SEBSFormer: A Spectral-Enhanced Bi-Stream Transformer for Robust EEG Decoding

Lin Zhang[1], **Shikui Tu**[1*], **Lei Xu**[1,2*]

[1]School of Computer Science, Shanghai Jiao Tong University, Shanghai, China
[2]Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Guangdong, China
linzhang@sjtu.edu.cn, tushikui@sjtu.edu.cn, leixu@sjtu.edu.cn

Abstract

Electroencephalography (EEG) plays a vital role in clinical and cognitive applications such as epilepsy diagnosis and emotion recognition. However, the low signal-to-noise ratio, inter-subject variability, and inherent non-stationarity of EEG signals present substantial modeling challenges. While recent Transformer-based models offer promising long-range modeling capabilities, their self-attention mechanism behaves as a low-pass filter, suppressing high-frequency neural patterns critical for decoding transient events. In this work, we provide the first formal analysis demonstrating this low-pass behavior in self-attention mechanisms when applied to EEG signals, revealing a fundamental limitation of deep attention-based EEG models. To address this, we propose SEBSFormer, a spectral-enhanced bi-Stream Transformer that jointly models temporal dependencies and spectral structures. SEBSFormer integrates three key modules: a spectral compensation module that restores high-frequency components via residual correction in the Fourier domain; a multi-scale temporal attention module for saliency-guided temporal compression; and a graph-guided dynamic fusion module for adaptive spatial aggregation across electrodes. Extensive experiments on three benchmark datasets—TUAB, TUEV, and SEED—demonstrate that SEBSFormer consistently outperforms existing state-of-the-art models across both clinical and affective tasks. Our findings establish a new paradigm for frequency-aware EEG modeling.

Code — https://github.com/CMACH508/SEBSFormer

Introduction

Electroencephalography (EEG) is a non-invasive technique that records spontaneous brain activity via scalp electrodes following the international 10–20 system (Niedermeyer and da Silva 2005). With its millisecond-level resolution and low cost, EEG is widely used in clinical neuroscience and brain-computer interface (BCI) research (Nicolas-Alonso and Gomez-Gil 2012), supporting tasks such as epilepsy diagnosis (Alotaiby et al. 2014; Ein Shoka et al. 2023; Zhang et al. 2023), emotion recognition (Jenke, Peer, and Buss 2014; Alarcao and Fonseca 2017), sleep staging(Nie, Tu, and Xu 2021; Zhu et al. 2023), and neural decoding (Saeidi et al. 2021; Mou et al. 2024). Among these, seizure detection and emotion monitoring are particularly critical, as accurate interpretation of neural dynamics directly impacts patient care and cognitive assessment (Naganur et al. 2022; Ibanez, Kringelbach, and Deco 2024). However, EEG presents significant challenges due to its complex temporal dynamics, low signal-to-noise ratio, and high variability across sessions, subjects, and devices, which hinder robust representation learning. Long-term monitoring also produces massive data, making manual annotation costly and error-prone; inter-rater variability among neurologists further introduces label noise (Yan et al. 2018). In this context, AI-based EEG analysis holds great promise to improve diagnostic efficiency, consistency, and accuracy.

Earlier approaches used 1D CNNs to capture local temporal dependencies (Nagabushanam et al. 2020; Dar et al. 2020; Jing et al. 2023), or combined STFT-based time–frequency representations with 2D CNNs for spectral feature extraction (Cui et al. 2020; Yang et al. 2022; Kumar and Upadhyay 2024). Hybrid CNN-RNN models further explored spatiotemporal dynamics (Yang et al. 2018; Supakar, Satvaya, and Chakrabarti 2022; Bdaqli et al. 2024), while multi-level attention, ensemble, and multi-stream architectures were also proposed (Wu and Liu 2022; Zhang et al. 2023; Deng et al. 2024; Nour, Senturk, and Polat 2024). More recently, Transformer-based models have shown strong performance in large-scale EEG modeling due to their capacity for long-range temporal and spatial dependency modeling (Yang, Westover, and Sun 2023; Zheng and Pan 2024; Jiang et al. 2025; Zhang et al. 2025). However, recent findings suggest that stacked self-attention layers inherently act as low-pass filters (Wang et al. 2022; Shin et al. 2024; Wi, Choi, and Park 2025), progressively attenuating high-frequency components crucial for decoding transient EEG events such as epileptic spikes or emotion bursts.

In this study, we provide the first rigorous theoretical analysis demonstrating that the self-attention mechanism inherently behaves as a low-pass filter when applied to EEG signals. Specifically, we prove that multi-layer self-attention exponentially attenuates high-frequency components , leading to an oversmoothing effect in which high-resolution temporal dynamics gradually vanish with increasing depth. This low-frequency bias causes the model to overemphasize smoother components, ultimately weakening its ability to

*Corresponding author

detect salient neural events. This finding not only explains the performance degradation observed in deep attention-based EEG models but also provides theoretical motivation for restoring spectral diversity.

To address spectral suppression and representational degradation in EEG Transformers, we propose SEBSFormer, a frequency-compensated bi-stream Transformer architecture. SEBSFormer consists of three core components: a Spectral Compensation Module (SCM) that decomposes and reweights signals in the Fourier domain to inject frequency-aware residuals and mitigate low-pass bias; a Multi-Scale Temporal Attention Module (MS-TAM) that combines local and global attention to adaptively compress temporal features and retain transient neural events; and a Graph-Guided Dynamic Fusion module (GDGF) that constructs data-driven electrode graphs to enhance spatial dependency modeling. These modules are jointly integrated into a frequency-aware, structure-adaptive spatiotemporal framework. The main contributions are summarized below:

- We provide the first formal proof that standard Transformers exhibit low-pass filtering behavior when applied to EEG signals, revealing a fundamental limitation in modeling high-frequency neural dynamics.
- We propose SEBSFormer, a bi-stream Transformer with spectral compensation, temporal saliency modeling, and dynamic spatial fusion to jointly preserve frequency information and capture spatiotemporal dependencies.
- Extensive experiments on public EEG datasets for seizure detection and emotion recognition show that SEBSFormer achieves state-of-the-art performance and strong generalization.

Self-Attention as a Low-Pass Filter in EEG Modeling

We theoretically analyze how self-attention mechanisms process raw EEG signals, demonstrating that they intrinsically act as low-pass filters in the frequency domain. Our analysis shows that multi-layer self-attention progressively suppresses high-frequency components, leading to over-smoothing—where fine-grained temporal neural patterns diminish with increasing depth.

Problem Setting and Notation

Let $\mathbf{X} \in \mathbb{R}^{C\times T}$ denote the raw EEG signal of C channels and T time steps. For simplicity, we assume each channel is independently processed by a self-attention mechanism over time. Hence, we focus on a single-channel sequence $\mathbf{x} \in \mathbb{R}^T$. We denote a self-attention transformation $\mathbf{A} \in \mathbb{R}^{T\times T}$, which is computed as, $\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)$, where $\mathbf{Q}, \mathbf{K} \in \mathbb{R}^{T\times d}$ are projections of input $\mathbf{x}$ via learned matrices.

Our goal is to analyze the behavior of $\mathbf{x}^{(L)} = \mathbf{A}^L\mathbf{x}$ under repeated self-attention application, and characterize how the energy of high-frequency components $\text{HFC}(\mathbf{x}^{(L)})$ vanishes while low-frequency components $\text{LFC}(\mathbf{x}^{(L)})$ dominate.

Fourier Basis and Frequency Decomposition

Let $\mathbf{F} \in \mathbb{C}^{T\times T}$ be the Discrete Fourier Transform (DFT) matrix such that for any signal $\mathbf{x} \in \mathbb{R}^T$, its frequency representation is $\hat{\mathbf{x}} = \mathbf{F}\mathbf{x}$. Denote:

- $\text{LFC}(\mathbf{x})$: inverse DFT of only low-frequency components (e.g., first c coefficients of $\hat{\mathbf{x}}$).
- $\text{HFC}(\mathbf{x})$: inverse DFT of remaining high-frequency components.

We define a transformation $f : \mathbb{R}^T \to \mathbb{R}^T$ to be a low-pass filter if:

$$\lim_{L\to\infty} \frac{\|\text{HFC}(f^L(\mathbf{x}))\|_2}{\|\text{LFC}(f^L(\mathbf{x}))\|_2} = 0. \tag{1}$$

In practice, we define $\text{LFC}(\mathbf{x})$ as the inverse DFT of the first c frequency components, where c is selected according to physiologically meaningful EEG bands (e.g., δ/θ waves < 8 Hz, $\beta/\gamma > 12$ Hz). While DFT basis vectors and the eigenbasis of $\mathbf{A}$ are not guaranteed to align exactly, we assume they are approximately aligned in the sense that high-frequency DFT components tend to correspond to non-principal eigenvectors of $\mathbf{A}$, which are suppressed during repeated application.

Main Theorem

We now state and prove the central theoretical result.

Theorem 1. *Let $\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right) \in \mathbb{R}^{T\times T}$ be a self-attention matrix with non-negative entries and row-normalized. Then for any EEG signal $\mathbf{x} \in \mathbb{R}^T$, repeated application of self-attention leads to low-pass behavior:*

$$\lim_{L\to\infty} \frac{\|\text{HFC}(\mathbf{A}^L\mathbf{x})\|_2}{\|\text{LFC}(\mathbf{A}^L\mathbf{x})\|_2} = 0. \tag{2}$$

That is, self-attention gradually erases high-frequency EEG patterns across layers.

Proof Sketch

Step 1: Perron-Frobenius Property. Since each row of $\mathbf{A}$ is a softmax output, $\mathbf{A}$ is a row-stochastic matrix with non-negative entries ($A_{ij} \geq 0$) and unit row sums ($\sum_j A_{ij} = 1$).

By the Perron-Frobenius theorem (He and Wai 2021; Meyer 2023), the largest eigenvalue of $\mathbf{A}$ is $\lambda_1 = 1$, with a unique dominant eigenvector $\mathbf{v}_1$. All other eigenvalues satisfy $|\lambda_i| < 1$. To rigorously apply the Perron-Frobenius theorem, we assume that $\mathbf{A}$ is irreducible and aperiodic. This ensures the uniqueness and dominance of the principal eigenvalue $\lambda_1 = 1$. In practice, these conditions are often satisfied for EEG signals due to their temporal continuity and dense temporal dependencies. To maintain irreducibility during modeling, it is critical to exclude padding tokens (e.g., zero inputs) from the attention computation, which can otherwise create disconnected components.

Step 2: Jordan Form Decomposition. Let $\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$, where $\mathbf{J}$ is block-diagonal with eigenvalues $\lambda_1, \lambda_2, \ldots$. Then:

$$\mathbf{A}^L\mathbf{x} = \mathbf{P}\mathbf{J}^L\mathbf{P}^{-1}\mathbf{x}. \tag{3}$$

As $L \to \infty$, $\mathbf{J}^L \to \text{diag}(1, 0, \ldots)$, so $\mathbf{x}^{(L)} \to \text{Proj}_{\mathbf{v}_1}(\mathbf{x})$.

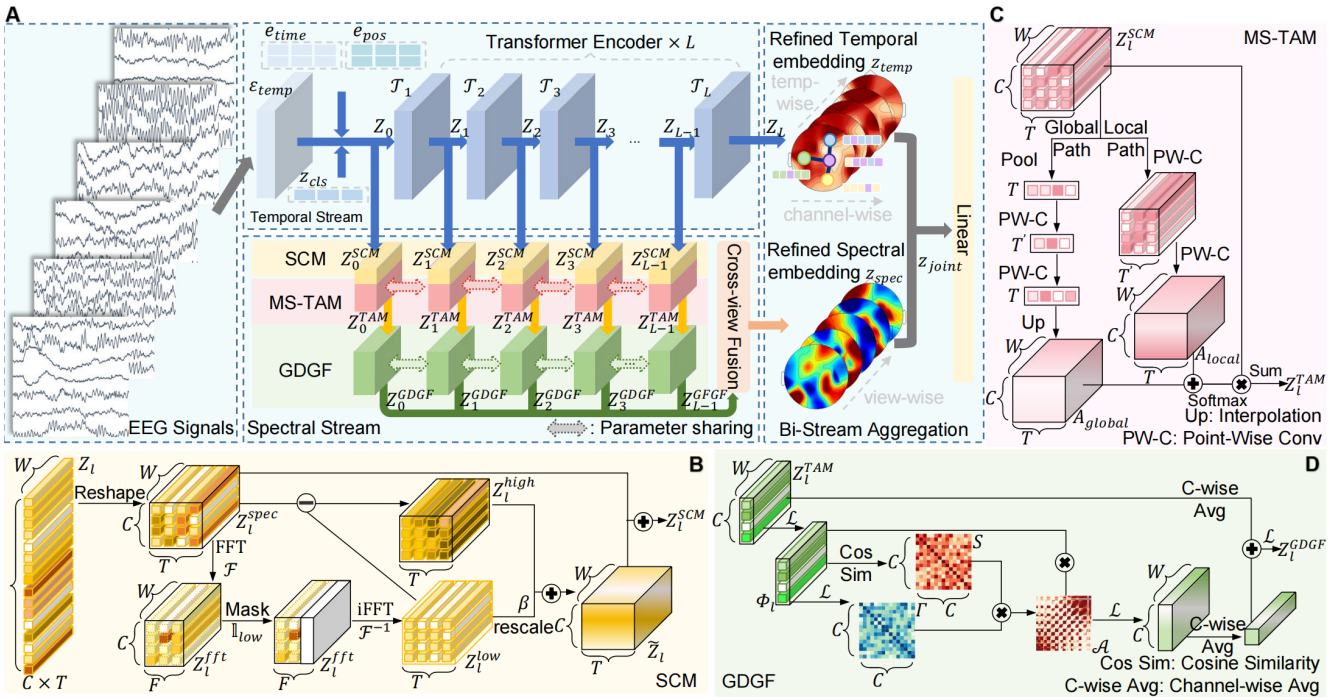


Figure 1: The architecture of SEBSFormer and its key components. (A) The overall bi-stream architecture of SEBSFormer, which integrates a Temporal Stream and a Spectral Stream for comprehensive EEG representation. (B) Spectral Compensation Module (SCM). (C) Multi-Scale Temporal Attention Module (MS-TAM). (D) Graph-Guided Dynamic Fusion (GDGF).

Step 3: Frequency Domain Interpretation. This asymptotic behavior corresponds to filtering out all non-dominant components — particularly, since high-frequency signals correspond to large sign changes, they cannot survive under such convergence. So:

$$\lim_{L\to\infty} \|\mathrm{HFC}(\mathbf{x}^{(L)})\|_2 = 0, \tag{4}$$

while $\|\mathrm{LFC}(\mathbf{x}^{(L)})\|_2$ remains non-zero. Moreover, the rate at which high-frequency components decay can be quantified. Let $\rho = \max_{i\geq 2} |\lambda_i|$ be the largest magnitude among non-principal eigenvalues of $\mathbf{A}$. Then for any $\mathbf{x}$ with nonzero high-frequency components, we have:

$$\|\mathrm{HFC}(\mathbf{A}^L\mathbf{x})\|_2 \leq \rho^L \|\mathrm{HFC}(\mathbf{x})\|_2. \tag{5}$$

This shows that high-frequency content decays exponentially with depth L, and ρ controls the rate of oversmoothing. This completes the proof.

Why It Matters: High-Frequency Suppression Degrades EEG Task Performance

While Transformer-based models have gained popularity in EEG analysis, the spectral characteristics of self-attention remain underexplored. Most existing studies focus on attention as a temporal modeling tool, overlooking its inherent frequency bias. However, EEG signals encode critical task-relevant information in higher frequency bands, particularly β (13–30 Hz), γ (30–100 Hz), and even higher-frequency oscillations (HFOs, $>$ 80 Hz), which have been shown to reflect transient neural events associated with epileptic activity tasks (Herrmann and Demiralp 2005; Zijlmans et al. 2012; Jacobs et al. 2012; Park and Hong 2019) and emotion-related activations or cognitive tasks (Lachaux et al. 2012; Yang et al. 2020; Roshanaei et al. 2025).

Neglecting high-frequency components can thus impair performance on clinically significant tasks. Our theoretical analysis is the first to demonstrate that multi-layer self-attention mechanisms function as low-pass filters when applied to EEG sequences. This provides a principled explanation for the performance degradation observed in deep Transformer models and underscores the need for spectral compensation mechanisms like SCM. Importantly, this low-pass filtering effect is not a modeling flaw but a natural consequence of the stochastic attention process. Our findings highlight a previously unrecognized limitation in attention-based EEG models and emphasize the importance of frequency-aware architectural designs.

Methodology

Overview

We introduce SEBSFormer, a novel dual-stream neural network designed to jointly model temporal dependencies and spectral–spatial patterns in EEG signals as shown in Figure 1. The architecture consists of two branches: the Temporal Modeling Stream and a Spectral Modeling Stream. The Temporal Modeling Stream captures long-range sequential dynamics through a combination of temporal convolution and transformer-based attention. The Spectral Modeling Stream processes the input through three specialized modules: the SCM preserves high-frequency compo-

nents suppressed by deep attention layers; the MS-TAM adaptively compresses temporal information by learning saliency-aware attention; and the GGDF module constructs a dynamic electrode graph to model spatial dependencies. By integrating these streams, SEBSFormer fuses global temporal trends with enhanced spectral–spatial representations. The final embedding aggregates both branches, enabling robust decoding of complex EEG signals. The following subsections detail each component of the temporal and spectral streams, as well as the final aggregation strategy.

Temporal Modeling Stream: Temporal Convolution and Transformer Encoding

To capture the sequential dynamics of EEG signals over time and across electrodes, SEBSFormer first constructs a temporal representation using a combination of temporal convolution and transformer-based attention. Given the input EEG tensor $X \in \mathbb{R}^{C\times T\times W}$, where C is the number of electrodes, T is the number of seconds, and W is the number of time points per segment. We first flatten the spatial and temporal axes, $X_{\text{flat}} \in \mathbb{R}^{(C\cdot T)\times W}$. The flattened sequence is then processed by a temporal convolutional encoder $\mathcal{E}_{\text{temp}}$, implemented as a three-layer 1D convolutional module, where each layer consists of a convolution operation followed by group normalization (GN) and GELU activation. Let $H_0 = X_{\text{flat}}$, formally, we define:

$$H_3 = \mathcal{E}_{\text{temp}}(X_{\text{flat}}) = f_3 \circ f_2 \circ f_1(X_{\text{flat}}), \tag{6}$$

where each transformation $f_i(\cdot)$ is defined as:

$$f_{i+1}(H_i) = \text{GELU}(\text{GN}(\text{Conv1D}(H_i))), \quad i = 0, 1, 2. \tag{7}$$

The convolutional parameters are configured as follows: the first layer f_1 uses a kernel size of 15, stride of 8, and padding of 7, to perform coarse temporal downsampling. The subsequent two layers, f_2 and f_3, both use a smaller kernel size of 3, stride of 1, and padding of 1, to refine the local temporal features while preserving resolution. The output $H_3 \in \mathbb{R}^{(C\cdot T)\times W}$ encodes temporally localized features for each EEG channel and time segment, serving as the input to the subsequent transformer encoder.

To model long-range dependencies, we augment the sequence with a learnable class token $z_{cls} \in \mathbb{R}^W$, time embeddings $e_{time} \in \mathbb{R}^{(C\cdot T)\times W}$, and positional embeddings $e_{pos} \in \mathbb{R}^{(C\cdot T+1)\times W}$. Specifically, we form the input as: $Z_0 = [z_{cls}; H_3 + e_{time}] + e_{pos}$, where $[;]$ denotes concatenation.

This sequence is then passed through a stack of L transformer block, each composed of multi-head self-attention and feedforward sublayers:

$$Z_\ell = \mathcal{T}_\ell(Z_{\ell-1}), \quad \ell = 1, \ldots, L. \tag{8}$$

Within each Transformer block, self-attention models long-range dependencies across time and channels. Given input $Z_{\ell-1}$, it is projected into query (Q), key (K), and value (V) matrices, and normalized before computing attention as $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{\text{LN}(Q)\cdot\text{LN}(K)^\top}{\sqrt{d_{\text{head}}}}\right)V$, where d_{head} is the dimension of each head and $\text{LN}(\cdot)$ denotes layer normalization. This formulation stabilizes the training by controlling the scale of attention logits, and facilitates better convergence.

We apply mean pooling over the token dimension to obtain the temporal representation:

$$z_{temp} = \text{LN}\left(\frac{1}{C\times T}\sum_{i=1}^{C\times T} Z_L^{(i)}\right), \tag{9}$$

where $Z_L^{(i)} \in \mathbb{R}^W$ denotes the i-th patch token at the final transformer layer. This summary aggregates context across all temporal and spatial segments, yielding a compact representation of the EEG input. This temporal branch forms the backbone of SEBSFormer, delivering an attention-enhanced encoding that captures long-range sequential dependencies in neural dynamics.

Spectral Modeling Stream: Preserving and Structuring Frequency-Aware Representations

Spectral Compensation Module: High-Frequency Enhancement via Residual Correction To preserve high-frequency oscillatory patterns that are often attenuated by deep attention modeling, SEBSFormer integrates a Spectral Compensation Module (SCM) before each transformer layer. SCM aims to explicitly separate and reweight low- and high-frequency components in the EEG representation, thereby enriching spectral diversity across layers.

Specifically, before each transformer block, we apply the spectral enhancement to the patch token sequence $Z_{\ell-1} \in \mathbb{R}^{(C\times T+1)\times W}$, where the first token is a class token and the remaining $C \times T$ tokens correspond to EEG patches. We exclude the class token and reshape the remaining tokens into a 3D tensor $Z_{\ell-1}^{\text{spec}} \in \mathbb{R}^{C\times T\times W}$.We then apply a real-valued Fast Fourier Transform (FFT) along the temporal axis:

$$Z_{\ell-1}^{\text{fft}} = \mathcal{F}(Z_{\ell-1}^{\text{spec}}) = \text{FFT}(Z_{\ell-1}^{\text{spec}}) \in \mathbb{C}^{C\times F\times W}, \tag{10}$$

where $F = T/2 + 1$ is the number of positive frequencies. To decompose the spectral components, we apply a binary mask $\mathbb{I}_{\text{low}}$ to selectively retain low-frequency content in the frequency domain. The inverse FFT then transforms the masked signal back to the time domain, yielding the low-frequency reconstruction. The high-frequency residual is obtained by subtracting this low-frequency component from the original signal:

$$Z_{\ell-1}^{\text{low}} = \mathcal{F}^{-1}(Z_{\ell-1}^{\text{fft}} \cdot \mathbb{I}_{\text{low}}), \quad Z_{\ell-1}^{\text{high}} = Z_{\ell-1}^{\text{spec}} - Z_{\ell-1}^{\text{low}}. \tag{11}$$

Here, Z_{low} retains slow neural oscillations, while Z_{high} captures rapid variations that are often attenuated by deep attention mechanisms. We then apply a learnable rescaling coefficient $\sqrt{\beta} \in \mathbb{R}^{1\times 1\times W}$ to the high-frequency part and reconstruct the sequence:

$$\widetilde{Z}_{\ell-1} = Z_{\ell-1}^{\text{low}} + \beta \cdot Z_{\ell-1}^{\text{high}}. \tag{12}$$

Finally, the output is normalized and fused with the original residual via LayerNorm and dropout:

$$Z_{\ell-1}^{\text{SCM}} = \text{LN}(\widetilde{Z}_{\ell-1} + Z_{\ell-1}^{\text{spec}}). \tag{13}$$

Motivated by our theoretical analysis that multi-layer self-attention inherently acts as a low-pass filter—gradually suppressing high-frequency components in EEG signals—we

introduce SCM to explicitly address this limitation. This module is independently applied before each transformer layer, injecting frequency-aware residual signals that recover high-frequency neural patterns prone to being erased, such as epileptic spikes or emotional bursts. By selectively enhancing these transient and discriminative features, SCM improves the model's sensitivity to fine-grained temporal structures. Acting as a residual connection in the frequency domain, this design preserves spectral diversity across layers and effectively counterbalances the low-pass bias of temporal attention, thereby enhancing the overall representational capacity of the transformer encoder.

Multi-Scale Temporal Attention Module: Saliency-Guided Time Compression To address the limitations of naive temporal averaging in preserving critical EEG dynamics, we introduce the Multi-Scale Temporal Attention Module (MS-TAM) as a component of SEBSFormer. MS-TAM performs adaptive temporal compression by learning data-driven attention weights across time steps, enabling selective emphasis on salient, short-lived neural events. In contrast to average pooling, which uniformly aggregates temporal information and may obscure transient yet informative patterns (e.g., epileptic spikes or emotional bursts), MS-TAM retains fine-grained temporal details by integrating local and global temporal cues in a unified attention framework.

Given the output from the spectral compensation module at layer $\ell - 1$, denoted as $Z_{\ell-1}^{\text{SCM}}$, MS-TAM learns attention weights across the temporal dimension to guide the compression process. We employ two complementary attention pathways to capture both local temporal bursts and global contextual patterns:

- **Local attention path.** This branch applies a bottleneck-style convolutional subnetwork to extract local temporal dependencies:

$$A_{\text{local}} = \text{Conv2D}(\text{ReLU}(\text{BN}(\text{Conv2D}(Z_{\ell-1}^{\text{SCM}})))), \tag{14}$$

  where Conv2D refers to a 1×1 kernel convolution (point-wise convolution) applied along the temporal axis.
- **Global attention path.** This branch captures long-range temporal context by global pooling followed by two convolutional transformations and upsampling:

$$H_{\text{global}} = \text{ReLU}(\text{BN}(\text{Conv2D}(\text{APool}(Z_{\ell-1}^{\text{SCM}})))), \tag{15}$$

$$A_{\text{global}} = \text{Up}\left(\text{BN}\left(\text{Conv2D}(H_{\text{global}})\right)\right), \tag{16}$$

  where APool denotes adaptive average pooling, and Up represents nearest-neighbor interpolation used to restore temporal resolution.

The local and global attention maps are then fused and normalized along the temporal dimension: $A = \text{Softmax}(A_{\text{local}} + A_{\text{global}})$, yielding a temporal attention matrix $A \in \mathbb{R}^{T\times C\times W}$ that reflects the relative importance of each time step. The final temporally compressed representation is obtained via weighted summation:

$$Z_{\ell-1}^{\text{TAM}} = \sum_{t=1}^{T} A_t \cdot Z_{\ell-1}^{\text{SCM}}(t), \tag{17}$$

where $Z_{\ell-1}^{\text{TAM}} \in \mathbb{R}^{C\times W}$ serves as the compact temporal feature map passed to the subsequent spatial modeling module.

To ensure stable learning and parameter efficiency, all MS-TAM modules across transformer layers share the same set of weights. This design provides a consistent temporal saliency criterion throughout the network and significantly improves sensitivity to high-importance temporal events that may otherwise be suppressed by uniform downsampling.

Graph-Guided Dynamic Fusion: Spatial Aggregation To capture spatial dependencies across EEG electrodes, we propose a lightweight Graph-Guided Dynamic Fusion (GDGF) module that adaptively learns inter-channel relationships through a dynamically constructed adjacency matrix. Given the temporally compressed input $Z_{\ell-1}^{\text{TAM}}$, we first apply a shared linear transformation $\mathcal{L}(\cdot)$ to obtain projected features $\Phi_{\ell-1} = \mathcal{L}(Z_{\ell-1}^{\text{TAM}})$. This transformation ensures dimensional consistency and enhances expressive capacity for graph modeling.

We then compute a soft affinity matrix $\tilde{A} \in \mathbb{R}^{C\times C}$ that captures pairwise relationships between electrodes. This is achieved by combining cosine similarity with a symmetric, learnable gating mechanism:

$$\Gamma_{ij} = \frac{1}{2}\left[\sigma(\mathcal{L}(\phi_{\ell-1,i})) + \sigma(\mathcal{L}(\phi_{\ell-1,j}))\right], \tag{18}$$

$$\tilde{A}_{ij} = \Gamma_{ij} \cdot S_{ij} = \Gamma_{ij} \cdot \left\langle \frac{\phi_{\ell-1,i}}{\|\phi_{\ell-1,i}\|_2}, \frac{\phi_{\ell-1,j}}{\|\phi_{\ell-1,j}\|_2} \right\rangle, \tag{19}$$

where $\phi_{\ell-1,i}$ denotes the projected feature vector of the i-th electrode, $\sigma(\cdot)$ denotes the sigmoid activation function, and S is the cosine similarity matrix. The resulting affinity matrix $\tilde{A} \in \mathbb{R}^{C\times C}$ is row-normalized to produce a valid propagation kernel $\mathcal{A}$, where each entry is computed as $\mathcal{A}ij = \tilde{A}ij/(\sum_k \tilde{A}_{ik} + \epsilon)$ with a small constant ϵ added for numerical stability.

We then perform graph-based spatial aggregation by multiplying the normalized adjacency matrix $\mathcal{A}$ with the projected features $\Phi_{\ell-1}$. The aggregated representation is compressed via global average pooling $\bar{\mathcal{L}}(\cdot)$, followed by a refinement through a linear transformation $\mathcal{L}(\cdot)$. To preserve the semantic consistency of the original input and stabilize training, we incorporate a residual connection using the mean of $Z_{\ell-1}^{\text{TAM}}$, followed by layer normalization:

$$Z_{\ell-1}^{\text{GDGF}} = \text{LN}\left(\mathcal{L}\left(\bar{\mathcal{L}}\left(\mathcal{A} \cdot \Phi_{\ell-1}\right)\right) + \bar{Z}_{\ell-1}^{\text{TAM}}\right). \tag{20}$$

This dynamic fusion strategy enables the model to flexibly adapt to varying spatial configurations and extract informative global representations without relying on predefined anatomical priors. Such flexibility is especially critical in EEG decoding tasks involving complex spatiotemporal dynamics, such as seizure propagation or emotion-induced cortical activation.

Dual-Stream Aggregation of Layer-Wise Representations for Final Embedding

To consolidate multi-depth representations, we aggregate features $\{Z_{\ell-1}^{\text{GDGF}}\}_{\ell=1}^{L}$ from all layers before the final Transformer block. Each $Z_{\ell-1}^{\text{GDGF}}$ already encodes spectral (SCM), temporal (MS-TAM), and spatial (GDGF) information. These features are concatenated along the view dimension

Dataset	Channel	Sampling Rate	Duration	Train	Val	Test	Task
TUAB	23	256 Hz	10 s	298,468	73,760	36,945	Binary classification
TUEV	23	256 Hz	5 s	68,656	14,414	28,305	6-class classification
SEED	62	1000 Hz	4 s	22,815	7,695	7,695	3-class classification

Table 1: Statistics of datasets used for evaluation.

Model	TUAB			TUEV			SEED		
	BAcc	PR	AUROC	BAcc	Kappa	W-F1	BAcc	Kappa	W-F1
ST-Trans (Song et al. 2021)	0.7966 ±0.0023	0.8521 ±0.0026	0.8707 ±0.0019	0.3984 ±0.0228	0.3765 ±0.0306	0.6823 ±0.0190	0.5479 ±0.0091	0.3261 ±0.0169	0.5505 ±0.0091
ContraWR (Yang et al. 2021)	0.7746 ±0.0041	0.8421 ±0.0104	0.8456 ±0.0074	0.4383 ±0.0349	0.3912 ±0.0237	0.6893 ±0.0136	0.6106 ±0.0078	0.4220 ±0.0129	0.6173 ±0.0085
CNN-Trans (Peh et al. 2022)	0.7777 ±0.0022	0.8433 ±0.0039	0.8461 ±0.0013	0.4087 ±0.0161	0.3815 ±0.0134	0.6854 ±0.0293	0.6161 ±0.0384	0.4262 ±0.0601	0.6150 ±0.0463
FFCL (Li et al. 2022)	0.7848 ±0.0038	0.8448 ±0.0065	0.8569 ±0.0051	0.3979 ±0.0104	0.3732 ±0.0188	0.6783 ±0.0120	0.5808 ±0.0322	0.3732 ±0.0462	0.5743 ±0.0402
SPaRCNet (Jing et al. 2023)	0.7896 ±0.0018	0.8414 ±0.0018	0.8676 ±0.0012	0.4161 ±0.0262	0.4233 ±0.0181	0.7024 ±0.0104	0.5596 ±0.0244	0.3464 ±0.0372	0.5585 ±0.0297
BIOT (Yang et al. 2023)	0.7959 ±0.0057	0.8792 ±0.0023	0.8815 ±0.0043	0.5281 ±0.0065	0.5273 ±0.0249	0.7492 ±0.0082	0.7097 ±0.0024	0.5682 ±0.0051	0.7143 ±0.0027
LaBraM (Jiang et al. 2024)	0.8140 ±0.0019	0.8965 ±0.0016	0.9022 ±0.0009	0.6409 ±0.0065	0.6637 ±0.0093	0.8312 ±0.0052	0.7318 ±0.0019	0.5994 ±0.0031	0.7354 ±0.0021
NeuroLM (Jiang et al. 2025)	0.7969 ±0.0091	0.7219 ±0.0082	0.7884 ±0.0194	0.4679 ±0.0356	0.4570 ±0.0498	0.7359 ±0.0219	0.6034 ±0.0010	0.4082 ±0.0036	0.6063 ±0.0030
SEBSFormer (Ours)	**0.8212 ±0.0025**	**0.9097 ±0.0016**	**0.9048 ±0.0026**	**0.6595 ±0.0069**	**0.6871 ±0.0082**	**0.8415 ±0.0055**	**0.7482 ±0.0039**	**0.6262 ±0.0064**	**0.7523 ±0.0040**

Table 2: Performance comparison on TUAB, TUEV, and SEED datasets (mean ± std). Best results are in **bold**; improvements of the proposed SEBSFormer over all baselines are statistically significant under a paired t-test ($p < 0.05$).

and refined with self-attention, followed by pooling and normalization:

$$z_{spec} = \text{LN}(\text{APool}(\text{Attention}([Z_0^{\text{GDGF}}; \ldots; Z_{L-1}^{\text{GDGF}}]))). \quad (21)$$

Finally, z_{spec} is concatenated with the temporal summary z_{temp} to from the final representation which serves as the input to the downstream classification head.

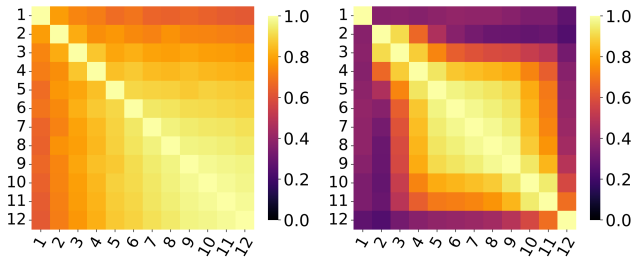
Experiments

Experimental Setups

We use binary cross-entropy for binary classification and categorical cross-entropy for multi-class tasks. Experiments are conducted on 8×NVIDIA RTX 3090 GPUs with Python 3.11.9, PyTorch 2.0.1, CUDA 11.8, and cuDNN 8.7. Models are trained on the training set, selected via validation, and evaluated on the test set. Results are averaged over three runs with different seeds (0, 42, 3407), and we report the mean and standard deviation of each metric.

Evaluation Datasets We evaluate our model on three public EEG datasets widely used in clinical and affective EEG analysis: TUAB, TUEV, and SEED, summarized in Table 1. **TUAB** (Obeid and Picone 2016) is a clinical EEG corpus with binary labels (normal vs. abnormal). **TUEV** (Harati et al. 2015) contains six segment-level annotations: SPSW, GPED, PLED, EYEM, ARTF, and BCKG. **SEED** (Zheng and Lu 2015) is an emotion dataset where subjects watch affective videos labeled as positive, neutral, or negative. We strictly follow the same data-splitting protocol as LaBraM (Jiang, Zhao, and Lu 2024) to ensure fair comparison and use the same preprocessing pipeline across all datasets rather than dataset-specific tuning.

Baselines and Evaluation Metrics We compare our method with representative EEG models: **ST-Trans** (Song et al. 2021): multi-level spatial-temporal transformer. **CNN-Trans** (Peh, Yao, and Dauwels 2022): hybrid of CNN and Transformer layers. **FFCL** (Li et al. 2022): fuses CNN and LSTM features at the representation level. **SPaRCNet** (Jing et al. 2023): 1D CNN enhanced with dense residual connections. **ContraWR** (Yang et al. 2021): self-supervised spectrogram-based model with 2D ResNet. **BIOT** (Yang, Westover, and Sun 2023): segments EEG into signal tokens and models them via lightweight Transformer. **LaBraM** (Jiang, Zhao, and Lu 2024): large-scale pre-trained Transformer with channel-token masking. **NeuroLM** (Jiang et al. 2025): hierarchical Transformer with multi-level masked modeling. We select evaluation metrics based on the specific classification task. For **binary classification**, we report Balanced Accuracy (BAcc), AUC-PR (PR), and AUROC. For **multi-class classification**, we report BAcc, Cohen's Kappa (Kappa), and Weighted F1-score (W-F1). BAcc handles class imbalance; PR and AUROC assess discrimination. Kappa measures agreement beyond chance; W-F1 weights precision and recall by class frequency.



(a) Transformer w/o SCM. (b) SEBSFormer with SCM.

Figure 2: Layer-wise cosine similarity on SEED. The horizontal and vertical axes represent different layers.

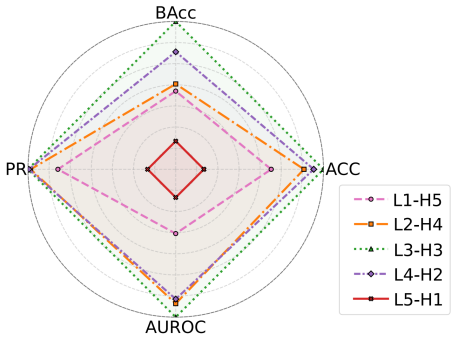


Figure 3: Effect of the mask $\mathbb{I}_{\text{low}}$ on TUAB.

Quantitative Results

We compare SEBSFormer with state-of-the-art EEG models on three widely used datasets (TUAB, TUEV, SEED) across multiple random seeds, as summarized in Table 2. SEBSFormer consistently outperforms all baselines on every metric and dataset, demonstrating strong generalization on both clinical and affective tasks. On TUAB, it yields better balanced accuracy, AUC-PR, and AUROC under class imbalance; on TUEV, it shows improved robustness to pathological variability; and on SEED, it achieves superior multi-class affective decoding. These results confirm that the spectral-enhanced bi-stream design effectively captures frequency-specific and temporal-context information, while the small standard deviations indicate stable training and convergence.

Inter-Layer Representation Similarity

To assess the impact of SCM, we compute the average cosine similarity between mean token representations at each Transformer layer on the SEED dataset. As shown in Figure 2(a), standard Transformers exhibit steadily increasing similarity across layers, indicating reduced representational diversity with depth. In contrast, SEBSFormer maintains lower inter-layer similarity (Figure 2(b)), suggesting that SCM helps preserve distinct representations by injecting frequency-aware residuals. These results support our theoretical analysis: SCM mitigates the homogenizing effect of attention and enables deeper layers to retain richer hierarchical features essential for decoding transient EEG patterns.

Hyperparameter Sensitivity

To assess the impact of the frequency mask $\mathbb{I}_{\text{low}}$, we performed sensitivity experiments on TUAB by varying the low-/high-frequency split across five settings (L1-H5 to L5-H1). As shown in Figure 3, balanced splits (L3-H3, L4-H2) yield the best results, suggesting that combining slow and fast oscillatory patterns enhances EEG modeling. Performance drops when the split is skewed, with L5-H1 performing the worst. Notably, L1-H5 outperforms L5-H1, indicating that retaining more high-frequency components is more beneficial than preserving only low-frequency ones. This aligns with our design of SCM to counteract the low-pass bias of attention layers and restore transient, discriminative patterns. We adopt a unified frequency-splitting strategy where the binary mask $\mathbb{I}_{\text{low}}$ assigns the first half of the bins as low-frequency and the rest as high. This standardization ensures consistency in spectral decomposition while adapting to dataset-specific temporal characteristics.

Ablation study

To assess the contribution of each component, we conduct an ablation study by individually removing the spectral stream (S1), SCM (S2), MS-TAM (S3), and GDGF (S4) from the full architecture. The results, summarized in Table 3, show that each module plays a critical role in enhancing model performance. In particular, MS-TAM and GDGF yield the most notable improvements across datasets. The complete Bi-Stream model consistently outperforms all ablated variants, achieving the best scores on all evaluation metrics.

Settings	TUAB			TUEV			SEED		
	BAcc	PR	ROC	BAcc	Kappa	W-F1	BAcc	Kappa	W-F1
S1	80.99	89.06	89.55	63.88	66.03	83.11	73.21	60.11	73.63
S2	81.35	89.20	89.05	64.26	66.63	83.14	73.91	61.08	74.29
S3	81.94	90.95	90.48	64.90	68.27	84.04	74.50	62.11	74.91
S4	82.08	90.51	90.47	65.87	67.02	83.60	74.42	61.97	74.85
Ours	**82.12**	**90.97**	**90.48**	**65.95**	**68.71**	**84.15**	**74.82**	**62.62**	**75.23**

Table 3: Ablation results on TUAB, TUEV, and SEED (%). Bold numbers indicate the best performance; gains of the full setting are significant (paired t-test, $p < 0.05$).

Conclusions

We present SEBSFormer, a spectral-enhanced bi-stream Transformer for EEG. We provide the first formal proof that self-attention acts as a low-pass filter, suppressing high-frequency components crucial for transient neural decoding, and introduce spectral compensation, saliency-guided temporal attention, and dynamic spatial fusion to preserve spectral diversity and structure. Experiments on TUAB, TUEV, and SEED datasets confirm its superiority over SOTA models in both clinical and affective tasks, demonstrating strong effectiveness and generalizability. This study paves the way for frequency-aware and interpretable EEG analysis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grants No. 62172273), the Science and Technology Commission of Shanghai Municipality (grants No. 24510714300), and the Shanghai Municipal Science and Technology Major Project, China (Grant No. 2021SHZDZX0102).

References

Alarcao, S. M.; and Fonseca, M. J. 2017. Emotions recognition using EEG signals: A survey. *IEEE transactions on affective computing*, 10(3): 374–393.

Alotaiby, T. N.; Alshebeili, S. A.; Alshawi, T.; Ahmad, I.; and Abd El-Samie, F. E. 2014. EEG seizure detection and prediction algorithms: a survey. *EURASIP Journal on Advances in Signal Processing*, 2014(1): 183.

Bdaqli, M.; Shoeibi, A.; Moridian, P.; Sadeghi, D.; Pouyani, M. F.; Shalbaf, A.; and Gorriz, J. M. 2024. Diagnosis of Parkinson disease from EEG signals using a CNN-LSTM model and explainable AI. In *International work-conference on the interplay between natural and artificial computation*, 128–138. Springer.

Cui, H.; Liu, A.; Zhang, X.; Chen, X.; Wang, K.; and Chen, X. 2020. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*, 205: 106243.

Dar, M. N.; Akram, M. U.; Khawaja, S. G.; and Pujari, A. N. 2020. CNN and LSTM-based emotion charting using physiological signals. *Sensors*, 20(16): 4551.

Deng, H.; Li, M.; Li, J.; Guo, M.; and Xu, G. 2024. A robust multi-branch multi-attention-mechanism EEGNet for motor imagery BCI decoding. *Journal of Neuroscience Methods*, 405: 110108.

Ein Shoka, A. A.; Dessouky, M. M.; El-Sayed, A.; and Hemdan, E. E.-D. 2023. EEG seizure detection: concepts, techniques, challenges, and future trends. *Multimedia Tools and Applications*, 82(27): 42021–42051.

Harati, A.; Golmohammadi, M.; Lopez, S.; Obeid, I.; and Picone, J. 2015. Improved EEG event classification using differential energy. In *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1–4. IEEE.

He, Y.; and Wai, H.-T. 2021. Identifying first-order lowpass graph signals using perron frobenius theorem. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5285–5289. IEEE.

Herrmann, C.; and Demiralp, T. 2005. Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical neurophysiology*, 116(12): 2719–2733.

Ibanez, A.; Kringelbach, M. L.; and Deco, G. 2024. A synergetic turn in cognitive neuroscience of brain diseases. *Trends in cognitive sciences*, 28(4): 319–338.

Jacobs, J.; Staba, R.; Asano, E.; Otsubo, H.; Wu, J.; Zijlmans, M.; Mohamed, I.; Kahane, P.; Dubeau, F.; Navarro, V.; et al. 2012. High-frequency oscillations (HFOs) in clinical epilepsy. *Progress in neurobiology*, 98(3): 302–315.

Jenke, R.; Peer, A.; and Buss, M. 2014. Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective computing*, 5(3): 327–339.

Jiang, W.-B.; Wang, Y.; Lu, B.-L.; and Li, D. 2025. NeuroLM: A Universal Multi-task Foundation Model for Bridging the Gap between Language and EEG Signals. In *International Conference on Learning Representations*.

Jiang, W.-B.; Zhao, L.-M.; and Lu, B.-L. 2024. Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI. In *International Conference on Learning Representations*.

Jing, J.; Ge, W.; Hong, S.; Fernandes, M. B.; Lin, Z.; Yang, C.; An, S.; Struck, A. F.; Herlopian, A.; Karakis, I.; et al. 2023. Development of Expert-Level Classification of Seizures and Rhythmic and Periodic Patterns During EEG Interpretation. *Neurology*, 100(17): e1750–e1762.

Kumar, P.; and Upadhyay, P. K. 2024. Seizure Detection with 2D Spectrogram Using CNN and SVM Integration. In *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*, 1–6. IEEE.

Lachaux, J.-P.; Axmacher, N.; Mormann, F.; Halgren, E.; and Crone, N. E. 2012. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Progress in neurobiology*, 98(3): 279–301.

Li, H.; Ding, M.; Zhang, R.; and Xiu, C. 2022. Motor Imagery EEG Classification Algorithm Based on CNN-LSTM Feature Fusion Network. *Biomedical Signal Processing and Control*, 72: 103342.

Meyer, C. D. 2023. *Matrix analysis and applied linear algebra*. SIAM.

Mou, X.; He, C.; Tan, L.; Yu, J.; Liang, H.; Zhang, J.; Tian, Y.; Yang, Y.-F.; Xu, T.; Wang, Q.; et al. 2024. ChineseEEG: A Chinese linguistic corpora EEG dataset for semantic alignment and neural decoding. *Scientific Data*, 11(1): 550.

Nagabushanam, P.; George, S. T.; Davu, P.; Bincy, P.; Naidu, M.; and Radha, S. 2020. Artifact Removal using Elliptic Filter and Classification using 1D-CNN for EEG signals. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 551–556. IEEE.

Naganur, V.; Sivathamboo, S.; Chen, Z.; Kusmakar, S.; Antonic-Baker, A.; O'Brien, T. J.; and Kwan, P. 2022. Automated seizure detection with noninvasive wearable devices: a systematic review and meta-analysis. *Epilepsia*, 63(8): 1930–1941.

Nicolas-Alonso, L. F.; and Gomez-Gil, J. 2012. Brain computer interfaces, a review. *sensors*, 12(2): 1211–1279.

Nie, H.; Tu, S.; and Xu, L. 2021. Recsleepnet: An automatic sleep staging model based on feature reconstruction. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1458–1461. IEEE.

Niedermeyer, E.; and da Silva, F. L. 2005. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Nour, M.; Senturk, U.; and Polat, K. 2024. A novel hybrid model in the diagnosis and classification of Alzheimer's disease using EEG signals: Deep ensemble learning (DEL) approach. *Biomedical Signal Processing and Control*, 89: 105751.

Obeid, I.; and Picone, J. 2016. The temple university hospital EEG data corpus. *Frontiers in neuroscience*, 10: 196.

Park, C. J.; and Hong, S. B. 2019. High frequency oscillations in epilepsy: detection methods and considerations in clinical application. *Journal of epilepsy research*, 9(1): 1.

Peh, W. Y.; Yao, Y.; and Dauwels, J. 2022. Transformer Convolutional Neural Networks for Automated Artifact Detection in Scalp EEG. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3599–3602. IEEE.

Roshanaei, M.; Norouzi, H.; Onton, J.; Makeig, S.; and Mohammadi, A. 2025. EEG-based functional and effective connectivity patterns during emotional episodes using graph theoretical analysis. *Scientific Reports*, 15(1): 2174.

Saeidi, M.; Karwowski, W.; Farahani, F. V.; Fiok, K.; Taiar, R.; Hancock, P. A.; and Al-Juaid, A. 2021. Neural decoding of EEG signals with machine learning: a systematic review. *Brain sciences*, 11(11): 1525.

Shin, Y.; Choi, J.; Wi, H.; and Park, N. 2024. An attentive inductive bias for sequential recommendation beyond the self-attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 8984–8992.

Song, Y.; Jia, X.; Yang, L.; and Xie, L. 2021. Transformer-based Spatial-Temporal Feature Learning for EEG Decoding. *arXiv preprint arXiv:2106.11170*.

Supakar, R.; Satvaya, P.; and Chakrabarti, P. 2022. A deep learning based model using RNN-LSTM for the detection of schizophrenia from EEG data. *Computers in Biology and Medicine*, 151: 106225.

Wang, P.; Zheng, W.; Chen, T.; and Wang, Z. 2022. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. *arXiv preprint arXiv:2203.05962*.

Wi, H.; Choi, J.; and Park, N. 2025. Learning Advanced Self-Attention for Linear Transformers in the Singular Value Domain. *arXiv preprint arXiv:2505.08516*.

Wu, H.; and Liu, J. 2022. A multi-stream deep learning model for EEG-based depression identification. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2029–2034. IEEE.

Yan, T.; Dong, X.; Mu, N.; Liu, T.; Chen, D.; Deng, L.; Wang, C.; and Zhao, L. 2018. Positive classification advantage: tracing the time course based on brain oscillation. *Frontiers in human neuroscience*, 11: 659.

Yang, C.; Qian, C.; Singh, N.; Xiao, C. D.; Westover, M.; Solomonik, E.; and Sun, J. 2022. ATD: Augmenting CP tensor decomposition by self supervision. *Advances in neural information processing systems*, 35: 32039–32052.

Yang, C.; Westover, M.; and Sun, J. 2023. Biot: Biosignal Transformer for Cross-data Learning in the Wild. *Advances in Neural Information Processing Systems*, 36: 78240–78260.

Yang, C.; Xiao, D.; Westover, M. B.; and Sun, J. 2021. Self-supervised EEG Representation Learning for Automatic Sleep Staging. *arXiv preprint arXiv:2110.15278*.

Yang, K.; Tong, L.; Shu, J.; Zhuang, N.; Yan, B.; and Zeng, Y. 2020. High gamma band EEG closely related to emotion: evidence from functional network. *Frontiers in human neuroscience*, 14: 89.

Yang, Y.; Wu, Q.; Qiu, M.; Wang, Y.; and Chen, X. 2018. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In *2018 international joint conference on neural networks (IJCNN)*, 1–7. IEEE.

Zhang, L.; Ren, Y.; Yuan, F.; Chen, X.; Tu, S.; and Xu, L. 2025. An EEG-based dual-stream spatial-spectral-temporal large model for self-limited epilepsy with centrotemporal spikes. *Expert Systems with Applications*, 129530.

Zhang, L.; Ren, Y.; Yuan, F.; Zhu, Y.; Tu, S.; Chen, Y.; and Xu, L. 2023. A Deep Learning Method with Multi-view Attention and Multi-branch GCN for BECT Diagnosis. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1718–1723. IEEE.

Zheng, W.; and Pan, B. 2024. A spatiotemporal symmetrical transformer structure for EEG emotion recognition. *Biomedical Signal Processing and Control*, 87: 105487.

Zheng, W.-L.; and Lu, B.-L. 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3): 162–175.

Zhu, Y.; Tu, S.; Zhang, L.; and Xu, L. 2023. Multi-source unsupervised domain-adaptation for automatic sleep staging. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2437–2440. IEEE.

Zijlmans, M.; Jiruska, P.; Zelmann, R.; Leijten, F. S.; Jefferys, J. G.; and Gotman, J. 2012. High-frequency oscillations as a new biomarker in epilepsy. *Annals of neurology*, 71(2): 169–178.