Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

TFDCA for EEG emotion recognition through dynamic channel aggregation

Yangfan Li
Anhui Agricultural University
Xinjun Zhou
Anhui Agricultural University
Hao Liu
Anhui Agricultural University
Xiule Wang
Anhui Agricultural University
Yubo Xing
Anhui Agricultural University
Jie He
Anhui Agricultural University
Tian Zhao
Anhui Agricultural University
Siwei Yang
Anhui Agricultural University
Yindong Dong
dongyindong66@163.com
Anhui Agricultural University

Research Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11215051/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

TFDCA for EEG emotion recognition through dynamic channel aggregation

Yangfan Li, Xinjun Zhou, Hao Liu, Xiule Wang, Yubo Xing, Jie He, Tian Zhao, Siwei Yang, Yindong Dong*

Anhui Agricultural University, Hefei 230036, Anhui, China

*Corresponding author: Yindong Dong, dongyindong66@163.com

Abstract

Electroencephalography (EEG) provides valuable information for emotion recognition owing to its high temporal resolution and sensitivity to emotional states. However, existing brain region methods often rely on predefined static partitions, which have limited capacity to capture dynamic channel interactions, functional overlap, and reorganization across regions. A temporal – frequency fusion method with dynamic channel aggregation (TFDCA) is proposed to characterize channel dependencies across emotional states. A fusion encoder extracts complementary temporal and frequency features: lightweight convolutional attention weights multiband differential entropy features, while gated recurrent units capture temporal dependencies. Adaptive weighting integrates these features into multiscale node representations. Channel similarities derived from self-attention are combined with electrode spatial topology under a sparsity constraint to construct an adjacency matrix. Differentiable soft sorting jointly reorders its rows and columns to aggregate similar channels into regions, incorporating local structure and global channel relationships. A brain region transformer further captures interregional dependencies for emotion classification. On SEED, TFDCA achieved accuracies of 96.96% and 85.31% under subject-dependent and subject-independent protocols, respectively. On DEAP, subject-dependent accuracies were 92.84% for valence and 93.02% for arousal, while subject-independent accuracies were 65.13% and 66.24%, respectively. Visualization analyses revealed emotion-related channel aggregation patterns and their variation, providing evidence for interpreting the learned representations. These findings support dynamic channel aggregation as an approach to representing EEG channel relationships for emotion recognition.

Keywords Electroencephalography · Emotion recognition · Temporal–frequency fusion · Soft sorting · Dynamic channel aggregation · Brain region transformer

1 Introduction

Emotion is a complex psychological and physiological state arising from the cognitive appraisal of external stimuli and influences behavioral regulation, decision-making, and social interaction [1]. Emotion recognition identifies emotional states computationally from multimodal information, including facial expressions, speech, and physiological signals, with applications in human–computer interaction and mental health monitoring [2].

Electroencephalography (EEG) provides an important source of information for emotion recognition because of its high temporal resolution and sensitivity to underlying emotional states [3]. However, EEG recordings are susceptible to interference from other physiological signals and environmental noise. Preprocessing and feature extraction are therefore necessary to reveal emotion-related information in EEG signals [4]. Common preprocessing steps include artifact removal, frequency filtering, and baseline correction. Feature extraction reduces dimensionality and emphasizes relevant information, lowering the computational burden of subsequent analysis and improving its efficiency and stability. Common EEG features describe either the temporal or the frequency domain. Temporal features are directly accessible [4] and include signal statistics, Hjorth parameters, and fractal dimensions. Frequency-domain features are commonly derived using the discrete Fourier transform and include power spectral density (PSD) and differential entropy (DE).

Neuropsychological research indicates that emotion generation and regulation involve coordinated activity across multiple regions, including the prefrontal cortex, limbic system, and parietal cortex. This functional differentiation suggests that EEG channels should contribute selectively and unequally to models of emotional activity. Accordingly, existing studies have grouped channels using electrode locations and functional connectivity, then constructed hierarchical aggregation structures to integrate information in a manner informed by brain organization. For example, the instance-adaptive graph method of Song et al. [5] selectively aggregates channels through spatial graph coarsening after learning graph connections; Gao et al. [6] progressively integrate local regions to model local and global relationships, improving recognition accuracy and feature representation; Jin et al. [7] combine electrode topology and multiscale structures for local, mesoscopic, and global aggregation; and the MD²GRL network of Tang et al. [8] models temporal and frequency features in parallel branches and performs region-level graph coarsening.

Despite these advances in brain region modeling, several issues remain unresolved.

Most existing studies extract temporal and frequency features separately and directly use the resulting features as graph node inputs. Although some approaches combine them sequentially or by concatenation, such static feature combinations cannot readily adapt to their changing dependencies and complementary information.

Most existing methods rely on static predefined brain region partitions, which cannot readily accommodate changing interactions between EEG channels during emotional processing. Individual differences, cognitive states, and external

stimuli produce temporal and interindividual variability, whereas fixed partitions restrict the representation of functional reorganization and individualized adaptation.

Conventional partitions are usually defined by anatomical boundaries or electrode positions and may not adequately capture functional overlap and dynamic reorganization across regions. This limitation can hinder the representation of coordinated activation and functional coupling, reducing the ability to characterize and interpret higher-level interactions between brain regions.

A method for EEG emotion recognition based on dynamic channel aggregation is proposed to characterize interactions between brain regions by adaptively modeling channel dependencies. First, a temporal–frequency fusion encoder is constructed to extract information from complementary signal domains. Its frequency branch separates EEG into frequency bands and computes differential entropy to represent activity at different rhythms. A lightweight convolutional attention (LCA) module models relationships between frequency bands, while a gated recurrent unit (GRU) models temporal dependencies associated with evolving emotional responses. Adaptive weighting then balances the contributions of temporal and frequency features, producing discriminative multiscale representations. Inspired by the ability of graph transformers [9] to model global node dependencies, self-attention based on node attributes is introduced to characterize channel similarity. The resulting similarity matrix is linearly combined with spatial topology and constrained for sparsity to construct an adaptive adjacency matrix. A soft-sorted dynamic channel aggregation (SDCA) module is then designed to integrate local brain region representations and global channel relationships. Locally, dynamic partitioning, within-region aggregation, and block-diagonal regularization organize emotion-related channels; globally, channel similarity and soft sorting align channel representations under a joint objective to capture dependencies across regions. The region-level features produced by SDCA are subsequently input to a brain region transformer (BRT) to model interactions between regions. The method was evaluated on the public SEED and DEAP emotion EEG datasets. The results demonstrated competitive emotion classification performance. Visualization revealed differences in channel aggregation across emotional states, providing evidence for interpreting the learned representations.

The methodological contributions comprise a matrix-sorting approach to dynamic channel aggregation and an adaptive temporal–frequency fusion encoder. A soft sorting matrix jointly reorders the rows and columns of the adjacency matrix to group channels with similar representations and characterize channel relationships. The fusion encoder integrates EEG temporal dynamics and spectral features through adaptive weighting to obtain multiscale representations. Evaluation on SEED and DEAP yielded competitive classification results, including an assessment of cross-subject performance under subject-independent protocols. Visualization analyses further revealed emotion-related channel aggregation patterns, providing evidence for interpreting the learned representations.

1.1 Related work

1.1.1 EEG features for emotion recognition

EEG is widely used for emotion recognition because of its high temporal resolution and noninvasive acquisition. Emotional states are associated with multiband neural oscillations, and changes in rhythms such as δ、θ、α、β、γ can reflect cognitive and emotional processing. EEG feature extraction is therefore a key determinant of emotion recognition performance.

Early studies primarily modeled a single feature domain. Temporal approaches use signal statistics, such as means, variances, and Hjorth parameters, or sequence models, such as recurrent neural networks and long short-term memory networks, to capture temporal dynamics; frequency approaches use PSD or DE to characterize spectral energy distributions and emotion-related patterns. However, single-domain approaches may not fully capture the coupling between temporal fluctuations and frequency modulation, limiting their representation of emotion-related neural dynamics.

Recent studies have consequently explored joint temporal–frequency modeling. Some obtain time–frequency representations using the short-time Fourier transform (STFT) or wavelet transforms to capture local variations in time and frequency, whereas others learn temporal and spectral representations directly from raw EEG using deep networks. These approaches improve feature discrimination, although the joint modeling of temporal–frequency dependencies and brain region interactions remains challenging. The previously developed TFF-MDCN model uses a gating mechanism to adaptively regulate the weights of temporal and frequency features [10], providing a basis for the fusion encoder developed here.

1.1.2 EEG representations in deep learning

EEG has spatial and temporal dimensions: channels record activity at different electrode locations, while time describes changes in neural electrical activity. Common representations include multichannel time series, time–frequency images, and graphs. For multichannel time series, temporal convolutions commonly extract information within channels, while spatial convolutions model relationships across channels. For time–frequency images, spectral or time–frequency features can be mapped onto a two-dimensional electrode layout, with different frequency bands or time windows stacked as model inputs.

Graph representations have attracted increasing attention because they explicitly model interactions between brain regions. When EEG is represented using a graph neural network (GNN), conventional constructions often quantify channel relationships through linear measures, such as Pearson correlation coefficients or linear Granger causality. These constructions are relatively straightforward, but fixed connectivity assumptions can limit their representation of temporal variation, nonlinear relationships, and individual differences. DGCNN treats the adjacency matrix as a learnable parameter and updates its entries through loss gradients, learning channel relationships that support discriminative EEG feature extraction [11]. Tang et al. [8] compute node similarity using self-attention, then combine the similarity and spatial matrices linearly with sparsification to obtain the adjacency matrix. These studies support adaptive EEG graph learning and motivate further modeling of global node interactions. The previously developed FCMGAT model uses a fuzzy adjacency matrix to smoothly adjust channel connection weights [12], providing a basis for adaptive channel relationship modeling.

Transformer-based graph networks further use self-attention to learn global dependencies between nodes. Graphormer, proposed by Ying et al. [9], incorporates structural encodings into a transformer so that node representations reflect graph topology. For brain network analysis, the Brain Network Transformer of Kan et al. [13] uses regional connectivity profiles as node features, models regional interactions through attention, and constructs graph-level representations through a clustering readout. Com-BrainTF, proposed by Bannadabhavi et al., uses a hierarchical local–global transformer to model relationships within and between functional magnetic resonance imaging (fMRI) communities [14]; BN-BrainTF, proposed by Kim et al., combines EEG source localization with a community-aware global–local transformer for emotion recognition [15]. These studies inform the modeling of coordinated relationships between brain regions. EEG emotion recognition also requires channels to be organized into functionally meaningful regions, linking within-region aggregation to between-region interaction modeling. Brain-region graph methods therefore provide a relevant direction for investigation.

1.1.3 Brain-region graph methods

Brain-region graph methods group and aggregate channels to learn regional representations and dependencies, as reviewed by Liu et al. [16]. Coordinated brain activity, such as prefrontal–amygdala interactions in emotion regulation [17], provides a neurophysiological basis for modeling regional relationships. The IAG method of Song et al. [5] organizes electrodes into 17 regions using the 10–20 system and averages channel embeddings within each region, providing a practical approach to region-level graph modeling. Two common implementations use predefined anatomical partitions or hierarchical graph learning.

With predefined anatomical partitions, a graph network first extracts channel-level features and then aggregates them according to the partition, constructing global representations before modeling local dependencies. The V-IAG extension developed by Song et al. [18] also uses predefined partitions. In MD²GRL, Tang et al. [8] apply predefined region grouping in both the temporal and frequency graph branches.

Hierarchical graph learning first groups channels into local graphs, learns within-region dependencies, and then treats local graphs as nodes of a global graph to model regional interactions. In contrast to the preceding processing order, local relationships are learned before global dependencies between regions. The LAG method of Gao et al. [6] uses a K-level self-adaptive ascending network (SALK) to progressively integrate regions, proceeding from local connections to global representations. The GMSS method of Li et al. [19] also uses hierarchical learning, combining spatial and spectral jigsaw tasks with multitask self-supervision and subsequently integrating regional information through contrastive learning.

These two approaches differ in processing order: predefined partitioning learns a global graph before local dependencies, whereas hierarchical methods learn local graphs before global regional relationships. Despite this distinction, both can depend heavily on static predefined regions, limiting the representation of changing channel interactions and functional overlap. The previously developed HMABRF model uses differentiable Gumbel–Softmax clustering under neuroscientific priors to learn adaptive brain region partitions [20], providing a basis for dynamic regional organization.

The present method further forms regional features through dynamic channel aggregation and inputs them to BRT, linking region construction to higher-level interaction modeling.

2 Methods

The TFDCA framework is illustrated in Fig. 1. It comprises three main modules: graph input construction, dynamic channel aggregation, and BRT. Graph input construction embeds temporal–frequency EEG features as nodes and adaptively builds the adjacency matrix from those attributes. Dynamic channel aggregation constructs a soft sorting matrix that jointly reorders adjacency rows and columns, grouping channels with similar representations to capture local and global structure. BRT further captures interactions between brain regions and refines their representations.

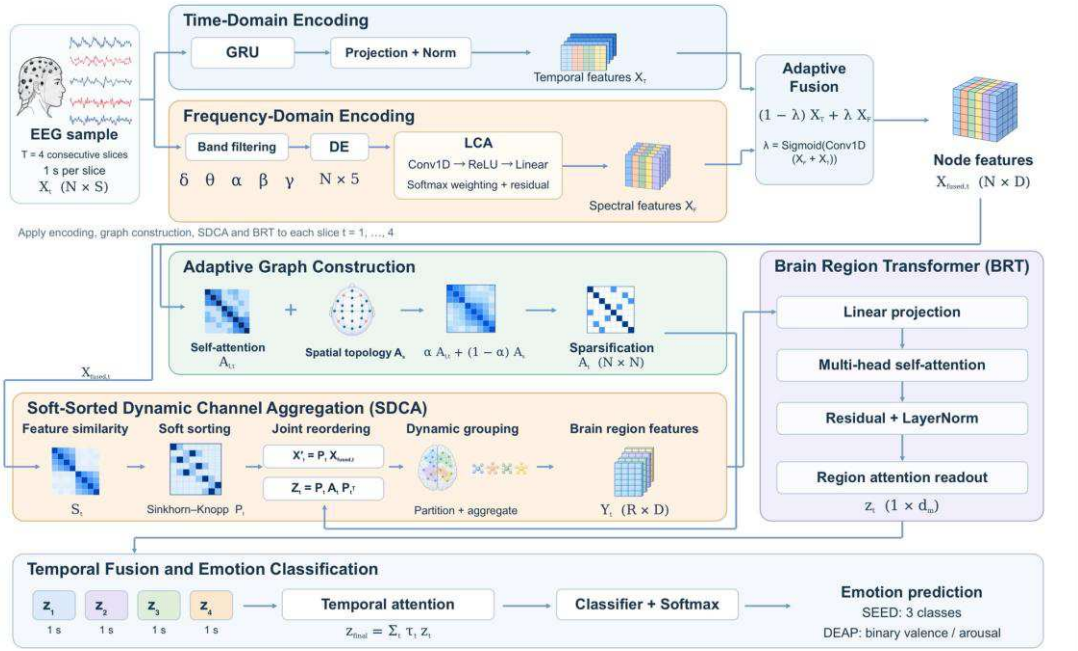


Fig. 1 Overall framework of TFDCA

2.1 Graph input construction

2.1.1 Temporal–frequency fusion encoding

Each EEG input sample consists of T consecutive temporal segments, denoted by $\mathcal{X} = \{X_1, X_2, \ldots, X_T\}$, where each segment X_t contains multichannel EEG within a 1-s window. T denotes the number of temporal segments in one input sample. Based on the parameter comparison, T = 4 is used, so each input comprises four consecutive 1-s segments and covers 4 s. The temporal–frequency encoder and graph learning modules process each segment separately, after which temporal attention integrates the segment-level representations.

(1) Frequency encoding: EEG is divided into five bands, δ (1–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–50 Hz), and DE is computed within each band to construct the frequency features $X_f \in \mathbb{R}^{N \times F}$. Here, N denotes the number of channels, and F denotes the number of frequency bands.

An LCA module is designed to model relationships among frequency features and improve spectral discrimination by adaptively adjusting the weights of different bands. First, a one-dimensional convolution captures local dependencies between bands:

$$X_c = \mathrm{Conv1D}(X_f; W_c) + b_c. \tag{1}$$

Here, W_c and b_c are the convolution kernel and bias, respectively. Nonlinear activation and normalization then convert the convolution output into frequency-band attention weights:

$$A_f = \mathrm{Softmax}(\mathrm{ReLU}(W_2 \cdot X_c)). \tag{2}$$

Here, $W_2 \in \mathbb{R}^{F \times F}$ is a learnable matrix, and the softmax function constrains the band weights to [0, 1] with a sum of 1.

The weighted frequency features are then computed as

$$X_F = A_f \odot X_f + W_r X_f, \tag{3}$$

where $\odot$ denotes elementwise multiplication, $W_r \in \mathbb{R}^{D\times F}$ is a linear transformation matrix, and D is the output feature dimension. The resulting $X_F \in \mathbb{R}^{N\times D}$ provides an enriched frequency representation for subsequent EEG processing.

(2) Temporal encoding: EEG exhibits dynamic behavior and long-term dependencies, and gated recurrent units provide efficient and stable sequence modeling. A GRU-based temporal encoder is therefore introduced. The GRU operations are defined as follows:

$$R_t = \sigma(W_R x_t + U_R h_{t-1} + b_R); \tag{4}$$

$$Z_t = \sigma(W_Z x_t + U_Z h_{t-1} + b_Z); \tag{5}$$

$$\tilde{h}_t = tanh(Wx_t + U(R_t \cdot h_{t-1}) + b); \tag{6}$$

$$h_t = Z_t \cdot h_{t-1} + (1 - Z_t) \cdot \tilde{h}_t. \tag{7}$$

Here, σ is the sigmoid function, and $tanh$ is the hyperbolic tangent function, both of which provide nonlinear activation. The reset gate R_t controls how much information from the previous hidden state contributes to the candidate hidden state. The update gate Z_t controls the contribution of the previous state to the updated hidden state. W and U denote trainable weight matrices, and b denotes the bias. The GRU update can be written compactly as

$$h_t = GRU(h_{t-1}, x_t), \tag{8}$$

where h_t and h_{t-1} are the hidden states at the current time step t and the preceding step, respectively, and x_t is the current input. For EEG input $X_t \in \mathbb{R}^{N\times S\times 1}$, S denotes the number of samples within a 1-s segment, and 1 denotes the added temporal feature dimension. Collecting the hidden outputs over time gives the temporal representation $X_T \in \mathbb{R}^{N\times S\times d}$, where d is the GRU hidden dimension. A fully connected layer and normalization transform X_T into $X_T \in \mathbb{R}^{N\times D}$, with D matching the frequency encoder output dimension for feature fusion. The GRU time index refers to sampling points within a segment; T counts segments and is distinct from the sampling-point count S. The subscript T in temporal feature symbols denotes the temporal domain, rather than the number of segments.

(3) Temporal–frequency fusion: an adaptive weighting mechanism is designed to integrate frequency features $X_F \in \mathbb{R}^{N\times D}$ and temporal features $X_T \in \mathbb{R}^{N\times D}$ at the channel level. The two representations are first added channelwise, then passed through a one-dimensional convolution and nonlinear activation to generate fusion weights:

$$\lambda = \sigma(\text{Conv1D}(X_F + X_T)). \tag{9}$$

Here, σ is the sigmoid function, which constrains the weights to [0, 1], and Conv1D captures local interactions between channels.

The temporal and frequency representations are then combined through adaptive weighting:

$$X_{\text{fused}} = \lambda \odot X_F + (1 - \lambda) \odot X_T. \tag{10}$$

Here, $\odot$ denotes elementwise multiplication, and $X_{\text{fused}} \in \mathbb{R}^{N\times D}$ is the fused feature matrix. The mechanism adjusts the contributions of the two domains according to channel characteristics, integrating complementary information across scales. Compared with a fixed linear combination, adaptive weighting can make more effective use of both domains and improve the discrimination and robustness of the fused representation.

2.1.2 Dynamic adjacency matrix construction

An adjacency construction strategy combines data-driven information with spatial priors to represent both functional channel relationships and the physical electrode layout.

Self-attention is introduced to construct a learnable adjacency matrix $A_l \in \mathbb{R}^{N\times N}$ that adapts to changing relationships between electrodes. Fixed connections based on predefined topology or similarity measures may not fully capture temporal changes in EEG. Accordingly, potential relationships between electrode pairs are estimated from node features using self-attention, allowing the adjacency matrix to adapt to feature variation. Given the input feature matrix $X_f \in \mathbb{R}^{N\times F}$, node similarity is computed as

$$A_l = \text{softmax}\left(\frac{(X_f W_Q)(X_f W_K)^T}{\sqrt{d}}\right), \tag{11}$$

where W_Q and W_K are trainable query and key projection matrices, respectively. The similarity matrix A_l represents potential functional relationships between electrodes and supplies dynamic structural information for subsequent feature propagation.

Electrode spatial relationships provide prior information for EEG modeling [21], and physical distances are used here to describe electrode distribution and construct edge features. Following distance-dependent brain network modeling [22], spatial connection weights are defined using an inverse-square distance form; for EEG with N channels, the entries a_{ij} of the spatial topology matrix $A_s \in \mathbb{R}^{N\times N}$ are given by

$$a_{i,j} = min(\frac{\delta}{{d_{i,j}}^2}, 1), \tag{12}$$

where $d_{i,j}$ is the distance between electrodes i and j in the 10–20 layout, and δ is a distance calibration constant. Following [23], $\delta = 5$ is used to obtain approximately 20% non-negligible connections.

The matrices A_l and A_s are linearly combined to obtain the final adjacency matrix $A \in \mathbb{R}^{N\times N}$:

$$A = \alpha \cdot A_l + (1-\alpha) \cdot A_s. \tag{13}$$

Here, $\alpha \in [0,1]$ balances the contributions of A_l and A_s. The weighted combination allows the adjacency matrix to reflect feature-derived relationships while retaining information about the electrode spatial topology. An L1 penalty is applied to the fused adjacency matrix A to encourage sparsity and limit redundant connections, with the aim of limiting computational complexity and mitigating overfitting.

$$\mathcal{L}_s = \|A\|_1. \tag{14}$$

Here, $\|\quad\|_1$ denotes the L1 norm, which encourages a sparse adjacency matrix A.

2.2 Soft-sorted dynamic channel aggregation

2.2.1 Soft sorting matrix construction

EEG models based on static partitions or predefined graphs have limited ability to represent functional reorganization and changing channel relationships across emotional states. A differentiable channel ordering mechanism is therefore introduced to learn channel similarities and potential coordination patterns, providing an adaptive structure for feature aggregation. Specifically, a channel similarity matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$ is defined using Euclidean distance:

$$s_{i,j} = e^{-\frac{{||x_i - x_j||_2}^2}{2\sigma^2}} + \delta_{i,j}. \tag{15}$$

Here, ${||x_i - x_j||_2}^2$ is the squared Euclidean distance between the feature vectors of channel nodes i and j. The bandwidth parameter σ controls the rate at which similarity decays. The Kronecker delta $\delta_{i,j}$ equals 1 when $i = j$ and 0 otherwise, retaining self-connections.

Ordering with a discrete permutation matrix is not differentiable and cannot readily be optimized through backpropagation. The Sinkhorn–Knopp algorithm is therefore used to construct a soft sorting matrix. $\mathbf{P}$ is approximately doubly stochastic, with row and column sums close to 1 and entries in $[0,1]$ representing channel assignments to ordering positions. The sorting matrix is initialized from the similarity matrix $\mathbf{S}$:

$$\mathbf{P}^{(0)} = \mathbf{S}. \tag{16}$$

The matrix $\mathbf{P}^{(0)}$ then undergoes K Sinkhorn iterations. Each iteration alternates row and column normalization to approximate a doubly stochastic matrix. At iteration k, normalization is expressed as

$${p_{i,j}}^{(k)} = \frac{{p_{i,j}}^{(k-1)}}{\sum_{s=1}^{N} {p_{i,s}}^{(k-1)}}; \tag{17}$$

$${p_{i,j}}^{(k)} = \frac{{p_{i,j}}^{(k-1)}}{\sum_{s=1}^{N} {p_{s,j}}^{(k-1)}}, \tag{18}$$

where Eq. (17) normalizes rows and Eq. (18) normalizes columns. After K iterations, the soft sorting matrix is obtained as

$$\mathbf{P} = \mathbf{P}^{(K)}. \tag{19}$$

The soft sorting matrix remains differentiable during training, allowing the ordering process and emotion recognition objective to be optimized jointly through backpropagation.

2.2.2 Dynamic brain region partitioning and aggregation

Emotion-related EEG activity can exhibit synchronization and reorganization among groups of channels. The ordered channels are therefore partitioned into dynamic regions, whose features are aggregated to support emotion discrimination. Using the learned sorting matrix **P**, the input feature matrix **X** and adjacency matrix **A** are reordered jointly:

$$\mathbf{X}' = \mathbf{PX}; \tag{20}$$

$$\mathbf{Z} = \mathbf{PAP}^{\mathbf{T}}. \tag{21}$$

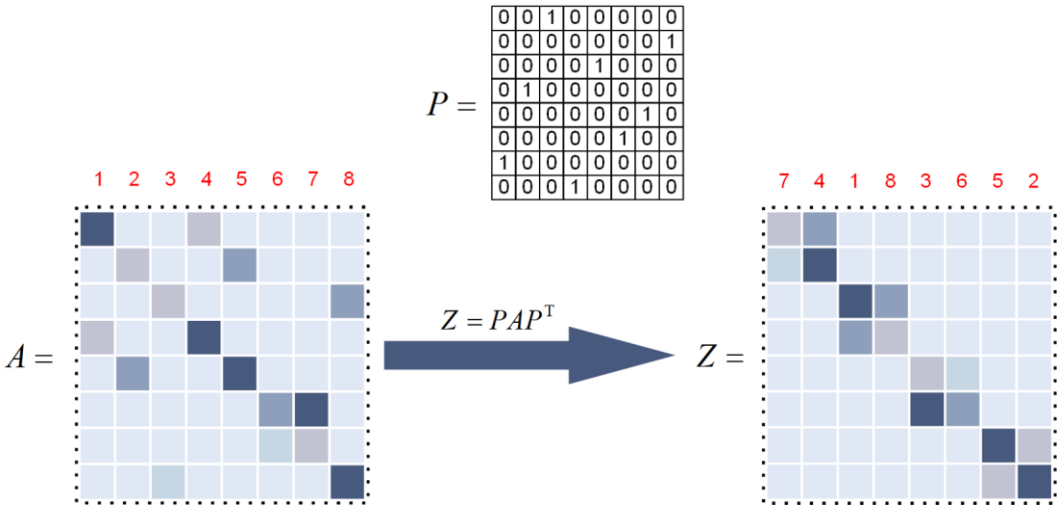


Fig. 2 Adjacency reordering illustrated with a permutation matrix P: jointly reordering the rows and columns of A produces Z; the method uses soft sorting for differentiable reordering

This transformation is intended to increase the local concentration of similar channels in the reordered matrix, facilitating block structure in subsequent aggregation.

A balance constraint is introduced when assigning N channels to R dynamic regions, encouraging similar region sizes to improve aggregation robustness and reduce sensitivity to individual differences. This constraint limits dependence on a small subset of channels and supports the modeling of patterns shared across subjects. Specifically, the regional channel counts $\{c_1, c_2, \ldots, c_R\}$ are determined by defining the count c_j for region j as follows:

$$c_j = \begin{cases} q+1, j > R-r \\ q, j \le R-r \end{cases}. \tag{22}$$

Here, $q = \left\lfloor \frac{N}{R} \right\rfloor$, and r is the remainder. The region channel counts define a mapping matrix $\mathbf{S_p} \in \mathbb{R}^{R \times N}$, whose entries $s_{j,i}$ are given by

$$s_{j,i} = \begin{cases} \frac{1}{c_j}, \sum_{n=1}^{j-1} c_n < i \le \sum_{n=1}^{j} c_n \\ 0, i \le \sum_{n=1}^{j-1} c_n \cup i > \sum_{n=1}^{j} c_n \end{cases}. \tag{23}$$

The matrix **Z** is then multiplied by the transpose of $\mathbf{S_p}$ to obtain channel-to-region relationships:

$$\mathbf{Z}' = \mathbf{Z}\mathbf{S_p}^{T}. \tag{24}$$

Here, $Z' \in \mathbb{R}^{N \times R}$ represents relationships between N channels and R regions. Channel features are linearly transformed and aggregated within each region:

$$\mathbf{Y} = \mathrm{ReLU}(\mathbf{Z}'\mathbf{X}'\mathbf{W} + b). \tag{25}$$

The aggregated region features are $\mathbf{Y} \in \mathbb{R}^{R \times D}$, and **W** and b are the trainable weight matrix and bias, respectively. This aggregation operation is formally similar to a simple graph convolution.

For each temporal segment t, SDCA adaptively generates a separate soft sorting matrix P_t and extracts a time-varying regional representation $Y_t \in \mathbb{R}^{R\times D}$, where t = 1, …, T and T = 4.

2.2.3 Block-structured regularization

A region aggregation loss encourages the reordered adjacency matrix **Z** to approach a block-diagonal structure, with strong within-region connections and weak between-region connections, supporting interpretable partitions. The total entry weight of the reordered matrix **Z** is first computed as

$$W_Z = \sum_i^N \sum_j^N z_{i,j} + \varepsilon. \tag{26}$$

Here, $z_{i,j}$ is the entry of **Z** at row i and column j, while ε ensures numerical stability when W_Z is used as a denominator. A block-diagonal mask **B** then distinguishes within-region from between-region connections, with **B** constructed from the region partition as follows:

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_{c_1\times c_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{c_2\times c_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{c_R\times c_R} \end{bmatrix}. \tag{27}$$

Here, $\mathbf{I}_{c_i\times c_i}$ is an all-ones matrix of size $c_i \times c_i$, and $\mathbf{0}$ denotes appropriately sized zero matrices, with $N = \sum_{i=1}^R c_i$. The region aggregation loss $\mathcal{L}_p$ is defined as

$$\mathcal{L}_p = \left(1 - \frac{\sum_i^N \sum_j^N (\mathbf{Z}\odot\mathbf{B})_{i,j}}{W_Z}\right)^2 + \left(\frac{\sum_i^N \sum_j^N (\mathbf{Z}\odot(1-\mathbf{B}))_{i,j}}{W_Z}\right)^2. \tag{28}$$

The region aggregation loss $\mathcal{L}_p$ contains two squared terms that constrain within-region and between-region connections, respectively. The first measures the deviation of the within-region weight fraction from its target value of 1. Minimizing this squared deviation increases the relative contribution of within-region connections. The second constrains the between-region weight fraction to approach 0. Minimizing this term suppresses between-region connections and encourages a clearer block-diagonal adjacency structure.

Jointly minimizing these terms encourages within-region connections and reduces between-region connections, producing a block-structured adjacency matrix.

Algorithm 1 Dynamic channel aggregation based on soft sorting

Soft-sorting-based dynamic channel aggregation
Input: node features $\mathbf{X} \in \mathbb{R}^{N\times D}$, adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, number of regions R, number of iterations K.
Output: aggregated region features $\mathbf{Y} \in \mathbb{R}^{R\times D}$.
Step 1: Construct the matrix $\mathbf{P}$.
Compute $\mathbf{S}$ using Eq. (15).
$\mathbf{P}^{(0)} \leftarrow \mathbf{S}$; initialize $\mathbf{P}$.
for $k = 1$ to K do
  Compute $\mathbf{P}^{(k)}$ using Eq. (17) for row normalization.
  Compute $\mathbf{P}^{(k)}$ using Eq. (18) for column normalization.
end for
$\mathbf{P} \leftarrow \mathbf{P}^{(K)}$; final soft sorting matrix.
Step 2: Reorder $\mathbf{X}$ and $\mathbf{A}$.
$\mathbf{X}' \leftarrow \mathbf{PX}$; $\mathbf{Z} \leftarrow \mathbf{PAP}^T$.
Step 3: Allocate channels to regions.
$q \leftarrow \left\lfloor \frac{N}{R} \right\rfloor$; $r \leftarrow N \bmod R$.
Compute $c \leftarrow \{c_1, c_2, \ldots, c_R\}$ using Eq. (22).
Step 4: Construct the region pooling matrix.
Initialize the pooling matrix $\mathbf{S_p} \in \mathbb{R}^{R\times N}$ to zero.
for each region j do
  if (node $i \in$ region j) then
    $s_{i,j} \leftarrow \frac{1}{c_j}$;
  end if
end for
$\mathbf{Z}' \leftarrow \mathbf{Z}{\mathbf{S_p}}^T$;
Step 5: Aggregate region features.
Compute Y using Eq. (25).
return $\mathbf{Y}$.

2.3 Brain region transformer

After SDCA aggregation and feature transformation, the EEG channels are represented by regional features $Y \in \mathbb{R}^{R\times D}$. Transformer multihead self-attention is introduced to learn dependencies between regions at this coarser spatial scale.

The regional feature matrix Y is linearly projected into the transformer hidden space to obtain the initial representation $H^{(0)} \in \mathbb{R}^{R\times d_m}$:

$$H^{(0)} = YW_{\text{in}}. \tag{29}$$

Here, $W_{\text{in}} \in \mathbb{R}^{D\times d_m}$ is the input projection matrix. Following the dense-network design of Brain Network Transformer [13], standard scaled dot-product attention is used after local topology has been incorporated through aggregation, enabling adaptive modeling of higher-order regional interactions.

For attention head m, the query, key, and value matrices are computed as follows:

$$Q^m = H^{(0)}W_Q^m,\ K^m = H^{(0)}W_K^m,\ V^m = H^{(0)}W_V^m; \tag{30}$$

$$h_m = \text{Softmax}\left(\frac{Q^m(K^m)^T}{\sqrt{d_k}}\right)V^m. \tag{31}$$

The learnable projection weights are $W_Q^m, W_K^m, W_V^m \in \mathbb{R}^{d_m\times d_k}$, with $d_k = d_m/M$, where M denotes the number of attention heads. The head outputs are concatenated and processed through linear projection, a residual connection, and layer normalization:

$$H_{\text{out}} = \text{LayerNorm}\big(H^{(0)} + [h_1 \parallel h_2 \parallel \cdots \parallel h_M]W_O\big). \tag{32}$$

Here, $W_O \in \mathbb{R}^{d_m\times d_m}$ is the output projection matrix. Given the updated regional representation H_{out}, an attention readout estimates the relative contribution of each region and aggregates them into a whole-brain feature vector z:

$$\alpha_r = \frac{\exp(H_{\text{out},r}w_r)}{\sum_{j=1}^{R}\exp(H_{\text{out},j}w_r)}, r = 1,2,\ldots,R; \tag{33}$$

$$z = \sum_{r=1}^{R}\alpha_r\, H_{\text{out},r} \in \mathbb{R}^{1\times d_m}. \tag{34}$$

Here, $w_r \in \mathbb{R}^{d_m}$ is a learnable readout vector, $H_{\text{out},r}$ is the output of region r, and α_r is its task-specific attention weight.

Whole-brain vectors are computed for the T temporal segments (T = 4), forming a brain-state sequence $[z_1, z_2, \ldots, z_T]$; temporal attention then weights the relevant time windows to produce the final emotion representation z_{final}:

$$\tau_t = \frac{\exp(z_t w_t)}{\sum_{k=1}^{T} \exp(z_k w_t)}; \tag{35}$$

$$z_{\text{final}} = \sum_{t=1}^{T} \tau_t \, z_t. \tag{36}$$

Here, $w_s, w_t \in \mathbb{R}^{d_m}$ denote the spatial and temporal readout vectors, respectively, and τ_t measures the contribution of time window t to the emotion decision. A classifier then produces the predicted label:

$$\hat{y} = \text{Softmax}(z_{\text{final}} W_c + b_c). \tag{37}$$

Here, $W_c \in \mathbb{R}^{d_m \times C}$ and $b_c \in \mathbb{R}^{C}$ are the classification parameters, and C is the number of emotion classes.

2.4 Datasets and preprocessing

TFDCA was evaluated on two publicly available datasets for EEG emotion recognition:

DEAP [24] contains EEG recordings and emotional self-ratings from 32 participants (16 men and 16 women; mean age, 26.9 years). EEG was recorded at 512 Hz from 32 electrodes arranged according to the 10–20 system. Each participant watched 40 one-minute video clips while EEG was recorded simultaneously. The released preprocessed recordings were downsampled to 128 Hz, corrected for ocular artifacts, and bandpass filtered at 4–45 Hz. DEAP also provides self-ratings of valence, arousal, dominance, and liking on scales from 1 to 9.

The 3-s baseline was removed, and only valence and arousal ratings were used as emotion labels. A threshold of 5 divided the ratings into high valence (HV), low valence (LV), high arousal (HA), and low arousal (LA), yielding separate DEAP valence (DEAP-V) and arousal (DEAP-A) tasks. Because the preprocessed DEAP signals exclude the δ band (0.5–4.0 Hz), DE features were computed only for the θ, α, β, and γ bands.

SEED [25] includes 15 Chinese participants (7 men and 8 women; mean age, 23.3 years), whose EEG was recorded at 1000 Hz using 62 electrodes arranged according to the 10–20 system. Participants watched 15 selected film clips representing positive, neutral, and negative emotions, each lasting approximately 4 min. Each clip was preceded by a 5-s cue and followed by 45 s of self-assessment and 15 s of rest to support emotion elicitation and recording. Each participant completed three sessions with 15 trials per session, giving 45 trials in total. The recordings were downsampled to 200 Hz and bandpass filtered at 0–75 Hz to reduce noise and artifacts.

Although SEED contains three repeated sessions, only the 15 trials from the first session were used. The DE features supplied with SEED, which were processed using a linear dynamical system (LDS), were used directly as frequency-domain inputs. Dataset characteristics are summarized in Table 1.

Table 1 Dataset characteristics

Dataset	SEED	DEAP-V	DEAP-A
Labels	Positive, neutral, negative	High/low valence	High/low arousal
Data count	15×3394	32×2400	32×2400
Channels	62	32	32
Sampling rate	200 Hz	128 Hz	128 Hz
Frequency bands	$\delta, \theta, \alpha, \beta, \gamma$	$\theta, \alpha, \beta, \gamma$	$\theta, \alpha, \beta, \gamma$
Participants	15	32	32
Trials per session	15	40	40

2.5 Evaluation protocols

Subject-dependent and subject-independent protocols were used to evaluate the proposed model from complementary perspectives.

(1) Subject-dependent evaluation: a separate model was trained and tested for each participant to assess recognition of individual emotional patterns.

For SEED, fivefold cross-validation divided each participant's 15 trials into five groups, using 12 trials for training and three for testing in each fold. For DEAP, tenfold cross-validation divided each participant's 40 trials into ten groups, using 36 trials for training and four for testing in each fold.

(2) Subject-independent evaluation: leave-one-subject-out (LOSO) cross-validation used one participant for testing and all remaining participants for training. This protocol evaluates generalization to unseen participants, where interindividual EEG differences typically make recognition more challenging.

For SEED, all 15 trials from the held-out participant formed the test set, and 210 trials from the remaining 14 participants formed the training set. Each participant was held out once, producing 15 evaluation runs. The same LOSO strategy was applied to DEAP, using 40 trials from one participant for testing and 1240 trials from the remaining 31 participants for training, producing 32 runs.

2.6 Training objective

The TFDCA training objective contains four components. The first is the cross-entropy loss, which measures the discrepancy between predictions and class labels:

$$\mathcal{L}_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\, log(p_{i,c}). \tag{38}$$

Here, N is the number of samples, C is the number of classes, $y_{i,c}$ is the ground-truth label, and $p_{i,c}$ is the predicted probability that sample i belongs to class c. The second component is weight decay, which penalizes parameter magnitude to limit model complexity:

$$\mathcal{L}_{w} = \|W\|_2{}^2. \tag{39}$$

Here, W denotes model parameters, and $\|\quad\|_2$ is the L2 norm. The complete objective $\mathcal{L}$ is given by

$$\mathcal{L} = \mathcal{L}_{ce} + \rho\mathcal{L}_{w} + \eta\mathcal{L}_{S} + \mu\mathcal{L}_{p}. \tag{40}$$

Here, $\mathcal{L}_S$ and $\mathcal{L}_p$ are the adjacency sparsity and region aggregation losses, respectively, while ρ, η, and μ weight parameter regularization, adjacency sparsity, and region aggregation, respectively.

2.7 Experimental environment

The experiments used an Intel Core i7-13700K processor and an NVIDIA GeForce RTX 4070 graphics card. The model was implemented and evaluated in PyTorch. The hardware and software configuration is provided in Table 2.

Table 2 Experimental environment

Component	Specification
Processor (CPU)	Intel Core i7-13700K
Graphics card (GPU)	NVIDIA GeForce RTX 4070
Memory (RAM)	32 GB
Operating system	Ubuntu 16.04
Programming language	Python 3.10
Development environment	PyCharm 2023.3.3
Deep learning framework	PyTorch 2.1.0
Optimization framework	Optuna 1.4.0

2.8 Parameter settings

Table 3 Model parameters of TFDCA

Parameter	Description	Setting
Time samples	Sampling points per 1-s segment	Dataset-specific
Frequency bands	Frequency bands	Dataset-specific
N	Channels	Dataset-specific
Classes	Emotion classes	Dataset-specific
Hidden dimension	Hidden dimension	64
dropout	Dropout probability	0.5
α	Adjacency mixing coefficient	0.8
T	Temporal segments per input	4
R	Aggregated regions	SEED: 16; DEAP: 12
μ	Region aggregation loss weight	0.5
K	Sinkhorn iterations	50
σ	Soft sorting bandwidth	1

Table 3 lists the principal TFDCA hyperparameters. The parameter "Time samples" denotes the sampling-point count S per 1-s segment at the preprocessed sampling rate; T = 4 specifies four segments (4 s) per input; "Frequency bands" denotes the band count; N is the electrode count; "Classes" denotes the class count; a hidden dimension of 64 controls feature extraction and projection; a dropout rate of 0.5 helps prevent overfitting; $\alpha = 0.8$ balances learned connectivity and spatial priors; the region count depends on the dataset; $\mu = 0.5$ weights the region aggregation constraint; $K = 50$ specifies the normalization iterations; and $\sigma = 1$ sets the soft sorting bandwidth. These settings were selected through grid-based tuning to optimize model performance.

Table 4 Training parameters

Parameter	Description	Setting
batch size	Batch size	64
learning rate	Learning rate	5e-5
optimizer	Optimizer	Adam
max epochs	Maximum epochs	300
early stop patience	Early stopping patience	20
ρ	L2 regularization weight	5e-3
η	L1 regularization weight	1e-3

The training configuration is summarized in Table 4. Training used a batch size of 64, a learning rate of 5×10^{-5} to promote stable convergence, and the Adam optimizer for up to 300 epochs to allow sufficient training; early stopping terminated training after 20 consecutive epochs without validation improvement to limit overfitting and unnecessary computation; L2 parameter regularization limited model complexity and overfitting, while L1 adjacency regularization encouraged sparsity and reduced redundant connections.

3 Results

3.1 Parameter sensitivity

The number of temporal segments T determines the duration covered by an input sample. T = 4 gave the best recognition performance among the compared settings, so four consecutive 1-s segments were used and integrated through temporal attention.

The effect of the number of aggregated regions on classification accuracy was examined. Under the subject-dependent protocol, model performance was evaluated while varying the region count R. The results are reported in Table 5.

Table 5 Effect of the number of aggregated regions on subject-dependent accuracy

Regions R	SEED Accuracy (%)	DEAP-V Accuracy (%)	DEAP-A Accuracy (%)
4	89.92	87.90	89.99
6	90.27	89.15	89.84
8	91.03	89.82	90.64
10	92.76	90.79	91.12
12	92.25	**91.84**	**92.62**
14	<u>93.98</u>	<u>91.03</u>	<u>92.57</u>
16	**94.69**	90.65	91.97
18	93.33	90.84	90.88
20	92.12	89.87	90.88

Table 5 shows that accuracy generally increased and then decreased as the region count R increased across the three tasks. SEED reached its highest accuracy of 94.69% at R = 16, falling to 93.33% and 92.12% when R increased to 18 and 20; DEAP-V peaked at 91.84% for R = 12, falling to 91.03% and 90.65% at R = 14 and 16. DEAP-A reached 92.62% at R = 12 and generally declined as R increased further. Accordingly, the best region counts among the tested settings were R = 16 for SEED and R = 12 for DEAP-V and DEAP-A. These results indicate that R affects performance: too few regions may restrict local representations, whereas too many may weaken aggregation.

3.2 Results on SEED

Table 6 Performance comparison on SEED

Method	Subject-dependent Accuracy (%)	Subject-independent Accuracy (%)
SVM [23]	83.99 ± 9.92	56.73 ± 16.29
DBN [25]	86.08 ± 8.34	—
BiDANN [26]	92.38 ± 7.04	—
DGCNN [11]	90.40 ± 8.49	79.95 ± 9.02
RGNN [23]	94.24 ± 5.95	85.30 ± 6.72
CU-GCN [27]	95.70 ± 5.32	**87.10** ± 5.44
PGCN [7]	<u>96.93</u> ± 5.11	84.59 ± 8.68
EEG-SWTNS [28]	94.83 ± 7.16	80.07 ± 10.75
CCMTL [29]	86.93 ± 6.29	57.44 ± 9.35
GCD-JFSE [30]	—	78.67 ± 14.67
ST-DADGAT [31]	—	84.68 ± —
TFDCA	**96.96** ± 3.93	<u>85.31</u> ± 5.03

Values are reported as mean ± standard deviation (SD), where available; — indicates an unreported value. SD definitions follow the respective studies.

Note: The table reproduces reported three-class SEED results from the cited studies; the methods were not reimplemented under a unified protocol; the symbol "—" indicates an unreported setting or standard deviation. The 87.11% reported for GCD-JFSE is a cross-session result and is therefore excluded from the subject-dependent column. TFDCA used Session 1, whereas the CCMTL cross-subject result used all three sessions; preprocessing and data splits also differed, so cross-study comparisons are descriptive.

Table 6 compares TFDCA with other methods for emotion recognition on SEED. The comparison includes SVM, DBN, and BiDANN; EEG-SWTNS [28], which uses EEG spectral images and attention; CCMTL [29], which combines convolutional channel modulation and sequence modeling; graph-based DGCNN, RGNN, CU-GCN, and PGCN; GCD-JFSE [30], which uses graph-based feature selection and ensemble learning; and ST-DADGAT [31], which combines dynamic graph attention with domain adaptation. Bold and underlined values indicate the highest and second-highest accuracies, respectively. Under the subject-dependent protocol, TFDCA achieved 96.96% accuracy with a standard deviation of 3.93%, giving the highest accuracy and lowest reported standard deviation; the accuracy exceeded DBN,

BiDANN, EEG-SWTNS, and CCMTL by 10.88, 4.58, 2.13, and 10.03 percentage points, respectively. Under the subject-independent protocol, TFDCA achieved 85.31% accuracy, below CU-GCN at 87.10% but above EEG-SWTNS at 80.07%, GCD-JFSE at 78.67%, and ST-DADGAT at 84.68%; its standard deviation of 5.03% was the lowest reported. Relative to SVM, TFDCA improved subject-dependent and subject-independent accuracy by 12.97 and 28.58 percentage points, respectively. These results indicate competitive performance under both protocols when temporal–frequency fusion is combined with SDCA. Differences in SEED sessions, preprocessing, and training/test partitions across studies limit direct comparison.

3.3 Results on DEAP

Table 7 Performance comparison on DEAP

Method	Subject-dependent Accuracy (%)		Subject-independent Accuracy (%)	
	Valence	Arousal	Valence	Arousal
SVM	53.76 ± 19.56	55.67 ± 20.91	55.62 ± 9.44	52.66 ± 14.47
ATDD-LSTM [32]	90.91 ± 12.95	90.87 ± 11.32	—	—
DGCNN [11]	86.32 ± 6.04	83.68 ± 5.68	58.46 ± 7.85	61.65 ± 13.34
ATGRNet [33]	78.22 ± 18.33	76.46 ± 19.48	68.65 ± 9.09	68.75 ± 7.85
DG-JCA [34]	89.25 ± —	92.42 ± —	57.41 ± —	64.38 ± —
AP-CapsNet [35]	93.89 ± —	95.04 ± —	62.71 ± —	63.51 ± —
MAS-DGAT-Net [36]	94.69 ± 4.11	95.21 ± 4.52	64.76 ± 5.02	62.86 ± 5.34
MD²GRL [8]	**96.51** ± 2.89	**95.77** ± 2.35	**92.58** ± 3.80	**92.04** ± 3.96
DGAT [37]	93.55 ± 3.89	93.19 ± 3.65	84.27 ± 13.74	83.84 ± 10.20
TFDCA	92.84 ± 5.05	93.02 ± 4.76	65.13 ± 7.99	66.24 ± 7.79

Values are reported as mean ± standard deviation (SD), where available; — indicates an unreported value. SD definitions follow the respective studies.

Table 7 summarizes binary valence and arousal classification on DEAP. TFDCA achieved 92.84% and 93.02% under the subject-dependent protocol and 65.13% and 66.24% under the subject-independent protocol, respectively, supporting the utility of temporal–frequency fusion and dynamic channel aggregation; results from other studies remain descriptive because their protocols differ.

4 Discussion

The contributions of the TFDCA modules were examined through ablation and comparison with static brain region partitions. SDCA outputs were also visualized to inspect the learned channel organization.

4.1 Ablation study

An ablation study was conducted to evaluate the contribution of each TFDCA module. All ablation experiments used SEED under the subject-dependent protocol.

Table 8 Ablation results under the subject-dependent protocol on SEED

Variant	Description	Accuracy (%)
TFDCA (w/o temp)	Without temporal features	90.70 ± 8.94
TFDCA (w/o freq)	Without frequency features	89.01 ± 7.24
TFDCA (w/o fusion)	Temporal and frequency features added without adaptive fusion	91.92 ± 8.86
TFDCA (w/o SDCA)	Without SDCA	87.56 ± 8.22
TFDCA (w/o BRT)	Without BRT	92.47 ± 6.90
TFDCA	Complete model	**96.96** ± 3.93

Values are reported as mean ± standard deviation (SD).

Table 8 presents the ablation configurations and their SEED performance. Removing temporal features (w/o temp) or frequency features (w/o freq) reduced accuracy to 90.70% and 89.01%, respectively, indicating that both domains contributed to recognition. The larger reduction after removing frequency features indicates their importance for

representing EEG energy and rhythmic characteristics. Replacing adaptive fusion (w/o fusion) reduced accuracy to 91.92%, supporting its role in integrating multiscale features and improving discrimination. Removing SDCA yielded the lowest accuracy, 87.56%, indicating that modeling channel dependencies and suppressing redundant features made a substantial contribution to overall performance. Removing BRT yielded 92.47%, higher than the other ablated variants but below the complete model, supporting the additional contribution of attention-based regional interactions. All modules contributed to TFDCA performance, with the largest accuracy reduction observed when SDCA was removed.

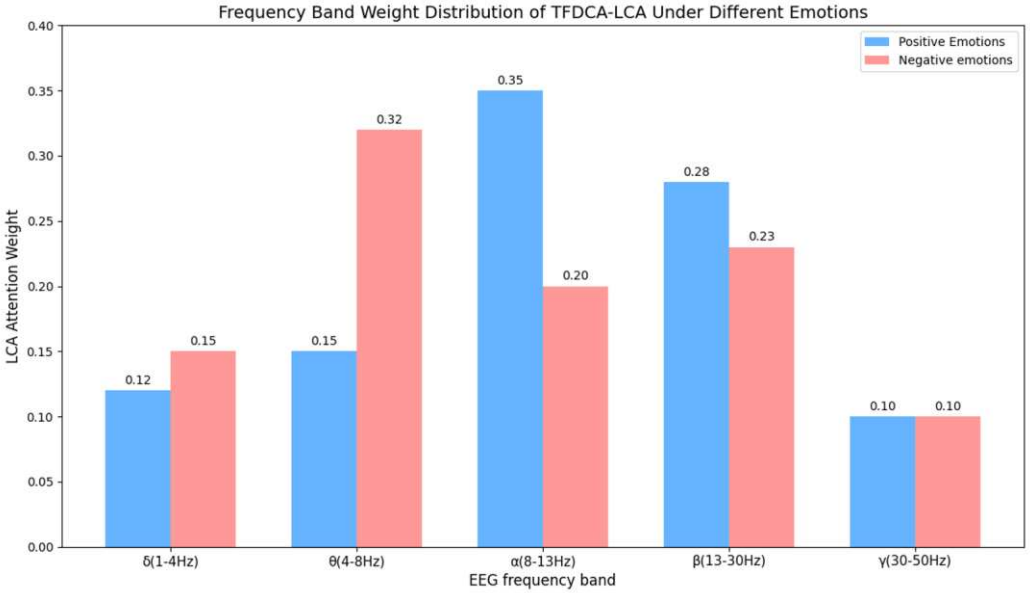


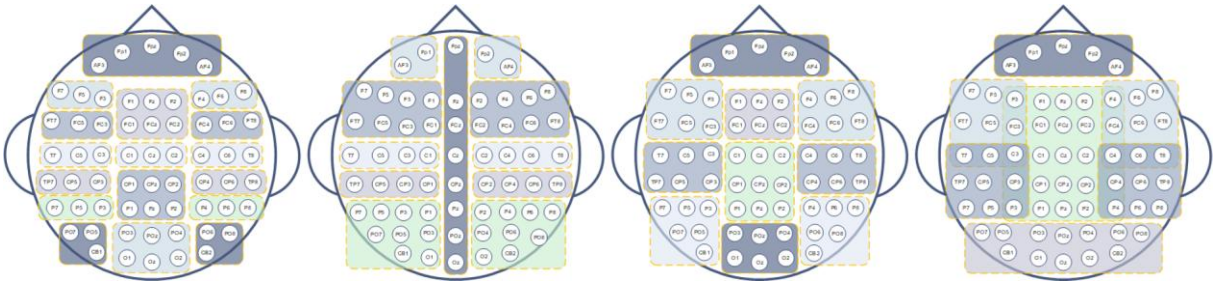
Fig. 3 Example LCA frequency-band weights for positive (blue) and negative (red) emotions

Together with the ablation results, Fig. 3 illustrates the distinct weights assigned by LCA to different frequency bands. Under positive emotion, the α band (8–13 Hz) had the highest weight, 0.35, broadly consistent with studies of emotion-related frequency activity [38]; under negative emotion, the θ band (4–8 Hz) had the highest weight, 0.32, and the δ band (1–4 Hz) received 0.15, indicating different spectral emphasis across states [39]. The β-band (13–30 Hz) weight decreased from 0.28 under positive emotion to 0.23 under negative emotion, while the γ-band (30–50 Hz) weight remained 0.10; these model weights should not be equated directly with physiological contributions [39]. These observations suggest that LCA adjusts spectral weighting to the input and supplies emotion-discriminative features for temporal–frequency fusion.

4.2 Comparison with static brain region partitions

To evaluate dynamic region partitioning, SDCA was compared with four static schemes on SEED and DEAP-V under both subject-dependent and subject-independent protocols.

Figure 4 presents four static partitions: (a) the 17-region nonoverlapping scheme of Song et al. [5], based on the 10–20 layout; (b) a custom 11-region nonoverlapping scheme preserving hemispheric separation but allowing region sizes from 2 to 9 electrodes; (c) the 10-region nonoverlapping scheme of Li et al. [19]; and (d) the 7-region overlapping scheme of Jin et al. [7].



(a) 17-region (b) 11-region (c) 10-region (d) 7-region

Fig. 4 Static brain region partitions: (a) 17 regions, Song et al. [5]; (b) 11 regions, defined here; (c) 10 regions, Li et al. [19]; (d) 7 regions, Jin et al. [7]

Table 9 Comparison of dynamic aggregation and static brain region partitions

Method	SEED Accuracy (%)		DEAP-V Accuracy (%)	
	Subject-dependent	Subject-independent	Subject-dependent	Subject-independent
TF+BRT	92.47 ± 6.90	81.94 ± 8.15	90.47 ± 4.70	63.09 ± 7.61
TF7R+BRT	93.89 ± 5.54	83.26 ± 6.62	91.36 ± 4.65	64.21 ± 7.42
TF10R+BRT	93.12 ± 6.91	82.47 ± 7.31	90.71 ± 5.08	63.45 ± 7.88
TF11R+BRT	92.97 ± 7.48	82.03 ± 7.67	90.38 ± 5.23	62.94 ± 8.10
TF17R+BRT	94.53 ± 5.72	83.95 ± 6.11	91.02 ± 4.92	64.79 ± 7.33
TFDCA	**96.96** ± 3.93	**85.31** ± 5.03	**92.84** ± 5.05	**65.13** ± 7.99

Values are reported as mean ± standard deviation (SD).

Table 9 presents the results on DEAP-V and SEED. TFDCA achieved the highest accuracy in every reported setting and lower standard deviations than the static schemes on SEED, although its DEAP-V standard deviations were not consistently the lowest. The overlapping TF7R+BRT scheme had the lowest standard deviation among the static schemes under the subject-dependent protocols on both datasets, suggesting that overlapping partitions may support stable representations. TF11R+BRT had lower accuracy in several settings, potentially related to unequal region sizes, although the cause cannot be established from these comparisons alone. The number and balance of regions therefore warrant further evaluation. Overall, soft sorting improved classification accuracy in these comparisons, whereas its stability advantage depended on the dataset and evaluation setting.

4.3 Visualization analysis

A dynamic region visualization procedure was designed to inspect how SDCA organizes channels. Soft sorting weights represent channel assignments to ordering positions, with larger values indicating stronger assignments. Displaying all weights would make the visualization excessively complex. Therefore, only the three largest entries in each row of P were retained to show the main assignments and reduce visual redundancy. Channels were then organized on the standard 10–20 scalp layout using the balanced partitioning rule in Eq. (22). SEED was divided into 16 regions and DEAP into 12 regions. The aggregated regions were numbered and mapped onto the scalp layout to display channel organization and its variation across emotional states.

The dynamic region visualizations show channel organization under different emotional states in SEED and DEAP. Each participant retained a relatively stable aggregation pattern across states, with some variation. Certain common aggregation characteristics were also observed across participants despite individual differences. Under positive emotion, aggregation involving frontal channels, such as Fp1 and Fp2, and central channels, such as C3 and C4, was prominent and can be considered alongside studies of emotion regulation [40, 41]. Under negative emotion, aggregation involving some temporal and parietal channels, such as T7 and P3, was prominent and may be interpreted in relation to emotion-processing networks [42], although model-derived groups do not themselves validate a neural mechanism. Channel allocation followed the balanced rule in Eq. (22) to limit excessive differences in region size. Functionally associated electrodes need not be spatially adjacent. In the visualized examples, an aggregated region could contain nonadjacent electrodes from both hemispheres or from frontal and occipital areas. Thus, feature-based aggregation can represent channel relationships beyond spatial proximity.

These visualizations suggest that dynamic channel aggregation adjusts channel organization according to input features, providing a data-driven representation of regional relationships. The contribution of SDCA, which combines temporal–frequency features with channel similarity, was also evident in ablation: removing it reduced SEED subject-dependent accuracy from 96.96% to 87.56%, a decrease of 9.40 percentage points, and increased the standard deviation from 3.93% to 8.22% (Table 8); relative to TF17R+BRT, the most accurate static scheme on SEED, TFDCA achieved higher accuracy and lower standard deviations under both protocols (Table 9).

After SDCA constructs regional representations, BRT further models interactions between regions. Its attention weights describe the relative contribution of regions to the current classification task. Feeding SDCA outputs into BRT connects channel organization with regional interaction modeling. Removing BRT reduced accuracy to 92.47% (Table 8), a smaller decrease than removing SDCA, indicating its complementary contribution. Temporal–frequency fusion is another important component, combining LCA spectral weighting and GRU temporal modeling into adaptively integrated features. Removing temporal or frequency information yielded accuracies of 90.70% and 89.01%, respectively; replacing

adaptive fusion with direct addition yielded 91.92%, which was 5.04 percentage points below the complete model (Table 8).

To inspect class distributions in the learned representations, t-distributed stochastic neighbor embedding (t-SNE) [43] projected high-dimensional features into two dimensions for comparison before and after model processing. Perplexity was set to 30 and the number of iterations to 1000, controlling local neighborhood modeling and the optimization procedure.

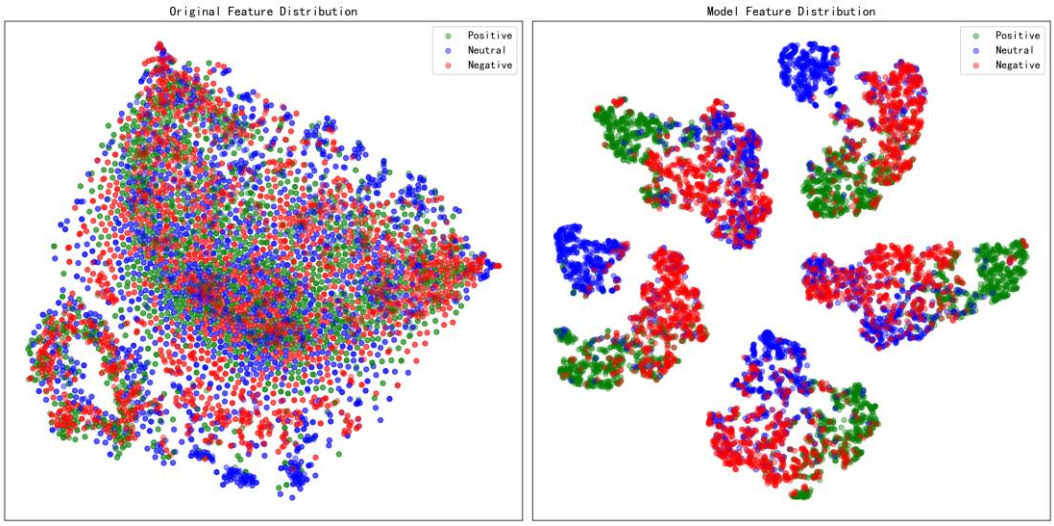


Fig. 5 t-SNE visualization for selected SEED subjects: input features (left) and model features (right); colors indicate emotion classes

Figure 5 shows t-SNE projections for selected SEED participants. The input features exhibited substantial overlap among the three emotion classes. The output features formed more concentrated within-class groups with clearer separation between classes.

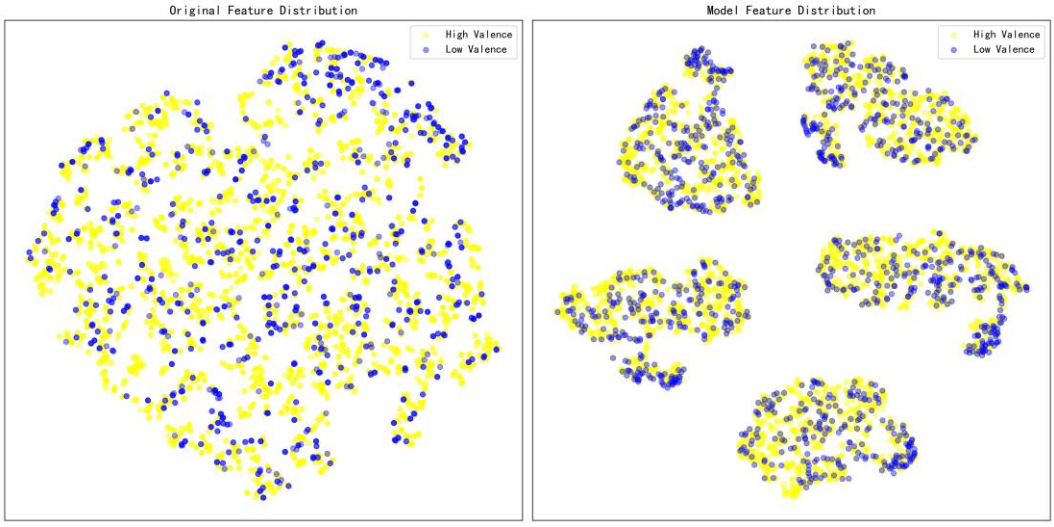
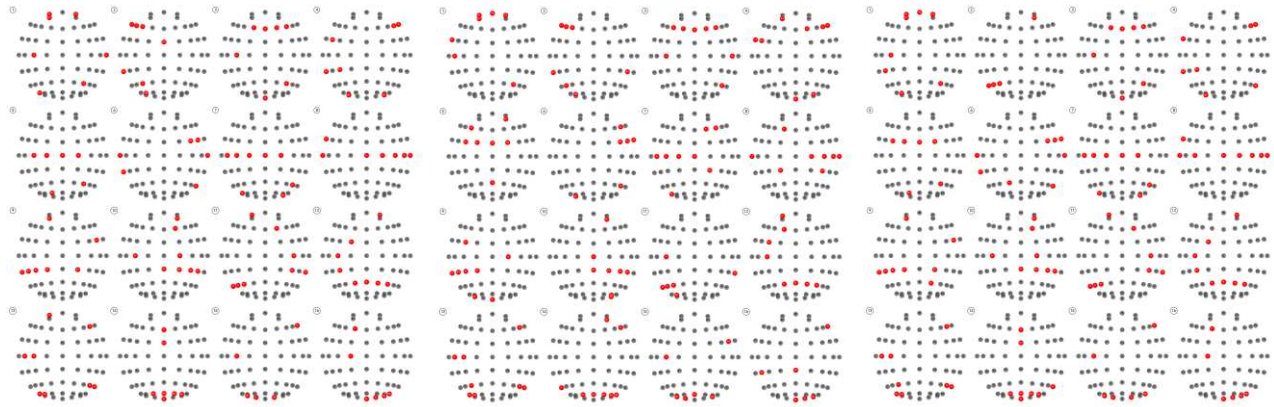


Fig. 6 t-SNE visualization for the valence task in selected DEAP subjects: input features (left) and model features (right); colors indicate valence classes

Figure 6 compares input and output feature distributions for selected DEAP participants. Input features from high and low valence were intermingled, with indistinct class boundaries. Output features showed some clustering, although separation was weaker than on SEED, with some overlap remaining in the valence task. This pattern is consistent with the lower DEAP accuracy, but visualization alone cannot attribute the difference to sampling rate.

Across the two datasets, SEED showed clearer class separation, whereas DEAP retained greater within-class variation. Differences in stimuli and experimental design may contribute: SEED uses longer film clips and DEAP shorter music videos, yielding different emotion elicitation conditions. Moreover, SEED uses three discrete emotion labels, whereas

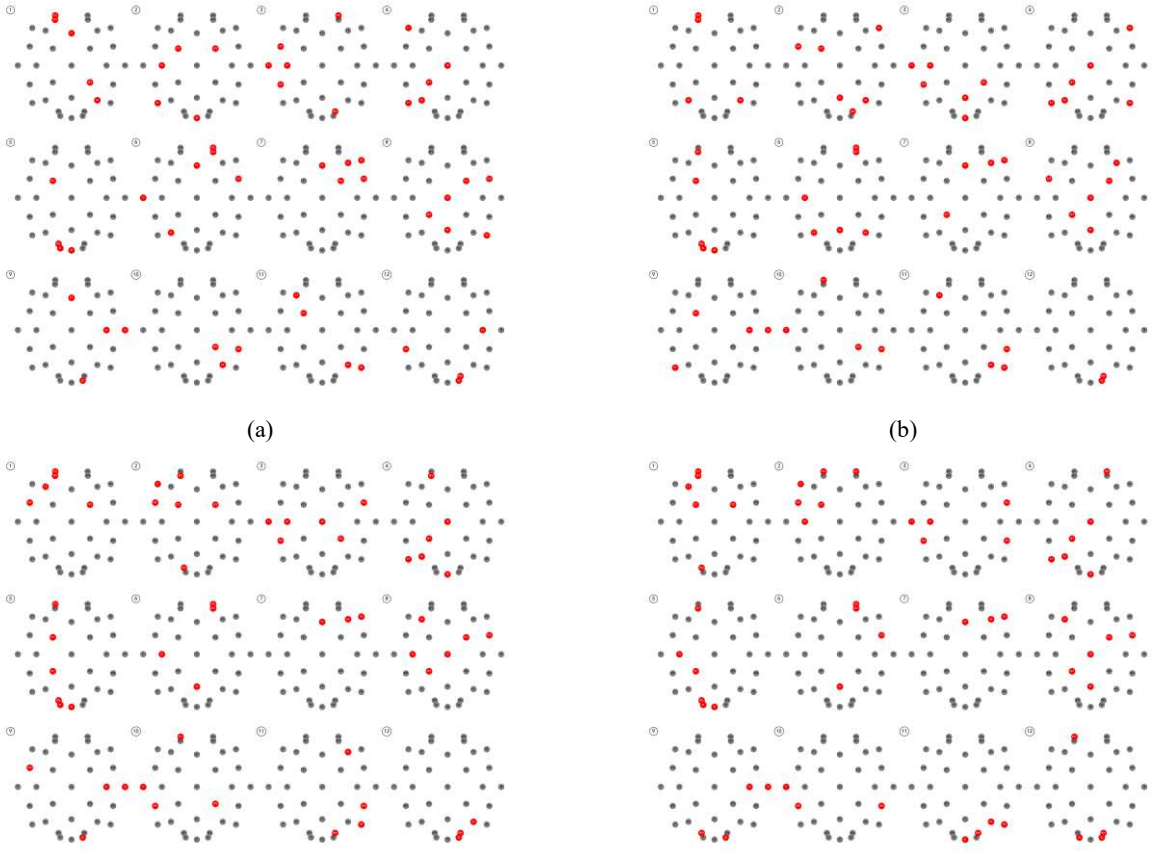
DEAP constructs binary labels from continuous self-ratings, potentially making samples near the threshold harder to distinguish.



(a) (b) (c)

Fig. 7 Dynamic brain region distributions for one SEED subject under (a) negative, (b) neutral, and (c) positive states

Figure 7(a–c) shows the dynamic region distributions of one SEED participant under negative, neutral, and positive states. The visualization shows 62 electrodes organized into 16 aggregated regions. The partitions were relatively consistent across the three states, suggesting that dynamic aggregation retained stable aspects of individual channel organization.



(a) (b) (c) (d)

Fig. 8 Dynamic brain region distributions for DEAP subject 21 under (a) low valence, (b) high valence, (c) low arousal, and (d) high arousal

Figure 8(a–d) shows the dynamic region distributions of DEAP participant 21 under low valence, high valence, low arousal, and high arousal. The partitions were relatively consistent both between high and low states and across valence and arousal tasks, suggesting a stable aggregation pattern for this participant.

Channel distributions in the SEED and DEAP visualizations generally followed the balanced partitioning rule, without excessive differences in aggregated region size. Balanced partitioning is intended to provide consistent region sizes for subsequent feature aggregation. Electrodes assigned to the same region were not always spatially adjacent. For example, some regions in Fig. 8(a) included both frontal and occipital electrodes, while some in Fig. 8(b) included nonadjacent electrodes from both hemispheres. These observations show that dynamic aggregation can represent relationships not fully described by electrode positions alone.

5 Conclusion

TFDCA is proposed for EEG emotion recognition to characterize channel interactions that are difficult to represent through static brain region partitions. Its graph input module constructs node representations from fused temporal and frequency features and combines node similarity with spatial priors to define adjacency. Dynamic aggregation jointly reorders adjacency rows and columns through soft sorting, organizing similar channels into regions while incorporating local structure and global channel relationships. BRT further models regional interactions to form representations for emotion classification. Experiments on SEED and DEAP demonstrated classification capability under subject-dependent and subject-independent protocols, although relative advantages depended on the dataset and comparison setting. Visualizations provided interpretive evidence for the learned emotion-related representations. Individual aggregation patterns remained relatively stable across emotional states while showing some variation, and certain characteristics were shared across participants. These findings support the use of dynamic channel aggregation for EEG emotion recognition, while the neurophysiological meaning of the learned relationships requires further validation.

Statements and declarations

Data availability

The data analyzed in this study were obtained from two existing datasets. The SJTU Emotion EEG Dataset (SEED) is maintained by the Brain-Computer Interface and Machine Intelligence laboratory at Shanghai Jiao Tong University and is available upon application and acceptance of its license agreement at https://bcmi.sjtu.edu.cn/home/seed/. The DEAP dataset (A Database for Emotion Analysis using Physiological Signals) is available from its dataset website at https://www.eecs.qmul.ac.uk/mmv/datasets/deap/ after completion of the provider’s access and license procedures. The original recordings are subject to the respective providers’ data-use conditions and cannot be redistributed by the authors.

Funding

This work was supported by the Key Project of Anhui Provincial Student Mental Health Education Research Program (Grant No. 25XLKT-GX004); and by the National College Students' Innovation and Entrepreneurship Training Program (Grant No. 202610364005).

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent

This study used the existing SEED and DEAP datasets, which are available to researchers under the respective data-use conditions. No new participants were recruited, and no new EEG data were collected. No additional ethics approval or informed consent was required for the present secondary analysis.

Author contributions

Y.L. conceived the study, developed the methodology and software, conducted the experiments, analyzed the results, and wrote the original draft. X.Z. contributed to methodology, software development, experiments, data analysis, and manuscript revision. H.L. revised the manuscript. X.W. contributed to software development, data processing, analysis, and validation. Y.X. conducted experiments and contributed to data processing and visualization. J.H. contributed to data processing and validation. T.Z. interpreted the results and revised the manuscript. S.Y. contributed to visualization and manuscript revision. Y.D. supervised the study, provided resources, administered the project, and revised the manuscript.

References

1. Gross JJ. Emotion regulation: current status and future prospects. Psychol Inq 2015;26(1):1–26. doi:10.1080/1047840X.2014.940781

2. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, et al. Emotion recognition in human-computer interaction. IEEE Signal Process Mag 2001;18(1):32–80. doi:10.1109/79.911197

3. Dzedzickis A, Kaklauskas A, Bucinskas V. Human Emotion Recognition: Review of Sensors and Methods. Sensors 2020;20(3):592. doi:10.3390/s20030592

4. Li J, Du X, Zhu Z, Deng X, Ma C, Wang H. Deep learning for EEG-based emotion recognition: a survey. J Softw 2023;34(1):255–276. doi:10.13328/j.cnki.jos.006420. In Chinese.

5. Song T, Liu S, Zheng W, Zong Y, Cui Z. Instance-Adaptive Graph for EEG Emotion Recognition. Proc AAAI Conf Artif Intell 2020;34(3):2701–2708. doi:10.1609/aaai.v34i03.5656

6. Gao D, Zhang H, Li P, Tang T, Liu S, Zhou Z, et al. A Local-Ascending-Global Learning Strategy for Brain-Computer Interface. Proc AAAI Conf Artif Intell 2024;38(9):10039–10047. doi:10.1609/aaai.v38i9.28867

7. Jin M, Du C, He H, Cai T, Li J. PGCN: Pyramidal Graph Convolutional Network for EEG Emotion Recognition. IEEE Trans Multimed 2024;26:9070–9082. doi:10.1109/tmm.2024.3385676

8. Tang H, Xie S, Xie X, Cui Y, Li B, Zheng D, et al. Multi-Domain Based Dynamic Graph Representation Learning for EEG Emotion Recognition. IEEE J Biomed Health Inform 2024;28(9):5227–5238. doi:10.1109/jbhi.2024.3415163

9. Ying C, Cai T, Luo S, Zheng S, Ke G, He D, et al. Do Transformers really perform badly for graph representation? Adv Neural Inf Process Syst 2021;34:28877–28888.

10. Zhou X, Tang Y, Yue Z, Liu H, Chen X, Dong Y. A multi-scale graph diffusion convolution method for EEG-based emotion recognition with temporal-frequency fusion. CCF Trans Pervasive Comput Interact 2026;8(3):520–540. doi:10.1007/s42486-026-00230-6

11. Song T, Zheng W, Song P, Cui Z. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. IEEE Trans Affect Comput 2020;11(3):532–541. doi:10.1109/taffc.2018.2817622

12. Cao M, Dong Y, Chen D, Wu G, Xu G, Zhang J. EEG-based emotion recognition model using fuzzy adjacency matrix combined with convolutional multi-head graph attention mechanism. Clust Comput 2025;28(4):274. doi:10.1007/s10586-024-04994-3

13. Kan X, Dai W, Cui H, Zhang Z, Guo Y, Yang C. Brain network transformer. Adv Neural Inf Process Syst 2022;35:25586–25599.

14. Bannadabhavi A, Lee S, Deng W, Ying R, Li X. Community-aware transformer for autism prediction in fMRI connectome. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Cham: Springer; 2023. p. 287–297. doi:10.1007/978-3-031-43993-3_28

15. Kim T, Wang G, Lee WH. BN-BrainTF: Brain Network Community-Aware Global–Local Transformer for EEG-Based Emotion Recognition. IEEE Access 2025;13:197967–197984. doi:10.1109/access.2025.3634141

16. Liu C, Deng Y, Wu Y, Yang R, Wang Z, Zhang L, et al. Graph neural networks in EEG-based emotion recognition: a survey. In: Data Science: Foundations and Applications. PAKDD 2026. Lecture Notes in Computer Science. Vol. 16618. Singapore: Springer; 2027. p. 581–598. doi:10.1007/978-981-92-1947-6_46

17. Salzman CD, Fusi S. Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex. Annu Rev Neurosci 2010;33(1):173–202. doi:10.1146/annurev.neuro.051508.135256

18. Song T, Liu S, Zheng W, Zong Y, Cui Z, Li Y, et al. Variational Instance-Adaptive Graph for EEG Emotion Recognition. IEEE Trans Affect Comput 2023;14(1):343–356. doi:10.1109/taffc.2021.3064940

19. Li Y, Chen J, Li F, Fu B, Wu H, Ji Y, et al. GMSS: Graph-Based Multi-Task Self-Supervised Learning for EEG Emotion Recognition. IEEE Trans Affect Comput 2023;14(3):2512–2525. doi:10.1109/taffc.2022.3170428

20. Liu H, Dong Y, Tang Y, Deng J, Ye Y, Xing Y, et al. A Hierarchical Meta-Learning Driven Adaptive Brain Region Flow Fusion Method for EEG-Based Emotion Recognition. Int J Cogn Comput Eng 2026;7:473–482. doi:10.1016/j.ijcce.2026.04.005

21. Klepl D, Wu M, He F. Graph Neural Network-Based EEG Classification: A Survey. IEEE Trans Neural Syst Rehabil Eng 2024;32:493–503. doi:10.1109/tnsre.2024.3355750

22. Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E. Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain. Cereb Cortex 2005;15(9):1332–1342. doi:10.1093/cercor/bhi016

23. Zhong P, Wang D, Miao C. EEG-Based Emotion Recognition Using Regularized Graph Neural Networks. IEEE Trans Affect Comput 2022;13(3):1290–1301. doi:10.1109/taffc.2020.2994159

24. Koelstra S, Mühl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, et al. DEAP: a database for emotion analysis using physiological signals. IEEE Trans Affect Comput 2012;3(1):18–31. doi:10.1109/t-affc.2011.15

25. Zheng WL, Lu BL. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. IEEE Trans Auton Ment Dev 2015;7(3):162–175. doi:10.1109/tamd.2015.2431497

26. Li Y, Zheng W, Cui Z, Zhang T, Zong Y. A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018. p. 1561–1567. doi:10.24963/ijcai.2018/216

27. Gao H, Wang X, Chen Z, Wu M, Cai Z, Zhao L, et al. Graph Convolutional Network With Connectivity Uncertainty for EEG-Based Emotion Recognition. IEEE J Biomed Health Inform 2024;28(10):5917–5928. doi:10.1109/jbhi.2024.3416944

28. Cai M, Chen J, Hua C, Wen G, Fu R. EEG emotion recognition using EEG-SWTNS neural network through EEG spectral image. Inf Sci 2024;680:121198. doi:10.1016/j.ins.2024.121198

29. Kang H, Choi JW, Kim BH. Convolutional channel modulator for transformer and LSTM networks in EEG-based emotion recognition. Biomed Eng Lett 2025;15(4):749–761. doi:10.1007/s13534-025-00475-7

30. Luo G, Han Y, Xie W, Tian F, Zhu L, Qian K, et al. GCD-JFSE: Graph-based class-domain knowledge joint feature selection and ensemble learning for EEG-based emotion recognition. Knowl Based Syst 2025;309:112770. doi:10.1016/j.knosys.2024.112770

31. Liu S, Wang X, Gu Z, An Y, Zhao S, Li B, et al. Cross-subject emotion recognition by EEG driven spatio-temporal hybrid network based on domain adaptation and dynamic graph attention. Biomed Signal Process Control 2025;110:108231. doi:10.1016/j.bspc.2025.108231

32. Du X, Ma C, Zhang G, Li J, Lai YK, Zhao G, et al. An Efficient LSTM Network for Emotion Recognition From Multichannel EEG Signals. IEEE Trans Affect Comput 2022;13(3):1528–1540. doi:10.1109/taffc.2020.3013711

33. Li C, Wang F, Zhao Z, Wang H, Schuller BW. Attention-Based Temporal Graph Representation Learning for EEG-Based Emotion Recognition. IEEE J Biomed Health Inform 2024;28(10):5755–5767. doi:10.1109/jbhi.2024.3395622

34. Cheng C, Liu W, Feng L, Jia Z. Dense Graph Convolutional With Joint Cross-Attention Network for Multimodal Emotion Recognition. IEEE Trans Comput Soc Syst 2024;11(5):6672–6683. doi:10.1109/tcss.2024.3412074

35. Liu S, Wang Z, An Y, Zhao J, Zhao Y, Zhang YD. EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network. Knowl Based Syst 2023;265:110372. doi:10.1016/j.knosys.2023.110372

36. Liu S, Wang X, Jiang M, An Y, Gu Z, Li B, et al. MAS-DGAT-Net: A dynamic graph attention network with multibranch feature extraction and staged fusion for EEG emotion recognition. Knowl Based Syst 2024;305:112599. doi:10.1016/j.knosys.2024.112599

37. Ding S, Wang K, Jiang W, Xu C, Bo H, Ma L, et al. DGAT: a dynamic graph attention neural network framework for EEG emotion recognition. Front Psychiatry 2025;16:1633860. doi:10.3389/fpsyt.2025.1633860

38. Aftanas LI, Varlamov AA, Pavlov SV, Makhnev VP, Reva NV. Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands. Int J Psychophysiol 2002;44(1):67–82. doi:10.1016/s0167-8760(01)00194-5

39. Alarcao SM, Fonseca MJ. Emotions Recognition Using EEG Signals: A Survey. IEEE Trans Affect Comput 2019;10(3):374–393. doi:10.1109/taffc.2017.2714671

40. Davidson RJ, Jackson DC, Kalin NH. Emotion, plasticity, context, and regulation: Perspectives from affective neuroscience. Psychol Bull 2000;126(6):890–909. doi:10.1037/0033-2909.126.6.890

41. Phan KL, Wager T, Taylor SF, Liberzon I. Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. Neuroimage 2002;16(2):331–348. doi:10.1006/nimg.2002.1087

42. Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. Neuroimage 2008;42(2):998–1031. doi:10.1016/j.neuroimage.2008.03.059

43. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–2605.