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.

SDH-LGCNet: Static–Dynamic Hybrid Spatio-Temporal Graph Learning With Factorized Spatial Priors for EEG Emotion Recognition

Xiule Wang
Anhui Agricultural University
Yindong Dong
dongyindong@ahau.edu.cn
Anhui Agricultural University
Yiming Tang
Hefei University of Technology
Yangfan Li
Anhui Agricultural University
Yubo Xing
Anhui Agricultural University
Jie He
Anhui Agricultural University
Tian Zhao
Anhui Agricultural University

Research Article

Keywords: Electroencephalogram (EEG) emotion recognition, factorized spatial priors, local-global collaborative learning, spatio-temporal graph convolution, static-dynamic hybrid graph

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11366891/v1

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

Additional Declarations: No competing interests reported.

SDH-LGCNet: Static–Dynamic Hybrid Spatio-Temporal Graph Learning With Factorized Spatial Priors for EEG Emotion Recognition

Xiule Wang[1†], Yindong Dong[1*†], Yiming Tang[2], Yangfan Li[1], Yubo Xing[1], Jie He[1], Tian Zhao[1]

[1]School of Artificial Intelligence, Anhui Agricultural University, Hefei, 230036, China.
[2]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, 230009, China.

*Corresponding author(s). E-mail(s): dongyindong@ahau.edu.cn;
†Xiule Wang and Yindong Dong contributed equally to this work and share first authorship.

Declarations

Funding and acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62576130, and in part by the Key Project of the Anhui Provincial Student Mental Health Education Research Program under Grant 25XLKT-GX004.

Author contributions

X.W. conceived the study, developed the methodology, implemented SDH-LGCNet and the core software, conducted the main and ablation experiments, analyzed and visualized the results, and wrote the original draft. Y.D. co-conceived the study and co-developed the methodology, designed the experimental and evaluation protocols, validated and interpreted the key findings, supervised the study, administered the project, contributed to the original draft, and critically revised the manuscript. Y.T. supervised the research, reviewed the overall technical approach, provided and coordinated research resources, contributed to project administration, and critically revised the manuscript. Y.L. implemented EEG preprocessing and supporting experimental software, curated the data, and validated the experimental results. Y.X. conducted part of the SEED and DEAP experiments, performed repeated experiments, and contributed to statistical analysis and validation across experimental settings. J.H. conducted part of the ablation experiments, analyzed model outputs, and prepared visualizations, including confusion matrices and attention weights. T.Z. organized the literature and comparison methods, curated and compiled experimental data, formatted the tables and manuscript, and contributed to manuscript revision and proofreading.

Declarations (continued)

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.

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.

Abstract

Electroencephalogram (EEG) signals can directly reflect neural activity and are an important source of information for affective computing. However, using only static or dynamic graphs makes it difficult to accommodate both stable sensor-space priors and input-adaptive dependencies. Electrode-level local modeling also lacks global integration and feedback guided by region and lateralization priors. We therefore propose SDH-LGCNet, a static-dynamic hybrid three-dimensional (3D) graph convolutional network with factorized prior-guided local-global collaborative learning. Within local temporal windows, the static-dynamic hybrid 3D graph convolution module (SDH-G3D) constructs a static spatio-temporal prior graph and an input-adaptive sparse dynamic graph. Through residual gating and multi-order propagation, SDH-G3D learns local spatio-temporal representations. The factorized prior-guided local-global collaborative learning module (LGCL) encodes five scalp-topographic regions and left, midline, and right lateralization positions as factorized prior tokens. Local-to-global aggregation, token context learning, and global-to-local feedback jointly integrate local details with global context. On the public Shanghai Jiao Tong University Emotion EEG Dataset (SEED) and Database for Emotion Analysis using Physiological Signals (DEAP), SDH-LGCNet achieves competitive recognition performance under both subject-dependent and subject-independent settings. Ablation experiments on SEED further validate the effectiveness of static-dynamic graph fusion, cross-time propagation, and joint modeling of region and lateralization priors. The approach provides a modeling framework for the collaborative learning of local spatio-temporal relations and global spatial information in EEG.

Keywords: Electroencephalogram (EEG) emotion recognition, factorized spatial priors, local-global collaborative learning, spatio-temporal graph convolution, static-dynamic hybrid graph

1 Introduction

Emotion recognition provides information on emotional states for human–computer interaction (Wang et al. 2024). Unlike observable cues such as facial expressions, speech, and posture, which are susceptible to voluntary modulation and environmental conditions, electroencephalogram (EEG) signals provide physiological information related to neural activity. EEG-based emotion recognition has received sustained attention with advances in public datasets and representation learning (Wang et al. 2024; Koelstra et al. 2012; Zheng et al. 2019). However, EEG signals are nonstationary and exhibit substantial inter-individual variability, while emotion-related information spans multiple electrodes and successive time positions. Jointly modeling spatial relations and temporal changes therefore remains a key issue (Jafari et al. 2023; Huang et al. 2024; Ding et al. 2024b; Zhang et al. 2024; Li et al. 2020).

Early methods combined handcrafted EEG features with shallow classifiers (Zheng et al. 2019; Jenke et al. 2014), whereas convolutional, temporal, and graph neural networks learn more discriminative representations (Jafari et al. 2023; Huang et al. 2024; Ding et al. 2024b; Klepl et al. 2024). By representing electrodes as nodes and inter-electrode relations as edges, graph neural networks directly describe the irregular channel topology (Klepl et al. 2024). Graph construction and its integration with temporal information are therefore central to EEG graph learning.

Existing methods describe electrode relations using static prior graphs, learnable graphs, and multi-graph fusion (Zhou et al. 2023; Li et al. 2025; Wu et al. 2025; Zhou et al. 2025; Song et al. 2020). Static graphs provide stable spatial references, whereas input-dependent dynamic graphs adapt to samples and temporal states. Prior studies have explored prior and data-driven fusion, graph learning with temporal modules, and dynamic cross-time graph construction (Huang et al. 2024; Ding et al. 2024b; Zhou et al. 2023; Li et al. 2025; Zhou et al. 2025; Yan et al. 2024; Cheng et al. 2024b). We focus on retaining stable prior paths within the same local spatio-temporal node domain while modulating dynamic relations according to the current window features, so that both types of connections participate in cross-time propagation.

Another issue concerns interaction between electrode-level representations and global information. Region-level and local-global models use grouping, hierarchical aggregation, or cross-level interaction (Ding et al. 2024a; Jin et al. 2024; Cheng et al. 2024a; Wang et al. 2025; Zheng et al. 2026), while hemispheric models exploit left–right spatial relations (Gong et al. 2024). Region membership and lateralization position describe distinct attributes: each electrode belongs to a scalp-topographic region and a left, midline, or right position. Encoding these attributes separately can preserve their membership constraints while enabling relations across groups and time positions to inform electrode-level representations.

We therefore propose SDH-LGCNet, comprising a static-dynamic hybrid 3D graph convolution module (SDH-G3D) and a factorized prior-guided local-global collaborative learning module (LGCL). SDH-G3D integrates static and input-adaptive relations in local spatio-temporal graphs, while LGCL combines electrode-level details with region and lateralization context through bidirectional interaction. The modules jointly form representations for emotion classification.

The main contributions are as follows:

1) We propose SDH-G3D to integrate static priors and input-dependent dynamic connections within a unified local spatio-temporal node domain. A window-specific scalar gate modulates the dynamic graph's contribution, while multi-order propagation learns inter-electrode and cross-time dependencies.

2) We propose factorized prior-guided LGCL, which separately encodes region and lateralization membership. Membership-constrained aggregation, prior context learning, and feature-wise gated feedback enable bidirectional interaction between electrode and prior representations.

3) We conduct subject-dependent and subject-independent experiments on SEED and DEAP. We use component ablation and attention visualization to examine static-dynamic fusion, cross-time propagation, and the roles of region and lateralization priors.

The remainder of this paper is organized as follows. Section 2 reviews related work, Section 3 presents the proposed method, and Section 4 describes experimental settings, results, and discussion. Section 5 concludes the paper.

2 Related Work

2.1 EEG-Based Emotion Recognition and Graph Learning

EEG emotion recognition has evolved from handcrafted features toward deep representation learning. Early studies combined differential entropy, power spectral density, band power, and asymmetry features with shallow classifiers (Wang et al. 2024; Zheng et al. 2019; Jenke et al. 2014). These methods have a clear signal-processing basis but depend on feature selection and domain expertise. Convolutional, recurrent, and temporal convolutional models learn spectral patterns and temporal dependencies more automatically (Jafari et al. 2023; Huang et al. 2024; Ding et al. 2024b), although their regular grids or predefined sequences limit direct representation of irregular electrode topologies. Graph neural networks instead use adjacency relations to propagate information among electrode nodes (Klepl et al. 2024; Song et al. 2020). Graph construction and representation scales extending beyond individual electrodes are therefore key modeling choices.

2.2 Static, Dynamic, and Spatio-Temporal Graph Modeling

Static graphs constructed from electrode distances, coordinates, or predefined relations provide stable sensor-space references (Klepl et al. 2024; Zhou et al. 2023), but cannot readily reflect changes across samples or temporal states. Dynamic approaches use learnable adjacency, feature similarity, or instance-dependent connections, encompassing both shared trainable topologies and sample- or time-varying graphs (Zhou et al. 2025; Song et al. 2020; Cheng et al. 2024b). FCMGAT smooths trainable adjacency through fuzzy processing and combines graph convolution with multi-head graph attention (Cao et al. 2025). Dual-graph and multi-graph frameworks combine spatial or cognitive priors with data-driven relations (Zhou et al. 2023; Li et al. 2025), while connection purification suppresses redundancy (Wu et al. 2025). TFF-MDCN integrates

spatial-distance priors and trainable adjacency with gated temporal-frequency features, bidirectional graph diffusion, and multi-scale attention (Zhou et al. 2026). These approaches exploit complementary relations while controlling their contributions.

Spatio-temporal methods combine electrode relations with dependencies over successive time positions. Spatio-temporal and space-aware temporal convolutions learn spatial and temporal features jointly (Huang et al. 2024; Ding et al. 2024b); graph-based models also use multi-scale Transformers (Yan et al. 2024) or cross-attention between dynamic-graph and temporal self-attention features (Cheng et al. 2024b). Spatio-temporal graph-aware networks combine graph encoding at successive slices with self-attention and sequence modules (Li and Chen 2026), while dynamic cross-time graphs incorporate different time positions directly into relation learning (Zhou et al. 2025). Although these studies advance static-dynamic fusion and temporal modeling, coordination of static spatio-temporal priors, input-dependent connections, and cross-time propagation within a unified local node domain remains limited.

2.3 Prior-Guided Local-Global EEG Representation Learning

Local-global models extend electrode representations to regional and whole-brain scales through local and global graphs in LGGNet and ALGGNet (Ding et al. 2024a; Wang et al. 2025), hierarchical aggregation in pyramidal graph networks and regional-to-global Transformers (Jin et al. 2024; Cheng et al. 2024a), and bidirectional projection with gated fusion in region–channel frameworks (Zheng et al. 2026). HMABRF combines adaptive brain-region flow partitioning with hierarchical meta-learning to coordinate local interactions and global temporal modeling (Liu et al. 2026). These methods show how regional aggregation and cross-level feedback complement electrode-level propagation.

Cognitive priors, functional-connectivity grouping, and hemispheric differences guide EEG representation learning (Li et al. 2025; Gong et al. 2024; Guo et al. 2024), while channel importance and mutual-information constraints adjust electrode and connection contributions (Yang et al. 2024, 2025). Region membership distinguishes frontal, central, temporal, parietal, and occipital scalp-topographic groups; lateralization distinguishes left, midline, and right positions. Drawing on region–channel interaction (Zheng et al. 2026) and hemispheric modeling (Gong et al. 2024), we encode these memberships separately and allow the resulting prior representations to interact before feeding context back to electrodes.

3 Method

SDH-LGCNet combines local spatio-temporal relation modeling with prior-guided global information exchange (Fig. 1). After frequency-band projection, SDH-G3D extracts electrode-level representations from static and input-adaptive relations. LGCL aggregates these representations along the region and lateralization dimensions and feeds prior-token context back to electrodes. Global average pooling and a classifier produce the emotion predictions.

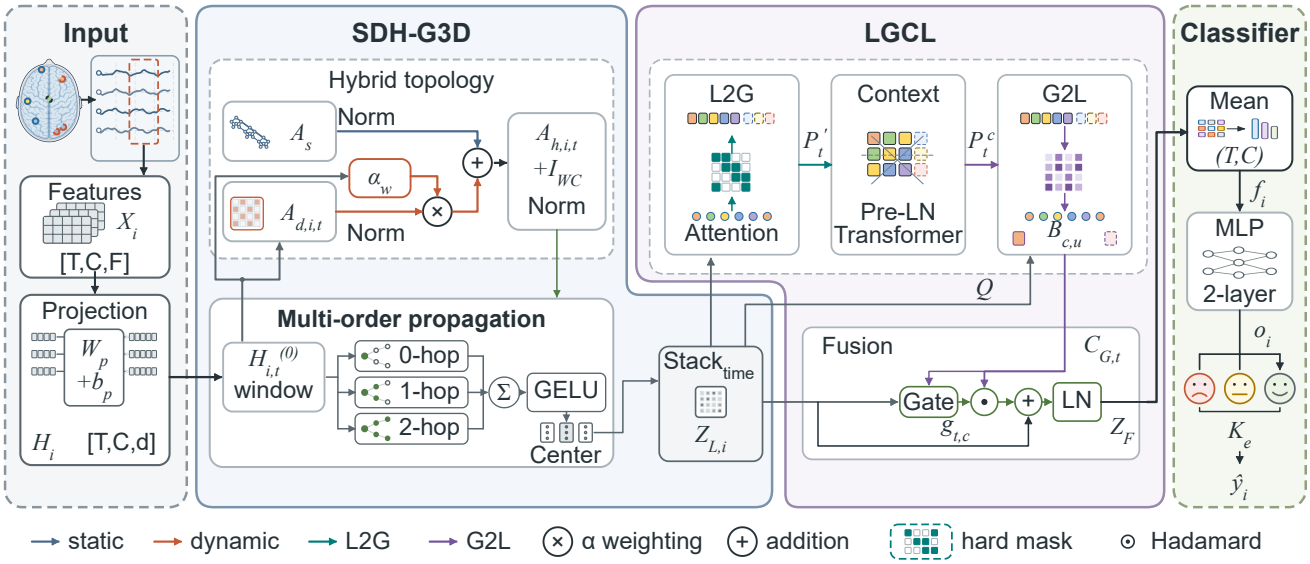


Fig. 1 Overall framework of SDH-LGCNet. SDH-G3D extracts local spatio-temporal representations, and LGCL enables local-global information exchange through region and lateralization priors; L2G, local-to-global; G2L, global-to-local; LN, layer normalization; MLP, multilayer perceptron; GELU, Gaussian error linear unit

3.1 Preliminaries

Consider an EEG emotion recognition dataset $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$, where N is the total number of samples and y_i is the emotion label of sample i. Let $X_i \in \mathbb{R}^{T\times C\times F}$ denote an input sample, where T, C, and F denote the numbers of time slices, EEG channels, and frequency-band features, respectively. A linear projection shared across all time positions and EEG channels maps the frequency-band features into a common hidden representation space:

$$H_i = X_i W_p + b_p. \tag{1}$$

Here, $W_p \in \mathbb{R}^{F\times d}$ and $b_p \in \mathbb{R}^{d}$ are the learnable projection matrix and bias vector, respectively, and d is the hidden feature dimension. The projected representation $H_i \in \mathbb{R}^{T\times C\times d}$ preserves the original temporal order and electrode-level spatial structure and serves as input to SDH-G3D.

3.2 Static-Dynamic Hybrid 3D Graph Convolution

SDH-G3D jointly models electrode relations and temporal changes within a local spatio-temporal graph. Inspired by unified spatio-temporal graph convolution (Liu et al. 2020), we pair local time positions with electrodes to form spatio-temporal nodes. A static prior graph and an input-adaptive dynamic graph are constructed over the same node domain. The two graphs are combined through gated fusion for multi-order propagation, as shown in Fig. 2. G3D refers to graph propagation performed directly in this spatio-temporal node domain.

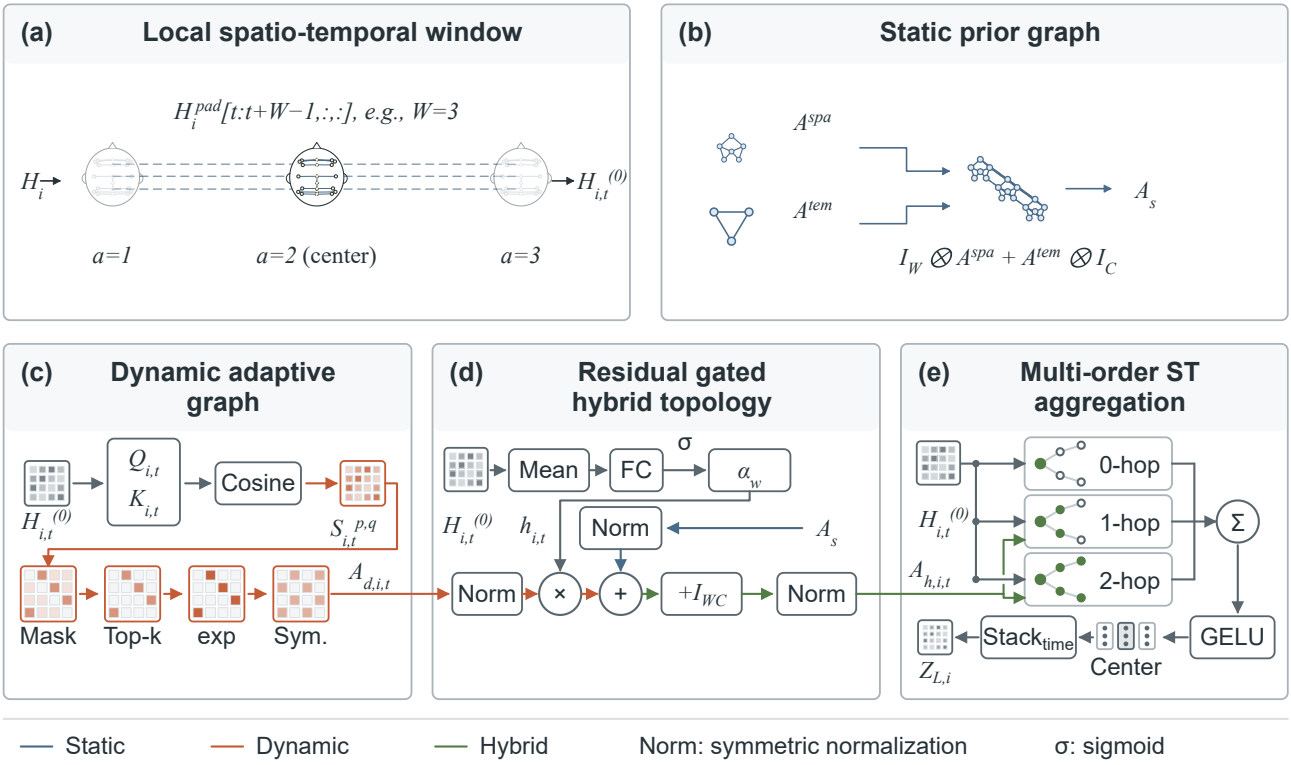


Fig. 2 Architecture of the SDH-G3D module: (a) local spatio-temporal window; (b) static prior graph; (c) dynamic adaptive graph; (d) residual gated hybrid topology; (e) multi-order spatio-temporal aggregation. Blue, orange, and green arrows denote static, dynamic, and hybrid paths, respectively. FC, fully connected; ST, spatio-temporal; Norm, symmetric normalization; Sym., symmetrization; GELU, Gaussian error linear unit; σ, sigmoid

3.2.1 Local Spatio-Temporal Node Construction

Reflection padding is applied along the temporal dimension of the projected input $H_i \in \mathbb{R}^{T\times C\times d}$ to obtain $H_i^{pad} = ReflectPad(H_i, r_w)$. Let r_w denote the window radius and $W = 2r_w+1$ the window length. For a center time position t, W consecutive positions from t to $t+W-1$ are selected from the padded input. These positions are rearranged into spatio-temporal nodes in time-major, electrode-minor order:

$$H_{i,t}^{(0)} = reshape_{time\text{-}major}\left(H_i^{pad}[t:t+W-1,:,:]\right) \tag{2}$$

where $H_{i,t}^{(0)} \in \mathbb{R}^{WC\times d}$. The interval in (2) includes both endpoints, and reflection padding does not repeat boundary values. Each time-position–electrode pair corresponds to one node, and adjacent center positions use overlapping windows. After propagation, only the C electrode nodes at the center time position are read out, preserving the original temporal length.

3.2.2 Static Spatio-Temporal Prior Graph

The static spatio-temporal prior graph provides fixed spatial and temporal propagation paths across inputs. We construct spatial adjacency from standard template-based 3D electrode coordinates that match the channel configuration (Oostenveld and Praamstra 2001; Jurcak et al. 2007). The coordinates are centered and normalized to the unit sphere before graph construction. For SEED, we use the 10–05 template, with CB1 and CB2 mapped to I1 and I2, respectively (Oostenveld and Praamstra 2001; Compumedics USA, Ltd. 2016). For DEAP, the template coordinates correspond to its 32 channels. Let $p_c \in \mathbb{R}^3$ denote the coordinates of electrode c, $\mathcal{N}_c^{spa}$ its original set of k_n nearest neighbors, and $d_{c,j} = \| p_c - p_j \|_2$. Undirected spatial edges are formed by taking the union of the nearest-neighbor relations and weighted using a Gaussian kernel:

$$A_{c,j}^{spa} = \begin{cases} exp\left(-\frac{d_{c,j}^2}{2\sigma_s^2}\right), \ j \in \mathcal{N}_c^{spa} \ or \ c \in \mathcal{N}_j^{spa}, \ c \neq j, \\ 0, \qquad\qquad otherwise, \end{cases} \tag{3}$$

Here, σ_s is the mean distance over all retained spatial edges. To represent temporal continuity within a window, let $\rho \in (0, 1]$ denote the temporal decay coefficient and τ_t the maximum range of direct temporal connections. The temporal adjacency matrix is defined as:

$$A_{a,b}^{tem} = \begin{cases} \rho^{|a-b|}, \ 0 < |a-b| \leq \tau_t, \\ 0, \qquad otherwise. \end{cases} \tag{4}$$

The spatial and temporal adjacency matrices are combined in the unified node domain to obtain A_s, which represents the static spatio-temporal prior graph:

$$A_s = I_W \otimes A^{spa} + A^{tem} \otimes I_C \tag{5}$$

Here, $\otimes$ denotes the Kronecker product, and I is the identity matrix of the corresponding dimension. The first term reuses spatial adjacency at each time position, while the second connects the same electrode across time positions.

3.2.3 Dynamic Adaptive Spatio-Temporal Graph

Fixed propagation paths may not fully capture state-dependent relations across samples and local windows. We therefore generate a sparse dynamic graph from local node features. Learnable projections produce the query and key representations, $Q_{i,t} = H_{i,t}^{(0)} W_Q$ and $K_{i,t} = H_{i,t}^{(0)} W_K$, where $W_Q, W_K \in \mathbb{R}^{d \times d_q}$. The normalized matching score between nodes p and q is:

$$S_{i,t}^{p,q} = \frac{Q_{i,t,p} K_{i,t,q}^{\top}}{\| Q_{i,t,p} \|_2 \ \| K_{i,t,q} \|_2 + \varepsilon} \tag{6}$$

Here, $\varepsilon > 0$ is a numerical stabilizer that prevents division by zero.

Let $p = (a, c)$ and $q = (b, j)$, where a, b are time positions within the window and c, j are electrode indices. At the same time position, candidate connections include all other electrodes to capture spatial relations beyond the nearest-neighbor range.

Across time positions, candidates are restricted to the same electrode and its original spatial nearest neighbors within the specified temporal range. The candidate set is:

$$\mathcal{C}_p = \left\{ q = (b, j) \;\middle|\; \begin{array}{c} a = b,\ c \neq j; \quad or \\ 0 < |a - b| \leq \tau_t,\ j \in \{c\} \cup \mathcal{N}_c^{spa} \end{array} \right\} \tag{7}$$

Within $\mathcal{C}_p$, the $k_p = min\left(k_d, |\mathcal{C}_p|\right)$ nodes with the highest matching scores are retained. Here, k_d is the predefined number of top-k neighbors in the dynamic graph, and $\mathcal{I}_{i,t,p}$ denotes the set of their indices. The directed dynamic adjacency is defined as:

$$A_{d,i,t}^{\rightarrow p,q} = \begin{cases} exp\left(S_{i,t}^{p,q}\right), & q \in \mathcal{I}_{i,t,p}, \\ 0, & otherwise. \end{cases} \tag{8}$$

Symmetrizing the directed adjacency yields the dynamic spatio-temporal adjacency matrix:

$$A_{d,i,t} = \frac{A_{d,i,t}^{\rightarrow} + \left(A_{d,i,t}^{\rightarrow}\right)^{\top}}{2} \tag{9}$$

3.2.4 Residual Gated Hybrid Topology

The static graph supplies fixed propagation paths, whereas the dynamic graph represents input-dependent node relations. Using the static graph as the base, we generate a window-specific scalar gate from the current sample's window features to modulate the dynamic graph's injection strength.

The two graphs first undergo separate symmetric normalization, $Norm(A) = D_A^{-1/2} A D_A^{-1/2}$, where D_A is the diagonal node-degree matrix of A. The normalized static and dynamic graphs are denoted by $\overline{A}_s$ and $\overline{A}_{d,i,t}$, respectively. Averaging the window's node features gives $h_{i,t} = \frac{1}{WC} \sum_{p=1}^{WC} H_{i,t,p}^{(0)}$, from which the gate is computed:

$$\alpha_{i,t} = Sigmoid\left(\mathbf{h}_{i,t} \mathbf{w}_\alpha + b_\alpha\right) \tag{10}$$

Here, $w_\alpha \in \mathbb{R}^d$ and $b_\alpha \in \mathbb{R}$ are learnable parameters. The scalar gate $\alpha_{i,t}$ is specific to the current sample and window and corresponds to α_w in Fig. 2. The weighted dynamic graph is added to the static graph, followed by self-loop addition and normalization:

$$\mathbf{A}_{h,i,t} = Norm\left(\overline{\mathbf{A}}_s + \alpha_{i,t} \overline{\mathbf{A}}_{d,i,t} + \mathbf{I}_{WC}\right) \tag{11}$$

3.2.5 Multi-Order Spatio-Temporal Graph Aggregation

Multi-order propagation over the hybrid topology aggregates information from local spatio-temporal neighborhoods of different ranges:

$$H_{i,t}^{(g)} = GELU\left(\sum_{k=0}^{K_g} A_{h,i,t}^{k} H_{i,t}^{(0)} W_k^{(g)}\right) \tag{12}$$

Here, K_g is the maximum graph propagation order, and $W_k^{(g)} \in \mathbb{R}^{d\times d}$ is the learnable propagation weight for order k. GELU denotes the Gaussian error linear unit activation.

The zeroth-order term preserves each node's own features, while higher-order terms aggregate information over different propagation ranges. The electrode representations at the center time position, $Z_{L,i,t} = Center\left(H_{i,t}^{(g)}\right)$, are then read out and stacked along the temporal axis to produce the SDH-G3D output:

$$Z_{L,i} = Stack_{time}\left[Z_{L,i,1}, Z_{L,i,2}, \ldots, Z_{L,i,T}\right] \tag{13}$$

The resulting $Z_{L,i} \in \mathbb{R}^{T\times C\times d}$ serves as the electrode-level local spatio-temporal representation for LGCL.

3.3 Factorized Prior-Guided Local-Global Collaborative Learning

LGCL aggregates electrode representations into region and lateralization prior tokens, learns inter-group and cross-time context, and feeds this context back to electrodes (Fig. 3). Membership-constrained local-to-global (L2G) aggregation preserves the spatial grouping of tokens, while global-to-local (G2L) feedback integrates their context with local representations. The sample index i is omitted below.

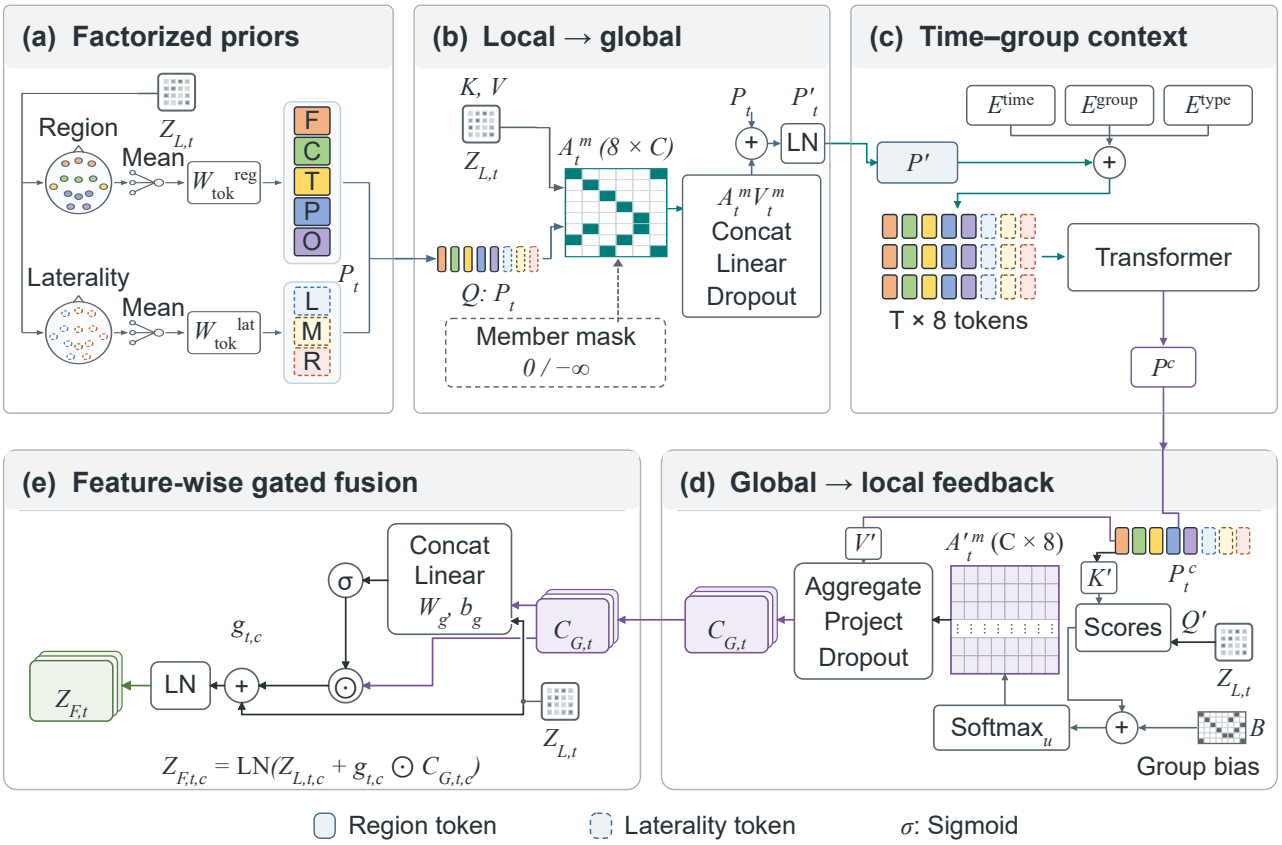


Fig. 3 Architecture of the factorized prior-guided LGCL module: (a) factorized region and lateralization priors; (b) local-to-global aggregation; (c) time-group context learning; (d) global-to-local feedback; (e) feature-wise gated fusion. Solid and dashed token borders denote region and lateralization tokens, respectively. F, frontal; C, central; T, temporal; P, parietal; O, occipital; L, left; M, midline; R, right; LN, layer normalization; Concat, concatenation; σ, sigmoid

3.3.1 Factorized Prior Grouping and Token Initialization

Following electrode naming and scalp-position conventions (Oostenveld and Praamstra 2001; Klem et al. 1999), we encode region membership and lateralization positions as two separate sets of prior tokens, forming factorized spatial priors. Each electrode belongs to one region group and one lateralization group. In Fig. 4, fill colors indicate the five scalp-topographic regions: Frontal, Central, Temporal, Parietal, and Occipital. Dashed borders indicate the three lateralization positions: Left, Midline, and Right.

Let $M_r \in \{0,1\}^{C\times 5}$ and $M_l \in \{0,1\}^{C\times 3}$ denote the region and lateralization membership matrices, respectively, each with one nonzero entry per row. Concatenating them along the group axis gives $M = [M_r, M_l] \in \{0,1\}^{C\times U}$, where $U = 8$. The region tokens, lateralization tokens, and their concatenation along the group axis are denoted by P_r, P_{lat}, $P = [P_r; P_{lat}]$. For token u, $s_u \in \{reg, lat\}$ denotes the prior type and

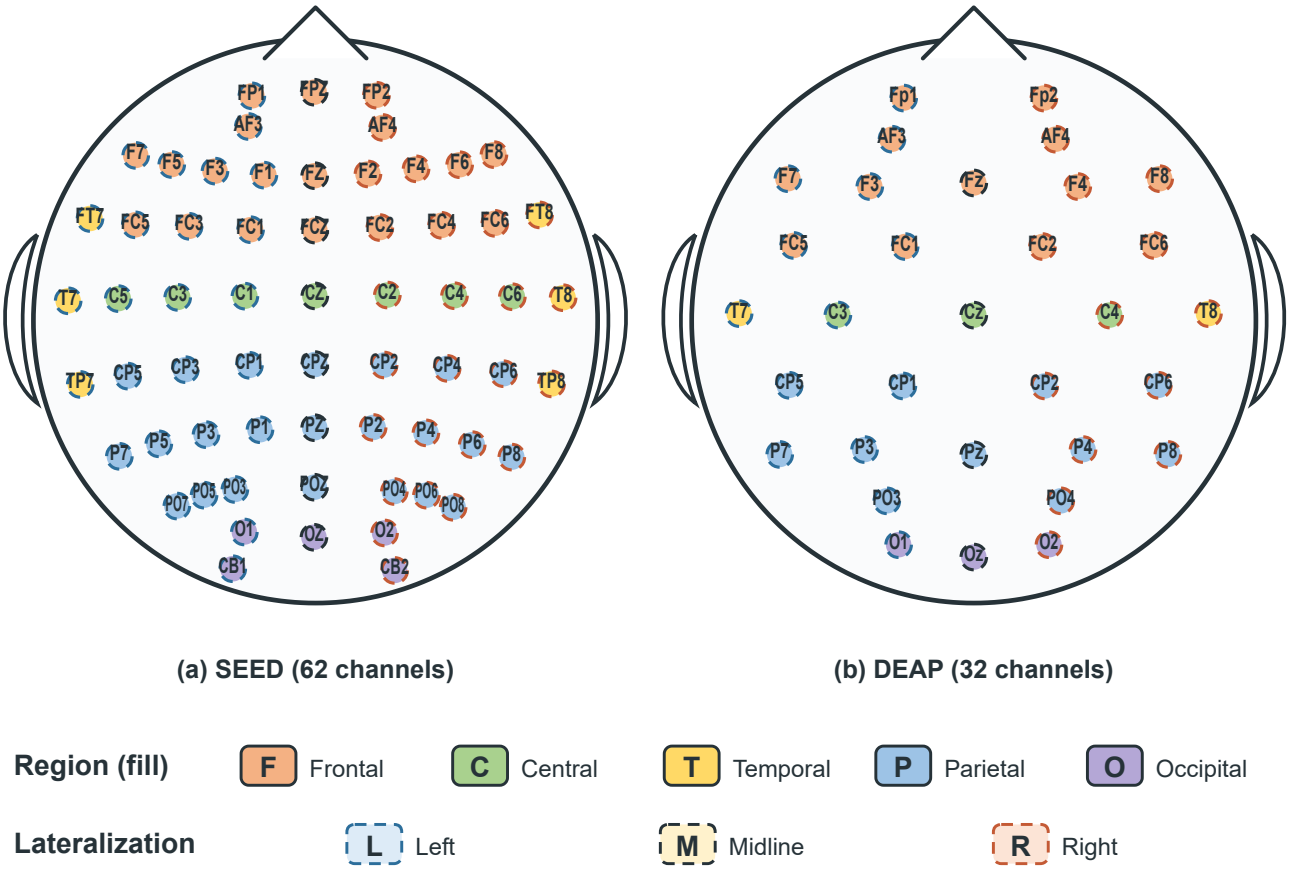


Fig. 4 Electrode grouping by region and lateralization position: (a) SEED (62 channels); (b) DEAP (32 channels). Fill colors denote frontal, central, temporal, parietal, and occipital regions. Colored dashed borders denote left, midline, and right lateralization positions, as identified in the legend

$\mathcal{C}_u = \{c \mid M_{c,u} = 1\}$ the set of member electrodes. Each token type uses its own projection matrix. The initial tokens are obtained by averaging the member electrode representations and projecting the averages:

$$\mathbf{P}_{t,u} = \frac{1}{|\mathcal{C}_u|} \sum_{c \in \mathcal{C}_u} \mathbf{Z}_{L,t,c} \mathbf{W}_{tok}^{s_u} \tag{14}$$

Here, $W_{tok}^{s_u} \in \mathbb{R}^{d \times d}$ projects member electrode features into a common token space. The token set at a single time position is $P_t \in \mathbb{R}^{U \times d}$, and stacking these sets along the temporal dimension gives $P \in \mathbb{R}^{T \times U \times d}$.

3.3.2 Local-to-Global Cross-Attention

Let M_h denote the number of attention heads in each of the L2G and G2L cross-attention operations, with head index $m = 1, \ldots, M_h$ and per-head feature dimension d_h.

L2G uses prior tokens as queries and electrode representations as keys and values. A membership mask restricts aggregation to member electrodes, preserving the spatial grouping of each token. For attention head m, let $Q_t^m = P_t W_{Q,L2G}^m$, $K_t^m = Z_{L,t} W_{K,L2G}^m$, and $V_t^m = Z_{L,t} W_{V,L2G}^m$. The mask $M_{u,c}^{mem}$ is 0 when $M_{c,u} = 1$ and is $-\infty$ otherwise. The attention weights are:

$$A_{t,u,c}^m = Softmax_c \left(\frac{\mathbf{Q}_{t,u}^m \left(\mathbf{K}_{t,c}^m\right)^\top}{\sqrt{d_h}} + M_{u,c}^{mem} \right) \tag{15}$$

The aggregated outputs from all attention heads are concatenated and projected. Dropout, a residual connection, and layer normalization then update the prior tokens:

$$\begin{aligned} \mathbf{P}'_t = \mathrm{LayerNorm}\Big(\mathbf{P}_t + \mathrm{Dropout}\Big(\\ \mathrm{Concat}_{m=1}^{M_h}\big(\mathbf{A}_t^m \mathbf{V}_t^m\big)\mathbf{W}_{o,L2G}\Big)\Big) \end{aligned} \tag{16}$$

For local-to-global cross-attention, the projection matrices satisfy $W_{Q,L2G}^m,\ W_{K,L2G}^m,\ W_{V,L2G}^m \in \mathbb{R}^{d \times d_h}$, and the output projection is $W_{o,L2G} \in \mathbb{R}^{M_h d_h \times d}$. Thus, ${P'}_t \in \mathbb{R}^{U \times d},\ P' \in \mathbb{R}^{T \times U \times d}$.

3.3.3 Prior-Token Spatio-Temporal Context Learning

After L2G aggregates information within each group, a Transformer encoder learns relations across prior groups and time positions.

Temporal embeddings E^{time}, group embeddings E^{group}, and type embeddings E^{type} are added to P' to encode time positions, specific groups, and region or lateralization types, respectively. The tokens are then flattened in time-major, group-minor order and passed through a non-causal pre-layer-normalized (Pre-LN) Transformer encoder (Vaswani et al. 2017; Xiong et al. 2020). The temporal and group axes are subsequently restored:

$$\begin{aligned} \mathbf{P}^c = \mathrm{Reshape}\Big(\mathrm{Encoder}\Big(\\ \mathrm{Flatten}_{t,u}\big(\mathbf{P}' + \mathbf{E}^{time} + \mathbf{E}^{group} + \mathbf{E}^{type}\big)\Big)\Big) \end{aligned} \tag{17}$$

After indexing and broadcasting, $E^{time},\ E^{group}$ and E^{type} have the same effective shape $\mathbb{R}^{T \times U \times d}$ as P'. The flattened token sequence has shape $\mathbb{R}^{TU \times d}$, and encoding followed by restoration of the temporal and group axes yields $P^c \in \mathbb{R}^{T \times U \times d}$.

3.3.4 Global-to-Local Feedback and Gated Fusion

G2L uses electrode representations as queries and contextualized tokens at the same time position as keys and values. These tokens have already undergone cross-time encoding, so the feedback carries spatio-temporal context from the input segment. For

attention head m, let $Q_t^{'m} = Z_{L,t}W_{Q,G2L}^m$, $K_t^{'m} = P_t^c W_{K,G2L}^m$, and $V_t^{'m} = P_t^c W_{V,G2L}^m$. Each electrode can attend to all prior tokens. Positive biases increase its attention scores for tokens corresponding to its own region and lateralization groups. The two bias strengths are $\beta_{reg} = \beta_{\max} Sigmoid\left(\widetilde{\beta}_{reg}\right)$ and $\beta_{lat} = \beta_{\max} Sigmoid\left(\widetilde{\beta}_{lat}\right)$, respectively, and the dual membership bias is:

$$B_{c,u} = \begin{cases} \beta_{reg} M_{c,u}, & u \in \mathcal{U}_{reg}, \\ \beta_{lat} M_{c,u}, & u \in \mathcal{U}_{lat}. \end{cases} \tag{18}$$

Here, $\mathcal{U}_{reg} = \{1, \ldots, 5\}$, $\mathcal{U}_{lat} = \{6, 7, 8\}$ are the index sets of region and lateralization tokens, respectively; $\beta_{\max} > 0$ is the maximum bias scale, and $\widetilde{\beta}_{reg}$ and $\widetilde{\beta}_{lat}$ are learnable parameters. Nonmember tokens have zero bias but still participate in attention. Thus, L2G aggregates by membership, whereas G2L provides cross-group feedback while retaining membership cues.

With the dual membership bias, the feedback attention weights are:

$$A_{t,c,u}^{'m} = Softmax_u\left(\frac{Q_{t,c}^{'m}\left(K_{t,u}^{'m}\right)^\top}{\sqrt{d_h}} + B_{c,u}\right) \tag{19}$$

The head outputs are concatenated, linearly projected, and passed through dropout to obtain a prior-guided global context message:

$$C_{G,t} = Dropout\left(Concat_{m=1}^{M_h}\left(A_t^{'m} V_t^{'m}\right) W_{o,G2L}\right) \tag{20}$$

For global-to-local cross-attention, the projection matrices satisfy $W_{Q,G2L}^m$, $W_{K,G2L}^m$, $W_{V,G2L}^m \in \mathbb{R}^{d \times d_h}$, and the output projection is $W_{o,G2L} \in \mathbb{R}^{M_h d_h \times d}$, giving the global context message $C_{G,t} \in \mathbb{R}^{C \times d}$.

A feature-wise gate is generated from the local representation and feedback message to modulate feedback strength separately for each feature dimension at each electrode:

$$g_{t,c} = Sigmoid\left(\left[Z_{L,t,c} \parallel C_{G,t,c}\right] W_g + b_g\right) \tag{21}$$

Here, $W_g \in \mathbb{R}^{2d \times d}$, $b_g \in \mathbb{R}^d$, $g_{t,c} \in \mathbb{R}^d$. The gate controls the injection strength of the global context message elementwise along the feature dimension.

The gated feedback message is added to the local representation and layer-normalized to obtain the fused representation:

$$Z_{F,t,c} = LayerNorm\left(Z_{L,t,c} + g_{t,c} \odot C_{G,t,c}\right) \tag{22}$$

The resulting fused representation $Z_F \in \mathbb{R}^{T \times C \times d}$ is used for subsequent emotion classification.

3.4 Classification and Objective Function

The sample index i is reintroduced in the classification stage. Global average pooling over the temporal and electrode dimensions of the fused representation gives $f_i = \frac{1}{TC}\sum_{t=1}^{T}\sum_{c=1}^{C} Z_{F,i,t,c} \in \mathbb{R}^d$. This vector, f_i, is passed to a two-layer multilayer perceptron (MLP) classifier to produce the classification logits:

$$o_i = GELU\left(f_i W_1 + b_1\right) W_2 + b_2 \in \mathbb{R}^{K_e} \tag{23}$$

Here, $W_1 \in \mathbb{R}^{d \times d_{mlp}}$, $W_2 \in \mathbb{R}^{d_{mlp} \times K_e}$; K_e denotes the number of emotion classes, and d_{mlp} is the hidden dimension of the classifier.

Applying softmax to the logits gives the class probabilities $p_i = Softmax(o_i)$, and the class with the highest probability is predicted. We train the model using the cross-entropy loss:

$$L_{CE} = -\frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \log p_{i,y_i} \tag{24}$$

The loss is computed over the training samples, where N_{tr} is the number of training samples included in the loss computation, $y_i \in \{1, \ldots, K_e\}$ is the ground-truth class label, and p_{i,y_i} is the predicted probability assigned to that class.

4 Experiments

We evaluate SDH-LGCNet on SEED and DEAP under subject-dependent and subject-independent settings. Additional analyses on SEED include component ablation, classwise evaluation, and attention visualization.

4.1 Datasets and Preprocessing

Experiments are conducted on two public EEG emotion datasets, SEED and DEAP. The tasks comprise three-class classification on SEED and separate binary classification of valence and arousal on DEAP. Differential entropy features are extracted using 2 s windows with a 0.5 s stride for both datasets.

SEED: The SEED dataset (Zheng et al. 2019; Zheng and Lu 2015) contains 62-channel EEG recordings from 15 subjects. The film stimuli elicit positive, neutral, and negative emotions. Our evaluation uses 15 trials per subject, with 5 trials per class. We extract differential entropy (DE) features from the officially preprocessed EEG signals in five frequency bands, δ (1–4 Hz), θ (4–8 Hz), α (8–14 Hz), β (14–31 Hz), and γ (31–50 Hz), for three-class emotion recognition. Inputs are generated using 20 s windows with a 4 s stride. We discard the last DE time point in each window, leaving 36 consecutive DE time points in each input.

DEAP: The DEAP dataset (Koelstra et al. 2012) contains EEG and peripheral physiological signals recorded from 32 subjects watching music videos. Each subject watched 40 videos lasting 1 min each and rated their emotions after each trial. We use the 32 EEG channels from the officially preprocessed data, sampled at 128 Hz. We remove the initial 3 s baseline from each trial and retain the subsequent 60 s stimulus recording. DE features are extracted from four frequency bands: θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–45 Hz). Two separate binary classification tasks are defined using subjects' self-ratings of valence and arousal on a 1–9 scale. Samples with ratings no greater than 5 are assigned to the low class, and those with ratings above 5 to the high class. Inputs are generated using 10 s windows with a 2 s stride. Each input contains 17 consecutive DE time points.

4.2 Experimental Protocol

Subject-dependent and subject-independent experiments evaluate within-subject emotion recognition performance and cross-subject generalization, respectively.

Subject-dependent evaluation uses trial-wise five-fold cross-validation, with training, validation, and test data drawn from different trials of the same subject. In each round, one fold serves as the test set, while the remaining four folds provide training and validation data. For SEED, evaluation is based on 15 trials per subject, with 9 trials used for training, 3 for validation, and 3 for testing in each round. The validation and test sets each contain one trial from each class. For DEAP, separate binary classification experiments are conducted for valence and arousal. Each subject's 40 trials are divided into 24 training, 8 validation, and 8 test trials per round. Each trial is used for testing once across the five rounds.

Subject-independent evaluation uses leave-one-subject-out (LOSO) cross-validation. In each fold, one subject is held out for testing, and all remaining subjects are used for training without a validation set. Standardization statistics are computed from the training data and applied to both training and test data.

In subject-dependent experiments, each trial is segmented using the window lengths and strides specified in Section 4.1. All segments from the same trial belong to the same set. Standardization statistics are computed from the training set and applied to the training, validation, and test sets.

Evaluation is performed at the EEG segment level. For subject-dependent experiments, test predictions from all five folds are pooled for each subject to calculate classification accuracy. For subject-independent experiments, accuracy is calculated from the pooled predictions for each held-out subject. Mean accuracy (ACC) and standard deviation (STD) across subjects are reported.

4.3 Implementation Details

The hidden dimension is 64. SDH-G3D uses a window length of 3, static nearest-neighbor count $k_n = 4$, dynamic top-k count $k_d = 6$, propagation order $K_g = 2$, and query/key dimensions of 32. The temporal decay coefficient is 0.8, and the maximum direct temporal range is 2. LGCL uses 4 heads in both L2G and G2L, a maximum membership-bias scale of 1.0, and 1 prior-token Transformer encoder layer with 2 heads and a feed-forward dimension of 64. The classifier hidden dimension is 128.

The model is trained using AdamW with an initial learning rate of 3×10^{-4} and a weight decay of 5×10^{-4}. The batch size is 64, the maximum number of training epochs is 150, and the dropout rate is 0.5. In subject-dependent experiments, the model with the highest segment-level validation accuracy is selected for testing. If validation accuracies are tied, the model from the earlier epoch is selected. For subject-independent experiments, each LOSO fold is trained once for 150 epochs, and the checkpoint from the final epoch is used for evaluation. The same hyperparameter settings are used across all folds, with the held-out subject excluded from hyperparameter tuning and model selection. The model is implemented in PyTorch and trained on an NVIDIA GeForce RTX 4060 Laptop GPU.

4.4 Comparison with Existing Methods

4.4.1 Subject-Dependent Results

Tables 1 and 2 compare SDH-LGCNet with basic classifiers and representative deep models. SEED comparisons include support vector machine (SVM) and deep belief network (DBN) (Zheng and Lu 2015), DGCNN (Song et al. 2020), RGNN (Zhong et al. 2022), HN-DGTS (Cheng et al. 2024b), BGAGCN-MT (Yan et al. 2024), and BF-GCN (Li et al. 2025).

DEAP comparisons include SVM and ACRNN (Tao et al. 2023), HN-DGTS (Cheng et al. 2024b), DGETN (Jin et al. 2025), and ATDD-LSTM (Du et al. 2022) on valence and arousal.

SEED: SDH-LGCNet achieves 95.93% mean accuracy, exceeding SVM, DBN, DGCNN, and RGNN but remaining below HN-DGTS, BGAGCN-MT, and BF-GCN (Table 1). Its 2.17% STD is the lowest reported among the compared methods, indicating relatively small inter-subject variation.

Table 1 Subject-Dependent Three-Class Classification Results on SEED

Method	**ACC±STD (%)**
SVM (Zheng and Lu 2015)	83.99 ± 9.72
DBN (Zheng and Lu 2015)	86.08 ± 8.34
DGCNN (Song et al. 2020)	90.40 ± 8.49
RGNN (Zhong et al. 2022)	94.24 ± 5.95
HN-DGTS (Cheng et al. 2024b)	97.53 ± —
BGAGCN-MT (Yan et al. 2024)	96.82 ± 4.85
BF-GCN (Li et al. 2025)	97.44 ± 3.89
SDH-LGCNet	95.93 ± 2.17

Note: "—" indicates that STD was not reported.

DEAP: SDH-LGCNet exceeds 92% accuracy on both tasks, outperforming SVM, ATDD-LSTM, and HN-DGTS while remaining below ACRNN and DGETN (Table 2). The two mean accuracies differ by 0.60 percentage points, and both STDs are below 3%, showing similar recognition performance across valence and arousal with relatively small inter-subject variation.

Table 2 Subject-Dependent Binary Classification Results on DEAP

Method	Valence ACC±STD (%)	Arousal ACC±STD (%)
SVM (Tao et al. 2023)	89.33 ± 7.41	89.99 ± 6.74
ATDD-LSTM (Du et al. 2022)	90.91 ± 12.95	90.87 ± 11.32
HN-DGTS (Cheng et al. 2024b)	88.20 ± —	87.04 ± —
ACRNN (Tao et al. 2023)	93.72 ± 3.21	93.38 ± 3.73
DGETN (Jin et al. 2025)	99.24 ± —	98.85 ± —
SDH-LGCNet	92.85 ± 2.19	92.25 ± 2.94

Note: Valence and Arousal denote the two affective dimensions; "—" indicates that STD was not reported.

4.4.2 Subject-Independent Results

Tables 3 and 4 report LOSO comparisons. SEED baselines include DGCNN (Song et al. 2020), ShareLink (Huang et al. 2025), TSGCN (Li et al. 2024), and BGAGCN-MT (Yan et al. 2024); DEAP baselines include DGCNN (Song et al. 2020; Zhang et al. 2023), SparseDGCNN (Zhang et al. 2023), and TSGCN (Li et al. 2024).

SEED: SDH-LGCNet achieves 87.14% mean accuracy, exceeding DGCNN, ShareLink, and TSGCN (Table 3). BGAGCN-MT (Yan et al. 2024) attains higher accuracy, whereas SDH-LGCNet has a lower STD (3.98% versus 4.72%).

Table 3 Subject-Independent Three-Class Classification Results on SEED (LOSO)

Method	ACC±STD (%)
DGCNN (Song et al. 2020)	79.95 ± 9.02
ShareLink (Huang et al. 2025)	80.61 ± 6.16
TSGCN (Li et al. 2024)	86.13 ± 5.29
BGAGCN-MT (Yan et al. 2024)	89.66 ± 4.72
SDH-LGCNet	87.14 ± 3.98

DEAP: SDH-LGCNet achieves the highest mean accuracy among the compared methods on both tasks (Table 4), improving over SparseDGCNN by 11.61 and 8.25 percentage points for valence and arousal, respectively. Inter-subject variation remains greater for arousal, despite its higher mean accuracy.

Table 4 Subject-Independent Binary Classification Results on DEAP (LOSO)

Method	Valence ACC±STD (%)	Arousal ACC±STD (%)
DGCNN (Song et al. 2020; Zhang et al. 2023)	58.46 ± 7.85	61.65 ± 13.34
SparseDGCNN (Zhang et al. 2023)	60.65 ± 6.24	65.39 ± 9.41
TSGCN (Li et al. 2024)	70.65 ± 4.19	72.54 ± 6.52
SDH-LGCNet	72.26 ± 6.29	73.64 ± 10.12

4.5 Ablation Study

We examine the contributions of SDH-G3D and LGCL through component ablation experiments on SEED, with additional analysis of key components under LOSO. All variants use the same evaluation settings as the full model. The results are presented in Tables 5–7.

4.5.1 Analysis of SDH-G3D

For SDH-G3D, we examine graph structure, dynamic information fusion, and cross-time propagation. Static-only and Dynamic-only retain only the static and dynamic graphs, respectively. Static + Dynamic w/o Adaptive Gate fixes the dynamic injection coefficient at 1, whereas Shared Dynamic Gate uses a shared learnable gate. The w/o Cross-time Edges variant removes cross-time connections from both graphs.

Table 5 SDH-G3D Ablation Results on SEED

Variant	ACC±STD (%)
Static-only	92.65 ± 4.64
Dynamic-only	89.42 ± 5.95
Static + Dynamic w/o Adaptive Gate	88.73 ± 6.31
Shared Dynamic Gate	92.27 ± 5.81
w/o Cross-time Edges	89.37 ± 5.77
Full SDH-LGCNet	95.93 ± 2.17

Table 5 addresses the complementarity of the two graphs: the full model exceeds Static-only and Dynamic-only by 3.28 and 6.51 percentage points, respectively. Static priors provide a useful base, while adaptive dynamic relations further improve local spatio-temporal representations.

Fusion requires control of the dynamic contribution. A shared gate improves on fixed unit-strength injection, but both remain below window-specific adaptive gating. Removing cross-time connections causes a 6.56-point accuracy decrease, supporting propagation across time positions within local windows.

4.5.2 Analysis of LGCL

We use the following variants to examine prior organization and local-global interaction in LGCL. Local-only removes the entire LGCL module. Region-only and Lateralization-only retain only region and lateralization priors, respectively. The w/o Prior Context Modeling variant removes position-related embeddings and Transformer encoding. L2G-only retains local-to-global aggregation and prior context encoding and uses token pooling for classification. We also remove the L2G membership mask, G2L dual membership bias, and feature-wise gate individually to examine spatial membership constraints and feedback fusion.

Table 6 LGCL Ablation Results on SEED

Variant	**ACC±STD (%)**
Local-only	79.48 ± 8.57
Region-only	88.17 ± 5.70
Lateralization-only	89.57 ± 7.45
w/o Prior Context Modeling	86.84 ± 6.51
L2G-only	88.35 ± 6.49
w/o L2G Membership Mask	93.37 ± 4.23
w/o G2L Dual Membership Bias	92.34 ± 5.15
w/o Feature-wise Gate	92.71 ± 4.00
Full SDH-LGCNet	95.93 ± 2.17

Table 6 examines prior organization: removing LGCL causes a 16.45-point accuracy decrease, while either Region-only or Lateralization-only improves on Local-only. Combining both priors performs best, supporting their complementary organization of electrode information.

Removing prior context modeling lowers accuracy, supporting the combined role of position-related embeddings and token encoding. L2G-only also underperforms the full model; because it uses token pooling for classification, this comparison reflects joint changes to the feedback path and classification readout.

Removing the L2G membership mask, G2L dual membership bias, or feature-wise gate each reduces accuracy. Together, these variants support membership-constrained aggregation and membership-aware, gated feedback for effective local-global interaction.

4.5.3 Key Components under LOSO

The trends persist under SEED LOSO (Table 7): Static-only and removal of cross-time edges both reduce accuracy, while Local-only produces the largest decrease (12.34 percentage points). Either spatial prior improves on Local-only, and their joint use performs best. The full model also has the lowest STD among these variants, supporting the combined local spatio-temporal and global design for unseen subjects.

Table 7 Key-Component Ablation under SEED LOSO

Variant	ACC±STD (%)
Static-only	84.68 ± 6.18
w/o Cross-time Edges	82.22 ± 6.91
Local-only	74.80 ± 8.42
Region-only	81.32 ± 7.08
Lateralization-only	82.37 ± 7.26
Full SDH-LGCNet	87.14 ± 3.98

4.6 Classwise Performance and Visualization

The confusion matrices in Fig. 5 show classwise recognition performance on SEED under the subject-dependent and LOSO settings. Both matrices are normalized by true class, so their diagonal entries represent classwise recall.

Under subject-dependent evaluation, all three class recalls exceed 94%, with positive emotion performing best. Positive-emotion recall is also best preserved under LOSO, whereas neutral emotion shows the largest decrease and is confused with both other classes. This identifies neutral-emotion recognition as a particular difficulty across subjects.

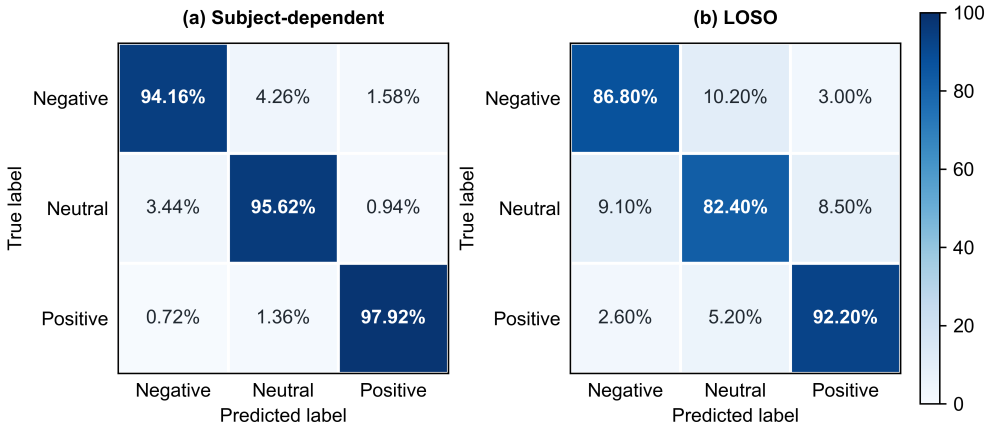


Fig. 5 SEED confusion matrices: (a) subject-dependent; (b) subject-independent (LOSO). Values are percentages normalized by true class

To analyze how LGCL aggregates information across emotion classes, we extract L2G attention weights for samples in each test fold of subject-dependent five-fold cross-validation on SEED and average them over the time and attention-head dimensions. For each electrode, we select the weights associated with its region token and lateralization token separately and multiply each weight by the number of electrodes in the corresponding group. This rescales uniform aggregation to a weight of 1 across prior groups of different sizes. We then group the samples by ground-truth emotion

label and sequentially average over windows within each trial, equally over trials of the same class within each subject, and equally over all 15 subjects. This procedure prevents differences in window counts from affecting the contributions of trials and subjects to the aggregate. Fig. 6 uses a shared color scale to facilitate comparison of the aggregation patterns of region and lateralization priors across emotion classes.

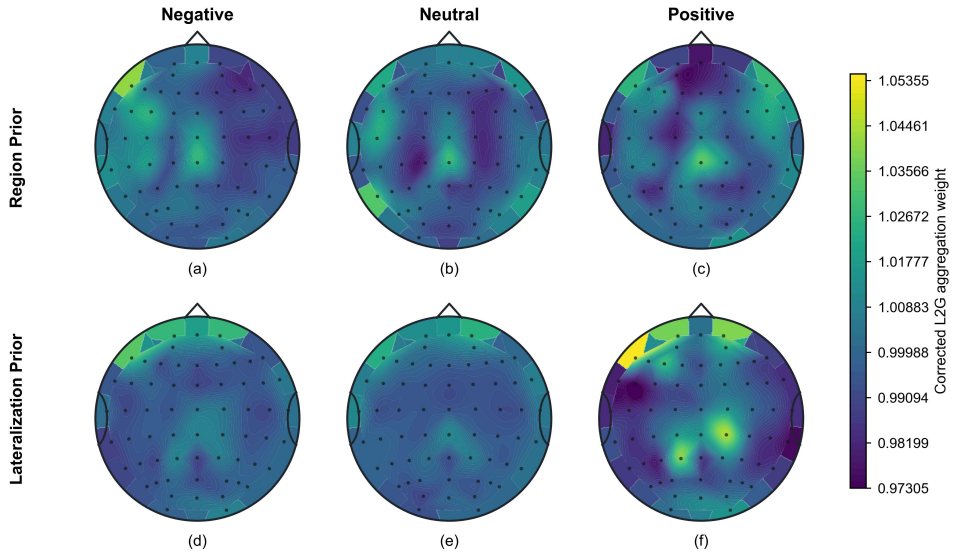


Fig. 6 Scalp distributions of L2G aggregation weights for three emotion classes under the subject-dependent setting on SEED. The top row (a–c) shows region priors, and the bottom row (d–f) shows lateralization priors. Columns correspond to negative, neutral, and positive emotions, respectively. Colors indicate L2G aggregation weights rescaled according to the number of member electrodes; all six panels share the same color scale

In Fig. 6, region and lateralization priors show different aggregation emphases across emotion classes. Lateralization weights exhibit more pronounced local differences for positive emotion and milder variation for neutral emotion. Values above 1 indicate greater aggregation weight than uniform within-group aggregation. These patterns help characterize the two priors, whose combined performance benefit is supported by Table 6.

4.7 Discussion

SDH-G3D combines a stable sensor-space reference with state-dependent relations in the same local spatio-temporal domain. Window-specific gating regulates the dynamic contribution without replacing the prior paths, while cross-time propagation connects temporal changes directly to electrode relations. The ablations support this coordination under both evaluation settings.

LGCL complements these local representations by organizing information along two distinct spatial attributes. Factorized region and lateralization tokens preserve

their respective memberships during aggregation and exchange context before feeding it back to electrodes. This connects spatial prior organization with global interaction, accounting for the consistent benefit of joint priors over either prior alone.

LOSO accuracy remains lower than subject-dependent accuracy on both SEED and DEAP, indicating that individual differences continue to limit cross-subject recognition. Future work will investigate adaptation strategies to improve generalization.

5 Conclusion

We propose SDH-LGCNet for EEG emotion recognition, combining static-dynamic local spatio-temporal modeling through SDH-G3D with region- and lateralization-guided local-global interaction through LGCL. Experiments on SEED and DEAP demonstrate its recognition performance under subject-dependent and subject-independent settings. SEED ablations support static-dynamic fusion, cross-time propagation, and joint spatial priors as complementary components of the model.

References

Cao M, Dong Y, Chen D, et al (2025) EEG-based emotion recognition model using fuzzy adjacency matrix combined with convolutional multi-head graph attention mechanism. Cluster Comput 28(4):274. https://doi.org/10.1007/s10586-024-04994-3

Cheng C, Liu W, Feng L, et al (2024a) Emotion recognition using hierarchical spatial–temporal learning Transformer from regional to global brain. Neural Netw 179:106624. https://doi.org/10.1016/j.neunet.2024.106624

Cheng C, Yu Z, Zhang Y, et al (2024b) Hybrid network using dynamic graph convolution and temporal self-attention for EEG-based emotion recognition. IEEE Trans Neural Netw Learn Syst 35(12):18565–18575. https://doi.org/10.1109/TNNLS.2023.3319315

Compumedics USA, Ltd. (2016) CURRY 7.0.12 Release Notes. URL https://www.compumedicsneuroscan.com/Uploads/Curry7012ReleaseNotes.pdf, document 6418, rev. L, February 19, 2016. Accessed: September 28, 2026

Ding Y, Robinson N, Tong C, et al (2024a) LGGNet: learning from local-global-graph representations for brain–computer interface. IEEE Trans Neural Netw Learn Syst 35(7):9773–9786. https://doi.org/10.1109/TNNLS.2023.3236635

Ding Y, Zhang S, Tang C, et al (2024b) MASA-TCN: multi-anchor space-aware temporal convolutional neural networks for continuous and discrete EEG emotion recognition. IEEE J Biomed Health Inform 28(7):3953–3964. https://doi.org/10.1109/JBHI.2024.3392564

Du X, et al (2022) An efficient LSTM network for emotion recognition from multichannel EEG signals. IEEE Trans Affect Comput 13(3):1528–1540. https://doi.org/10.1109/TAFFC.2020.3013711

Gong L, Chen W, Zhang D (2024) An attention-based multi-domain bi-hemisphere discrepancy feature fusion model for EEG emotion recognition. IEEE J Biomed Health Inform 28(10):5890–5903. https://doi.org/10.1109/JBHI.2024.3418010

Guo W, Li Y, Liu M, et al (2024) Functional connectivity-enhanced feature-grouped attention network for cross-subject EEG emotion recognition. Knowl-Based Syst 283:111199. https://doi.org/10.1016/j.knosys.2023.111199

Huang W, Wang W, Li Y, et al (2024) FBSTCNet: a spatio-temporal convolutional network integrating power and connectivity features for EEG-based emotion decoding. IEEE Trans Affect Comput 15(4):1906–1918. https://doi.org/10.1109/TAFFC.2024.3385651

Huang Z, Kong L, Ao L, et al (2025) ShareLink: Neuro-inspired EEG-based cross-subject emotion recognition via shared bi-hemisphere. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, Lecture Notes in Computer Science, vol 15971. Springer, Cham, Switzerland, pp 605–615, https://doi.org/10.1007/978-3-032-05162-2_58

Jafari M, et al (2023) Emotion recognition in EEG signals using deep learning methods: a review. Comput Biol Med 165:107450. https://doi.org/10.1016/j.compbiomed.2023.107450

Jenke R, Peer A, Buss M (2014) Feature extraction and selection for emotion recognition from EEG. IEEE Trans Affect Comput 5(3):327–339. https://doi.org/10.1109/TAFFC.2014.2339834

Jin M, Du C, He H, et al (2024) PGCN: pyramidal graph convolutional network for EEG emotion recognition. IEEE Trans Multimedia 26:9070–9082. https://doi.org/10.1109/TMM.2024.3385676

Jin X, Zhu F, Shen Y, et al (2025) Data-driven dynamic graph convolution Transformer network model for EEG emotion recognition under IoMT environment. Big Data Min Anal 8(3):712–725. https://doi.org/10.26599/BDMA.2024.9020071

Jurcak V, Tsuzuki D, Dan I (2007) 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. NeuroImage 34(4):1600–1611. https://doi.org/10.1016/j.neuroimage.2006.09.024

Klem GH, Lüders HO, Jasper HH, et al (1999) The ten-twenty electrode system of the International Federation. Electroencephalogr Clin Neurophysiol Suppl 52:3–6

Klepl D, Wu M, He F (2024) Graph neural network-based EEG classification: a survey. IEEE Trans Neural Syst Rehabil Eng 32:493–503. https://doi.org/10.1109/TNSRE.2024.3355750

Koelstra S, et al (2012) DEAP: a database for emotion analysis using physiological signals. IEEE Trans Affect Comput 3(1):18–31. https://doi.org/10.1109/T-AFFC.2011.15

Li C, et al (2025) An efficient graph learning system for emotion recognition inspired by the cognitive prior graph of EEG brain network. IEEE Trans Neural Netw Learn Syst 36(4):7130–7144. https://doi.org/10.1109/TNNLS.2024.3405663

Li J, Qiu S, Shen YY, et al (2020) Multisource transfer learning for cross-subject EEG emotion recognition. IEEE Trans Cybern 50(7):3281–3293. https://doi.org/10.1109/TCYB.2019.2904052

Li L, Chen W (2026) EEG-based emotion recognition using spatial-temporal graph-aware network with channel selection. IEEE J Biomed Health Inform 30(8):6565–6578. https://doi.org/10.1109/JBHI.2026.3665517

Li R, Yang X, Lou J, et al (2024) A temporal-spectral graph convolutional neural network model for EEG emotion recognition within and across subjects. Brain Inform 11(1):30. https://doi.org/10.1186/s40708-024-00242-x

Liu H, et al (2026) A hierarchical meta-learning driven adaptive brain region flow fusion method for EEG-based emotion recognition. Int J Cogn Comput Eng 7:473–482. https://doi.org/10.1016/j.ijcce.2026.04.005

Liu Z, Zhang H, Chen Z, et al (2020) Disentangling and unifying graph convolutions for skeleton-based action recognition. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp 143–152, https://doi.org/10.1109/CVPR42600.2020.00022

Oostenveld R, Praamstra P (2001) The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophysiol 112(4):713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Song T, Zheng W, Song P, et al (2020) EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Trans Affect Comput 11(3):532–541. https://doi.org/10.1109/TAFFC.2018.2817622

Tao W, et al (2023) EEG-based emotion recognition via channel-wise attention and self attention. IEEE Trans Affect Comput 14(1):382–393. https://doi.org/10.1109/TAFFC.2020.3025777

Vaswani A, et al (2017) Attention is all you need. In: Adv. Neural Inf. Process. Syst., pp 5998–6008

Wang W, et al (2025) ALGGNet: an adaptive local-global-graph representation network for brain-computer interfaces. Knowl-Based Syst 311:113096. https://doi.org/10.1016/j.knosys.2025.113096

Wang Y, Zhang B, Di L (2024) Research progress of EEG-based emotion recognition: a survey. ACM Comput Surv 56(11):1–49. https://doi.org/10.1145/3666002

Wu M, Chen CLP, Chen B, et al (2025) Grop: graph orthogonal purification network for EEG emotion recognition. IEEE Trans Affect Comput 16(1):319–332. https://doi.org/10.1109/TAFFC.2024.3433613

Xiong R, et al (2020) On layer normalization in the Transformer architecture. In: Proc. 37th Int. Conf. Mach. Learn. (ICML), pp 10524–10533

Yan H, Guo K, Xing X, et al (2024) Bridge graph attention based graph convolution network with multi-scale Transformer for EEG emotion recognition. IEEE Trans Affect Comput 15(4):2042–2054. https://doi.org/10.1109/TAFFC.2024.3394873

Yang H, Chen CLP, Chen B, et al (2025) Improving the interpretability through maximizing mutual information for EEG emotion recognition. IEEE Trans Affect Comput 16(2):744–757. https://doi.org/10.1109/TAFFC.2024.3463469

Yang K, et al (2024) Automatically extracting and utilizing EEG channel importance based on graph convolutional network for emotion recognition. IEEE J Biomed Health Inform 28(8):4588–4598. https://doi.org/10.1109/JBHI.2024.3404146

Zhang G, Yu M, Liu YJ, et al (2023) SparseDGCNN: recognizing emotion from multichannel EEG signals. IEEE Trans Affect Comput 14(1):537–548. https://doi.org/10.1109/TAFFC.2021.3051332

Zhang H, Zuo T, Chen Z, et al (2024) Evolutionary ensemble learning for EEG-based cross-subject emotion recognition. IEEE J Biomed Health Inform 28(7):3872–3881. https://doi.org/10.1109/JBHI.2024.3384816

Zheng WL, Lu BL (2015) Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans Auton Ment Develop 7(3):162–175. https://doi.org/10.1109/TAMD.2015.2431497

Zheng WL, Zhu JY, Lu BL (2019) Identifying stable patterns over time for emotion recognition from EEG. IEEE Trans Affect Comput 10(3):417–429. https://doi.org/10.1109/TAFFC.2017.2712143

Zheng X, Yao Z, Li X, et al (2026) Cross-level topological framework: learning explainable region-channel representations from EEG signals for emotional decoding. IEEE J Biomed Health Inform https://doi.org/10.1109/JBHI.2026.3705480, early access, June 22, 2026

Zhong P, Wang D, Miao C (2022) EEG-based emotion recognition using regularized graph neural networks. IEEE Trans Affect Comput 13(3):1290–1301. https://doi.org/10.1109/TAFFC.2020.2994159

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

Zhou Y, Zhang X, Sun Y, et al (2025) EEG emotion recognition based on dynamic temporal causal graph convolutional network. Knowl-Based Syst 323:113752. https://doi.org/10.1016/j.knosys.2025.113752

Zhou Y, et al (2023) Progressive graph convolution network for EEG emotion recognition. Neurocomputing 544:126262. https://doi.org/10.1016/j.neucom.2023.126262