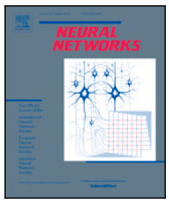
Neural Networks 184 (2025) 107127

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



A Fine-grained Hemispheric Asymmetry Network for accurate and interpretable EEG-based emotion classification ☆

Ruofan Yan [a,b], Na Lu [a,*], Yuxuan Yan [a], Xu Niu [a], Jibin Wu [b,c,**]

[a] Systems Engineering Institute, School of Automation Science and Engineering, Xi'an Jiaotong University, People's Republic of China
[b] Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[c] Department of Computing, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Emotion classification
EEG emotion interpretability
Brain signal analysis

ABSTRACT

In this work, we propose a Fine-grained Hemispheric Asymmetry Network (FG-HANet), an end-to-end deep learning model that leverages hemispheric asymmetry features within 2-Hz narrow frequency bands for accurate and interpretable emotion classification over raw EEG data. In particular, the FG-HANet extracts features not only from original inputs but also from their mirrored versions, and applies Finite Impulse Response (FIR) filters at a granularity as fine as 2-Hz to acquire fine-grained spectral information. Furthermore, to guarantee sufficient attention to hemispheric asymmetry features, we tailor a three-stage training pipeline for the FG-HANet to further boost its performance. We conduct extensive evaluations on two public datasets, SEED and SEED-IV, and experimental results well demonstrate the superior performance of the proposed FG-HANet, i.e. 97.11% and 85.70% accuracy, respectively, building a new state-of-the-art. Our results also reveal the hemispheric dominance under different emotional states and the hemisphere asymmetry within 2-Hz frequency bands in individuals. These not only align with previous findings in neuroscience but also provide new insights into underlying emotion generation mechanisms.

1. Introduction

Compared with non-neural signals like facial expressions, body language and voices (Atanassov, Pilev, Tomova, & Kuzmanova, 2021; Gupta & Mishra, 2023; Karnati, Seal, Bhattacharjee, Yazidi, & Krejcar, 2023), electroencephalograph (EEG) signals are directly generated by the human nervous system and thus can reflect human emotional states reliably (Ekman, 1992). EEG-based emotion classification has been extensively researched (Kamble & Sengupta, 2023; Vempati & Sharma, 2023), with wide applications in various human–computer interaction (HCI) systems, such as mental e-healthcare (Ali, Mosa, Al Machot, & Kyamakya, 2016), mental workload monitoring (Kothe & Makeig, 2011), etc. This task is generally tackled by exploiting the distinct patterns of asymmetric neural activity in the human brain. As revealed in previous neuroscience findings (Balconi & Mazza, 2009; Huang, Guan, Ang, Zhang, & Pan, 2012; Li & Lu, 2009), different areas of the brain (e.g., temporal and frontal lobes) exhibit asymmetrical energy changes in EEG signals across various frequency bands during emotional experiences, like alpha (8–13 Hz), beta (13–30 Hz), and gamma (> 40 Hz), etc., which can be used as fundamental cues in the development of emotion classification models.

Traditional methods for emotion classification typically rely on manually crafted features that capture hemispheric asymmetry properties, and analyze these features with statistical models (Hinrikus et al., 2009; Orgo, Bachmann, Lass, & Hinrikus, 2015). However, they often require substantial human effort due to the labor-intensive nature of feature engineering, and moreover, tend to underperform due to the limited representation power of the statistical models. In recent years, deep learning techniques (Chen, Yang, Huang, Li, Lu, & Wang, 2024; Chen, Yang, Huang, Wang, & Liu, 2024) have been applied for EEG-based emotion classification. Deep learning methods (Liu et al., 2020; Xiao et al., 2022) mostly utilize a manually crafted feature called Differential Entropy (DE) as input to deep neural networks, which captures crucial frequency domain information that is essential for accurately classifying different emotional states. Recent works (Ahmed, Sinha, Phadikar, & Ghaderpour, 2022; Huang et al., 2021; Zheng, Hu, Zhang, Li, & Zheng, 2021) extend the DE feature to incorporate prior

☆ This work is supported by National Natural Science Foundation of China under Grant No. 62476213, Natural Science Basic Research Program of Shaanxi Province under Grant 2024JC-YBMS-486.

* Corresponding author.
** Correspondence to: PQ825, The Hong Kong Polytechnic University Hung Hom, Kowloon, Hong Kong.
E-mail addresses: lvna2009@mail.xjtu.edu.cn (N. Lu), jibin.wu@polyu.edu.hk (J. Wu).

https://doi.org/10.1016/j.neunet.2025.107127

Available online 8 January 2025

knowledge of hemispheric asymmetry characteristics during emotional experiences, achieving improved performance. However, the development of these handcrafted features demands extensive expertise. Also, these features mostly mine from the coarse-grained frequency bands, like delta (0.1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–40 Hz), and gamma (> 40 Hz), and seldom consider the fine-grained frequency information.

Beyond the improvements in terms of input representation, some works (Ding, Robinson, Zhang, Zeng, & Guan, 2022; Li et al., 2020; Zhong, Gu, Luo, Zeng, & Liu, 2023) also focused on designing novel neural architectures to automatically extract hemispheric asymmetry features during network processing, which however remains a challenging task though with promising initial results. It is worth noting that handful end-to-end deep learning approaches (Cui et al., 2020; Yan, Lu, Niu, & Yan, 2022) have been proposed, demonstrating significant potential in extracting useful feature representations from raw EEG signals, yet their performance still lags behind the above-mentioned approaches. This performance gap can be attributed to the lack of effective preprocessing techniques to address the noise captured during EEG signal recording, and the limited availability of training data to enhance the model generalization. Despite the promising results, the deep learning models tend to lack interpretability, which may lead to concerns and distrust when applied in medical domains. Some efforts have been made to enhance the interpretability of these models (Ahmed et al., 2022; Ding et al., 2022; Zhong, Wang, & Miao, 2020), mainly focusing on coarse-grained brain regions or frequency bands.

In this work, we propose an interpretable end-to-end deep learning model that leverages fine-grained hemispheric asymmetry features within 2-Hz narrow frequency bands for emotion classification. The proposed model, named Fine-grained Hemispheric Asymmetry Network (FG-HANet), learns discriminative representations directly from raw EEG signals end-to-end, which does not need to be driven with expertise as those manually crafting asymmetric features. The FG-HAnet applies a Mirror Flip Block to generate mirrored versions of the original EEG samples, which are used as dual inputs to two spatiotemporal feature extractors to automatically extract deep features, and a Hemispheric Asymmetry Discriminator (HAD) to measure the discrepancy of the hemispheric features. Notably, in feature extraction, we apply Finite Impulse Response (FIR) filters with narrow bandpass of 2-Hz, enabling the FG-HANet to focus on fine-grained spectral features. We also devise a three-stage training pipeline to optimize the learning of fine-grained hemispheric asymmetry features. During the whole optimization process, spatial and frequency bands information is consistently maintained, benefiting fine-grained interpretability of the proposed model.

We conduct comprehensive experiments on two public datasets, namely SEED (Zheng & Lu, 2015) and SEED-IV (Zheng, Liu, Lu, Lu, & Cichocki, 2018), to evaluate the performance of the proposed FG-HANet. We first compare it with several strong baselines under the same settings, and it is demonstrated that our method yields 97.11% and 85.70% accuracy on SEED and SEED-IV, respectively, establishing new state-of-the-arts. We also conduct ablation studies on the three-stage training pipeline, validating the contribution of each stage. Moreover, we conduct interpretable analysis of our model through visualizations. In particular, we propose two new metrics to quantitatively measure the hemispheric dominance level under different emotional states. We also investigate hemispheric asymmetry in 2-Hz narrow frequency bands obtained by the HAD, compared with the previous wide range approach (including alpha, beta, gamma bands). The above interpretability analysis not only contributes to better understanding of the network decision process, but also provides important insights into the emotion generation mechanisms.

We summarize the contributions of this work as follows:

(1) We propose an end-to-end deep learning framework, named the FG-HANet, which is capable of extracting distinctive hemispheric asymmetry features for EEG-based emotion classification. It has been shown through extensive experiments that the FG-HANet beats strong baselines with new state-of-the-art performance.
(2) We propose to use FIR filters at a fine granularity of 2-Hz bandwidth to initialize the spectral filters within the FG-HANet. The resultant fine-grained spectral features greatly benefit the classification performance.
(3) We propose a three-stage training pipeline to enhance the learning of fine-grained hemispheric asymmetry features for effective emotion classification.
(4) We devise two interpretability metrics to quantitatively analyze the inter-band hemispheric activity and dominance level, providing new insights regarding emotion generation mechanisms.

2. Related work

2.1. Deep learning-based emotion classification

Deep learning has greatly advanced the progress in EEG-based emotion classification. Due to the research scope in this work, our review will concentrate on deep learning methods that explore and utilize hemispheric asymmetry cues for emotion classification, which demonstrate impressive performance. Generally, existing studies follow two directions. On one hand, some works design handcrafted features that capture hemispheric asymmetry. For example, Zheng et al. (2021) proposed hemispheric asymmetry features derived through subtraction and division of features from both hemispheres, which are then processed by a graph neural network (GNN) model for emotion classification; Huang et al. (2021) constructed EEG asymmetric feature matrices and utilized a bi-hemisphere discrepancy convolutional neural network model (BiDCNN) to classify emotions; Ahmed et al. (2022) introduced the asymmetric map (AsMap) feature, which are generated from DE features, as an effective representation. Though effective, the design of these handcrafted features requires substantial expertise and often faces issues in terms of generalization across different subjects.

On the other hand, some works focus on developing novel neural architectures capable of automatically learning hemispheric asymmetric features. For instance, Li et al. (2020) proposed a bi-hemispheric discrepancy (BiHDM) model, which incorporates a pairwise subnetwork to capture hemispheric discrepancy information based on DE feature inputs; Zhong et al. (2023) introduced a bi-headed attention mechanism that acquires discriminative attention weights for broad frequency bands in both hemispheres. In these methods, DE features are usually used as advanced hemispheric asymmetry representations, which are often calculated from wide frequency bands and then combined for classification, well preserving frequency bands information. Ding et al. (2022) presented TSception, a model featuring multi-scale convolutional layers to learn hemispheric asymmetry representations for emotion classification, which however still requires manual removal of artifacts during preprocessing. In addition, a few end-to-end networks have been developed, which do not rely on human expertise or intervention in preprocessing. For example, Cui et al. (2020) proposed an end-to-end regional-asymmetric convolutional neural network for binary emotion classification, in which an asymmetric differential layer is employed to capture distinct features.

While existing works have made notable progress in leveraging hemispheric asymmetry for EEG-based emotion classification, they primarily rely on pre-computed DE features from broad frequency bands, potentially missing critical fine-grained information. Additionally, many approaches depend on manual pre-processing or artifact removal, limiting their scalability. To address these limitations, we propose FG-HANet, an end-to-end framework that automatically captures fine-grained asymmetry patterns within 2-Hz frequency bands through a well -designed architecture. To further enhance its performance, a three-stage training strategy is introduced, effectively optimizing the learning of hemispheric asymmetry features essential for accurate emotion classification.

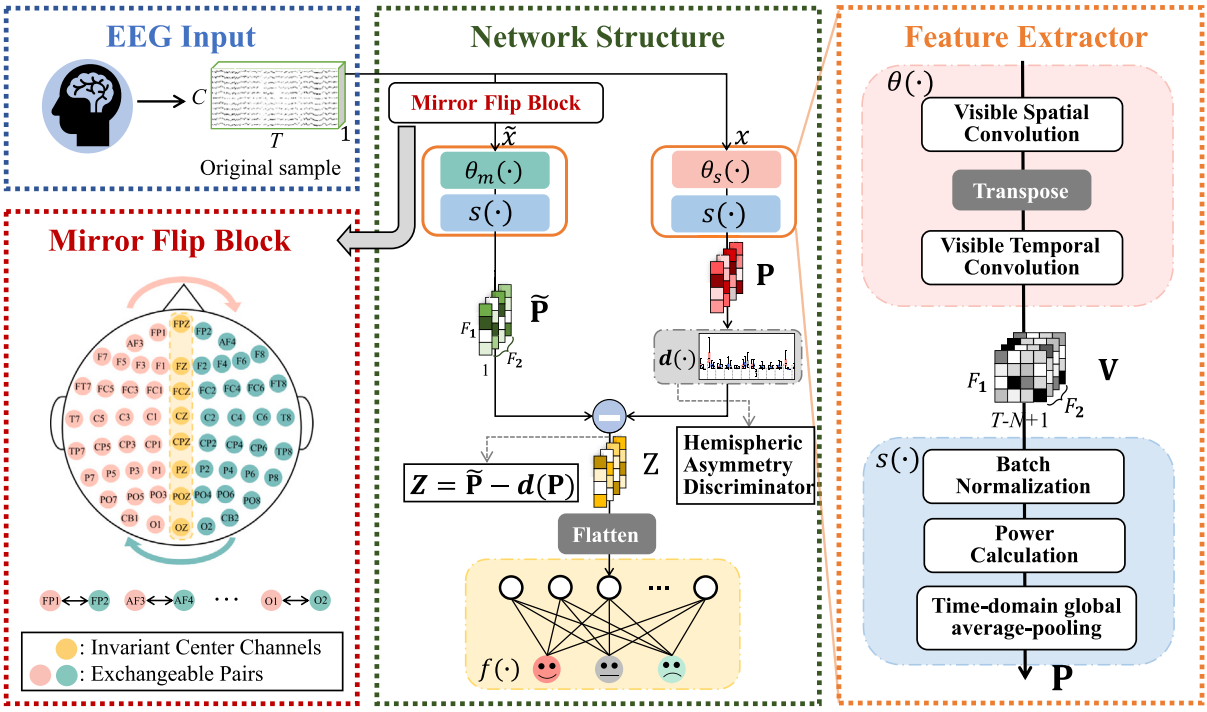


Fig. 1. An illustration of the overall framework of FG-HANet. The original EEG samples X are first fed into the Mirror Flip Block to generate channel mirrored samples $\tilde{X}$. Two parallel feature extractors are then used to extract features from both inputs independently. Subsequently, a HAD extracts discriminative hemispheric asymmetry features across narrow frequency bands. Finally, the HAD's output feature is flattened and classified by the classifier.

2.2. Interpretability of EEG-based deep emotion classification

Interpretability plays a vital role in advancing EEG-based emotion classification models, as it helps uncover the underlying neural mechanisms of emotion generation and enhances the plausibility of model predictions. Recent years have witnessed increasing efforts devoted to improving the interpretability of the EEG-based deep learning models. To name a few, Li et al. (2020) mapped deep feature representations onto electrode activity maps, uncovering pronounced hemispheric asymmetry regions within the frontal and temporal lobes during emotion expression; Ye, Chen, and Zhang (2022) performed interpretable analysis based on the dynamic spatial adjacency matrix learned by their proposed deep learning model, demonstrating distinct patterns in specific brain regions and frequency bands in emotion activities, such as strong connections in gamma and alpha bands distributed in occipital and parietal lobe sites; Ding et al. (2022) employed saliency maps to visualize informative regions identified by the neural network for emotion classification. These works provide qualitative analysis to reveal the activation patterns in specific brain regions or wide frequency bands during emotion generation. Currently, there has few studies providing fine-grained analysis over brain regions and narrow frequency bands simultaneously, which however has practical significance, e.g., applications in personalized treatment of emotional disorders using transcranial magnetic stimulation (TMS) (Dumitru et al., 2020; Luo, Che, & Li, 2023).

In this work, we delve into fine-grained interpretability and identify hemispheric asymmetry in specific brain regions and 2-Hz EEG frequency narrow-band. We also design metrics to quantitatively evaluate the inter-hemispheric activity in narrow-band frequencies, and the dominance of hemispheres across narrow frequency bands. These results and analyses not only revalidate previous findings in neuroscience, but also provide new insights into the underlying emotional mechanisms of humans, greatly enhancing the plausibility of the proposed model.

2.3. Pseudo-labeling for EEG-based emotion classification

In this work, we use the pseudo-label algorithm to align the learning of the two spatiotemporal extractors in the FG-HANet. Here, we review previous work on pseudo-labeling. At the core of pseudo-label algorithms is the use of the model itself to derive pseudo labels for unlabeled data, and the obtained pseudo labels can be integrated with nearly all neural network models and training methods to enhance decision boundary and model robustness (Chapelle & Zien, 2005). Pseudo-labeling has found wide applications in several EEG signal classification tasks. For example, Zhang et al. (2023) partitioned the sleep EEG training dataset artificially into labeled and unlabeled data, and used a model trained on labeled data to generate pseudo-labels, which are then utilized to optimize the model together with labeled data; Meng, Hu, Gao, Kong, and Luo (2022) proposed a Deep Subdomain Associate Adaptation Network (DSAAN) for EEG emotion classification, leveraging ground truth labels in the source domain and predicted pseudo labels in the target domain to improve model performance.

Building on this foundation, our approach integrates pseudo-labeling into the three-stage training pipeline to align the learned representations across network branches and enhance the extraction of hemispheric asymmetry features.

3. Method

Our proposed Fine-grained Hemispheric Asymmetry Network (FG-HANet) is an end-to-end deep learning model that leverages hemispheric asymmetry features within 2-Hz narrow frequency bands for accurate and interpretable emotion classification. As illustrated in Fig. 1, it first takes both the original EEG samples and their mirrored versions as inputs, and then uses two feature extractors to extract spatial and temporal features from both inputs separately. In particular, a 2-Hz bandpass FIR filter bank, covering frequency bands relevant to EEG emotion activities, is used to initialize specific layers in the dual feature extractors, embedding prior knowledge into the FG-HANet.

Table 1
List of notations used in the FG-HANet..

Notation	Definition
X	Raw EEG matrix
N	Number of EEG samples
E	Number of electrodes
T	Number of sampling points
Y	Emotion label set
c	Number of emotion classes
F_1	Number of Spatial filters
F_2	Number of FIR filters
L	Temporal convolution kernel size
P	EEG PSD feature
Z	Fine-grained hemispheric asymmetry features
$\theta(\cdot)$	Spatiotemporal encoder
$s(\cdot)$	Power calculation function
$d(\cdot)$	Hemispheric asymmetry discriminator
$f(\cdot)$	Classifier
$L(\cdot)$	Loss function for network optimization

A hemispheric asymmetry discriminator (HAD) is subsequently employed to extract fine-grained hemispheric asymmetry information. The output feature from the HAD is then flattened and classified by a single-layer fully-connected classifier. Notably, we devise a three-stage training pipeline to ensure that the proposed FG-HANet is trained to extract hemispheric asymmetry features with sufficient discrimination. We elaborate on key model components and training pipeline of our FG-HANet in below subsections.

For notations, throughout the paper, we follow the definitions summarized in Table 1. Specifically, we use $X = \{\boldsymbol{x}_i\}_{i=1}^{n}$, $\boldsymbol{x}_i \in \mathbb{R}^{E\times T}$ to represent the raw EEG data, where E denotes the number of electrodes and T denotes the number of sampling points. We use $Y = \{y_i\}_{i=1}^{n}$, $y_i \in \{0, 1, \ldots, c\}$ to represent the labels corresponding to the raw EEG data, where c represents the number of emotion classes.

3.1. Mirror flip block

Our FG-HANet uses a Mirror Flip Block to generate the mirrored versions of the input EEG signals to facilitate the extraction of hemispheric asymmetry information. As shown in Fig. 1, the electrodes of the input signal are configured using an ESI NeuroScan System according to the international 10–20 system with 62 channels in total. We flip the original EEG samples symmetrically along the middle column to obtain the mirrored samples. Specifically, the eight yellow electrodes in the middle column are kept unchanged, and the remaining 27 pairs of electrode channels are flipped. Through this Mirror Flip Block, for the raw EEG data X, we obtain the mirrored sample $\tilde{X} = \{\tilde{\boldsymbol{x}}_i\}_{i=1}^{N}$, where $\tilde{\boldsymbol{x}}_i \in \mathbb{R}^{E\times T}$. Both X and $\tilde{X}$ are transposed to the size of $(1, E, T)$ before fed to the following Feature Extractor Block.

3.2. Feature extractor

The FG-HANet has two parallel feature extractors to take in the original and mirrored samples separately as inputs. Each feature extractor comprises an independent spatiotemporal encoder, denoted as $\theta_s(\bullet)$ or $\theta_m(\bullet)$ for the right and left branches in Fig. 1 respectively, along with a power computation (PC) module. The spatiotemporal encoder incorporates a visible spatial convolution (VSConv) layer and a visible temporal convolution (VTConv) layer to extract features from the spatial and temporal domains, respectively. The VSConv layer performs spatial filtering on X and $\tilde{X}$ through spatial convolution, with goals of spatial source separation and spatial noise suppression. The VSConv layer consists of F_1 spatial convolution kernels, facilitating the derivation of F_1 visualizable spatial filtered components. The output size of the VSConv layer is $(F_1, 1, T)$, which is transposed to the size $(1, F_1, T)$ before passing to the VTConv layer.

On the other hand, the VTConv layer performs temporal convolution to extract useful spectral features. In order to incorporate the prior knowledge of emotion-related frequency information and capture fine-grained features within narrow frequency bands, we utilize F_2 FIR band-pass filters, with length denoted as L, which meanwhile equals the kernels' size on the time axis. These filters are designed based on the window function method to initialize the convolution kernels of the VTConv layer, which has been demonstrated effective for motor imagery classification (Niu, Lu, Kang, & Cui, 2022).

Specifically, a band-pass FIR filter $h[n]$ truncated by a rectangular window function can be formulated as

$$h[n] = \begin{cases} \frac{\sin\left[\omega_h\left(n-\frac{L-1}{2}\right)\right]-\sin\left[\omega_l\left(n-\frac{L-1}{2}\right)\right]}{\pi\left(n-\frac{L-1}{2}\right)}, & \text{if} \quad n \neq \frac{L-1}{2} \\ \frac{\omega_h-\omega_l}{\pi}, & \text{if} \quad n = \frac{L-1}{2} \end{cases} \tag{1}$$

where $0 \le n \le L-1$, and ω_l and ω_h denote the lower and upper cutoff frequencies of the band-pass filter, respectively. To cover the entire frequency range associated with emotion expression (Alarcao & Fonseca, 2017), the pass band of the FIR filters is set from 0.1-Hz to 76-Hz, spanning across the alpha, beta, theta, delta, and gamma bands. The 0.1-Hz is set as the lower bound to filter out the direct current signal, and the 76-Hz is used as the upper bound to filter the noise and remove the unrelated artifacts according to Zheng et al. (2018), Zheng and Lu (2015). By designing narrower pass bands for the filters, we can extract more fine-grained spectral features. To achieve a good balance between spectral granularity and computational cost, the width of the band-pass filter bank is set to 2-Hz. Thus the number of kernels F_2 is equal to $76/2 = 38$. The output of the VTConv is denoted as $V = \{\boldsymbol{v}_{i,j}\}_{i=1,\ j=1}^{F_2,\ F_1}$, where $\boldsymbol{v}_{i,j} \in \mathbb{R}^{(T-N+1)}$ represents the ith sub-band component of the jth spatial-filtered signal.

3.3. Power calculation module

The extracted features V are then input to a Power Calculation (PC) Module to calculate the average power spectral density (PSD) feature. The PC module consists of a batch normalization layer (BN), a square calculation (SQ) layer, and a time-domain global average pooling layer. The BN layer is used to normalize the input, which can improve the model generalization. The SQ layer calculates the square of the input itself, and then the time-domain global average pooling layer performs the average operation by setting the kernel size and stride to $(1, T)$ and T, respectively.

Specifically, suppose $V_f[k]$ represents the Fourier transform of the discrete-time signal $\boldsymbol{v}_{i,j}[t]$. According to the Parseval theory for Discrete Fourier Transform (DFT) (Brigham, 1988), the energy of digital EEG signals is proportional to the sum of the square of all amplitude values. Thus the average PSD of $\boldsymbol{v}_{i,j}$ in a given frequency band can be formulated as

$$PSD = \frac{1}{T}\sum_{k=0}^{T-1}\left|V_f[k]\right|^2 = \sum_{t=0}^{T-1}\left|\boldsymbol{v}_{i,j}[t]\right|^2. \tag{2}$$

The output of $s(\bullet)$ is proportional to the energy of the sub-band component $\boldsymbol{v}_{i,j}$, which can be formulated as $s(\boldsymbol{v}_{i,j}) \propto \sum_{t=0}^{T-1}|\boldsymbol{v}_{i,j}[t]|^2$. Then, the PC Module can effectively implement power calculation similarly as in Eq. (2). The output of the PC module in the right network branch, as shown in Fig. 1, denoted as $\boldsymbol{p}_i = s(\boldsymbol{v}_{i,j})$, is the average PSD feature extracted by $s(\bullet)$. Similarly, the output for the left branch can be formulated as $\widetilde{\boldsymbol{p}}_i = s(\widetilde{\boldsymbol{v}}_{i,j})$.

3.4. Hemispheric asymmetry discriminator

The HAD is designed to enhance the network's capability in extracting hemispheric asymmetry features. Specifically, the PSD feature $P = \{\boldsymbol{p}_i\}_{i=1}^{N}$ is fed to the HAD $d(\bullet)$. HAD performs a group convolution operation (Cohen & Welling, 2016), where the number of groups is set to F_2, i.e., the number of FIR filters. Thus $\boldsymbol{p}_i$ is divided into F_2 groups, and each group learns weights specific to narrow frequency bands at

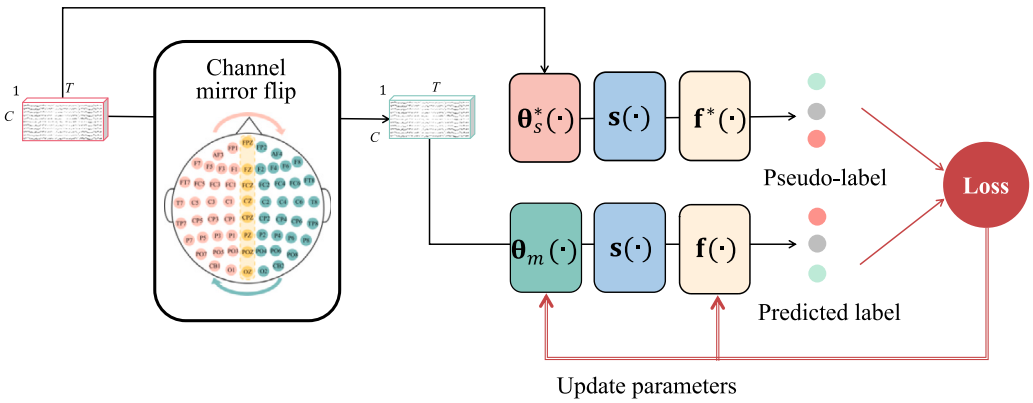


Fig. 2. An illustration of Stage 2 in the proposed training pipeline. Stage 2 is to align features of left and right feature extractors. Original samples are fed into the original branch (right branch in Fig. 1) to obtain pseudo-labels which then guide the training of the mirrored branch (left branch in Fig. 1).

a particular spatial location. Unlike traditional convolutional layers where all input channels are convolved with all output channels, the input and output channels are divided into groups in group convolution, and convolution is performed independently within each group. By making each group's convolution kernels process specific frequency bands independently, the network is able to capture more distinctive features from different frequency bands. In addition, group convolution effectively preserves the spatial and frequency band structure of the feature $\boldsymbol{p}_i$. This property facilitates interpretability by shedding light on how the model learns features across spatial and frequency domains in emotion classification. The final feature $\boldsymbol{z}_i \in \mathbb{R}^{F_2 \times F_1}$, which combines the features $\boldsymbol{p}_i$ and $\widetilde{\boldsymbol{p}}_i$, is formulated as

$$\boldsymbol{z}_i = \widetilde{\boldsymbol{p}}_i - d(\boldsymbol{p}_i) = \widetilde{\boldsymbol{p}}_i - \sum_{g=1}^{F_2} \boldsymbol{\omega}^g * \boldsymbol{p}_i^{g,\bullet}, \tag{3}$$

where the weight vector of $d(\bullet)$ is denoted as $\{\boldsymbol{\omega}^g\}_{g=1}^{F_2}$, with $\boldsymbol{\omega}^g \in \mathbb{R}^{F_1}$, $\boldsymbol{p}_i^{g,\bullet} \in \mathbb{R}^{F_1}$ is the gth row vector of $\boldsymbol{p}_i$, and $*$ denotes the convolution operation.

We then proceed to explain how $d(\bullet)$ works to extract hemispheric asymmetry features. Assume there exists a pair of PSD features $p_i^{a,h}$ and $\widetilde{p}_i^{a,h}$ in the asymmetric narrow band and a pair of PSD features $p_i^{u,h}$ and $\widetilde{p}_i^{u,h}$ in another less important narrow band. Due to hemispheric asymmetry in emotional activities, unilateral brain energy changes more significantly in the ath narrow band, and the following formula holds:

$$\left|p_i^{a,h} - \widetilde{p}_i^{a,h}\right| \gg \left|p_i^{u,h} - \widetilde{p}_i^{u,h}\right|. \tag{4}$$

As a result, the ath dimension of the feature $\boldsymbol{z}_i$ (i.e., $z_i^{a,h} = \widetilde{p}_i^{a,h} - \omega_h^a p_i^{a,h}$) is more discriminative for emotion classification than the uth dimension of the feature (i.e., $z_i^{u,h} = \widetilde{p}_i^{u,h} - \omega_h^u p_i^{u,h}$). ω_h^a denotes the hth element of $\boldsymbol{\omega}^a$. As the PSD feature is non-negative, the network will tend to assign a negative value for ω_h^a, since the negative value of the asymmetric coefficient will pull the PSD features of two branches further apart during the feature fusion process, expressed as

$$\widetilde{p}_i^{a,h} - \omega_h^a p_i^{a,h} > \left|p_i^{a,h} - \widetilde{p}_i^{a,h}\right| \gg \left|p_i^{u,h} - \widetilde{p}_i^{u,h}\right|, \text{ if } \omega_h^a < 0. \tag{5}$$

3.5. Training pipeline

The proposed FG-HANet seeks to extract discriminative hemispheric asymmetry features to enhance emotion classification performance. However, directly training the model end-to-end may not guarantee sufficient attention to hemispheric asymmetry features. To address this issue, we propose a three-stage training pipeline as summarized in Algorithm 1.

For Stage 1, we train the right feature extractor (right branch in Fig. 1) along with the classifier using the original samples $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^n$.

Algorithm 1: The three-stage training pipeline used in FG-HANet.

Input: Raw dataset $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^N$, training epochs for three stages I_1, I_2, and I_3
Output: θ_s^*, θ_m^*, d^*, f^*
Stage 1:
for *epoch* $= 1$ **to** I_1 **do**
  Random sample a batch from $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^N$;
  for *each* $\boldsymbol{x}_i$ *in batch* **do**
    $\theta_s^*, f^* \leftarrow$ Eq. (6);
  end
end
Stage 2:
for *epoch* $= 1$ **to** I_2 **do**
  Random sample a batch from $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^N$;
  Generate mirror dataset $\{\widetilde{\boldsymbol{x}}_i\}_{i=1}^N$;
  for *each* $\boldsymbol{x}_i$ *in batch* **do**
    $\theta_m^*, f_m^* \leftarrow$ Eq. (7);
  end
end
Stage 3:
for *epoch* $= 1$ **to** I_3 **do**
  Random sample a batch from $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^N$;
  Generate mirror dataset $\{\widetilde{\boldsymbol{x}}_i\}_{i=1}^N$;
  for *each* $\boldsymbol{x}_i$ *in batch* **do**
    $\theta_s^*, \theta_m^*, d^*, f^* \leftarrow$ Eq. (8);
  end
end

This provides a good initialization for the feature extractor. For Stage 2, we train the left feature extractor (left branch in Fig. 1) using the mirrored samples $\tilde{X} = \{\tilde{\boldsymbol{x}}_i\}_{i=1}^n$. The main objective in this stage is to align the feature representation of the mirror samples with that of the original samples, ensuring the network to focus solely on extracting hemispheric symmetry features. For Stage 3, we incorporate the hemispheric asymmetry discriminator into the network and fine-tune the entire network end-to-end, including the model parameters obtained in the previous two stages. This stage ensures that the model is trained to extract hemispheric asymmetry features, thereby enhancing its performance in emotion classification.

More specifically, in Stage 1, the original samples with ground truth labels $\{\boldsymbol{x}_i,\ y_i\}_{i=1}^n$ are fed as input to the right feature extractor and the classifier. The optimization at this stage can be formulated as

$$\theta_s^*, f_s^* = \arg\min_{\theta_s, f_s} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\Big(f_s\big(s\big(\theta_s(\boldsymbol{x}_i)\big)\big), y_i\Big), \tag{6}$$

where $\mathcal{L}(\bullet)$ is the cross-entropy (CE) loss, N denotes the total training samples, and θ_s^*, f_s^* are optimal parameters of the network.

In Stage 2, the objective is to align the representation of the left feature extractor with that of the right feature extractor. To achieve this, the original samples with their corresponding ground truth labels, denoted as $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ are input into the right feature extractor, generating pseudo labels that are necessary for training the left feature extractor. It is worth noting that, at this stage, the network parameters of the right feature extractor (θ_s^* and f_s^*) are kept frozen to serve as the anchor feature space. Additionally, the mirror samples are fed into the left feature extractor, and the predicted labels, together with the pseudo labels, are utilized jointly to drive the model training. The complete training pipeline is depicted in Fig. 2, and the optimization objective during this stage is formulated as

$$\theta_m^*, f_m^* = \arg\min_{\theta_m, f_m} \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}\left(f_m(s(\theta_m(\widetilde{\boldsymbol{x}}_i))), f_s^*(s(\theta_s^*(\boldsymbol{x}_i)))\right). \tag{7}$$

In Stage 3, we introduce the hemispheric asymmetry discriminator into the network and perform end-to-end fine-tuning of the entire model. The optimal parameters obtained from the earlier stages, denoted as θ_s^* and θ_m^*, are used to initialize $\theta_s(\bullet)$ and $\theta_m(\bullet)$, respectively. However, to ensure the consistency of spectral filters that are independent of the location, the two VTConv layers within $\theta_s(\bullet)$ and $\theta_m(\bullet)$ are frozen and kept unchanged. This freezing mechanism prevents any modifications to the spectral filters during the fine-tuning process. The optimization during this stage can be expressed as follows:

$$\theta_s^*, \theta_m^*, d^*, f^* = \arg\min_{\theta_s, \theta_m, d, f}\left\{\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}\left(f\left(s(\theta_m(\widetilde{\boldsymbol{x}}_i))\right) - d\left(s(\theta_s(\boldsymbol{x}_i))\right), y_i\right)\right\}. \tag{8}$$

4. Experimental results

In this section, we report experimental results of the proposed FG-HANet. We first compare our method with strong baselines to validate its effectiveness, and then we conduct ablation study to validate the effectiveness of each stage within our training pipeline. Meanwhile, we also compare the computational cost of our model against that of the baselines.

4.1. Datasets

We conduct experiments on two public datasets, i.e., SEED (Petrantonakis & Hadjileontiadis, 2011) and SEED-IV (Zheng et al., 2018), which are released by the BCMI Lab of Shanghai Jiaotong University. The two datasets are collected by requiring participants to comfortably sit in front of a monitor and watch emotional video clips, during which EEG signals are recorded from 62 electrode channels using ESI NeuroScan at a sampling rate of 1000-Hz, with electrode placement following the international 10–20 system. The collected EEG signals are then downsampled to 200-Hz, and cropped by second to generate the samples. The **SEED** dataset contains 15 subjects, each with three sessions. The film clips for participants to watch contain three types of emotions, i.e., happy, neutral, and sad, and each emotion is displayed with 5 film clips. There are 15 trails, and each trail contains 185–238 samples for each session per subject. Thus there are totally about 3400 samples in one session. The **SEED-IV** dataset also includes 15 subjects with three sessions per subject. But it includes an extra emotion (i.e., fear) compared with SEED, and it offers 6 film clips per emotion. There are 24 trails, each with 12–64 samples for each session per subject. Thus there are totally about 830 samples in one session.

4.2. Experiment settings

For SEED, we set the spatial convolution kernel size of our FG-HANet (F_1) to 41, number of FIR filters (F_2) to 38, kernel size of the temporal convolution layer (L) to 41. For SEED-IV, F_1 is set to 81, F_2 to 38, and L to 41. Adam optimizer is employed to minimize the loss function. The first-moment decay term and the second-moment decay term within the Adam optimizer are set to 0.9 and 0.999, respectively. The learning rate is set as 1×10^{-4} for the Stage 1 and Stage 2 and changes to 0.5×10^{-4} in Stage 3. Training is conducted in three stages, with each stage iterating for 700 epochs. The batch sizes for each stage are 32, 32, and 100, respectively. We use Pytorch library for implementation. All experiments are performed on a computer server with one NVIDIA TITAN V GPU card.

4.3. Superior emotion classification capability

We first compare the performance of our FG-HANet with previous methods to verify its effectiveness. We use five-fold cross-validation as in Kuang and Michoski (2023), Liu et al. (2024) on each subject to ensure the rigor of our evaluation. The model is trained five times, with four folds used for training and the remaining fold serving as the test set to evaluate model performance. For fair comparisons, we reproduce the results of the baselines under the same experimental settings. Specifically, we first compare our FG-HANet with six deep learning methods specially designed for EEG emotion classification, including deep belief network (DBN) (Zheng & Lu, 2015), bi-hemisphere domain adversarial neural network (BiDANN) (Li, Zheng et al., 2018), regularized graph neural network (RGNN) (Zhong et al., 2020), pyramidal graph convolutional network (PGCN) (Jin et al., 2023), dynamical graph convolutional neural network (DGCNN) (Song et al., 2018) and bi-hemispheric discrepancy model (BiHDM) (Li et al., 2020). We also compare with two generic end-to-end deep convolutional networks, i.e., compact convolutional network (EEGNET) (Lawhern et al., 2018) and deep convolutional network (DEEPNET) (Schirrmeister et al., 2017). All the comparison results are summarized in Table 2, where we report the mean accuracy (ACC) and standard deviation (STD) across five folds. The types of features used by each method (DE feature or raw EEG) are also reported. Additionally, the preprocessing techniques employed by each method are detailed. "Artificial Removal + Feature Smoothing" indicates that the method involves manual artifact removal from the EEG signals and applies a feature smoothing technique, such as Linear Dynamical Systems (LDS), while "None" signifies that the algorithm operates directly on the raw EEG without extra pre-processing beyond cropping the EEG samples. We conduct a paired t-test between our FG-HANet and each of the baselines at a significance level of 0.05. Furthermore, to facilitate a more effective comparison of model complexity, we report the number of model parameters and Floating Point Operations (FLOPS) of the key methods in Table 3.

Our FG-HANet achieves an accuracy of 97.11% on the SEED dataset and 85.70% on the SEED-IV dataset, showing great effectiveness. Such performance statistically surpasses that of all the baseline methods ($p < 0.05$), except BiHDM on the SEED dataset ($p = 0.17$). In particular, compared with the end-to-end baselines, i.e., EEGNet and DEEPNet, our FG-HANet obtains higher ACC with significant margins up to 8.79%, 6.82% on the SEED dataset and 14.02%, 9.96% on the SEED-IV dataset, respectively. Although these two baselines and our FG-HANet are all end-to-end networks, they utilize spatial and temporal convolution layers to extract EEG features for general EEG tasks, while FG-HANet stands out by incorporating prior frequency knowledge of EEG emotion in VTConv and capturing the hemispheric asymmetry features through HAD. These results highlight the superiority of FG-HANet among end-to-end methods for EEG emotion classification, which can be attributed to the usage of emotional hemispheric asymmetry. Compared with BiHDM, which also consider the hemispheric asymmetry for EEG emotion classification, our FG-HANet increases the accuracy by 1.13% on

Table 2
Accuracy and standard deviation (ACC/STD(%)) of different methods on SEED and SEED-IV datasets.

Method	Feature	Pre-processing	SEED	SEED-IV
DBN (Zheng & Lu, 2015)	DE	Artificial Removal + Feature Smoothing	73.55/10.64	57.82/10.56
BiDANN (Li, Zheng, Zong, Cui, Zhang & Zhou, 2018)	DE	Artificial Removal + Feature Smoothing	91.15/7.50	78.88/8.86
RGNN (Zhong et al., 2020)	DE	Artificial Removal + Feature Smoothing	91.87/4.83	61.96/12.34
PGCN (Jin, Zhu, Du, He, & Li, 2023)	DE	Artificial Removal + Feature Smoothing	87.06/7.29	75.29/9.02
DGCNN (Song, Zheng, Song, & Cui, 2018)	DE	Artificial Removal + Feature Smoothing	90.12/5.19	78.58/9.01
BiHDM (Li et al., 2020)	DE	Artificial Removal + Feature Smoothing	95.98/2.72	80.38/6.67
EEGNET (Lawhern et al., 2018)	Raw	None	88.32/2.41	71.68/11.83
DEEPNET (Schirrmeister et al., 2017)	Raw	None	90.29/7.30	75.74/6.01
FG-HANet	**Raw**	**None**	**97.11**/1.49	**85.70**/7.71

Table 3
Model Parameters and FLOPS on SEED and SEED-IV datasets.

Method	SEED		SEED-IV	
	Params (k)	FLOPS (M)	Params (k)	FLOPS (M)
RGNN	2.5	0.1	2.6	0.1
BiHDM	13.3	4.4	13.4	4.4
EEGNET	23.3	495.0	23.4	495.0
DEEPNET	235.6	3039.0	235.7	3039.0
BiDANN	3348.5	1151.6	3348.7	1151.6
FG-HANet	**8.7**	**2730.1**	**12.3**	**3092.7**

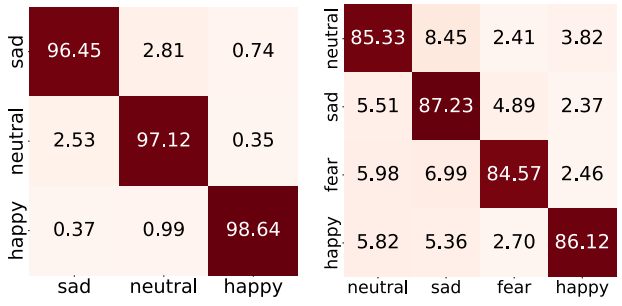


Fig. 3. Confusion matrices of our FG-HANet on (Left) SEED dataset and (Right) SEED-IV dataset.

SEED dataset, and 5.32% on SEED-IV dataset. The main difference between BiHDM and our FG-HANet lies in that BiHDM utilizes coarse-grained hemispheric asymmetry features handcrafted from DE features as network input, while FG-HANet embed FIR filter banks with 2-Hz band-pass to extract fine-grained ones to improve the classification performance. Besides, we plot the confusion matrices of our FG-HANet in Fig. 3, which demonstrate stable performance across different emotion classes.

4.4. Ablation study on three-stage training pipeline

We further conduct an ablation study to evaluate the validity of each training stage in our proposed three-stage training pipeline. For fairness, all the experiments are performed under the same experimental settings. Since Stage 2 aims to align the representations of the dual extractors in FG-HANet, which serves as the foundation for Stage 3, we did not include the ablation experiment for the combination of Stages 1 and 2. We present the results in Tables 4 and 5 for all subjects in SEED and SEED-IV datasets, respectively.

It can be seen that on both datasets, model training with Stage 1+2+3 achieves the best average accuracy, followed by Stage 1+3 and Stage 1. Specifically, the average accuracy of Stage 1+3 is 2.57% and 4.39% higher than Stage 1 on SEED and SEED-IV datasets, respectively. Compared to training with only Stage 1, the added Stage 3 fuses the two branches' representations to capture discriminative asymmetry features from both hemispheres. The above results well validate the effectiveness of Stage 3, which adopts the dual branches and HAD to capture hemispheric asymmetry features. Compared with Stage 1+3, the average accuracy of Stage 1+2+3 is improved by 1.31% on SEED dataset and 1.12% on SEED-IV dataset. This highlights the contribution of Stage 2, which aligns the features of the dual extractors using guidance from the generated pseudo labels. In summary, these ablation results confirm that each training stage contributes to the overall effectiveness of our method.

5. Visualization analysis

In this section, we perform visualization analysis on deep features extracted from each block of the proposed FG-HANet. Specifically, we analyze the spectrum features obtained from VTConv, the topographic map obtained from VSConv, hemispheric energy obtained from PC modules, and hemispheric asymmetry feature from HAD. Two new metrics, named inter-band dispersion and inter-band hemispheric dominance, are defined to quantitatively analyze the experimental results. We also discuss how the new insights gained from these analyses can be used to further enhance model design.

5.1. EEG-based emotional spectrum

To assess our model's performance across different filter banks for initializing the VTConv layer, we conduct experiments on SEED and SEED-IV with filter banks spanning the commonly used broad frequency bands of delta, theta, alpha, beta, gamma, and their combination, as summarized in Table 6. Each filter bandwidth is set to 2-Hz. For both datasets, the performance with filter banks in high-frequency bands (beta and gamma bands) outperform that in low-frequency bands, which is consistent with previous findings (Li, Zhang, & He, 2018; Li, Zheng et al., 2018; Seal et al., 2020; Zheng & Lu, 2015). This indicates that high-frequency components in EEG signals are more informative for emotion classification. Notably, filter banks covering the gamma frequency band demonstrate the highest accuracy on both datasets, indicating that gamma band signals are the most discriminative for our model.

To further inspect features in the critical gamma band (40–76 Hz) learned by the VTConv layer, we visualize the convolution kernels of the FG-HANet trained on Subject 4 in SEED and SEED-IV datasets respectively, as shown in Fig. 4. Compared to the original filters' amplitude–frequency response curves in blue lines, these red curves with changes, obtained after network updates, well demonstrate the VTConv layer's ability to adaptively learn spectral features.

Specifically, some of the amplitude–frequency response curves of learned filters shift along the frequency axis, indicating that the filters have captured emotion-related information present within narrow frequency bands. For example, in Fig. 4(b), curves of the 1st, 2nd, 9th, 12th, and 16th filters have obvious shifting along the frequency axis. In addition, for the amplitude–frequency response curves of the 2nd, 4th, 7th, 9th, and 11th filters in Fig. 4(b), the energy is more concentrated in the main lobe, and the ratio of the main lobe's energy to the side lobes' energy is increased. This indicates that spectral leakage has been effectively reduced, and the amplitude–frequency response of filters is improved, thereby automatically enhancing the performance of the filter. The 1st, 6th and 10th, 17th filters in Fig. 4(a), and the 10th filter in Fig. 4(b) almost disappear, indicating that the corresponding frequency bands may contain little information related to emotion classification.

Table 4
Classification accuracy for different training stage combinations on SEED dataset.

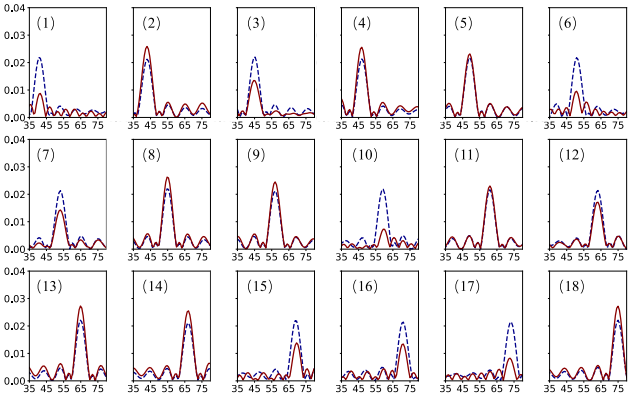
Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	Average
Stage 1	89.16	90.26	94.44	90.84	92.16	95.40	92.90	94.57	96.42	95.30	97.68	94.55	97.19	87.71	96.07	93.23
Stage 1+3	95.62	94.85	95.71	94.42	93.79	96.92	94.94	96.02	97.61	95.30	97.68	94.62	97.19	98.66	99.69	95.80
Stage 1+2+3	97.64	95.72	96.31	96.77	94.62	98.96	97.49	98.23	98.64	98.53	97.05	96.77	98.37	93.84	97.78	**97.11**

Table 5
Classification accuracy for different training stage combinations on SEED-IV dataset.

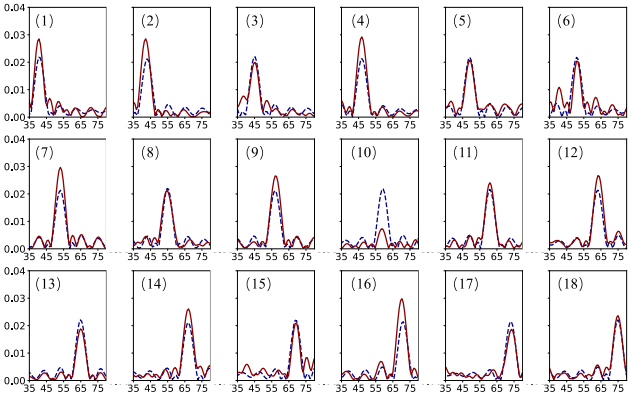
Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	Average
Stage 1	69.14	80.86	73.16	81.12	61.81	84.86	83.14	86.90	83.99	79.77	86.61	75.11	88.36	86.75	81.23	80.19
Stage 1+3	71.80	85.64	79.89	85.17	65.32	86.88	84.72	86.36	87.71	89.89	91.41	86.89	89.12	92.78	85.14	84.58
Stage 1+2+3	72.54	89.68	82.89	87.39	64.22	88.82	88.25	88.78	82.97	91.60	93.84	88.93	89.79	87.93	87.80	**85.70**

Table 6
Classification Accuracy and Standard Deviation (ACC/STD (%)) of FG-HANet with filter banks in different frequency bands on SEED and SEED-IV datasets.

Method	Delta band	Theta band	Alpha band	Beta band	Gamma band	All bands
SEED	81.37/6.02	82.52/6.05	84.17/5.30	93.63/3.76	**95.80**/3.15	94.17/3.56
SEED-IV	71.68/11.45	73.88/7.61	72.32/11.11	81.06/9.87	**84.58**/9.56	83.56/7.09



(a) SEED



(b) SEED-IV

Fig. 4. An illustration of amplitude–frequency maps of 18 filter banks (ranging from 40–76 Hz) with 2-Hz bandwidth after Fourier transform. The x-axis represents frequency (Hz), and the y-axis represents amplitude–frequency response. Blue dotted lines represent the initial filter states; red solid lines depict filters learned by our FG-HANet.

5.2. EEG-based emotional topographic maps

To illustrate the correlation between different brain regions and emotion classification, we present the typical brain topographic maps of the learned kernels in the VSConv layers over SEED dataset, as shown in Fig. 5. The dark areas of the first three components concentrate on the prefrontal, temporal, frontal lobes and central areas, and those of the last component concentrate on the occipital and temporal regions, indicating these are important brain regions related to emotion classification.

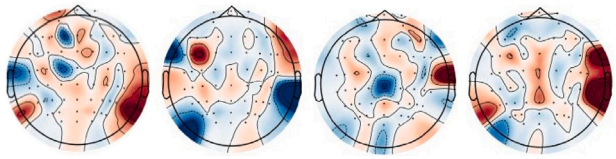


Fig. 5. An illustration of brain topographic maps. Black dots represent positions of 60 electrodes. Red and blue colors represent positive and negative weights, respectively. The darker the color is, the greater the absolute value will be. Two edge electrodes (CB1, CB2) are removed due to insignificance.

Since EEG artifacts have not been manually removed in advance, such as eye electricity and muscle spots, the dark areas in the last component that concentrate at the edge of the electrode are possibly the spatial noise distribution learned by the network. The change of energy in unilateral regions occurs in the temporal and frontal lobes, which is consistent with conclusions in neuroscience (Li & Lu, 2009; Palmiero & Piccardi, 2017). It is indicated that the VSConv layer have learned noise distribution and hemispherical asymmetric information simultaneously. In addition, the third spatial filtered component indicates that the central area of the brain also contains effective features for emotion classification, which has been rarely reported in previous studies.

5.3. EEG emotional hemispheric energy

To further investigate the differences in energy distribution across important brain regions in the left and right hemispheres during typical emotional states, we depict the energy box diagrams of representative spatial filtered components learned by the FG-HANet (i.e., frontal and temporal regions of Subject 3) under positive and negative emotions in Fig. 6 respectively. To analyze fine-grained narrow frequency bands, we utilize 38 filter banks spanning frequencies from 0-Hz to 76-Hz, with each filter spanning a width of 2-Hz. For clear visualization on the energy distribution of both hemispheres, unlike previous network using inputs of all electrode channels, we input EEG signals from the left and right hemispheres separately to the two branch networks of our model to discern their distinct activity patterns across different emotional states. The PSD features obtained from PC module of dual

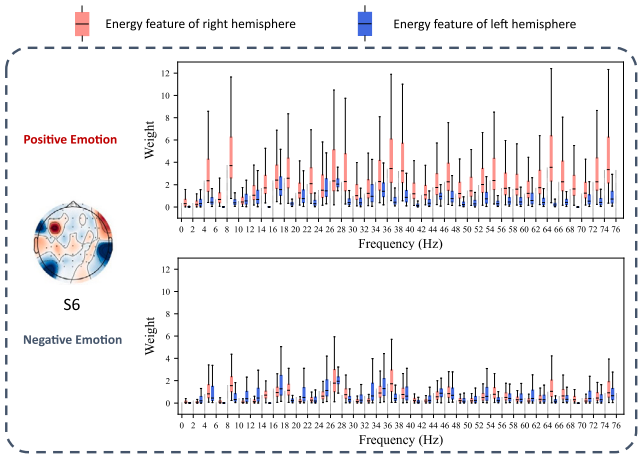


Fig. 6. Energy box diagram of the mainly frontal lobe of S6 under negative and positive emotions. 38 filters are applied, ranging from 0.1–76 Hz. Blue bars represent the relative energy distribution of the left hemisphere and red bars represent that of the right.

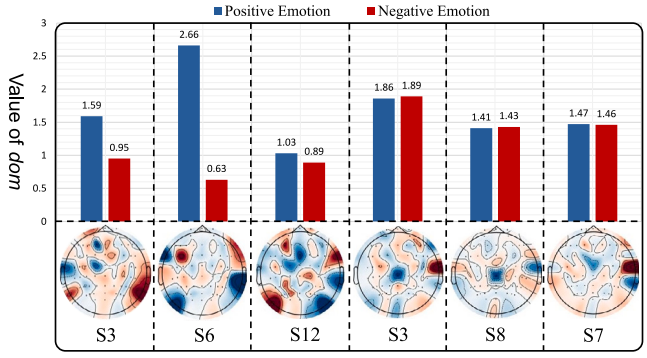


Fig. 7. Values of dom of typical components on different subjects under positive and negative emotions.

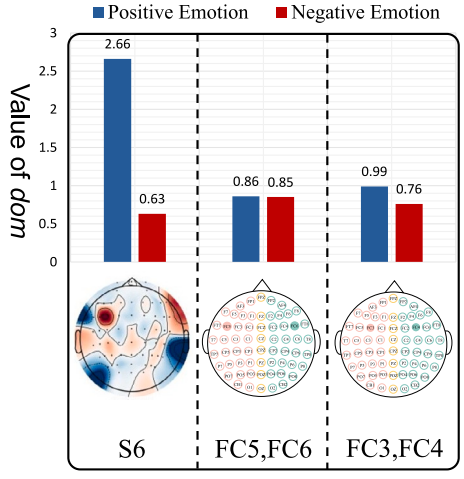


Fig. 8. Comparison of dom values for spatial filtered components of S6 under positive and negative emotions with paired electrodes (FC5, FC6), (FC3, FC4) as input without VSConv layers.

branches serve as proxies for the energy levels of the left and right brain hemispheres.

From Fig. 6, a preliminary observation can be made that under positive emotions, the energy distribution in the left hemisphere appears more dispersed compared to that under negative emotions. This dispersity indicates a higher level of mutual information exchange among different frequency bands, suggesting increased brain activity. Moreover, it is noteworthy that the PSD features extracted by our FG-HANet from the left and right hemispheres are influenced by spatial filters in different network branches, resulting in inconsistent energy scales between the two hemispheres. To align and quantitatively measure the level of energy dispersion across different frequency bands between hemispheres, we introduce a new fine-grained metric termed "inter-band dispersity", denoted as τ. The PSD features of the left hemisphere of the samples under specific emotions in the ith frequency band is denoted as $P_{l,i} = \{p^n_{l,i}\}^N_{n=1}$, and so are $P_{r,i} = \{p^n_{r,i}\}^N_{n=1}$, where N denotes the number of samples and F_2 denotes the number of frequency bands. The inter-band dispersity of the left hemisphere is formulated as

$$\tau_l = \frac{\sigma\left(\{P_{l,i}\}^{F_2}_{i=1}\right)}{\left(\frac{1}{F_2}\sum^{F_2}_{i=1} P_{l,i}\right)} \tag{9}$$

where $\sigma(\cdot)$ represents taking the standard deviation. We posit that the ratio of the standard deviation of inter-band average energy to the overall mean can provide a more accurate assessment of the overall activity level of the brain hemisphere and mitigate differences caused by spatial filters in the FG-HANet. Similarly, the inter-band dispersity of the right hemisphere τ_r is obtained. To further investigate the dominance of the left and right brain for different emotions, we define the ratio of τ_l and τ_r as the inter-band hemispheric dominance, denoted as dom. It can be expressed as

$$dom = \frac{\tau_l}{\tau_r}. \tag{10}$$

If $dom > 1$, it indicates that for this type of emotion, the left brain is more activated, signifying the left brain's dominance in perceiving emotions. Conversely, if dom < 1, it indicates that the right brain is more activated.

To analyze hemispheric dominance under different emotional states, we calculate the values of dom for six typical spatial filtered components obtained from the VSConv layer under positive and negative emotions in Fig. 7. Among the six spatial filtered components, the activated regions of first three components on S3, S6, S12 are frontal and temporal lobes, typically with emotional hemispheric asymmetry, while the activated regions of the rest three on S3, S8, S7 are located in the central region of the human brain. From Fig. 7, we can observe that the first three spatial filtered components have a $dom > 1$ under positive emotions and a $dom < 1$ under negative emotions, indicating a relative activation of the right frontal lobe is associated with negative emotions, which is consistent with the conclusion in Alarcao and Fonseca (2017). For the last three spatial filtered components, the values of dom under the two types of emotions in the central area are relatively closer.

In Fig. 7, the second component on S6 shows the largest difference in dom values between the two types of emotions, which prompts us to conduct further investigation on this prominent component. We observe that on this component the dark regions are concentrated around specific electrodes, such as FC3 and FC5, while the corresponding electrodes in the opposite hemisphere do not show dark regions. This suggests that the hemispheric asymmetry is notably presented at these electrode pairs' locations, like the two pairs (FC5, FC6) and (FC3, FC4). Thus we focus on the two pairs to explore the impact of the VSConv layer in the FG-HANet. Specifically, we select the EEG inputs from these specific electrode pairs and feed them into the network. Then we visualize and compare the dom values obtained by our FG-HANet with and without the VSConv layer, as shown in Fig. 8. It can be observed that the difference in dom values between positive and negative emotions is amplified when the VSConv layer is used. This suggests that the VSConv layer enhances the network's ability to distinguish different categories. Besides, this layer improves the model's interpretability.

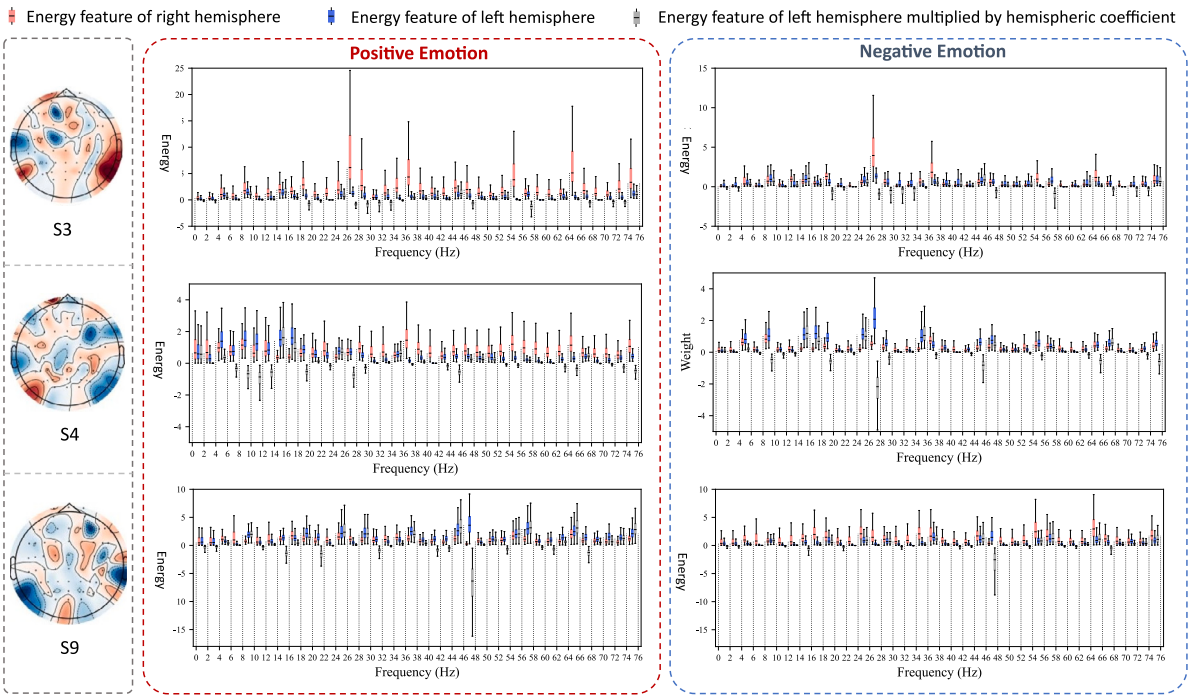


Fig. 9. Energy box diagram for 2-Hz narrow-bands analysis of S3, S4, and S9. Left panel shows samples under positive emotions and right shows samples under negative emotions. 38 filters are applied, ranging from 0.1–76 Hz. Blue and red bars represent the relative energy distribution of the left and right hemisphere, respectively, and gray bars represent the relative energy distribution of the left hemisphere multiplied by the hemispheric coefficient.

5.4. Hemispheric asymmetry discriminators

To investigate the hemispheric asymmetry on narrow-bands of each individual subject, we multiply the weight of the HAD and the PSD feature of the right branch network, and visualize it in Fig. 9 based on the energy box diagram in Fig. 6. In Section 3.4, we have analyzed that the discriminator assigns weights to the PSD features of each narrow-band, and a negative value indicates it is a more crucial hemispheric asymmetry feature in the corresponding frequency narrow band. From the narrow bands with negative values by the discriminator, it can be observed that the distance between the left hemisphere energy multiplied by the asymmetric coefficient and the right hemisphere energy is further pulled apart, verifying the critical role of the HAD.

Previous studies (Balconi & Mazza, 2009) have shown that positive and negative emotions are associated with energy hemispheric asymmetry in the alpha range of frontal lobes. In addition, the energy changes within the beta range have been confirmed to be related to emotional valence (Jatupaiboon, Pan-Ngum, & Israsena, 2013), and the changes in gamma waves on both sides of the temporal lobe are also related to happy and sad emotions (Park et al., 2011). From Fig. 9, high-frequency bands, such as the beta and gamma bands, stronger asymmetric activation can be observed. The narrow-bands where hemisphere differences are maximized through the discriminator on S9, S3, and S4 are located in the 24th filter (46–48 Hz, gamma band), 29th filter (56–58 Hz, gamma band), and 14th filter (26–28 Hz, beta band), respectively. There is also significant hemispherical asymmetry in the alpha band of S4, specifically in the 3rd, 4th, 5th, and 6th filters. These observations are consistent with the wide-band findings in previous studies. It is worth noting that our work is the first to analyze the hemispheric asymmetry regarding emotions in the fine-grained narrow-band of 2-Hz, to our best knowledge.

Based on the energy box diagram of each subject, posterior knowledge regarding the critical narrow bands can be obtained, which can be used to modify the initialization of the VSConv in our FG-HANet. In particular, we initialize the VSConv using 38 filters with a bandwidth of 2-Hz ranging from 0.1–76 Hz, and also using filters selected through the HAD on three subjects, where the VSConv is frozen during the training. We compare the accuracy of the three subjects obtained with the FG-HANet, as shown in Fig. 10. It can be seen that, on S9, S3, and S4, utilizing a small number of filters for the initialization of VTConv reduces the accuracy by 0.54%, 0.17%, and 0.94%, respectively, compared to using all 38 filters. The performance of using selected fewer filters is comparable to that initializing all 38 filters in the FG-HANet. In addition, using fewer filters indicates a reduction in network parameters. The effectiveness of asymmetric emotional information within narrow bands and the discriminative ability of the discriminators for asymmetric features have been demonstrated.

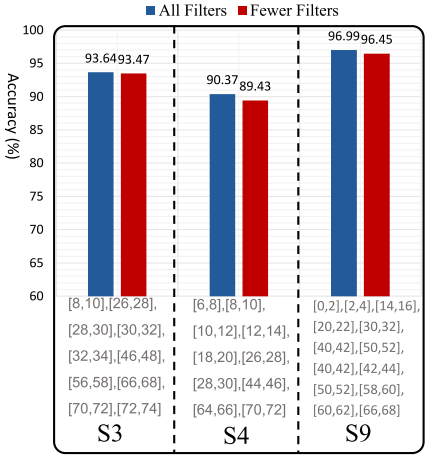


Fig. 10. Accuracy of all 38 filters (0.1–76 Hz) and few filters selected by HAD for initialization of VSConv in FG-HANet.

6. Conclusion

This paper introduces FG-HANet, an interpretable end-to-end model design for EEG emotion classification, inspired by neuroscience findings about hemispheric asymmetry patterns in emotion expression. Emotion-related prior knowledge in the frequency domain is embedded into the network to implement fine-grained features extraction. Moreover, a three-stage training pipeline is applied to enhance the model performance. Through extensive experiments on two datasets, it is demonstrated that the FG-HANet can achieve state-of-the-art performance in emotion classification, due to its strong ability to capture fine-grained asymmetry information between hemispheres. This work further explores the hemispheric asymmetry regarding emotions over critical narrow frequency bands, and inter-band hemispheric activity and dominance under different emotions, providing valuable insights into network design improvement and emotion generation mechanisms.

Besides its excellent performance on classifying EEG emotion states, our method also holds favorable interpretability. These advantages promise its real-world applications, such as human–computer interaction and mental health monitoring. Furthermore, the model's ability to differentiate fine-grained hemispheric asymmetry opens up possibilities for advancing non-invasive emotion regulation techniques, such as transcranial magnetic stimulation (TMS) implemented on specific frequency bands. This could be particularly beneficial in the treatment of emotion-related disorders like depression and anxiety.

CRediT authorship contribution statement

Ruofan Yan: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Na Lu:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis. **Yuxuan Yan:** Methodology, Investigation, Formal analysis. **Xu Niu:** Writing – review & editing, Investigation, Formal analysis, Conceptualization. **Jibin Wu:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ahmed, Md Zaved Iqubal, Sinha, Nidul, Phadikar, Souvik, & Ghaderpour, Ebrahim (2022). Automated feature extraction on AsMap for emotion classification using EEG. *Sensors*, *22*(6), 2346.

Alarcao, Soraia M., & Fonseca, Manuel J. (2017). Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, *10*(3), 374–393.

Ali, Mouhannad, Mosa, Ahmad Haj, Al Machot, Fadi, & Kyamakya, Kyandoghere (2016). EEG-based emotion recognition approach for e-healthcare applications. In *2016 eighth international conference on ubiquitous and future networks* (pp. 946–950). IEEE.

Atanassov, Atanas V., Pilev, Dimitar I., Tomova, Fani N., & Kuzmanova, Vanya D. (2021). Hybrid system for emotion recognition based on facial expressions and body gesture recognition. In *2021 international conference automatics and informatics* (pp. 135–140). http://dx.doi.org/10.1109/ICAI52893.2021.9639829.

Balconi, Michela, & Mazza, Guido (2009). Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues.: ERS/ERD and coherence measures of alpha band. *International Journal of Psychophysiology*, *74*(2), 158–165.

Brigham, E. Oran (1988). *The fast Fourier transform and its applications*. Prentice-Hall, Inc.

Chapelle, Olivier, & Zien, Alexander (2005). Semi-supervised classification by low density separation. In *International workshop on artificial intelligence and statistics* (pp. 57–64). PMLR.

Chen, Zhige, Yang, Rui, Huang, Mengjie, Li, Fumin, Lu, Guoping, & Wang, Zidong (2024). EEGProgress: A fast and lightweight progressive convolution architecture for EEG classification. *Computers in Biology and Medicine*, *169*, Article 107901.

Chen, Zhige, Yang, Rui, Huang, Mengjie, Wang, Zidong, & Liu, Xiaohui (2024). Electrode domain adaptation network: Minimizing the difference across electrodes in single-source to single-target motor imagery classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

Cohen, Taco, & Welling, Max (2016). Group equivariant convolutional networks. In *International conference on machine learning* (pp. 2990–2999). PMLR.

Cui, Heng, Liu, Aiping, Zhang, Xu, Chen, Xiang, Wang, Kongqiao, & Chen, Xun (2020). EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*, *205*, Article 106243.

Ding, Yi, Robinson, Neethu, Zhang, Su, Zeng, Qiuhao, & Guan, Cuntai (2022). Tsception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. *IEEE Transactions on Affective Computing*.

Dumitru, Ana, Rocchi, Lorenzo, Saini, Fedal, Rothwell, John C, Roiser, Jonathan P, David, Anthony S, et al. (2020). Influence of theta-burst transcranial magnetic stimulation over the dorsolateral prefrontal cortex on emotion processing in healthy volunteers. *Cognitive, Affective, & Behavioral Neuroscience*, *20*, 1278–1293.

Ekman, Paul (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200.

Gupta, Anjali, & Mishra, Divya (2023). Sentimental voice recognition: An approach to analyse the emotion by voice. In *2023 international conference on electrical, electronics, communication and computers* (pp. 1–6). http://dx.doi.org/10.1109/ELEXCOM58812.2023.10370064.

Hinrikus, Hiie, Suhhova, Anna, Bachmann, Maie, Aadamsoo, Kaire, Võhma, Ülle, Lass, Jaanus, et al. (2009). Electroencephalographic spectral asymmetry index for detection of depression. *Medical & Biological Engineering & Computing*, *47*, 1291–1299.

Huang, Dongmin, Chen, Sentao, Liu, Cheng, Zheng, Lin, Tian, Zhihang, & Jiang, Dazhi (2021). Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for EEG emotion recognition. *Neurocomputing*, *448*, 140–151.

Huang, Dong, Guan, Cuntai, Ang, Kai Keng, Zhang, Haihong, & Pan, Yaozhang (2012). Asymmetric spatial pattern for EEG-based emotion detection. In *The 2012 international joint conference on neural networks* (pp. 1–7). IEEE.

Jatupaiboon, Noppadon, Pan-Ngum, Setha, & Israsena, Pasin (2013). Emotion classification using minimal EEG channels and frequency bands. In *The 2013 10th international joint conference on computer science and software engineering* (pp. 21–24). IEEE.

Jin, Ming, Zhu, Enwei, Du, Changde, He, Huiguang, & Li, Jinpeng (2023). PGCN: Pyramidal graph convolutional network for EEG emotion recognition. arXiv preprint arXiv:2302.02520.

Kamble, Kranti, & Sengupta, Joydeep (2023). A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals. *Multimedia Tools and Applications*, *82*(18), 27269–27304.

Karnati, Mohan, Seal, Ayan, Bhattacharjee, Debotosh, Yazidi, Anis, & Krejcar, Ondrej (2023). Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–31. http://dx.doi.org/10.1109/TIM.2023.3243661.

Kothe, Christian A., & Makeig, Scott (2011). Estimation of task workload from EEG data: new and current tools and perspectives. In *2011 annual international conference of the IEEE engineering in medicine and biology society* (pp. 6547–6551). IEEE.

Kuang, Dongyang, & Michoski, Craig (2023). SEER-net: Simple EEG-based recognition network. *Biomedical Signal Processing and Control*, *83*, Article 104620.

Lawhern, Vernon J, Solon, Amelia J, Waytowich, Nicholas R, Gordon, Stephen M, Hung, Chou P, & Lance, Brent J (2018). Eegnet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), Article 056013.

Li, Mu, & Lu, Bao-Liang (2009). Emotion classification based on gamma-band EEG. In *2009 annual international conference of the IEEE engineering in medicine and biology society* (pp. 1223–1226). IEEE.

Li, Yang, Wang, Lei, Zheng, Wenming, Zong, Yuan, Qi, Lei, Cui, Zhen, et al. (2020). A novel bi-hemispheric discrepancy model for EEG emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(2), 354–367.

Li, Jinpeng, Zhang, Zhaoxiang, & He, Huiguang (2018). Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cognitive Computation*, *10*, 368–380.

Li, Yang, Zheng, Wenming, Zong, Yuan, Cui, Zhen, Zhang, Tong, & Zhou, Xiaoyan (2018). A bi-hemisphere domain adversarial neural network model for EEG emotion recognition. *IEEE Transactions on Affective Computing*, *12*(2), 494–504.

Liu, Junxiu, Wu, Guopei, Luo, Yuling, Qiu, Senhui, Yang, Su, Li, Wei, et al. (2020). EEG-based emotion classification using a deep neural network and sparse autoencoder. *Frontiers in Systems Neuroscience*, *14*, 43.

Liu, Huan, Yang, Shusen, Zhang, Yuzhe, Wang, Mengze, Gong, Fanyu, Xie, Chengxi, et al. (2024). LibEER: A comprehensive benchmark and algorithm library for EEG-based emotion recognition. arXiv preprint arXiv:2410.09767.

Luo, Xi, Che, Xianwei, & Li, Hong (2023). Concurrent TMS-eeg and EEG reveal neuroplastic and oscillatory changes associated with self-compassion and negative emotions. *International Journal of Clinical and Health Psychology*, *23*(1), Article 100343.

Meng, Ming, Hu, Jiahao, Gao, Yunyuan, Kong, Wanzeng, & Luo, Zhizeng (2022). A deep subdomain associate adaptation network for cross-session and cross-subject EEG emotion recognition. *Biomedical Signal Processing and Control*, *78*, Article 103873.

Niu, Xu, Lu, Na, Kang, Jianghong, & Cui, Zhiyan (2022). Knowledge-driven feature component interpretable network for motor imagery classification. *Journal of Neural Engineering*, *19*(1), Article 016032.

Orgo, Laura, Bachmann, Maie, Lass, Jaanus, & Hinrikus, Hiie (2015). Effect of negative and positive emotions on EEG spectral asymmetry. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society* (pp. 8107–8110). IEEE.

Palmiero, Massimiliano, & Piccardi, Laura (2017). Frontal EEG asymmetry of mood: a mini-review. *Frontiers in Behavioral Neuroscience*, *11*, 224.

Park, Kwang Shin, Choi, Hyun, Lee, Kuem Ju, Lee, Jae Yun, An, Kwang Ok, & Kim, Eun Ju (2011). Emotion recognition based on the asymmetric left and right activation. *International Journal of Medicine and Medical Sciences*, *3*(6), 201–209.

Petrantonakis, Panagiotis C., & Hadjileontiadis, Leontios J. (2011). A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition. *IEEE Transactions on Information Technology in Biomedicine*, *15*(5), 737–746.

Schirrmeister, Robin Tibor, Springenberg, Jost Tobias, Fiederer, Lukas Dominique Josef, Glasstetter, Martin, Eggensperger, Katharina, Tangermann, Michael, et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, *38*(11), 5391–5420.

Seal, Ayan, Reddy, Puthi Prem Nivesh, Chaithanya, Pingali, Meghana, Arramada, Jahnavi, Kamireddy, Krejcar, Ondrej, et al. (2020). An EEG database and its initial benchmark emotion classification performance. *Computational and Mathematical Methods in Medicine*, *2020*(1), Article 8303465.

Song, Tengfei, Zheng, Wenming, Song, Peng, & Cui, Zhen (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, *11*(3), 532–541.

Vempati, Raveendrababu, & Sharma, Lakhan Dev (2023). A systematic review on automated human emotion recognition using electroencephalogram signals and artificial intelligence. *Results in Engineering*, *18*, Article 101027.

Xiao, Guowen, Shi, Meng, Ye, Mengwen, Xu, Bowen, Chen, Zhendi, & Ren, Quansheng (2022). 4D attention-based neural network for EEG emotion recognition. *Cognitive Neurodynamics*, 1–14.

Yan, Ruofan, Lu, Na, Niu, Xu, & Yan, Yuxuan (2022). Hemispheric asymmetry measurement network for emotion classification. In *Chinese conference on biometric recognition* (pp. 307–314). Springer.

Ye, Mengqing, Chen, C. L. Philip, & Zhang, Tong (2022). Hierarchical dynamic graph convolutional network with interpretability for EEG-based emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhang, Yongqing, Cao, Wenpeng, Feng, Lixiao, Wang, Manqing, Geng, Tianyu, Zhou, Jiliu, et al. (2023). SHNN: A single-channel EEG sleep staging model based on semi-supervised learning. *Expert Systems with Applications*, *213*, Article 119288.

Zheng, Fa, Hu, Bin, Zhang, Shilin, Li, Yalin, & Zheng, Xiangwei (2021). EEG emotion recognition based on hierarchy graph convolution network. In *2021 IEEE international conference on bioinformatics and biomedicine* (pp. 1628–1632). IEEE.

Zheng, Wei-Long, Liu, Wei, Lu, Yifei, Lu, Bao-Liang, & Cichocki, Andrzej (2018). Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, *49*(3), 1110–1122.

Zheng, Wei-Long, & Lu, Bao-Liang (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175.

Zhong, Xinyue, Gu, Yun, Luo, Yutong, Zeng, Xiaomei, & Liu, Guangyuan (2023). Bihemisphere asymmetric attention network: recognizing emotion from EEG signals based on the transformer. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *53*(12), 15278–15294.

Zhong, Peixiang, Wang, Di, & Miao, Chunyan (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, *13*(3), 1290–1301.