Towards Recognizing Spatial-temporal Collaboration of EEG Phase Brain Networks for Emotion Understanding

Jiangfeng Sun[1], **Kaiwen Xue**[1], **Qika Lin**[2], **Yufei Qiao**[3], **Yifan Zhu**[1*], **Zhonghong Ou**[1] and **Meina Song**[1,4]

[1]School of Computer Science, Beijing University of Posts and Telecommunications
[2]Saw Swee Hock School of Public Health, National University of Singapore
[3]Department of Otorhinolaryngology, Peking Union Medical College Hospital
[4]China University of Petroleum - Beijing at Karamay
{sun2017, xkw, yifan_zhu, zhonghong.ou, mnsong}@bupt.edu.cn, qikalin@foxmail.com, qiaoyufei1219@163.com

Abstract

Emotion recognition from EEG signals is crucial for understanding complex brain dynamics. Existing methods typically rely on static frequency bands and graph convolutional networks (GCNs) to model brain connectivity. However, EEG signals are inherently non-stationary and exhibit substantial individual variability, making static-band approaches inadequate for capturing their dynamic properties. Moreover, spatial-temporal dependencies in EEG often lead to feature degradation during node aggregation, ultimately limiting recognition performance. To address these challenges, we propose the **S**patial-**T**emporal **E**lectroencephalograph Co**lla**boration framework (**Stella**). Our approach introduces an Adaptive Bands Selection module (ABS) that dynamically extracts low- and high-frequency components, generating dual-path features comprising phase brain networks for connectivity modeling and time-series representations for local dynamics. To further mitigate feature degradation, the Fourier Graph Operator (FGO) operates in the spectral domain, while the Spatial-Temporal Encoder (STE) enhances representation stability and density. Extensive experiments on benchmark EEG datasets demonstrate that Stella achieves state-of-the-art performance in emotion recognition, offering valuable insights for graph-based modeling of non-stationary neural signals. The code is available at https://github.com/sun2017bupt/EEGBrainNetwork.

1 Introduction

Decoding emotional states from brain activity is a fundamental challenge at the intersection of neuroscience and artificial intelligence [Pessoa and Adolphs, 2010; Huang *et al.*, 2017; Kragel and LaBar, 2016]. Non-invasive neuroimaging techniques such as EEG, fMRI, and MEG have enabled fine-grained analysis of brain dynamics associated with emotional and cognitive processes [Meer *et al.*, 2020; He and Zhang, 2024; Flournoy *et al.*, 2024]. Among these, EEG stands out due to its high temporal resolution, which is crucial for capturing the rapid neural fluctuations underlying affective states. Nevertheless, accurately recognizing emotions from EEG signals remains a formidable task, as it requires advanced models capable of capturing complex spatial-temporal interactions within brain networks.

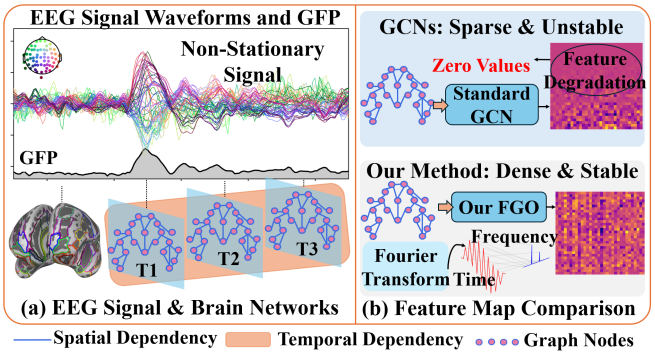

Figure 1: EEG signals and brain networks. (a) Non-stationary EEG signals can be represented as time-series brain networks. The Global Field Potential (GFP) summarizes fluctuating brain activity. (b) Standard GCNs suffer from feature degradation, while the Fourier Graph Operator (FGO) preserves frequency-domain features, producing stable and dense feature maps.

Current EEG-based emotion recognition methods typically employ Region-of-Interest (ROI) strategies, which partition the brain into predefined regions and frequency bands. While widely used, such approaches often fail to capture the dynamic and collaborative nature of emotional neural processes [Pessoa, 2017; Pessoa, 2018; Zheng *et al.*, 2023]. Although state-of-the-art time-series models have shown strong capability in capturing local temporal dynamics [Ju *et al.*, 2024; Si *et al.*, 2024], they are inherently limited in modeling inter-regional synchronization—an essential feature for decoding emotion-related brain activity. This limitation aligns with our core motivation: going beyond sequential modeling

* Corresponding author.

to explicitly account for spatial-temporal dependencies across brain regions. As a result, existing approaches still struggle to fully leverage the structure of brain networks, substantially restricting their performance in emotion decoding.

Brain network-based methods have shown strong potential in cognitive prediction and neurological diagnosis [Alarcao and Fonseca, 2017; Zhu *et al.*, 2021; Shi *et al.*, 2023]. However, applying them to EEG-based emotion recognition presents two key challenges. First, constructing effective brain networks requires capturing the dynamic, non-stationary nature of EEG signals. Most existing methods rely on static frequency bands, which constrain electrode interactions to fixed ranges [Miao *et al.*, 2023] and overlook individual spatial-temporal variability (Fig.1-(a)). Adaptive frequency selection, guided by physiological priors, offers a promising alternative[Liu *et al.*, 2024a]. Second, critical spatial-temporal dependencies often degrade during node aggregation, especially when temporal asynchronies between brain regions—vital for emotional processing—are not well preserved (Fig.1-(b)). Addressing these challenges calls for advanced techniques that can jointly model spatial-temporal interactions[Liu *et al.*, 2024b].

To address these challenges, we propose the **S**patial-**T**emporal **E**lectroencephalograph Co**lla**boration (**Stella**) framework. Specifically, Stella introduces Adaptive Bands Selection (ABS) to dynamically partition EEG signals into optimal frequency bands, effectively constructing Phase Brain Networks (PBNs) to capture inter-region synchronization. Additionally, we design a dual-path architecture integrating a Fourier Graph Operator (FGO) to mitigate feature degradation, and a Spatial-Temporal Encoder (STE) to strengthen spatial-temporal representations. Comprehensive experiments on benchmark datasets (DEAP and FACED) demonstrate that Stella achieves state-of-the-art performance. Our key contributions are:

- **Adaptive Bands Selection (ABS)**: A data-driven approach that dynamically divides EEG signals into base and harmonic frequency bands, enabling robust Phase Brain Network (PBN) construction through phase synchronization.
- **Stella Framework**: A dual-path architecture integrating the Fourier Graph Operator (FGO) to reduce feature degradation and the Spatial-Temporal Encoder (STE) to strengthen spatial-temporal representations.
- **Comprehensive Evaluation**: Extensive experiments on public EEG datasets show that Stella achieves state-of-the-art emotion recognition, supported by detailed ablation and sensitivity analyses.

2 Related Work

In this section, we review key research areas relevant to our approach, including frequency band division, graph learning from brain networks, and spatial-temporal learning from EEG signals. These methods have been widely explored to enhance the analysis of EEG signals, with each contributing to the understanding of brain activity and emotion recognition.

Frequency band division. Existing methods are mainly categorized into filter-based methods driven by physiological priors and data-driven filter methods. Filter-based methods driven by physiological priors rely on well-defined physiological frequency bands, such as Delta (0.5-4Hz), Theta (4-8Hz), Alpha (8-13Hz), Beta (13-30Hz), and Gamma (above 30Hz) [Jenke *et al.*, 2014]. Common implementations include bandpass filters [Li *et al.*, 2023], wavelet transforms [Akin, 2002], and Fourier transforms [Lin *et al.*, 2010], which effectively extract the corresponding features through predefined frequency bands. Although static frequency band divisions align with physiological paradigms, they may not fully account for individual differences, limiting the model's adaptability. On the other hand, data-driven methods, while capable of handling non-stationary signals and capturing dynamic variations, do not rely on physiological plausibility, which can affect their interpretability and alignment with known brain functions.

Graph learning from EEG brain networks. Graph Convolutional Network (GCN) extend traditional Convolutional Neural Network (CNN) by leveraging spectral graph theory, enabling the direct processing of non-Euclidean data, such as graphs [Kipf and Welling, 2016]. Despite the impressive successes of GCN in emotion recognition, the static nature of their adjacency matrices limits their ability to adapt dynamically. Dynamic Graph Convolutional Neural Networks (DGCN) [Song *et al.*, 2018] is a GCN-based method that models multichannel EEG features using a dynamic adjacency matrix, enabling the dynamic learning of intrinsic relationships between EEG channels for enhanced emotion recognition. [Wang *et al.*, 2018] integrate the strengths of dynamic adjacency matrix and Broad Learning Systems to design a novel architecture. [Zhang *et al.*, 2021] impose sparsity constraints on graph to enhance emotion recognition. [Gu *et al.*, 2023] use a Generative Adversarial Network to generate latent representations of EEG signals while combining a GCN and a Long Short-Term Memory (LSTM) network to recognize emotions from EEG signals. [Li *et al.*, 2023] combine single-channel Differential Entropy (DE) features with cross-channel functional connectivity features to simultaneously extract temporal variations and spatial topology information from EEG data. However, existing approaches tend to ignore the feature degradation problem and underestimate the effect of channel heterogeneity in feature engineering, reducing the efficiency of graph learning.

Spatial-Temporal Learning from EEG Signals. Many EEG-based emotion recognition methods adopt hybrid architectures (e.g., GCN + LSTM, CNN + RNN) to model spatial-temporal patterns [Li *et al.*, 2016]. Yang et al. [Yang *et al.*, 2018] and Ma et al. [Ma *et al.*, 2019] integrate CNNs and RNNs for joint feature learning, with Ma further introducing residual shortcuts. Tao et al. [Tao *et al.*, 2020] enhance spatial focus via channel attention, while Yin et al. [Yin *et al.*, 2021] and Liu et al. [Liu *et al.*, 2024a] fuse GCNs with RNNs to better capture spatial-temporal dependencies. However, these composite designs often apply temporal modeling post hoc to graph-based features, limiting direct temporal reasoning at the raw signal level. Recent work such as DeepCN [Yi *et al.*, 2024] emphasizes joint modeling of intra- and inter-series couplings, offering deeper spatial-temporal insights. Moreover, research beyond EEG, including OptIForest [Xi-

Algorithm 1 Adaptive Bands Selection

Input: Original EEG signal $x[n]$
Output: Selected bands features $\mathcal{D}$

Initialize: window function w = *Hamming*, window length L, number of windows $\mathcal{S}$, harmonic order h
Apply STFT to EEG signal $x[n]$ to obtain $\mathcal{X}(f)$
for each channel i **do**
  for each window $l \in L$ **do**
    $fb_i^l \leftarrow \arg\max_f |\mathcal{X}_i(f)^l|$
  end for
  $fb_i^s \leftarrow \frac{1}{L}\sum_{l=1}^{L} fb_i^l$, $fh_i^s \leftarrow (fb_i^s)^h$
  $\mu_{b_i}, \sigma_{b_i} \leftarrow$ mean and std of fb_i^s
  $\mu_{h_i}, \sigma_{h_i} \leftarrow$ mean and std of fh_i^s
  Compute frequency boundaries using μ and σ: $fb_{low,i}, fb_{high,i}, fh_{low,i}, fh_{high,i}$
  if $fh_{low,i} < fb_{high,i}$ **then**
    Adjust boundaries: $fh_{low,i} \leftarrow fb_{high,i}$ if necessary
  end if
end for
return $\mathcal{D}$ for the selected bands across all channels

ang *et al.*, 2023] and federated isolation forests [Xiang *et al.*, 2024], highlights the importance of capturing multi-source, distributed, and cross-domain dependencies in time series data. These perspectives inspire more integrated and system-level approaches to EEG-based spatial-temporal modeling.

3 The Stella Framework

3.1 Overall Architecture

As illustrated in Fig. 2, we introduce the Stella framework to effectively model spatial-temporal features in EEG signals. The proposed method comprises two key components: Adaptive Bands Selection (ABS) and a dual-path architecture integrating the Fourier Graph Operator (FGO) and Spatial-Temporal Encoder (STE). Specifically, ABS adaptively divides EEG signals into optimal Base and Harmonic frequency bands, addressing the limitation of static band definitions. The Base Band is utilized by FGO to construct robust Phase Brain Networks (PBNs), capturing stable inter-regional connectivity, while the Harmonic Band is processed by STE to extract detailed spatial-temporal dynamics, effectively mitigating feature degradation.

3.2 Adaptive Bands Selection

Frequency Filter Method. While traditional approaches often rely on static frequency bands such as the γ band for emotion recognition, emerging evidence indicates that dynamic, individualized frequency bands are critical for accurately capturing neural dynamics, highlighting the limitations of a fixed-band strategy.

The Adaptive Bands Selection (ABS) module leverages signal decomposition theory to adaptively identify frequency bands across EEG channels and individuals. Using the Short-Time Fourier Transform (STFT), the raw EEG signal is decomposed into the frequency domain. For each electrode, we extract the dominant frequency (base frequency) from each windowed segment and identify its harmonics as integer multiples. This decomposition enables personalized frequency band selection that captures individual-specific spectral features. The formal process is defined in Eq. 1.

$$\begin{aligned} \mathcal{X}_m(f) &= \sum_n x[n]\, W[n-m]\, e^{-j\frac{2\pi}{N}fn}, \\ fb_i^l &= \arg\max_f \left|\mathcal{X}_i^l(f)\right|, \\ fb_i^s &= \frac{1}{L}\sum_{l=1}^{L} fb_i^l, \end{aligned} \tag{1}$$

where $x[n]$ denotes the raw EEG signal, $W(\cdot)$ is a window function centered at time m, and $f \in \{0, \ldots, N-1\}$ is the frequency bin. $\mathcal{X}_i^l(f)$ is the STFT of electrode i at window l, and fb_i^l denotes its dominant frequency. The final base frequency fb_i^s is obtained by averaging across all L windows.

As detailed in Alg. 1, due to inter-channel variability in EEG signals, base and harmonic frequencies are computed on a per-channel basis. For each channel i, the average base frequency is calculated as $\mu_{b_i} = \frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}} fb_i^s$, with standard deviation $\sigma_{b_i} = \frac{1}{\mathcal{S}}\sqrt{\sum_{s=1}^{\mathcal{S}}(fb_i^s - \mu_{b_i})^2}$. The base band boundaries are defined as $fb_{low,i} = |\mu_{b_i} - \sigma_{b_i}|$ and $fb_{high,i} = |\mu_{b_i} + \sigma_{b_i}|$. Similarly, the harmonic frequency is computed as $fh_i^s = (fb_i^s)^h$, with harmonic band boundaries $fh_{low,i} = |\mu_{h_i} - \sigma_{h_i}|$ and $fh_{high,i} = |\mu_{h_i} + \sigma_{h_i}|$. This adaptive, channel-specific filtering ensures personalized and robust frequency band extraction.

Phase Brain Networks. Given the non-periodic nature of EEG signals, we construct brain networks using a sliding window strategy. To capture the topological relationships between brain regions, we compute the Phase Locking Value (PLV), which quantifies the functional connectivity between electrode pairs. Phase synchronization has been widely recognized in neuroscience as a robust indicator of functional brain connectivity, particularly in the context of cognitive and emotional coordination.

The phase synchronization matrix (PSM) $\mathcal{G}$ serves as the adjacency matrix of a weighted undirected graph, where stronger phase synchronization corresponds to higher edge weights. In accordance with common practice in EEG-based analysis, we use Differential Entropy (DE) features as a compact and effective representation of EEG signals in each frequency band. DE quantifies the information content of a continuous-valued signal and has been widely adopted as a standard pre-processing step in EEG emotion recognition due to its superior discriminability [Zheng and Lu, 2015]. Let $\mathcal{D}_p(\mathcal{X})$ and $\mathcal{D}_q(\mathcal{X})$ denote the DE features extracted from channels p and q, respectively. Within time window s, the instantaneous phase of each DE time series is computed via the Hilbert transform, denoted as $\phi_p(s, \mathcal{D}_p)$ and $\phi_q(s, \mathcal{D}_q)$. The phase synchronization between these two electrodes in

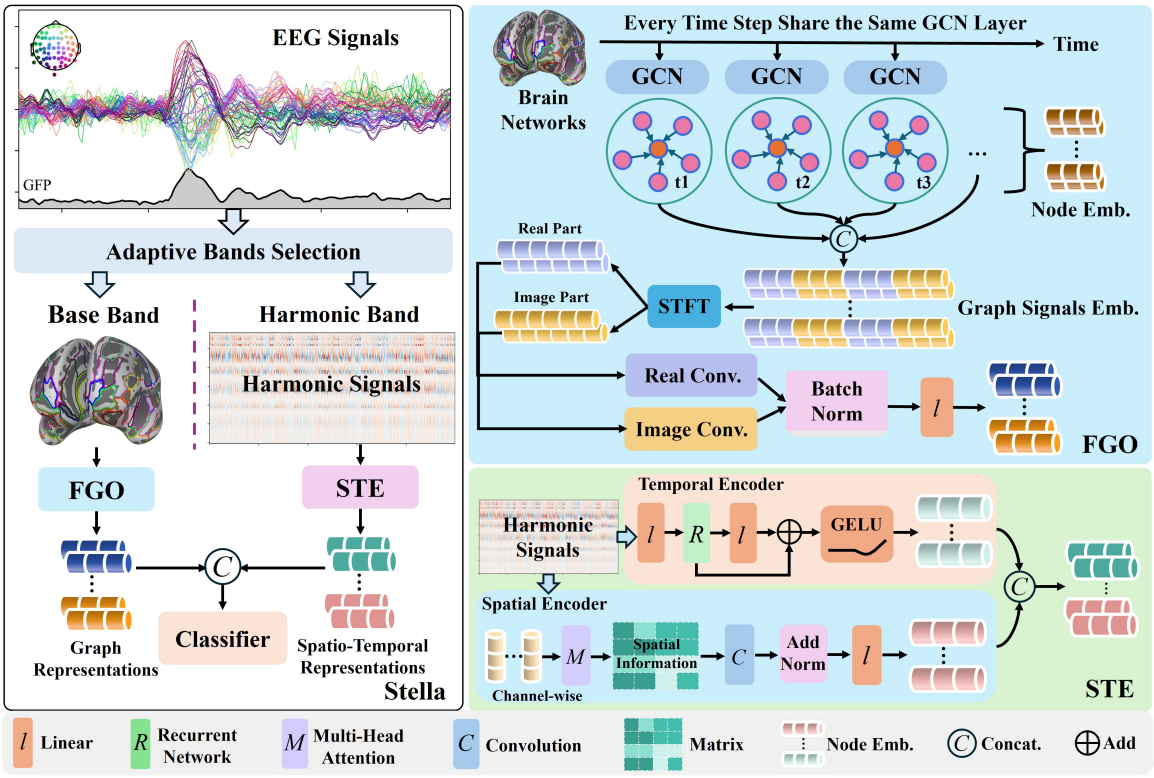

Figure 2: Overview of Stella framework. EEG signals are processed with Adaptive Bands Selection (ABS) to filter Base Band and Harmonic Band. The Fourier Graph Operator (FGO) uses the Base Band to capture graph representations, while the Spatial-Temporal Encoder (STE) processes the Harmonic Band to extract spatial-temporal features. The outputs of both components are combined and passed through a classifier for emotion recognition.

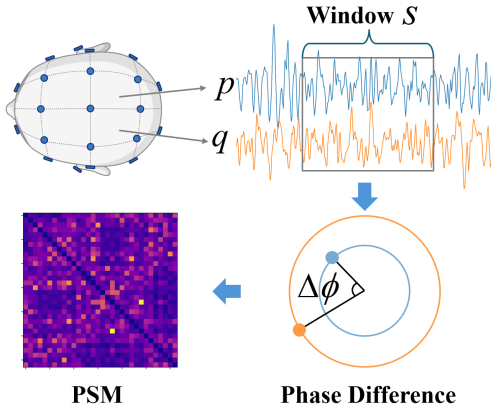

Figure 3: The overall process of constructing the phase synchronization matrix between channels as brain networks using EEG signals.

window s is then computed as:

$$\begin{aligned}\Delta\phi^s_{pq}(s,\mathcal{D}) &= \phi_p(s,\mathcal{D}) - \phi_q(s,\mathcal{D}) \\ &= \angle\left(e^{j(\mathcal{D}^s_p - \mathcal{D}^s_q)}\right), \\ PLV^s_{pq}(f) &= \left|\Im\left(\frac{1}{N}\sum_{n=1}^{N} e^{j\Delta\phi^s_{pq}(s,\mathcal{D})}\right)\right|.\end{aligned} \tag{2}$$

Here, $\angle(X)$ denotes the phase of X, $e^{j\theta}$ represents a complex exponential with phase θ, and $\Im(X)$ denotes the imaginary part of X. To prevent numerical instability, we apply logarithmic normalization on the PLV with a small constant α, as expressed in Eq.3.

$$NormalizedPLV^s_{pq} = \log(PLV^s_{pa}(f) + \alpha)/\log(0.5). \tag{3}$$

The base frequency, as the dominant and more stable component of EEG signals, is less affected by noise and is thus used to construct the phase synchronization matrix $\mathcal{G}$ to reflect functional connectivity between brain regions. Harmonic components, being more noise-sensitive, are excluded from this construction. Analysis on the DEAP dataset shows that the extracted base frequency band ranges from $[7, 16]$ Hz, covering traditional θ, α, and β bands—while harmonics span $[41, 58]$ Hz, corresponding to γ activity. These bands are associated with distinct cognitive and emotional states: θ with relaxation and light sleep, α with calm wakefulness, β with active thinking, and γ with high-level cognition and sensory processing. The ABS module naturally aligns with this five-band theory, providing a principled basis for adaptive frequency band selection in EEG analysis.

3.3 Model Design

As illustrated in Fig. 2, we design a dual-path deep learning model that integrates Fourier Graph Operator (FGO) with

a Spatial-Temporal encoder (STE) to produce the stable and dense representations.

Fourier Graph Operator. In the temporal graph model, we use the constructed temporal phase brain networks $\{\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_n\}$ as input and approximate the Laplacian matrix using Chebyshev polynomials for graph convolution. The recursive definition of Chebyshev polynomials $T_k(x)$ is defined as Eq. 4. We adopt Chebyshev-approximated spectral GCNs rather than GATs for both theoretical and practical reasons. First, our PLV-based brain graphs are inherently defined in the frequency domain, making spectral methods like GCNs more aligned with the underlying graph construction. In contrast, GATs operate in the spatial domain and rely on learned attention weights across node pairs, which can be less interpretable and harder to constrain physiologically in brain network analysis. Second, Chebyshev polynomial approximation allows efficient localized filtering, avoids eigendecomposition, and enhances scalability.

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x). \tag{4}$$

The Chebyshev polynomial approximation of the graph Laplacian matrix is given by Eq.5.

$$g_\theta(\hat{L}) \simeq \sum_{k=0}^{K} \theta_k T_k(\hat{L}), \tag{5}$$

where θ represents the coefficients of the Chebyshev polynomials, and K denotes the order of the polynomial. In this paper, a 5th-order Chebyshev polynomial is used to approximate the Laplacian matrix. The graph convolution operation is expressed as Eq.6.

$$\hat{X} = g_\theta(\hat{L})X \simeq \sum_{k=0}^{K} \theta_k T_k(\hat{L})X. \tag{6}$$

After graph convolution, the resulting temporal features, $\hat{X}_1, \hat{X}_2, ..., \hat{X}_t$, suffer from feature degradation, a typical issue with standard graph convolution due to the heterogeneity of EEG signals. To address this, we apply Short-Time Fourier Transform (STFT) to process these signals in the complex frequency domain, mitigating the degradation problem. The STFT operation is given by Eq.7.

$$STFT(\hat{X}_t)(\tau, f) = \sum_{n=-\infty}^{\infty} \hat{X}_n \cdot \omega(n-\tau) \cdot e^{-jfn}, \tag{7}$$

where $\hat{X}_n$ is the time series signal at time n, $\omega(\cdot)$ is the window function, τ is the central time point in this window and f is the frequency variable.

To perform frequency domain analysis, we apply a STFT to transition the temporal features to the frequency domain, yielding real and imaginary components. The complex convolution is then applied to these components, with the input signal $x = x_{real} + jx_{imag}$ and the convolution kernel $h = h_{real} + jh_{imag}$. The convolution operation is $y = (x_{real}+jx_{imag})*(h_{real}+jh_{imag})$, where real and imaginary parts are convolved independently using real-valued convolutions, with kernels h_{real} and h_{imag}. Then, a batch norm is applied followed by a linear projection W to produce the graph representation, as shown in Eq.8.

$$Output_{\text{graph}} = \text{BatchNorm}(y)W \tag{8}$$

Spatial-Temporal Encoder. The STE processes the EEG signals to capture both spatial and temporal dependencies. Channel-wise Multi-Head Attention is applied to model spatial relationships in the harmonic EEG signals. Each channel is treated independently to focus on the inter-channel relationships, allowing the model to capture connectivity between brain regions. To further enhance spatial feature extraction, convolution is applied to these channel-wise representations, capturing multi-scale spatial relationships across the EEG channels. The operations are as follows: for the input matrix $H \in R^{B\times D\times C}$, where B is the batch size, C is the number of channels, and D is the feature dimension, and W is the linear projection, then we compute the attention scores for each channel using $Q_i = XW_Q^i$, $K_i = XW_K^i$, $V_i = XW_V^i$. The attention scores for each channel are then computed as Eq.9.

$$e_i = \text{softmax}\left(\frac{Q_i(K_i)^T}{\sqrt{d_k}}\right) \tag{9}$$

and the output for each attention head is $head_i = e_iV_i$. The outputs of all heads are concatenated and passed through a linear transformation as Eq.10.

$$Output_{\text{attn}} = \text{Concat}(head_1, ..., head_n)W_o. \tag{10}$$

As expressed in Eq.11, the attention output is then processed by a convolution operation to further capture the spatial features, providing the output of the spatial encoding.

$$Output_{\text{s}} = \text{Conv}(Output_{\text{attn}}). \tag{11}$$

The temporal dependencies are captured mainly using a 3-layer RNN, which models the sequential nature of the EEG signal over time. The RNN is implemented using an off-the-shelf implementation in PyTorch, and details are omitted for brevity. The output of temporal encoding is given by Eq.12.

$$Output_{\text{t}} = \text{GELU}(W_{\text{output}}(\text{RNN}(XW_{\text{project}}))). \tag{12}$$

Finally, the spatial and temporal features are concatenated, and a linear layer is applied to fuse them into a single representation, as shown in Eq.13.

$$Output_{\text{st}} = \text{Concat}(Output_{\text{s}}, Output_{\text{t}})W_{\text{fusion}} \tag{13}$$

Above all, by combining outputs of FGO and STE modules, we obtain features enriched with brain networks and spatial-temporal information from the EEG signals. As shown in Eq.14, fusing these features results in a stable, dense representation that significantly enhances emotion recognition efficiency.

$$Output_{\text{fusion}} = \text{Concat}(Output_{\text{graph}}, Output_{\text{st}}) \tag{14}$$

Dataset	DEAP				FACED			
Metrics	Valence		Arousal		Valence		Arousal	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
ECLGCN	90.45%±3.09%	90.01%±3.13%	90.60%±2.63%	90.01%±3.33%	88.94%±4.19%	88.01%±18.18%	87.91%±4.74%	87.87%±16.61%
PCRNN	90.80%±3.08%	89.89%±2.92%	91.03%±2.99%	90.03%±3.07%	89.87%±5.67%	88.64%±18.59%	88.92%±4.24%	87.02%±17.67%
MM-ResLSTM	92.30%±1.55%	91.01%±4.23%	92.87%±2.11%	90.95%±3.28%	87.88%±6.25%	85.59%±27.32%	88.23%±6.65%	88.05%±24.12%
ACRNN	93.72%±3.21%	92.64%±3.33%	93.38%±3.73%	92.57%±3.64%	89.92%±5.31%	88.54%±24.21%	88.20%±4.31%	87.92%±22.39%
BiDCNN	94.38%±2.61%	93.25%±3.32%	94.72%±2.57%	93.83%±3.13%	91.25%±4.44%	90.02%±16.23%	90.98%±5.21%	90.01%±17.75%
ST-GCLSTM	95.52%±0.96%	93.08%±3.43%	95.04%±0.95%	90.94%±2.53%	94.52%±3.35%	92.05%±15.42%	92.32%±4.78%	92.21%±10.16%
STFCGAT	95.70%±3.36%	94.53%±4.22%	94.24%±3.95%	93.94%±4.57%	93.82%±2.15%	91.65%±19.82%	92.97%±3.98%	92.74%±19.16%
BF-GCN	96.33%±2.23%	95.69%±3.01%	97.69%±2.56%	97.01%±2.74%	91.62%±3.26%	92.12%±21.43%	92.75%±3.99%	92.86%±20.83%
DGGN*	96.98%±2.23%	96.72%±3.01%	97.19%±2.56%	97.01%±2.74%	95.12%±3.26%	92.12%±21.43%	93.87%±3.99%	92.86%±20.83%
Stella (Ours)	**97.44%±1.74%**†	**97.31%±1.96%**†	**98.11%±1.07%**†	**98.02%±1.08%**†	**96.98%±2.70%**††	**93.54%±17.40%**††	**95.83%±2.92%**††	**94.01%±12.49%**††
Acc ↑/ Std ↓	**0.46% / 0.49%**	**0.59% / 1.05%**	**0.91% / 1.49%**	**1.10% / 1.66%**	**1.85% / 0.56%**	**1.42% / 4.03%**	**1.96% / 1.07%**	**1.15% / 8.34%**

Table 1: Overall performance of Stella and baselines. † indicates the statistical significance (†: $p\text{-}value < 0.05$; ††: $p\text{-}value < 0.01$). And * is the baseline model we selected for the main comparison. Arrow ↑ / ↓ means higher/lower is better.

F	DEAP		FACED	
	Valence	Arousal	Valence	Arousal
-	92.16% ± 9.04%	89.03% ± 14.09%	69.54% ± 22.12%	66.34% ± 23.71%
✓	**97.44%±1.74%**	**98.11% ± 1.07%**	**96.98% ± 2.70%**	**95.83% ± 2.92%**

Table 2: Ablation Study in FGO. (**F** means STFT, - means removing STFT from GCN, ✓ means using STFT after GCN.)

3.4 Loss Function

EEG data is inherently noisy and often limited in quantity, making it prone to overfitting under standard cross-entropy loss, which relies on hard labels. To improve generalization, we adopt a label smoothing strategy that incorporates a smoothing factor $\epsilon \in [0, 1]$, transforming hard labels into soft ones. The overall loss consists of two components: the standard cross-entropy loss $\mathcal{L}_1$, capturing model's confidence in the correct class, and a smoothing term $\mathcal{L}_2$, promoting uniformity across all classes. The total loss is defined as:

$$\begin{aligned} \mathcal{L}_1 &= -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i, \\ \mathcal{L}_2 &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C} \log \hat{y}_{i,j}, \\ \mathcal{L}(\hat{y}, y) &= (1-\epsilon)\mathcal{L}_1 + \frac{\epsilon}{C}\mathcal{L}_2, \end{aligned} \tag{15}$$

where n is batch size, C is number of classes (e.g., $C = 2$ for valence and arousal classification), $\hat{y}_{i,j}$ is the predicted probability for class j, and y_i is the true label of sample i.

4 Experiments

4.1 Datasets

This section presents extensive experiments on two widely used EEG emotion recognition datasets:

- **DEAP** [Koelstra *et al.*, 2011]: EEG (40 channels, 128Hz) from 32 participants watching 40 music videos, with ratings on arousal, valence, dominance, liking, and familiarity, plus facial and physiological data.
- **FACED** [Chen *et al.*, 2023]: EEG (32 channels, 1000Hz) from 123 participants watching 28 emotion-eliciting videos, covering nine emotion categories, with ratings on arousal, valence, familiarity, and liking.

4.2 Baselines

We compare the performance of Stella against the following EEG-based emotion recognition baselines:

- **Graph models**: ECLGCN [Yin *et al.*, 2021], ST-GCLSTM [Feng *et al.*, 2022], STFCGAT [Li *et al.*, 2023], DGGN [Gu *et al.*, 2023], and BF-GCN [Li, 2024], which use Graph-based models to capture spatial-temporal EEG features.
- **Spatial-temporal models**: PCRNN [Yang *et al.*, 2018], ACRNN [Tao *et al.*, 2020], MM-ResLSTM [Ma *et al.*, 2019], and BiDCNN [Huang *et al.*, 2021], which leverage RNN or CNN architectures, with attention, multimodal integration, or hemispheric modeling.

4.3 Main Experimental Results

We compare Stella against strong baselines on the DEAP and FACED datasets for valence and arousal classification. As shown in Table 1, Stella consistently outperforms the best baseline, DGGN, improving valence accuracy by 0.46% and arousal accuracy by 0.91% on DEAP, and by 1.85% and 1.96% on FACED. It also achieves notable gains in F1 scores and reduces standard deviations across metrics. These improvements are statistically significant ($p < 0.05$ on DEAP, $p < 0.01$ on FACED), demonstrating both superior performance and stability. While some baselines show lower variance, their average accuracy is markedly lower, underscoring Stella's robustness. Furthermore, the smaller performance gap between DEAP and FACED indicates strong generalization across datasets, effectively handling FACED's higher heterogeneity.

4.4 Ablation Study

We conducted an ablation study to evaluate the contributions of the Fourier Graph Operator (FGO) and the Spatial-Temporal Encoder (STE). As shown in Table 3, removing FGO leads to substantial performance degradation—on DEAP, valence and arousal accuracy drop from 97.44%/98.11% to 85.75%/93.56%; on FACED, from

FGO	STE	DEAP				FACED			
		Valence		Arousal		Valence		Arousal	
		Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
-	✓	75.39%±9.91%	0.7453±0.1057	81.07%±8.02%	0.8093±0.0811	83.36%±13.64%	0.7979±0.2002	84.53%±13.94%	0.8272±0.1754
✓	-	85.75%±5.78%	0.8546±0.0616	93.56%±4.11%	0.9350±0.0417	94.65%±4.95%	0.9124±0.1746	92.90%±4.85%	0.9118±0.1267
✓	✓	**97.44%±1.74%**	**0.9731±0.0196**	**98.11%±1.07%**	**0.9812±0.0107**	**96.98%±2.70%**	**0.9354±0.1740**	**95.83%±2.92%**	**0.9401±0.1249**

Table 3: Ablation Study for FGO and STE using the same metrics as in Table.1.

Freq. Bands	DEAP				FACED			
	Valence		Arousal		Valence		Arousal	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
ABS	**97.44%±1.74%**	**0.9731±0.0196**	**98.11%±1.07%**	**0.9812±0.0107**	**96.98%±2.70%**	**0.9354±0.1740**	**95.83%±2.92%**	**0.9401±0.1249**
δ	95.29%±1.36%	0.9386±0.0241	96.17%±1.27%	0.9418±0.0349	93.48%±1.71%	0.9245±0.0195	93.97%±2.19%	0.8791±0.1596
θ	91.46%±2.01%	0.8806±0.0661	90.69%±3.01%	0.8618±0.0635	93.35%±1.69%	0.923±0.0190	93.91%±2.05%	0.8745±0.1716
α	91.98%±1.82%	0.8906±0.0502	90.78%±3.0%	0.8747±0.0460	93.47%±1.73%	0.9237±0.0208	94.01%±2.52%	0.8841±0.1546
β	93.11%±1.48%	0.9163±0.0291	91.44%±2.89%	0.8804±0.0518	93.26%±1.78%	0.9224±0.0199	94.04%±2.06%	0.8878±0.1496
γ	92.51%±2.22%	0.8954±0.0576	92.56%±2.21%	0.8946±0.0429	93.86%±1.80%	0.9292±0.0207	94.07%±2.30%	0.8769±0.1693

Table 4: ABS Bands VS. Classical Bands Using The Same Metrics as in Table.1.

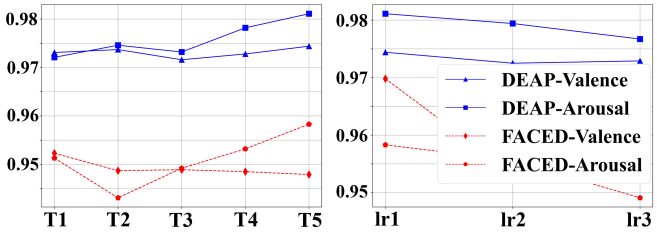

Figure 4: Model accuracy for DEAP and FACED variation curves under different history lengths and learning rates.

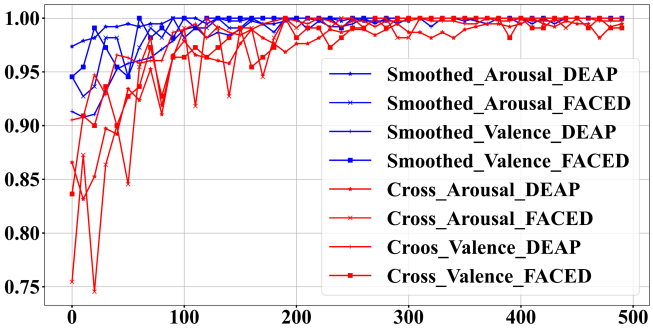

Figure 5: The accuracy curve under smoothed and non-smoothed cross entropy loss function with same environment and model configuration.

96.98%/95.83% to 94.65%/92.90%. This confirms the importance of modeling inter-regional brain synchronization. Removing STE also reduces performance, though less severely (e.g., DEAP valence down to 75.39%, FACED arousal to 84.53%). Notably, FGO alone performs better than STE alone across all metrics, suggesting that low-level brain coordination is more critical than local temporal encoding. However, the best performance is achieved only when both modules are combined, demonstrating their complementarity and the necessity of joint spatial-temporal modeling for accurate emotion decoding.

4.5 Parameter Sensitivity Analysis

To assess the framework's robustness, we analyze its sensitivity to key hyperparameters, such as the historical time period T (DE feature sequence length) and learning rate lr. Experiments on FACED ($T = 11, 13, 15, 17, 19$) and DEAP ($T = 30, 32, 34, 36, 38$) with $lr = 1e\text{-}5, 3e\text{-}5, 5e\text{-}5$ demonstrate stable performance across settings (Fig. 4), confirming its robustness. Additionally, as illustrated in Fig. 5, we compare the standard and smoothed cross-entropy loss, with the latter showing faster convergence, higher accuracy, and improved training stability.

4.6 Bands Selection Analysis

To assess the effectiveness of ABS, we compare it with the classic 5-band approach by using a single frequency band (e.g., $\mathcal{G}_\alpha$) to construct the phase brain network for FGO, while the remaining bands' DE features are input to STE, with all other settings identical to Stella. As shown in Table 4, STE consistently outperforms the 5-band method across all metrics—for instance, improving DEAP valence and arousal accuracy by 2.14% and 1.93%, and FACED by 3.50% and 1.85% over the δ band. These results demonstrate that accounting for channel differences provides superior performance, supporting the value of adaptive frequency selection in emotion recognition.

5 Conclusion

We present Stella, a framework that combines ABS to differentiate base and harmonic band signals for phase brain networks construction, FGO to address the feature degradation problem in graph convolution, and STE to enhance spatial-temporal information with stable and dense EEG signal representations. Stella effectively builds brain networks, extracts deep spatial-temporal features, and stabilizes the learning process, achieving state-of-the-art emotion recognition performance. It also offers valuable insights for applying graph learning to non-stationary signals like EEG, from feature construction to model design. Beyond emotion recognition, the proposed framework is scalable and potentially applicable to other EEG-based tasks such as cognitive load prediction and neurological disorder detection.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant 62406036, the National Key Research and Development Program of China under Grant 2024YFC3308500, and the Key Laboratory of Target Cognition and Application Technology under Grant 2023-CXPT-LC-005. It is also sponsored by the SMP-Zhipu.AI Large Model Cross-Disciplinary Fund under Grant ZPCG20241029322, and supported by the State Key Laboratory of Networking and Switching Technology under Grant NST20250110.

References

[Akin, 2002] M. Akin. Comparison of wavelet transform and fft methods in the analysis of eeg signals. *Journal of Medical Systems*, page 241–247, Jan 2002.

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

[Chen *et al.*, 2023] Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023.

[Feng *et al.*, 2022] Lin Feng, Cheng Cheng, Mingyan Zhao, Huiyuan Deng, and Yong Zhang. Eeg-based emotion recognition using spatial-temporal graph convolutional lstm with attention mechanism. *IEEE Journal of Biomedical and Health Informatics*, 26(11):5406–5417, 2022.

[Flournoy *et al.*, 2024] John C Flournoy, Nessa V Bryce, Meg J Dennison, Alexandra M Rodman, Elizabeth A McNeilly, Lucy A Lurie, Debbie Bitran, Azure Reid-Russell, Constanza M Vidal Bustamante, Tara Madhyastha, et al. A precision neuroscience approach to estimating reliability of neural responses during emotion processing: Implications for task-fmri. *NeuroImage*, 285:120503, 2024.

[Gu *et al.*, 2023] Yun Gu, Xinyue Zhong, Cheng Qu, Chuanjun Liu, and Bin Chen. A domain generative graph network for eeg-based emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 27(5):2377–2386, 2023.

[He and Zhang, 2024] Zeming He and Gaoyan Zhang. Ednet: A continuous emotion detection network for naturalistic stimuli using meg signals. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2001–2005. IEEE, 2024.

[Huang *et al.*, 2017] Xiaohua Huang, Su-Jing Wang, Xin Liu, Guoying Zhao, Xiaoyi Feng, and Matti Pietikäinen. Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 10(1):32–47, 2017.

[Huang *et al.*, 2021] Dongmin Huang, Sentao Chen, Cheng Liu, Lin Zheng, Zhihang Tian, and Dazhi Jiang. Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for eeg emotion recognition. *Neurocomputing*, 448:140–151, 2021.

[Jenke *et al.*, 2014] Robert Jenke, Angelika Peer, and Martin Buss. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective computing*, 5(3):327–339, 2014.

[Ju *et al.*, 2024] Xiangyu Ju, Ming Li, Wenli Tian, and Dewen Hu. Eeg-based emotion recognition using a temporal-difference minimizing neural network. *Cognitive Neurodynamics*, 18(2):405–416, 2024.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Koelstra *et al.*, 2011] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[Kragel and LaBar, 2016] Philip A Kragel and Kevin S LaBar. Decoding the nature of emotion in the brain. *Trends in cognitive sciences*, 20(6):444–455, 2016.

[Li *et al.*, 2016] Xiang Li, Dawei Song, Peng Zhang, Guangliang Yu, Yuexian Hou, and Bin Hu. Emotion recognition from multi-channel eeg data through convolutional recurrent neural network. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 352–359. IEEE, 2016.

[Li *et al.*, 2023] Zhongjie Li, Gaoyan Zhang, Longbiao Wang, Jianguo Wei, and Jianwu Dang. Emotion recognition using spatial-temporal eeg features through convolutional graph attention network. *Journal of Neural Engineering*, 20(1):016046, 2023.

[Li, 2024] et al. Li. An efficient graph learning system for emotion recognition inspired by the cognitive prior graph of eeg brain network. *IEEE TNNLS*, 2024.

[Lin *et al.*, 2010] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, page 1798–1806, Jul 2010.

[Liu *et al.*, 2024a] Chenyu Liu, Xinliang Zhou, Jiaping Xiao, Zhengri Zhu, Liming Zhai, Ziyu Jia, and Yang Liu. Vsgt: variational spatial and gaussian temporal graph models for eeg-based emotion recognition. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 3078–3086, 2024.

[Liu *et al.*, 2024b] Jinduo Liu, Feipeng Wang, and Junzhong Ji. Concept-level causal explanation method for brain function network classification. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 3087–3096, 2024.

[Ma *et al.*, 2019] Jiaxin Ma, Hao Tang, Wei-Long Zheng, and Bao-Liang Lu. Emotion recognition using multimodal residual lstm network. In *Proceedings of the 27th ACM*

international conference on multimedia, pages 176–183, 2019.

[Meer *et al.*, 2020] Johan N van der Meer, Michael Breakspear, Luke J Chang, Saurabh Sonkusare, and Luca Cocchi. Movie viewing elicits rich and reliable brain state dynamics. *Nature communications*, 11(1):5004, 2020.

[Miao *et al.*, 2023] Minmin Miao, Longxin Zheng, Baoguo Xu, Zhong Yang, and Wenjun Hu. A multiple frequency bands parallel spatial–temporal 3d deep residual learning framework for eeg-based emotion recognition. *Biomedical Signal Processing and Control*, 79:104141, 2023.

[Pessoa and Adolphs, 2010] Luiz Pessoa and Ralph Adolphs. Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, 11(11):773–782, 2010.

[Pessoa, 2017] Luiz Pessoa. A network model of the emotional brain. *Trends in cognitive sciences*, 21(5):357–371, 2017.

[Pessoa, 2018] Luiz Pessoa. Understanding emotion with brain networks. *Current opinion in behavioral sciences*, 19:19–25, 2018.

[Shi *et al.*, 2023] Gen Shi, Yifan Zhu, Jian K Liu, and Xuesong Li. HeGCL: Advance self-supervised learning in heterogeneous graph-level representation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Si *et al.*, 2024] Xiaopeng Si, Dong Huang, Zhen Liang, Yulin Sun, He Huang, Qile Liu, Zhuobin Yang, and Dong Ming. Temporal aware mixed attention-based convolution and transformer network for cross-subject eeg emotion recognition. *Computers in Biology and Medicine*, 181:108973, 2024.

[Song *et al.*, 2018] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2018.

[Tao *et al.*, 2020] Wei Tao, Chang Li, Rencheng Song, Juan Cheng, Yu Liu, Feng Wan, and Xun Chen. Eeg-based emotion recognition via channel-wise attention and self attention. *IEEE Transactions on Affective Computing*, 14(1):382–393, 2020.

[Wang *et al.*, 2018] Xue-han Wang, Tong Zhang, Xiang-min Xu, Long Chen, Xiao-fen Xing, and CL Philip Chen. Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1240–1244. IEEE, 2018.

[Xiang *et al.*, 2023] Haolong Xiang, Xuyun Zhang, Hongsheng Hu, Lianyong Qi, Wanchun Dou, Mark Dras, Amin Beheshti, and Xiaolong Xu. Optiforest: Optimal isolation forest for anomaly detection. *arXiv preprint arXiv:2306.12703*, 2023.

[Xiang *et al.*, 2024] Haolong Xiang, Xuyun Zhang, Xiaolong Xu, Amin Beheshti, Lianyong Qi, Yujie Hong, and Wanchun Dou. Federated learning-based anomaly detection with isolation forest in the iot-edge continuum. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.

[Yang *et al.*, 2018] Yilong Yang, Qingfeng Wu, Ming Qiu, Yingdong Wang, and Xiaowei Chen. Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network. In *2018 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE, 2018.

[Yi *et al.*, 2024] Kun Yi, Qi Zhang, Hui He, Kaize Shi, Liang Hu, Ning An, and Zhendong Niu. Deep coupling network for multivariate time series forecasting. *ACM Transactions on Information Systems*, 42(5):1–28, 2024.

[Yin *et al.*, 2021] Yongqiang Yin, Xiangwei Zheng, Bin Hu, Yuang Zhang, and Xinchun Cui. Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm. *Applied Soft Computing*, 100:106954, 2021.

[Zhang *et al.*, 2021] Guanhua Zhang, Minjing Yu, Yong-Jin Liu, Guozhen Zhao, Dan Zhang, and Wenming Zheng. Sparsedgcnn: Recognizing emotion from multichannel eeg signals. *IEEE Transactions on Affective Computing*, 14(1):537–548, 2021.

[Zheng and Lu, 2015] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

[Zheng *et al.*, 2023] Chuhang Zheng, Wei Shao, Daoqiang Zhang, and Qi Zhu. Prior-driven dynamic brain networks for multi-modal emotion recognition. In *26th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2023*, volume 14227, pages 389–398. Springer, 2023.

[Zhu *et al.*, 2021] Yifan Zhu, Xuesong Li, Yang Sun, Haixu Wang, Hua Guo, and Jie Sui. Investigating neural substrates of individual independence and interdependence orientations via efficiency-based dynamic functional connectivity: A machine learning approach. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2):761–771, 2021.