSTM network. EURASIP Journal on Advances in Signal Processing. 2024; 2024: 49. https://doi.org/10.1186/s13634-024-01146-y.

[15] Zhang E, Zong H, Li X, Feng M, Ren J. ICSF: Integrating intermodal and cross-modal learning framework for self-supervised heterogeneous change detection. IEEE Transactions on Geoscience and Remote Sensing. 2025; 63: 5501516.

[16] Tolie HF, Ren J, Elyan E. DICAM: Deep inception and channel-wise attention modules for underwater image enhancement. Neurocomputing. 2024; 584: 127585. https://doi.org/10.1016/j.neucom.2024.127585.

[17] Cheng Z, Bu X, Wang Q, Yang T, Tu J. EEG-based emotion recognition using multi-scale dynamic CNN and gated transformer. Scientific Reports. 2024; 14: 31319. https://doi.org/10.1038/s41598-024-82705-z.

[18] Mahmoud A, Amin K, Al Rahhal MM, Elkilani WS, Mekhalfi ML, Ibrahim M. A CNN approach for emotion recognition via EEG. Symmetry. 2023; 15: 1822. https://doi.org/10.3390/sym15101822.

[19] Mai G, Minett JW, Wang WSY. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. NeuroImage. 2016; 133: 516–528. https://doi.org/10.1016/j.neuroimage.2016.02.064.

[20] Yu X, Li Z, Zang Z, Liu Y. Real-Time EEG-Based Emotion Recognition. Sensors (Basel, Switzerland). 2023; 23: 7853. https://doi.org/10.3390/s23187853.

[21] Kosonogov V, Ntoumanis I, Hajiyeva G, Jääskeläinen I. The role of engagement and arousal in emotion regulation: an EEG study. Experimental Brain Research. 2024; 242: 179–193. https://doi.org/10.1007/s00221-023-06741-3.

[22] Mennella R, Patron E, Palomba D. Frontal alpha asymmetry neurofeedback for the reduction of negative affect and anxiety. Behaviour Research and Therapy. 2017; 92: 32–40. https://doi.org/10.1016/j.brat.2017.02.002.

[23] Allegretta RA, Rovelli K, Balconi M. The Role of Emotion Regulation and Awareness in Psychosocial Stress: An EEG-Psychometric Correlational Study. Healthcare (Basel, Switzerland). 2024; 12: 1491. https://doi.org/10.3390/healthcare12151491.

[24] Dar MN, Akram MU, Subhani AR, Khawaja SG, Reyes-Aldasoro CC, Gul S. Insights from EEG analysis of evoked memory recalls using deep learning for emotion charting. Scientific Reports. 2024; 14: 17080. https://doi.org/10.1038/s41598-024-61832-7.

[25] Chen J, Li H, Ma L, Bo H, Soong F, Shi Y. Dual-Threshold-Based Microstate Analysis on Characterizing Temporal Dynamics of Affective Process and Emotion Recognition From EEG Signals. Frontiers in Neuroscience. 2021; 15: 689791. https://doi.org/10.3389/fnins.2021.689791.

[26] Subasi A, Tuncer T, Dogan S, Tanko D, Sakoglu U. EEG-based emotion recognition using tunable Q wavelet transform and rotation forest ensemble classifier. Biomedical Signal Processing and Control. 2021; 68: 102648. https://doi.org/10.1016/j.bspc.2021.102648.

[27] Tuncer T, Dogan S, Subasi A. A new fractal pattern feature generation function based emotion recognition method using EEG. Chaos, Solitons & Fractals. 2021; 144: 110671. https://doi.org/10.1016/j.chaos.2021.110671.

[28] Salankar N, Mishra P, Garg L. Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot. Biomedical Signal Processing and Control. 2021; 65: 102389. https://doi.org/10.1016/j.bspc.2020.102389.

[29] Sarma P, Barma S. Emotion recognition by distinguishing appropriate EEG segments based on random matrix theory. Biomedical Signal Processing and Control. 2021; 70: 102991. https://doi.org/10.1016/j.bspc.2021.102991.

[30] Guo L, Li N, Zhang T. EEG-based emotion recognition via improved evolutionary convolutional neural network. International Journal of Bio-Inspired Computation. 2024; 23: 203–213. https://doi.org/10.1504/IJBIC.2024.139268.

[31] Cao L, Zhao W, Sun B. Emotion recognition using multi-scale EEG features through graph convolutional attention network. Neural Networks. 2025; 184: 107060. https://doi.org/10.1016/j.neunet.2024.107060.

[32] Huang Z, Ma Y, Wang R, Li W, Dai Y. A model for EEG-based emotion recognition: CNN-Bi-LSTM with attention mechanism. Electronics. 2023; 12: 3188. https://doi.org/10.3390/electronics12143188.

[33] Zhang Y, Zhang Y, Wang S. An attention-based hybrid deep learning model for EEG emotion recognition. Signal, Image and Video Processing. 2023; 17: 2305–2313. https://doi.org/10.1007/s11760-022-02447-1.

[34] Yang H, Han J, Min K. A Multi-Column CNN Model for Emotion Recognition from EEG Signals. Sensors (Basel, Switzerland). 2019; 19: 4736. https://doi.org/10.3390/s19214736.

[35] Hwang S, Hong K, Son G, Byun H. Learning CNN features from DE features for EEG-based emotion recognition. Pattern Analysis and Applications. 2020; 23: 1323–1335. https://doi.org/10.1007/s10044-019-00860-w.

[36] Cui F, Wang R, Ding W, Chen Y, Huang L. A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition. Mathematics. 2022; 10: 582. https://doi.org/10.3390/math10040582.

[37] Wang X, Ma Y, Cammon J, Fang F, Gao Y, Zhang Y. Self-Supervised EEG Emotion Recognition Models Based on CNN. IEEE Transactions on Neural Systems and Rehabilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society. 2023; 31: 1952–1962. https://doi.org/10.1109/

TNSRE.2023.3263570.
[38] Yao X, Li T, Ding P, Wang F, Zhao L, Gong A, *et al*. Emotion Classification Based on Transformer and CNN for EEG Spatial-Temporal Feature Learning. Brain Sciences. 2024; 14: 268. https://doi.org/10.3390/brainsci14030268.
[39] Lu K, Gu Z, Qi F, Sun C, Guo H, Sun L. CMLP-Net: A convolution-multilayer perceptron network for EEG-based emotion recognition. Biomedical Signal Processing and Control. 2024; 96: 106620. https://doi.org/10.1016/j.bspc.2024.106620.
[40] Qiao Y, Mu J, Xie J, Hu B, Liu G. Music emotion recognition based on temporal convolutional attention network using EEG. Frontiers in Human Neuroscience. 2024; 18: 1324897. https://doi.org/10.3389/fnhum.2024.1324897.
[41] Ezilarasan MR, Leung MF. An efficient EEG signal analysis for emotion recognition Using FPGA. Information. 2024; 15: 301. https://doi.org/10.3390/info15060301.
[42] Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, *et al*. DEAP: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing. 2012; 3: 18–31.
[43] Mohammadi Z, Frounchi J, Amiri M. Wavelet-based emotion recognition system using EEG signal. Neural Computation. 2017; 28: 1985–1990. https://doi.org/10.1007/s00521-015-2149-8.
[44] Liu S, Wang X, Zhao L, Li B, Hu W, Yu J, *et al*. 3DCANN: A Spatio-Temporal Convolution Attention Neural Network for EEG Emotion Recognition. IEEE Journal of Biomedical and Health Informatics. 2022; 26: 5321–5331. https://doi.org/10.1109/JBHI.2021.3083525.
[45] Tarailis P, Koenig T, Michel CM, Griškova-Bulanova I. The Functional Aspects of Resting EEG Microstates: A Systematic Review. Brain Topography. 2024; 37: 181–217. https://doi.org/10.1007/s10548-023-00958-9.
[46] Abdel-Hamid L. An Efficient Machine Learning-Based Emotional Valence Recognition Approach Towards Wearable EEG. Sensors (Basel, Switzerland). 2023; 23: 1255. https://doi.org/10.3390/s23031255.
[47] Mishra S, Srinivasan N, Tiwary US. Dynamic Functional Connectivity of Emotion Processing in Beta Band with Naturalistic Emotion Stimuli. Brain Sciences. 2022; 12: 1106. https://doi.org/10.3390/brainsci12081106.
[48] Wang W. Brain network features based on theta-gamma cross-frequency coupling connections in EEG for emotion recognition. Neuroscience Letters. 2021; 761: 136106. https://doi.org/10.1016/j.neulet.2021.136106.
[49] Gaillard C, Ben Hamed S. The neural bases of spatial attention and perceptual rhythms. The European Journal of Neuroscience. 2022; 55: 3209–3223. https://doi.org/10.1111/ejn.15044.
[50] Liu TY, Chen YS, Hsieh JC, Chen LF. Asymmetric engagement of amygdala and its gamma connectivity in early emotional face processing. PloS One. 2015; 10: e0115677. https://doi.org/10.1371/journal.pone.0115677.
[51] Farokhah L, Sarno R, Fatichah C. Simplified 2D CNN architecture with channel selection for emotion recognition using EEG spectrogram. IEEE Access. 2023; 11: 46330–46343.
[52] Lin X, Chen J, Ma W, Tang W, Wang Y. EEG emotion recognition using improved graph neural network with channel selection. Computer Methods and Programs in Biomedicine. 2023; 231: 107380. https://doi.org/10.1016/j.cmpb.2023.107380.
[53] Al-Asadi AW, Salehpour P, Aghdasi HS. A novel semi-supervised deep learning method for enhancing discriminability and diversity in EEG-based emotion recognition task. Physica Scripta. 2024; 99: 075030.
[54] Yilmaz BH, Kose C, Yilmaz CM. A novel multimodal EEG-image fusion approach for emotion recognition: Introducing a multimodal KMED dataset. Neural Computing and Applications. 2025; 37: 5187–5202. https://doi.org/10.1007/s00521-024-10925-5.
[55] Wan C, Xu C, Chen D, Wei D, Li X. Emotion recognition based on a limited number of multimodal physiological signals channels. Measurement. 2025; 242: 115940. https://doi.org/10.1016/j.measurement.2024.115940.