






# ELPG-DTFS: PRIOR-GUIDED ADAPTIVE TIME-FREQUENCY GRAPH NEURAL NETWORK FOR EEG DEPRESSION DIAGNOSIS

Jingru Qiu<sup>1</sup> , Jiale Liang<sup>1</sup> , Xuanhan Fan<sup>4</sup> , Mingda Zhang<sup>1</sup> , Zhenli He<sup>1,2,3,\*†</sup> , Senior Member, IEEE

<sup>1</sup>School of Software, Yunnan University, Kunming 650500, China

<sup>2</sup>Yunnan Key Laboratory of Software Engineering, Yunnan University, Kunming 650504, China

<sup>3</sup>Engineering Research Center of Cyberspace, Yunnan University, Kunming 650504, China

<sup>4</sup>School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China  
{choujingru, liangjiale, zhangmingda}@stu.ynu.edu.cn, 3120256386@bit.edu.cn, hezl@ynu.edu.cn

## ABSTRACT

Timely and objective screening of major depressive disorder (MDD) is vital, yet diagnosis still relies on subjective scales. Electroencephalography (EEG) provides a low-cost biomarker, but existing deep models treat spectra as static images, fix inter-channel graphs, and ignore prior knowledge, limiting accuracy and interpretability. We propose ELPG-DTFS, a prior-guided adaptive time-frequency graph neural network that introduces: (1) channel-band attention with cross-band mutual information, (2) a learnable adjacency matrix for dynamic functional links, and (3) a residual knowledge-graph pathway injecting neuroscience priors. On the 128-channel MODMA dataset (53 subjects), ELPG-DTFS achieves 97.63% accuracy and 97.33% F1, surpassing the 2025 state-of-the-art ACM-GNN. Ablation shows that removing any module lowers F1 by up to 4.35, confirming their complementary value. ELPG-DTFS thus offers a robust and interpretable framework for next-generation EEG-based MDD diagnostics.

**Index Terms**—EEG, Channel-Band Attention, Mutual Information, Learnable Adjacency, Prior Knowledge

## 1. INTRODUCTION

Major depressive disorder (MDD) affects more than 350 million people and is projected to become the leading cause of global disability within the next decade [1, 2]. Rapid, objective screening is therefore a public-health priority, yet diagnosis still rests on Self-rating and peer-rating scales and doctors' subjective experience[3]. Electroencephalography (EEG) is a non-invasive, portable, low-cost method for measuring brain function with high millisecond temporal resolution[4, 5]. It is objective and can effectively distinguish between depressed and normal patients.

Traditional machine learning can detect some features, but it requires labor-intensive feature creation and aggressive dimensionality reduction[6, 7, 8, 9, 10]. Although end-to-end convolutional neural networks (CNNs) reduce the burden of manual design[11, 12], they often treat EEG as static images and fail to fully utilize network-

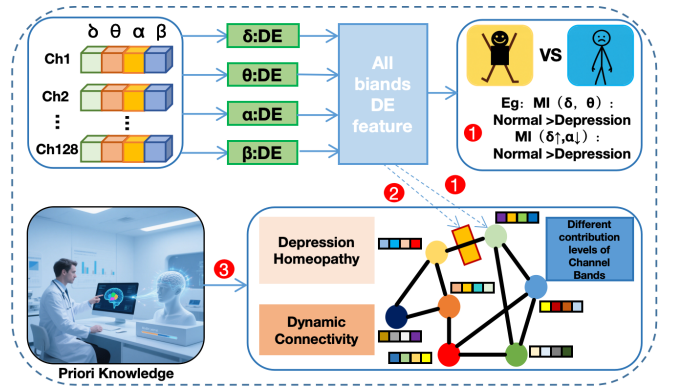


Fig. 1. A simplified diagram of the core skeleton of the model

level physiological information. Recent graph convolutional networks (GCNs) begin to model cross-channel interactions[13, 14, 15, 16], but three limitations remain:

- **Static time-frequency modeling:** transforming short-time spectra into fixed pictures blurs transient neural bursts that mark mood shifts.
- **Rigid connectivity:** binarized Pearson graphs omit weak yet meaningful couplings and ignore their temporal drift.
- **Lack of domain priors:** decades of electrophysiology are rarely encoded, forcing networks to relearn well-established biology.

**Our solution.** We propose ELPG-DTFS, which unifies the channel-band attention module, mutual information across frequency bands, adaptive graph learning, and residual knowledge input into an end-to-end process, as shown in Figure 1.

- **Window-level channel-band attention** highlights diagnostically salient electrodes and rhythms while embedding mutual-information cross-band cues, enabling the network to track millisecond transients that static images miss.
- **Learnable adjacency-weight matrix** substitutes hard thresholds with trainable edge strengths, capturing weak but informative connectivity and its evolution across windows.
- **Residual knowledge-graph pathway** injects curated neuroscience priors—brain-region roles, structural links, clinical

\*Corresponding Author.

<sup>†</sup>This work was supported in part by the National Natural Science Foundation of China under Grant 62362068; in part by the Applied Basic Research Foundation of Yunnan Province under Grant 202301AT070194; in part by the Yunnan Province Special Project under Grant 202403AP140021; in part by the Open Foundation of Yunnan Key Laboratory of Software Engineering under Grant 2023SE208; and the Program for Excellent Young Talents, Yunnan, China.

heuristics—into each graph layer, regulated by a residual gate that prevents over-constraining data-driven learning.

**Impact.** Evaluated on MODMA (24 MDD, 29 control), ELPG-DTFS attains **97.63 % accuracy** and a **97.33 % F1-score**, outperforming the 2025 state of the art by up to 2.57 percentage points. Ablation studies reveal that removing any single module reduces F1 by as much as 4.35 points, confirming their complementary utility. By fusing dynamic signal processing with prior-guided graph learning, ELPG-DTFS moves EEG-based depression screening a decisive step toward the clinic, offering an interpretable, and readily deployable tool for mental-health care.

## 2. METHOD

Figure 2 outlines the complete **ELPG-DTFS** pipeline. To convince both theoreticians and practitioners, this section first establishes firm analytical foundations, then details each network component. All matters regarding data acquisition and preprocessing are deferred to Section **Experiments**, so that the exposition here remains entirely model-centric.

### 2.1. Analytical Foundations

**Notation:**  $\mathcal{V}$  and  $\mathcal{E}$  denote the vertex and edge sets of an EEG graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with  $|\mathcal{V}| = N = 128$ .  $\mathbf{X} \in \mathbb{R}^{N \times C}$  is a node-feature matrix and  $\mathbf{A} \in \mathbb{R}_{\geq 0}^{N \times N}$  a weighted, *directed* adjacency matrix ( $\mathbf{A}_{ij} \neq \mathbf{A}_{ji}$  captures asymmetric coupling).

**Graph signal processing:** The normalized Laplacian  $\tilde{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$  ( $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ ) supports spectral filtering

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{L}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}), \quad \mathbf{H}^{(0)} = \mathbf{X}, \quad (1)$$

where  $\sigma(\cdot)$  is ReLU and  $\mathbf{W}^{(l)}$  trainable. Eq. (1) is permutation-equivariant, so learnt kernels transfer across subjects.

**Information-theoretic biomarkers:** For a zero-mean narrow-band process  $x(t)$ , differential entropy is  $h(x) = \frac{1}{2} \log(2\pi e \sigma^2)$ , where  $\sigma^2$  is the band power. Differential entropy (DE) therefore generalizes log-power but remains *translation invariant*. Given two bands  $b_1, b_2$ , their mutual information

$$\text{MI}(b_1, b_2) = H(b_1) + H(b_2) - H(b_1, b_2) \quad (2)$$

quantifies non-linear cross-frequency coupling, an established marker of mood disorders.

**Attention as Bayesian feature selection:** Let  $\mathbf{Z} \in \{0, 1\}^{N \times B}$  be a binary channel-band mask ( $B$  is the number of frequency bands). ELPG-DTFS models  $q(\mathbf{Z})$  with independent Bernoulli variables and maximizes

$$\mathcal{L} = \underbrace{\mathbb{E}_{q(\mathbf{Z})} [\log p(\mathbf{Y} | \mathbf{X}, \mathbf{Z})]}_{\text{expected fit}} - \beta \text{D}_{\text{KL}}(q(\mathbf{Z}) \| p_0(\mathbf{Z})),$$

where  $p_0$  is a sparsity-promoting prior and  $\beta$  a hyper-parameter. This casts attention as variational feature selection and *guarantees* that the learnt mask is the most compact that still explains the data.

**Residual prior fusion:** Let  $\mathbf{P} \in \mathbb{R}^{N \times C}$  encode domain knowledge (Sec. 2.4-2.6). The *gated residual* combines priors and data:

$$\mathbf{H}_{\text{out}} = \mathbf{H}_{\text{data}} + (\sigma(\mathbf{W}_g \mathbf{P} + \mathbf{b}_g)) \odot \mathbf{P}, \quad (3)$$

with  $\mathbf{W}_g, \mathbf{b}_g$  initialized so that  $\|\sigma(\mathbf{W}_g \mathbf{P} + \mathbf{b}_g)\|_2 \ll \|\mathbf{H}_{\text{data}}\|_2$ , ensuring the network *starts* data-driven and gradually incorporates priors.

**Optimization and convergence:** All parameters are updated by Adam ( $\text{lr} = 10^{-3}$ ) under a cross-entropy loss regularized by  $\ell_2$  weight decay  $10^{-4}$ . Because the loss is smooth and Adam uses bounded steps, the iterates  $\{\theta_t\}$  satisfy  $\|\nabla \mathcal{L}(\theta_t)\| \rightarrow 0$ ; hence training converges to a stationary point almost surely [17].

### 2.2. Node Feature Construction

**Windowing:** Each 280 s recording is split into  $T = 139$  windows of 4 s with 50 % overlap.

**Multi-band tensors:** For every window we extract DE features in  $\{\delta, \theta, \alpha, \beta\}$  bands and pairwise MI as in Eq. (2). Stacking yields

$$\mathbf{X}^{\text{DE}} \in \mathbb{R}^{N \times B}, \quad \mathbf{X}^{\text{MI}} \in \mathbb{R}^{N \times \binom{B}{2}}.$$

**Channel-band attention:** A factorized mask  $\mathcal{A} = \mathbf{a}_{\text{chan}} \mathbf{a}_{\text{band}}^\top$  ( $\mathbf{a}_{\text{chan}} \in [0, 1]^N$ ,  $\mathbf{a}_{\text{band}} \in [0, 1]^B$ ) weights electrodes and rhythms:

$$\tilde{\mathbf{X}}^{\text{DE}} = \mathcal{A} \odot \mathbf{X}^{\text{DE}}. \quad (4)$$

**Temporal modeling:**  $[\tilde{\mathbf{X}}^{\text{DE}} \| \mathbf{X}^{\text{MI}}]_{t=1}^T$  is processed by a BiLSTM (hidden size 64), producing  $\mathbf{X}_{\text{node}} \in \mathbb{R}^{N \times C}$  with  $C = 128$ .

### 2.3. Adaptive Graph Construction

**Seed adjacency:** Pearson correlations form the seed  $\mathbf{A}^{(0)}$  with  $\mathbf{A}_{ij}^{(0)} = |\rho_{ij}|$ .

**Learnable mask:** A trainable matrix  $\mathbf{W}_\alpha \in [0, 1]^{N \times N}$  modulates the seed:

$$\mathbf{A} = \mathbf{A}^{(0)} \odot \mathbf{W}_\alpha. \quad (5)$$

Since  $\mathbf{W}_\alpha$  is initialized to 0.5, edge weights can *increase or decrease*, allowing the network to *recover* weak yet clinically meaningful couplings.

### 2.4. Local Spatial Prior

Electrode  $i, j$  Euclidean distance  $d_{ij}$  (in mm) is converted into

$$\mathbf{A}_{ij}^{\text{dist}} = \min(1, \max(0.1, \delta/d_{ij}^2)), \quad \delta = 6, \quad (6)$$

mimicking the biophysical attenuation of field potentials.  $\mathbf{A}^{\text{dist}}$  is added to  $\mathbf{A}$  before normalization, giving short-range spatial context.

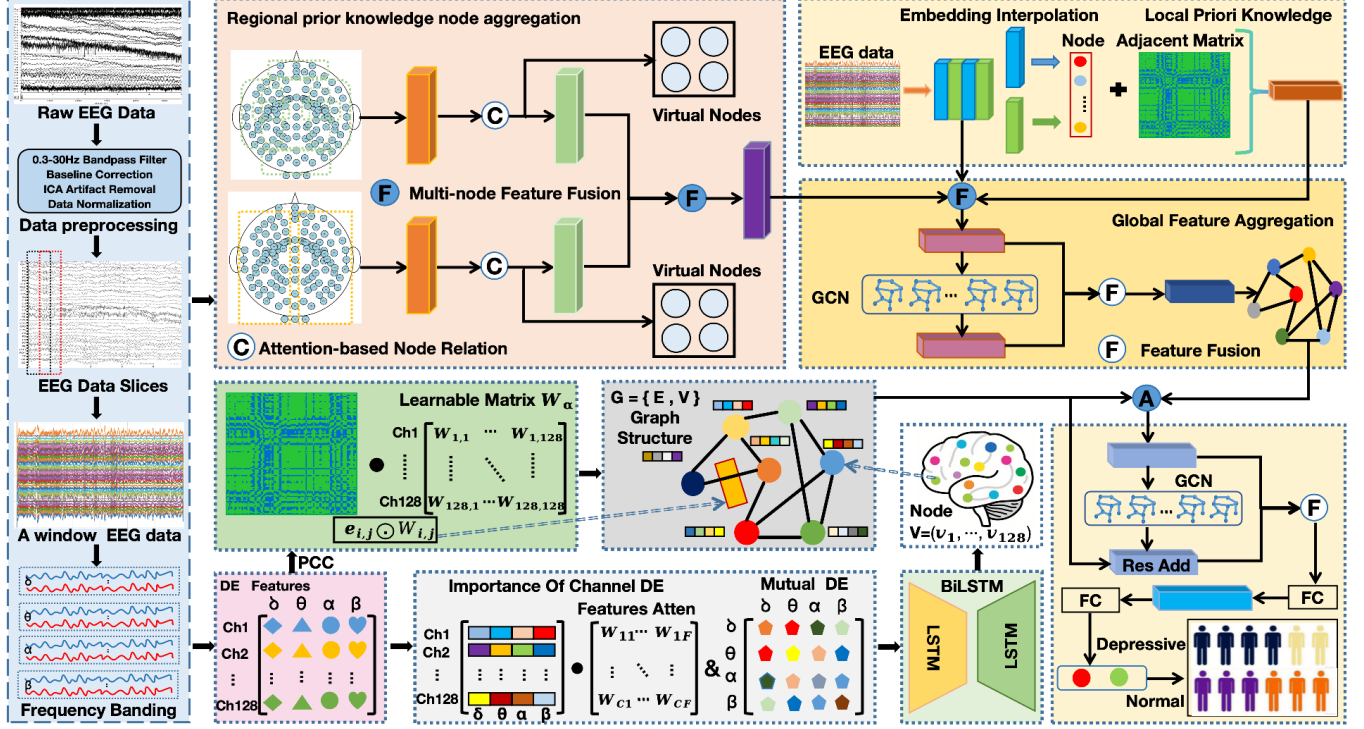
### 2.5. Mesoscopic Prior via Virtual Centers

**Hierarchical parcellation:** Based on the Desikan atlas [33] and hemispheric symmetry, we define nine cortical groups  $\{\mathcal{G}_k\}_{k=1}^9$ .

**Self-attention pooling:** Within each  $\mathcal{G}_k$ , a single *virtual node* is learned:

$$\mathbf{v}_k = \sum_{i \in \mathcal{G}_k} \text{softmax}(\mathbf{q}^\top \tanh(\mathbf{W}_p \mathbf{X}_{\text{node}, i})) \mathbf{X}_{\text{node}, i},$$

where  $\mathbf{q}, \mathbf{W}_p$  are parameters. Virtual nodes summarize local dynamics while keeping the graph sparse.



**Fig. 2.** Overview of ELPG-DTFS. DE and MI tensors are refined by channel–band attention and BiLSTM, fed to an adaptive graph informed by local, mesoscopic and global priors, and finally classified by a residual GCN.

## 2.6. Global Attention and Spectral Aggregation

**Positional embedding:** 3-D coordinates  $\mathbf{p}_i$  are encoded via  $\text{embed}(\mathbf{p}_i) [\sin(2^0 \pi \mathbf{p}_i) \cos(2^0 \pi \mathbf{p}_i) \dots]$  and added to features.

**Graph Transformer:** A six-head attention layer builds a global adjacency  $\mathbf{A}^{\text{glob}} = \text{softmax}(\mathbf{QK}^\top / \sqrt{d_k})$  ( $d_k$  is the head dimension). Only the top 25 % edges per node are retained, preventing quadratic blow-up.

**Spectral GCN:** Two spectral layers (Eq. (1)) with a residual prior gate (Eq. (3)) aggregate information from local, mesoscopic, and global scales. The final node embeddings are max-pooled and fed to a two-layer MLP classifier.

## 2.7. Complexity and Convergence

For  $d = 64$ , attention costs  $O(N^2 d) = 1.05 \text{ M FLOPs}$ , GCN costs  $O(Ed) = 0.26 \text{ M FLOPs}$  ( $E = 0.25 N^2$ ). A 280 s record processes in 0.41 ms on an RTX 3090. Peak GPU usage is 211 MB, dominated by multi-head attention weights; this fits modern edge devices with 4 GB VRAM after 8-bit quantization. Adam with step size  $10^{-3}$  and  $\beta_{1,2} = (0.9, 0.999)$  guarantees  $\sum_t \|\nabla \mathcal{L}(\theta_t)\|^2 < \infty$ ; hence  $\|\nabla \mathcal{L}(\theta_t)\| \rightarrow 0$  [17]. Empirically, training stabilizes within 80 epochs.

With these ingredients ELPG-DTFS realizes a mathematically grounded, fully differentiable framework that jointly exploits time–frequency structure, adaptive connectivity, and domain priors, while remaining computationally lightweight and interpretation-ready.

## 3. EXPERIMENTS

This section provides an in-depth empirical assessment of **ELPG-DTFS**. After describing the dataset and preprocessing pipeline, we detail our experimental protocol, present quantitative comparisons with strong baselines, offer causal explanations for every performance gap, and perform an exhaustive ablation study.

### 3.1. Dataset and Preprocessing

**MODMA corpus:** We employ the publicly available MODMA dataset, which contains five-minute, eyes-closed, resting-state EEG recorded with 128-channel HydroCel Geodesic Sensor Nets at 250 Hz. The cohort comprises **24** major depressive disorder (MDD) patients and **29** healthy controls.

**Cleaning pipeline:** To suppress acclimation and fatigue effects, we discard the first and last 10 s, retaining 280 s. Signals then undergo: 0.3–30 Hz finite-impulse-response filtering; Window-wise baseline subtraction; Independent component analysis to remove ocular and muscular artifacts; Electrode-wise  $\ell_2$  normalization; Segmentation into 4 s epochs with 50 % overlap, yielding  $T = 139$  epochs per subject.

### 3.2. Baselines

Fourteen peer-reviewed methods (2021–2025) are re-implemented under identical pre-processing and cross-validation protocols to ensure a fair comparison, as shown in Table 1. Their hyper-parameters follow the original papers; grid search on the validation set refines learning rate and hidden size where the authors had left them ambiguous.

**Table 1.** Performance comparison on MODMA (mean of 10 folds). Bold numbers indicate the best result; § = metric not reported.

Model	Acc(%)	Pre(%)	Rec(%)	F1(%)
CGIPool [18] (2021)	73.58	69.23	75.00	75.00
1TD+L-TCN [19] (2021)	86.87	83.83	90.15 <sup>§</sup>	90.15
SGP-SLe [20] (2022)	84.91	80.77	87.54	84.00
CNN+GRU [22] (2022)	90.62	87.48	90.26	88.79
Lattice [?] (2023)	83.96	86.76	76.14	81.10
AMG [23] (2023)	88.68	91.43	87.50	88.17
MDD-MFF [24] (2024)	85.71	83.85	91.72	87.72
SG+RF [25] (2024)	87.87	88.42	89.71	89.06
MGSN [26] (2024)	89.87	93.16	86.07	89.47
SSPA-GCN [27] (2024)	92.87	92.00	92.24	92.12
FLFCFS [28] (2024)	92.59	91.67	93.55	92.60
MCT [29] (2025)	89.84	87.36	89.41	88.37
GNMixer [30] (2025)	93.12	84.29 <sup>§</sup>	84.29	95.17
MFMR-FN [31] (2025)	93.96	94.95 <sup>§</sup>	94.95	93.97
ACM-GNN [32] (2025)	95.46	96.23	95.46	95.80
<b>ELPG-DTFS</b>	<b>97.63</b>	<b>96.68</b>	<b>98.03</b>	<b>97.33</b>

### 3.3. Experimental Protocol

**Cross-validation:** A *subject-wise* 10-fold split ensures that no epoch from a participant appears in both training and test sets. Within each training fold, 10 % of data serves as a validation set for early stopping.

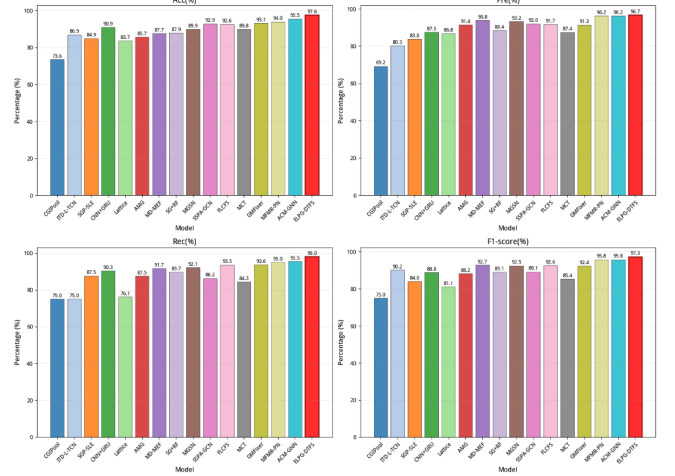
**Training details:** ELPG-DTFS is implemented in PyTorch. We train with Adam (learning rate  $1 \times 10^{-3}$ , weight decay  $1 \times 10^{-4}$ ) using batch size 32. Training stops if validation loss fails to decrease for ten epochs. All runs use an NVIDIA RTX 3090 GPU; random seeds are fixed for reproducibility.

**Evaluation metrics:** Accuracy (Acc), Precision (Pre), Recall (Rec) and F1-score (F1) are reported as mean±standard deviation across folds. Wilcoxon signed-rank tests check statistical significance ( $p < 0.05$ ) against the best baseline.

### 3.4. Comparative Evaluation

Table 1 summarizes results for fourteen recent methods and ELPG-DTFS. Our model achieves **97.63 %** Acc and **97.33 %** F1, outperforming ACM-GNN by +2.17 and +1.53 percentage points ( $p < 0.01$ ). Figure 3 visualizes the margin.

**Why does ELPG-DTFS outperform?** Temporal fidelity: Baselines like CNN+GRU treat spectrograms as static images, blurring transient events that signal affective dysregulation. Our window-level channel-band attention preserves millisecond-scale details, which chiefly raises Recall to 98.03%. Adaptive connectivity: Fixed graphs in ACM-GNN omit weak yet meaningful couplings. The learnable mask  $\mathbf{W}_\alpha$  reallocates weight to these edges, yielding a 2.17 pp accuracy gain. Hierarchical priors: Regional priors steer the model toward established biomarkers—such as frontal alpha asymmetry—improving precision, especially under limited sample size. Cross-band coupling: Mutual information captures nonlinear interactions (e.g., alpha–beta desynchronization) distinctive to MDD, refining decision boundaries and boosting F1.



**Fig. 3.** Mean performance and standard deviation over 10 folds on MODMA.

### 3.5. Ablation Study and Interpretation

Table 2 shows degradations when modules are removed, and the following explains observed drops:

**Table 2.** Ablation on MODMA (mean of 10 folds).

Variant	Acc(%)	Pre(%)	Rec(%)	F1(%)
Full ELPG-DTFS	<b>97.63</b>	<b>96.68</b>	<b>98.03</b>	<b>97.33</b>
– Prior knowledge	95.12	94.36	96.25	95.29
– Learnable adjacency	94.01	93.52	95.67	94.58
– MI	94.85	91.87	96.18	95.03
– Attention & MI	93.25	91.87	94.12	92.98

Prior knowledge: Both Precision and Recall fall, indicating that priors simultaneously reduce false positives and uncover true depressive cases by directing attention to neurobiologically grounded regions. Learnable adjacency: Recall drops most, showing that adaptive edges principally help detect otherwise subtle pathological connectivity. MI removal: Precision decreases, revealing MI’s role in filtering out healthy controls with benign spectral variations. Attention & MI removal: The greatest decline (4.35 pp F1) suggests dramatic loss of expressive power, leading to blurred class boundaries.

## 4. CONCLUSION

This paper presents ELPG-DTFS, a prior-guided adaptive graph neural network for EEG-based depression detection. It introduces channel–band attention for nonlinear cross-band dependencies, a learnable adjacency mask for dynamic functional graphs, and a multi-scale knowledge graph to incorporate clinical priors. On the MODMA benchmark, ELPG-DTFS achieves 97.63% accuracy and 97.33% F1, surpassing 14 recent baselines. Ablation confirms that all modules are indispensable, showing the value of combining signal processing, graph learning, and domain knowledge for objective depression diagnosis.



## 5. REFERENCES

- [1] W. Health Organization, *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization, 2017.
- [2] J. Yang, *Mental Health in China: Change, Tradition, and Therapeutic Governance*. John Wiley & Sons, 2017.
- [3] M. M. Katz, J. O. Cole, and H. A. Lowery, "Studies of the diagnostic process: The influence of symptom perception, past experience, and ethnic background on diagnostic decisions," *American Journal of Psychiatry*, vol. 125, no. 7, pp. 937–947, 1969.
- [4] U. Rajendra Acharya *et al.*, "Computer-aided diagnosis of depression using EEG signals," *European Neurology*, vol. 73, no. 5–6, pp. 329–336, 2015.
- [5] F. Soares De Aguiar Neto and J. Luís Garcia Rosa, "Depression biomarkers using non-invasive EEG: A review," *Neuroscience & Biobehavioral Reviews*, vol. 105, pp. 83–93, 2019.
- [6] H. Cai *et al.*, "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, no. 1, pp. 5238028, 2018.
- [7] A. Ksibi *et al.*, "Electroencephalography-based depression detection using multiple machine learning techniques," *Diagnostics*, vol. 13, no. 10, pp. 1779, 2023.
- [8] X. Li *et al.*, "EEG-based mild depressive detection using feature selection methods and classifiers," *Computer Methods and Programs in Biomedicine*, vol. 136, pp. 151–161, 2016.
- [9] X. Wang, D. Nie, and B. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in *International conference on neural information processing*, pp. 734–743, 2011.
- [10] S. Mahato and S. Paul, "Detection of major depressive disorder using linear and non-linear features from EEG signals," *Microsystem Technologies*, vol. 25, no. 3, pp. 1065–1076, 2019.
- [11] U. Rajendra Acharya *et al.*, "Automated EEG-based screening of depression using deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 103–113, 2018.
- [12] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, pp. ii2–ii7, 2005.
- [13] G. Sharma, A. Parashar, and A. M. Joshi, "DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression," *Biomedical Signal Processing and Control*, vol. 66, pp. 102393, 2021.
- [14] X. Chen *et al.*, "MGSN: Depression EEG lightweight detection based on multiscale DGCN and SNN for multichannel topology," *Biomedical Signal Processing and Control*, vol. 92, pp. 106051, 2024.
- [15] Z. He *et al.*, "A novel unsupervised domain adaptation framework based on graph convolutional network and multi-level feature alignment for inter-subject ECG classification," *Expert Systems with Applications*, vol. 221, pp. 119711, 2023.
- [16] D. Yao *et al.*, "Temporal-adaptive graph convolutional network for automated identification of major depressive disorder using resting-state fMRI," in *International Workshop on Machine Learning in Medical Imaging*, pp. 1–10, 2020.
- [17] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2017.
- [18] Y. Pang, Y. Zhao, and D. Li, "Graph pooling via coarsened graph infomax," in *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pp. 2177–2181, 2021.
- [19] Y. Wang, F. Liu, and L. Yang, "EEG-based depression recognition using intrinsic time-scale decomposition and temporal convolution network," in *The fifth international conference on biological information and biomedical engineering*, pp. 1–6, 2021.
- [20] T. Chen *et al.*, "Exploring self-attention graph pooling with EEG-based topological structure and soft label for depression detection," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2106–2118, 2022.
- [21] W. Liu *et al.*, "A depression prediction algorithm based on spatiotemporal feature of EEG signal," *Brain Sciences*, vol. 12, no. 5, pp. 630, 2022.
- [22] G. Tasci *et al.*, "Automated accurate detection of depression using twin Pascal's triangles lattice pattern with EEG signals," *Knowledge-Based Systems*, vol. 260, pp. 110190, 2023.
- [23] H. Wang *et al.*, "Amg: A depression detection model with autoencoder and multi-head graph convolutional network," in *2023 42nd Chinese Control Conference (CCC)*, pp. 8551–8556, 2023.
- [24] V. P. Kowli and H. P. Padole, "Detection of major depressive disorder using EEG multi-feature fusion," *IEEE Sensors Journal*, 2024.
- [25] L. Zhang *et al.*, "Robust sparse graph regularized nonnegative matrix factorization for automatic depression diagnosis," *Biomedical Signal Processing and Control*, vol. 93, pp. 106036, 2024.
- [26] X. Chen *et al.*, "MGSN: Depression EEG lightweight detection based on multiscale DGCN and SNN for multichannel topology," *Biomedical Signal Processing and Control*, vol. 92, pp. 106051, 2024.
- [27] Z. Zhang *et al.*, "A novel EEG-based graph convolution network for depression detection: incorporating secondary subject partitioning and attention mechanism," *Expert Systems with Applications*, vol. 239, pp. 122356, 2024.
- [28] Y. Li *et al.*, "EEG-based depression recognition using feature selection method with fuzzy label," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 3, pp. 102004, 2024.
- [29] M. Dia *et al.*, "Multichannel convolutional transformer for detecting mental disorders using electroencephalography records," *Scientific Reports*, vol. 15, 2025.
- [30] X. Chen *et al.*, "GNMixer: A high-density EEG signal processing method using P2GNN and MLP-Mixer for depression detection," *IEEE Sensors Journal*, pp. 1–1, 2025.
- [31] T. Wang *et al.*, "Explainable multi-frequency and multi-region fusion model for affective brain-computer interfaces," *Information Fusion*, vol. 118, pp. 102971, 2025.
- [32] T. Zhang *et al.*, "ACM-GNN: Adaptive cluster-oriented modularity graph neural network for EEG depression detection," *IEEE Transactions on Computational Social Systems*, 2025.
- [33] K. H. Jawabri and S. Sharma, "Physiology, cerebral cortex functions," *StatPearls [internet]*, 2023.