

A Dual-Stream Spatio-Temporal Network with Bidirectional Rectification for Differential Diagnosis of Dementia using EEG

Lingjia Yu

Hangzhou Dianzi University

Ming Meng

mnming@hdu.edu.cn

Hangzhou Dianzi University

Yunyuan Gao

Hangzhou Dianzi University

Xugang Xi

Hangzhou Dianzi University

Research Article

Keywords: Electroencephalogram (EEG), Alzheimer's disease, Frontotemporal dementia, Spatio-temporal network, Bidirectional rectification, Knowledge distillation

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10724309/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Dual-Stream Spatio-Temporal Network with Bidirectional Rectification for Differential Diagnosis of Dementia using EEG

Lingjia Yu¹, Ming Meng^{1,2*}, Yunyuan Gao^{1,2}, Xugang Xi¹

¹School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China.

²International Joint Research Laboratory for Autonomous Robotic Systems, Hangzhou Dianzi University, Hangzhou, 310018, China.

*Corresponding author(s). E-mail(s): mnming@hdu.edu.cn;

Abstract

Accurate differential diagnosis between Alzheimer’s disease (AD) and frontotemporal dementia (FTD) is of critical importance for the early screening of neurodegenerative diseases. Scalp electroencephalogram (EEG), as a non-invasive neurophysiological detection tool with excellent temporal resolution, has been widely adopted in dementia screening research. However, the inherent low signal-to-noise ratio and non-stationarity of scalp EEG signals, combined with the highly overlapping pathological rhythms between AD and FTD patients, render existing single-domain representation methods incapable of achieving precise pathological differentiation. To address this challenge, this paper proposes a dual-stream spatio-temporal network. In this framework, multiscale temporal features and effective connectivity graph features are extracted in parallel, followed by their integration. However, directly fusing the features from the two streams tends to amplify their respective noises. Therefore, the Spatio-Temporal Feature Fusion Based on Bidirectional Rectification module is proposed to execute dynamic noise suppression on the features and subsequently performs deep feature fusion via a cross-attention mechanism. Furthermore, this paper applies knowledge distillation using attention maps to address the issue of imbalanced sample distributions in the dataset. Experimental results on the AD/FTD-RS dataset show that the proposed framework yields a subject-level accuracy of 60.00% and an F1-score of 57.44%, showing measurable improvements over several baseline methods. This work offers a feasible approach for the computer-aided differential diagnosis of dementia based on resting-state EEG.

Keywords: Electroencephalogram (EEG), Alzheimer’s disease, Frontotemporal dementia, Spatio-temporal network, Bidirectional rectification, Knowledge distillation

1 Introduction

Early screening and precise differential diagnosis of neurodegenerative diseases hold profound practical value in modern clinical diagnostics [1]. With the intensification of population aging, the incidence of Alzheimer’s disease (AD) and frontotemporal dementia (FTD) has been rising year by year [2]. However, in complex clinical scenarios, distinguishing between AD and FTD remains a severe challenge due to their highly overlapping symptoms and EEG waveforms [3]. As a non-invasive neurophysiological measurement tool with excellent temporal resolution [4], electroencephalography (EEG) has been widely applied in early dementia screening research [5, 6]. Nevertheless, due to the inherent high non-stationarity, extremely low signal-to-noise ratio, and subtle inter-class differences of scalp EEG signals, traditional feature analysis methods often fail to satisfy the requirements for precise classification [7],

necessitating the introduction of more advanced deep representation learning paradigms [8].

In recent years, deep neural networks (DNNs), leveraging end-to-end feature extraction mechanisms, have demonstrated remarkable performance advantages in processing multi-channel continuous EEG signals [9, 10]. From the perspective of the temporal domain, models based on the Transformer architecture [11] (e.g., Medformer) can effectively capture long-range transient dependencies in EEG signals [12]. However, pure temporal models treat EEG as a collection of independent time series, thereby neglecting the brain network alterations accompanying neurodegenerative pathology, which results in insufficient discriminative power when confronting highly similar EEG features of FTD and AD [13, 14]. From the perspective of the spatial domain, researchers have utilized graph neural networks (GNNs) to compensate for the lack of topological modeling, efficiently characterizing whole-brain causal abnormalities in neurodegenerative diseases [15, 16]. Although

directed graph models possess robust spatial representation capabilities, estimating effective connectivity from highly non-stationary, short-term EEG data generates a substantial number of spurious edges, leaving severe challenges regarding graph construction reliability and dynamic information retention [17, 18]. Despite the respective limitations of single-domain models, temporal rhythms and spatial topologies both have irreplaceable advantages when representing neurodegenerative diseases [19, 20]. Therefore, exploring the joint modeling of spatio-temporal features has become a crucial approach. However, the direct concatenation of spatio-temporal features often leads to feature misalignment and noise amplification. Specifically, temporal features extracted from raw EEG contain physiological artifacts, while spatial graphs constructed via effective connectivity frequently include spurious edges. Directly fusing these two types of features causes noises to cross-contaminate and amplify, thereby masking pathological features and reducing diagnostic performance. Furthermore, clinical EEG datasets frequently suffer from severe sample imbalance. The scarcity of minority class data (e.g., FTD) often biases conventional data-driven models toward majority classes, making it extremely difficult to construct robust classification decision boundaries.

To address the aforementioned challenges, this paper proposes a dual-stream spatio-temporal network. To fully exploit the spatio-temporal complementarity of EEG signals, the model extracts multiscale temporal patching features and effective connectivity graph spatial features in parallel. To address the noise cross-contamination caused by the direct concatenation of spatio-temporal features, this paper introduces a spatio-temporal feature fusion module based on bidirectional rectification. Prior to feature fusion, this module executes dynamic noise suppression in both domains, ensuring that the subsequent cross-attention mechanism can obtain relatively precise spatio-temporal features. Furthermore, to overcome the classification difficulties caused by the severe imbalance of clinical data, this paper proposes a mask-guided knowledge distillation strategy. This strategy utilizes a pre-trained binary teacher model to guide the attention distribution of the three-class student model, thereby stabilizing the complex classification decision boundaries.

The contributions of this work can be summarized as follows:

- We propose a dual-stream spatio-temporal feature extraction architecture. By capturing multiscale temporal features and effective connectivity graphs in parallel, this architecture overcomes the representation bottleneck of single-domain models.
- We design a spatio-temporal feature fusion module based on bidirectional rectification. By executing bidirectional noise suppression prior to cross-attention fusion, this module effectively prevents noise amplification and captures robust spatio-temporal representations.

- We introduce a mask-guided attention distillation strategy. This strategy utilizes a pre-trained teacher model to stabilize decision boundaries and effectively handles the three-class classification task under class imbalance.

2 Related Work

2.1 Deep Learning-Based Automated Dementia Assessment via EEG

Driven by the necessity to decode complex neural dynamics, deep learning architectures have progressively replaced traditional feature engineering as the dominant paradigm for EEG-based dementia assessment. Specifically, Khalfallah et al. [21] evaluated the effectiveness of deep neural networks, demonstrating their significant advantages in preserving non-stationary signal characteristics compared to classical machine learning methods. To extract deeper representations, Kowshiga et al. [22] proposed a hybrid architecture combining convolutional and recurrent neural networks, establishing its efficacy in early Alzheimer’s disease screening. Similarly, Zini et al. [23] developed a lightweight network integrating multiscale temporal convolutions with recurrent layers to capture both short- and long-term EEG patterns for dementia classification. Furthermore, targeting the specific challenge of differentiating FTD from AD, Khan et al. [24] designed an explainable deep learning framework, while Acharya et al. [25] introduced the EEGConvNeXt model, which leverages advanced computer vision designs to enhance feature extraction from subtly varying waveforms. Nevertheless, these convolution-based architectures are inherently limited by local grid receptive fields, making it difficult to establish sufficiently robust decision boundaries when confronting highly overlapping inter-class representations.

To overcome the limitations of localized convolutional modeling, Transformer architectures based on self-attention mechanisms have subsequently emerged as a powerful tool in medical time-series analysis. For instance, Wang et al. [12] proposed Medformer, a multi-granularity patching Transformer designed specifically for medical signals, and demonstrated through extensive experiments that it significantly outperforms conventional networks in capturing long-range transient dependencies within EEG series. Although such self-attention models perform exceptionally well in capturing temporal rhythms, they are fundamentally constrained by Euclidean space assumptions and fail to explicitly characterize the information interactions across distinct brain regions. Consequently, when confronting complex pathologies that involve global cognitive network degeneration, including AD and FTD, these pure temporal models face difficulties in extracting neurophysiological features with sufficient discriminative power.

2.2 Brain Network Connectivity and Graph Neural Networks

To compensate for the lack of spatial topological modeling in pure temporal models, brain network connectivity analysis combined with GNNs has progressively emerged as a core paradigm for neurodegenerative disease assessment. For instance, Xu et al. [26] proposed a multi-graph convolutional network by fusing structural and functional connectivity matrices across multiple EEG frequency bands, demonstrating that spatial topological features significantly outperform isolated channel-wise features in capturing brain alterations associated with Alzheimer’s disease. Concurrently, Cao et al. [27] constructed EEG networks utilizing phase synchronization metrics, such as the phase-locking value (PLV), to extract robust spatial characteristics for dementia assessment. However, these approaches predominantly rely on undirected or symmetric connectivity graphs, which fail to explicitly reveal the “causal-driven” directional and asymmetric information flow between abnormal brain regions. To address this limitation, Cao et al. [28] directly transformed the effective connectivity of EEG signals into directed structural graphs, enabling GNNs to efficiently characterize whole-brain causal topological abnormalities in neurodegenerative diseases.

Despite the robust spatial representation capabilities of directed graph models, their application to highly non-stationary clinical EEG signals encounters severe reliability bottlenecks. Specifically, estimating effective connectivity from short-term EEG data inevitably introduces a substantial number of spurious causal edges. Although researchers such as Dong et al. [29] have introduced unidirectional denoising frameworks incorporating causal validation and temporal convolutions to improve graph construction reliability, effectively mitigating topological noise within highly complex dementia signals remains a persistent challenge. Consequently, relying solely on isolated graph networks not only loses high-frequency transient temporal dynamics but also leaves the model highly susceptible to erroneous graph topologies generated by signal non-stationarity, thereby underscoring the urgent need for a mechanism to dynamically verify and rectify spatial features.

2.3 Spatio-Temporal Feature Fusion

To achieve a more comprehensive representation of brain states, fusing features from distinct domains has become a mainstream paradigm in medical time-series analysis. Early fusion frameworks typically relied on direct feature concatenation. However, as research progressed, it became evident that simple concatenation often leads to feature misalignment and information redundancy. To address these issues, Zhang et al. [30] proposed a cross-attention mechanism utilizing global context to dynamically filter local noise, demonstrating the importance of deep feature alignment. Subsequently, Yan et al. [31] introduced multi-view cross-attention to

effectively enhance inter-domain interactions in physiological signals. Furthermore, Shi et al. [32] designed multi-level bidirectional interaction structures, verifying that dynamic cross-filtering significantly outperforms conventional concatenation in resolving conflicts between heterogeneous representations.

Despite these advancements, conventional fusion architectures lack a dynamic verification step tailored for non-Euclidean brain network topologies. When confronted with raw temporal dynamics containing high-frequency artifacts and graph structures burdened with spurious edges, these methods fail to achieve deep cross-filtering. This deficiency leaves existing fusion strategies highly susceptible to noise cascades, demanding a more sophisticated bidirectional rectification mechanism prior to deep integration.

In summary, although recent studies have made significant progress in the extraction and learning of spatio-temporal features, specific limitations remain in current EEG-based dementia diagnostics. First, existing spatio-temporal fusion architectures lack a dynamic verification mechanism for features, making them highly susceptible to noise interference. Second, in the EEG classification problem involving Alzheimer’s disease (AD), frontotemporal dementia (FTD), and healthy control (HC) cohorts, constructing robust decision boundaries under clinical data constraints remains a severe challenge due to the extremely subtle inter-class variances. These limitations represent the technical bottlenecks that current computer-aided diagnosis must overcome.

3 Proposed methodology

The overall architecture of the proposed dual-stream spatio-temporal framework is illustrated in Fig. 1. Given the preprocessed EEG signal $X \in \mathbb{R}^{B \times C \times T}$ (where B represents the batch size, C denotes the number of channels, and T indicates the time steps), the model extracts multiscale temporal patching features and effective connectivity graph spatial features in parallel. Subsequently, a spatio-temporal feature fusion module based on bidirectional rectification is employed to execute dynamic bidirectional filtering and feature alignment. This base network is initially optimized on a binary classification task to learn reliable diagnostic manifolds, establishing the foundation for subsequent complex classification.

3.1 Dual-Stream Spatio-Temporal Feature Extraction

3.1.1 Multiscale Temporal Patching Features

This branch is dedicated to capturing the multiscale temporal dynamics and cross-channel collaborative relationships of EEG signals. Given the preprocessed EEG signal $X \in \mathbb{R}^{B \times C \times T}$, the model uses sliding windows of different lengths L_i to segment the series, synchronously aggregating the local features across

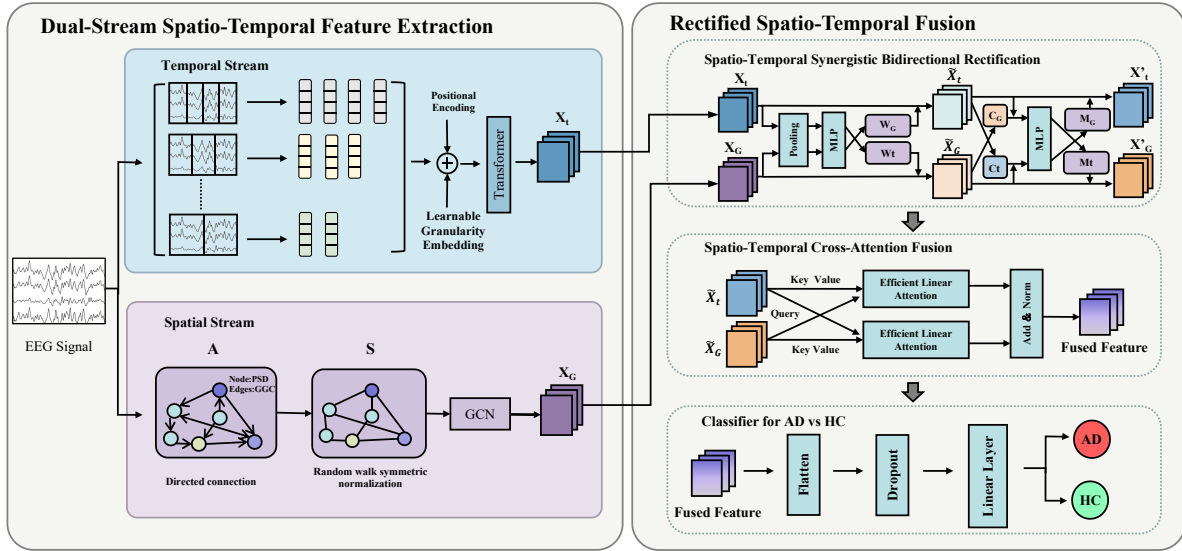


Fig. 1 The overall architecture of the dual-stream spatio-temporal framework. (1) Dual-stream spatio-temporal feature extraction module that extracts multi-scale temporal patching features and effective connectivity graph spatial features in parallel; (2) Spatio-temporal synergistic bidirectional rectification and fusion module that achieves denoising and alignment of heterogeneous features via cross-filtering and linear cross-attention.

whole-brain channels and linearly projecting them into D dimensional representation vectors. The vectors from each scale branch are then superimposed with positional encodings and learnable granularity embeddings to generate preliminary multiscale temporal patching representations. Finally, to capture long-range transient dependencies, the model feeds these representations into a multi-layer Transformer encoder. During this stage, the branches corresponding to different temporal granularities execute completely independent and parallel multi-head self-attention computations, enabling the model to adaptively and deeply focus on abnormal EEG rhythms. The final output is the high-level temporal features X_t .

3.1.2 Effective Connectivity Graph Spatial Features

The initial node features X_f are initialized using the logarithmic power spectral density (log-PSD) of each channel signal. Subsequently, a multivariate autoregressive (MVAR) model is utilized to calculate the Granger causality directed adjacency matrix A , and its random-walk symmetric normalization matrix S is computed to construct the effective connectivity graph as follows:

$$S = \frac{1}{2} \left(D_{deg}^{-1} A + A^T D_{deg}^{-1} + I \right) \quad (1)$$

where D_{deg} represents the degree matrix. Finally, spatial information is aggregated through a multi-layer graph convolutional network (GCN) to output the graph-perspective features $X_G \in \mathbb{R}^{B \times C \times D}$.

3.2 Spatio-Temporal Feature Fusion Based on Bidirectional Rectification

Directly fusing the temporal features X_t and graph spatial features X_G can easily amplify EEG noise. Inspired by cross-domain fusion architectures, this paper proposes a spatio-temporal cross-attention fusion module to perform bidirectional dynamic filtering.

3.2.1 Spatio-Temporal Synergistic Bidirectional Rectification

First, channel-level rectification is performed: the global average pooling and max pooling representations of the temporal features X_t and the graph spatial features X_G are extracted respectively. These representations are concatenated and passed through a multi-layer perceptron (MLP) to generate channel-level suppression weights W_t and W_G , yielding the initial channel-rectified features $X_t = X_t \odot W_t$ and $X_G = X_G \odot W_G$, where $\odot$ denotes broadcast multiplication. Subsequently, spatial-level synergistic rectification is executed: the global average context vectors C_t and C_G of the initial rectified features are extracted. The local features of each modality are concatenated with the global context of the other modality and fed into an MLP to jointly generate spatial cross-masks M_t and M_G :

$$M_t = \sigma(\text{MLP}(X_t \parallel \text{Expand}(C_G))) \quad (2)$$

$$M_G = \sigma(\text{MLP}(X_G \parallel \text{Expand}(C_t))) \quad (3)$$

Finally, the bidirectionally rectified features are obtained as $X'_t = X_t \odot M_t$ and $X'_G = X_G \odot M_G$.

3.2.2 Spatio-Temporal Cross-Attention Fusion

To reduce the computational complexity associated with long sequences, a linear cross-attention mechanism is adopted to interact X'_t and X'_G :

$$\text{Attention}(Q, K, V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_{\text{model}}}} \right) V \quad (4)$$

This mechanism executes a bidirectional query process: it uses the temporal features as the Query, and the graph features as the Key and Value for aggregation; conversely, it performs the same aggregation using the graph features as the Query and the temporal features as the Key and Value. Finally, the aggregated features from both directions are concatenated along the feature

dimension, achieving efficient, deep semantic alignment and fusion from a dual spatio-temporal perspective.

After completing the feature fusion, the resulting joint feature vector undergoes a flattening operation, and Dropout regularization is introduced to prevent overfitting. Finally, a fully connected layer (linear layer) maps the high-dimensional features to the diagnostic category space to yield the final classification prediction.

3.3 Knowledge Distillation via Attention Maps

To optimize the final three-class classification boundaries, a mask-guided attention distillation strategy is utilized, as illustrated in Fig. 2.

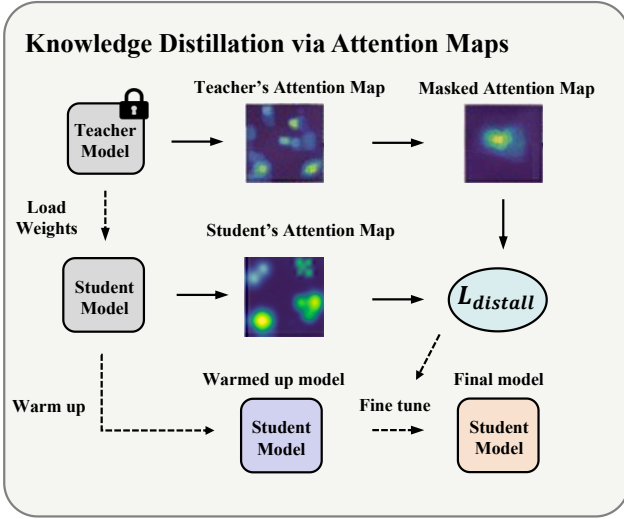


Fig. 2 The framework of knowledge distillation via attention maps. Where a pre-trained binary teacher model guides the optimization of the three-class student model via a masked distillation $\mathcal{L}_{distill}$.

First, a binary expert model (Teacher) is pretrained on the AD and HC data subsets to learn the baseline low-dimensional manifolds of neurodegenerative diseases, and its encoder weights are migrated to initialize the three-class student model (Student). Subsequently, a masked attention distillation $\mathcal{L}_{distill}$ is employed to utilize the attention distribution of the Teacher model as a soft label to guide the Student model:

$$\mathcal{L}_{distill} = \frac{1}{\sum_{i=1}^B V_i + \epsilon} \sum_{i=1}^B \|A_s^{(i)} - A_t^{(i)}\|_2^2 \cdot V_i \quad (5)$$

where V_i represents the effective mask. Since the Teacher model has not learned FTD samples, $V_i = 0$ when the ground-truth label is FTD; otherwise, $V_i = 1$. This mechanism forces the Student model to align with the Teacher model on AD/HC samples while enabling autonomous feature learning on FTD samples. The total classification loss is defined as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{distill} \quad (6)$$

where $\mathcal{L}_{CE}$ represents the cross-entropy loss with class-balanced weights, and λ is a hyperparameter regulating the distillation intensity.

To prevent the randomly initialized three-class decision head from generating excessively large anomalous

gradients during the early stages of training and potentially corrupting the robust feature manifolds transferred from the Teacher model, this paper adopts a two-stage fine-tuning strategy. In the first N epochs of training (the warm-up phase), the model freezes all spatio-temporal embedding layers and backbone encoders, optimizing only the classifier layers. Once the classification head establishes a preliminary decision surface, the model unfreezes all network parameters and simultaneously introduces the masked attention distillation $\mathcal{L}_{distill}$ for end-to-end collaborative optimization.

4 Experiments and analysis

4.1 Dataset and Preprocessing

The proposed method is evaluated on a publicly available clinical resting-state scalp EEG dataset (accessible via OpenNeuro under accession number ds004504). Collected in a routine clinical setting, the dataset comprises eyes-closed resting-state EEG recordings from 88 participants, including 36 patients diagnosed with AD, 23 patients with FTD, and 29 healthy controls (HC). In total, the dataset contains 485.5 minutes of AD recordings, 276.5 minutes of FTD recordings, and 402 minutes of HC recordings.

To standardize the data format and minimize computational overhead, all raw EEG signals are first resampled from their original 500 Hz to 128 Hz. Following resampling, 19 channels from the international standard 10–20 system (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz) are extracted for subsequent analysis. Considering the inherent non-stationarity of EEG signals and the substantial data requirements of deep learning models, a sliding window strategy is adopted to segment the continuous EEG time series. Specifically, the window length is set to 1 second (corresponding to 128 time steps), with an overlap rate of 50% between adjacent windows.

4.2 Experimental Setup and Evaluation Metrics

All experiments are implemented using the PyTorch framework. The AdamW optimizer is employed, and the initial learning rate is dynamically decayed using a cosine annealing strategy (Cosine Annealing LR) to guarantee stable convergence during the later stages of training. To mitigate overfitting, an early stopping mechanism based on the validation set F1-score is incorporated into the training process. To ensure a rigorous evaluation and clinical generalizability, the dataset partitioning strictly adheres to the subject-independent principle; that is, slice samples belonging to the same subject never appear simultaneously in both the training and testing sets.

Regarding performance evaluation, accuracy and F1-score are adopted as the core evaluation metrics. Furthermore, to align with practical clinical diagnostic scenarios, a subject-level evaluation mechanism is developed in addition to the sample-level evaluation.

Specifically, a majority voting scheme is applied to the prediction results of all sample slices belonging to a single subject to determine the final diagnostic label for that subject.

4.3 Baseline Comparisons

To comprehensively and objectively evaluate the diagnostic efficacy of the proposed dual-stream spatio-temporal framework in dementia EEG recognition tasks, five highly representative baseline models are selected for comparison under a strictly unified subject-independent partitioning standard. These baselines encompass the mainstream paradigms of current medical time-series analysis, including: handcrafted feature engineering (Manual Feature), compact and multi-scale convolutional neural networks (EEGNet, EEGInception), a multi-resolution graph neural network (MedGNN), and a pure temporal global attention architecture (Transformer). The quantitative evaluation results at both the segment and subject granularities, averaged over 5 independent randomized runs on the ADFTD-RS dataset, are summarized in Table 1, with the corresponding multi-metric profile visually contrasted in Fig. 3.

Comparative analysis reveals that traditional approaches and single-modality deep learning architectures exhibit clear representation limitations when handling the high non-stationarity and subtle inter-class variances of the ADFTD-RS dataset. As a traditional baseline, the Manual Feature method achieves an F1-score of 45.18%, illustrating the deficiency of handcrafted prior-based feature extraction workflows in capturing high-frequency dynamic temporal details. Among end-to-end frameworks, the classic convolutional model, EEGNet, yields a lower F1-score (42.62%) accompanied by a relatively large evaluation variance ($\pm 8.61\%$). Although EEGInception improves the F1-score to 47.39% by introducing multiscale convolutional kernels, purely convolutional architectures remain constrained by local receptive fields, rendering them deficient in perceiving large-scale global networks.

As baseline representatives of spatial topological and temporal sequence modeling, the multi-resolution graph network MedGNN (49.76%) and the pure temporal Transformer (50.57%) exhibit improved F1-scores, yet both suffer from high evaluation variances ($\pm 9.05\%$ and $\pm 8.76\%$, respectively). This phenomenon indicates that when confronting AD and FTD, which feature highly similar waveforms and clinical symptoms, relying solely on graph networks or global attention mechanisms renders the models susceptible to interference from transient noise or redundant connectivity edges, thereby generating large prediction fluctuations during the optimization of complex decision boundaries.

In contrast, the proposed dual-stream architecture (Ours) achieves the highest diagnostic performance among all compared models, with a subject-level accuracy of 60.00% and an F1-score of 57.44%. Compared to the top-performing pure temporal baseline, the Transformer (F1-score of 50.57%), the proposed model

improves the F1-score by 6.87% while converging the prediction variance to $\pm 7.38\%$. This improvement in performance metrics validates our core motivation: the proposed bidirectional rectification mechanism alleviates the noise amplification and representation bottlenecks of single-domain features, while the attention-guided distillation strategy enhances the stability of complex decision boundaries under imbalanced clinical data constraints, thereby jointly providing a more reliable differential diagnosis.

4.4 Ablation Studies

To systematically verify the functional contribution of each core component within the proposed dual-stream spatio-temporal network, comprehensive ablation experiments are conducted on the ADFTD-RS dataset. Specifically, five distinct architectural variants are constructed to evaluate the individual modules: (1) w/o Graph, which omits the effective connectivity graph spatial branch; (2) w/o STSBR, which removes the spatio-temporal synergistic bidirectional rectification mechanism; (3) w/o STCAF, which replaces the cross-attention fusion module with a conventional concatenation strategy; (4) w/o Distill, which excludes the mask-guided attention distillation strategy; and (5) w/o Warmup, which bypasses the two-stage warm-up fine-tuning strategy. As summarized in Table 2, all these architectural variants are evaluated under a strictly unified hyperparameter configuration and subject-level majority voting to guarantee fairness and clinical instructiveness.

To validate the necessity of spatio-temporal complementarity, we first isolated the temporal and spatial representations. Compared to complete model (Accuracy 60.00%), both the pure spatial model (w/o Temporal, Accuracy 56.84%) and the pure temporal model (w/o Graph, Accuracy 58.95%) exhibit overall diagnostic degradation. This accuracy reduction explicitly demonstrates the representation limitations of single-domain architectures: relying solely on either the temporal or spatial stream inevitably loses partial complementary information and leaves the model susceptible to domain-specific interference. Notably, while the w/o Graph variant yields a lower overall accuracy, its F1-score (58.09%) is marginally higher than that of the full model (57.44%). In clinical classification tasks, the F1-score serves as a rigorous metric for evaluating a model’s balanced diagnostic capability across all categories, heavily penalizing prediction biases toward majority classes. This metric discrepancy indicates that the pure temporal architecture possesses a distinct advantage in extracting fine-grained, localized rhythmic anomalies, which effectively maintains unbiased decision boundaries and preserves the F1-score. However, the absence of spatial graph topology limits the model’s capacity to characterize large-scale global network degenerations. Since these global features are crucial for accurately recognizing the majority classes (AD and HC), their absence consequently reduces the overall accuracy.

Table 1 Performance comparison between the proposed method and mainstream baseline models (Note: The experimental results are reported as mean $\pm$ standard deviation over 5 independent runs).

Model	Sample Accuracy (%)	Sample F1-score (%)	Subject Accuracy (%)	Subject F1-score (%)
Manual Feature	42.23 $\pm$ 3.34	40.59 $\pm$ 2.52	49.47 $\pm$ 2.58	45.18 $\pm$ 4.75
EEGNet	49.13 $\pm$ 4.93	40.96 $\pm$ 3.93	52.63 $\pm$ 9.42	42.62 $\pm$ 8.61
EEGInception	48.07 $\pm$ 1.39	45.86 $\pm$ 1.87	51.58 $\pm$ 6.14	47.39 $\pm$ 5.16
Transformer	45.87 $\pm$ 2.76	43.41 $\pm$ 2.87	53.68 $\pm$ 9.06	50.57 $\pm$ 8.76
MedGNN	46.31 $\pm$ 3.45	44.34 $\pm$ 3.34	52.63 $\pm$ 9.42	49.76 $\pm$ 9.05
Ours	51.87 $\pm$ 5.48	49.84 $\pm$ 5.97	60.00 $\pm$ 6.32	57.44 $\pm$ 7.38

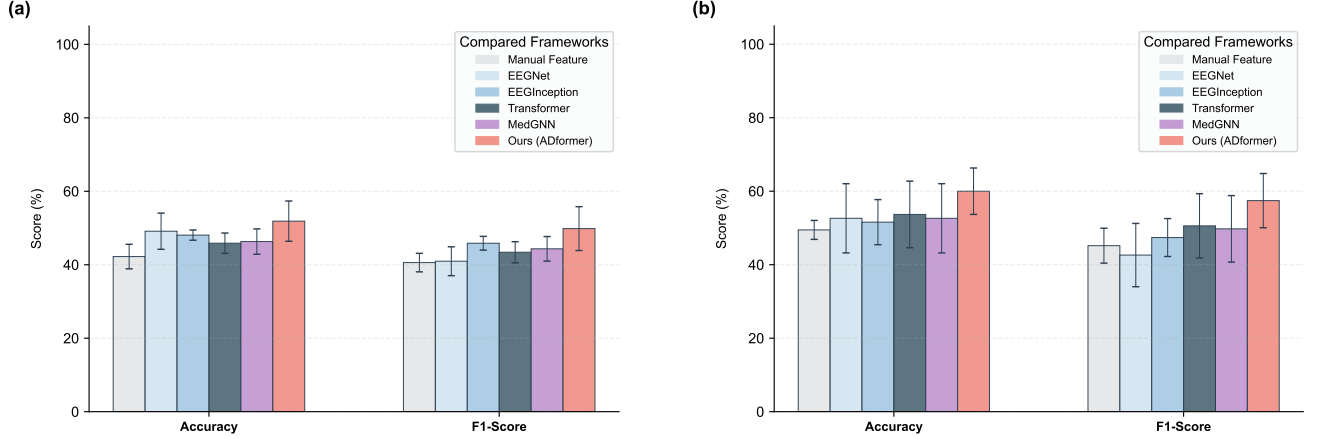


Fig. 3 Quantitative performance profiling of the proposed method against mainstream baselines on the ADFTD-RS dataset across 5 independent randomized runs. (a) Sample-level classification scores evaluated over 1-second EEG temporal slices; (b) Subject-level diagnostic performance metrics optimized via the majority voting scheme. Error bars denote the standard deviation, capturing the stability and statistical superiority sustained by our dual-stream framework.

We further evaluated the effectiveness of the feature fusion mechanisms. Removing the spatio-temporal sequential bidirectional rectification (w/o STSBR) leads to a noticeable performance degradation, with the subject-level F1-score decreasing to 42.34%. This indicates that directly combining spatial and temporal features without prior calibration introduces feature conflicts. This is likely because the inherent noise in both the temporal and spatial features mutually interferes. The STSBR module mitigates this issue by utilizing cross-stream context for mutual feature recalibration before integration. Furthermore, substituting the spatio-temporal cross-attention fusion with a standard concatenation strategy (w/o STCAF) reduces the F1-score to 56.51%. This demonstrates that the cross-attention mechanism is more effective than simple vector concatenation.

Finally, we evaluated the proposed distillation strategy. Removing this distillation module (w/o Distill) yields an F1-score of 54.64%. This indicates that the distillation strategy provides additional supervision, which helps mitigate overfitting on the imbalanced dataset and effectively improves classification performance. Additionally, omitting the initial warmup phase (w/o Warmup) reduces the F1-score to 52.96%. This suggests that freezing the backbone during early training prevents the learned representations from being disrupted by uncalibrated gradients, thereby aiding in stable model convergence.

4.5 Neurophysiological Interpretability Analysis

As demonstrated in the Section 4.4 ablation study, both the spatial stream and the STSBR module play an important role in capturing complementary network topologies and filtering inherent estimation artifacts. To explicitly investigate the intrinsic causal patterns captured by the spatial stream and visually validate these architectural contributions, continuous spatial topographic profiles were generated. Specifically, effective connectivity adjacency matrices were first estimated using a 5th-order MVAR model under the Granger causality framework. Node-level causal out-degree vectors were subsequently derived via column-summing and Min-Max normalization, and then projected onto the 19 channels of the international 10–20 system using cubic spline interpolation, as illustrated in Fig. 4.

The spatial distributions present connectivity variations among the categories, corresponding to their respective pathological mechanisms. Specifically, the AD group shows higher causal outflow along the centroparietal midline axis (channels Cz and Pz), with a relative reduction in the lateral frontotemporal regions (F7, F8, T4, and T6). This posterior distribution relates to the impairment of default mode network (DMN) hubs, including the precuneus and posterior cingulate cortex, during AD progression. In this context, this outflow is interpreted as compensatory synchronization, where the degenerating neural network maintains

Table 2 Ablation study quantitative results on the ADFTD-RS dataset (reported as mean $\pm$ standard deviation across 5 independent runs).

Model	Sample Accuracy (%)	Sample F1-score (%)	Subject Accuracy (%)	Subject F1-score (%)
w/o Temporal	51.06 $\pm$ 4.06	49.33 $\pm$ 4.58	56.84 $\pm$ 6.14	55.24 $\pm$ 6.80
w/o Graph	49.69 $\pm$ 3.20	48.42 $\pm$ 3.21	58.95 $\pm$ 5.16	58.09 $\pm$ 4.09
w/o STSBR	39.71 $\pm$ 3.34	39.48 $\pm$ 3.49	43.16 $\pm$ 10.21	42.34 $\pm$ 10.63
w/o STCAF	49.80 $\pm$ 3.23	48.99 $\pm$ 3.33	56.84 $\pm$ 6.98	56.51 $\pm$ 6.25
w/o Distill	49.68 $\pm$ 3.36	48.24 $\pm$ 3.36	55.79 $\pm$ 2.58	54.64 $\pm$ 3.10
w/o Warmup	46.08 $\pm$ 4.20	44.99 $\pm$ 4.49	53.68 $\pm$ 6.14	52.96 $\pm$ 6.55
Ours	51.87 $\pm$ 5.48	49.84 $\pm$ 5.97	60.00 $\pm$ 6.32	57.44 $\pm$ 7.38

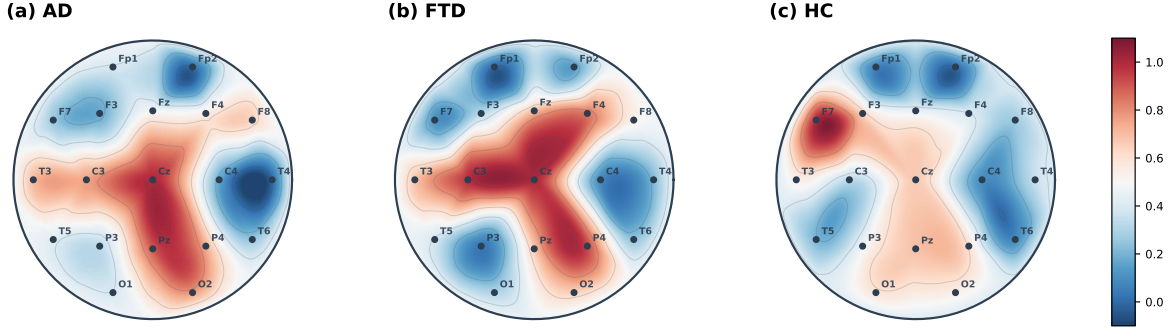


Fig. 4 Topographic visualization of the average normalized causal outflow intensity across (a) AD, (b) FTD, and (c) HC cohorts. The spatial heatmaps represent the interpolated node-level causal out-degree distributions, outlining distinct configurations across the three groups.

local synchrony to support cross-regional communication despite structural disconnection. The FTD group demonstrates an anteriorly shifted information outflow. This involves a frontocentral cluster (Fz, Cz, F3, and F4) and relative lateralization in the left temporal region. In clinical neurophysiology, this frontocentral pattern reflects cortical hyperexcitability and disinhibition. These alterations follow the selective loss of local GABAergic inhibitory interneurons during frontotemporal lobar degeneration (FTLD), leading to the synchronization of the remaining pyramidal neuronal populations. Furthermore, the left temporal observation corresponds to the asymmetric cortical atrophy associated with behavioral or semantic variants of frontotemporal dementia. In contrast, the HC group shows a uniform global profile, with localized activity in the left anterior frontotemporal region at channel F7. This reflects physiological top-down regulation and network efficiency within the frontoparietal control network in the absence of neurodegeneration.

These topographic variations provide a neurophysiological context for the quantitative findings in the ablation study. Since the AD and FTD groups present distinct effective network configurations despite overlapping temporal rhythms, modeling the spatial topology via the spatial stream mitigates the limitations of single-domain representations.

5 Discussion

This paper suggests that integrating temporal sequences and spatial topologies can provide complementary information for the differential diagnosis of AD and FTD. While single-domain deep learning models are widely

used in EEG analysis, our ablation findings indicate potential limitations when relying on single-domain representations to differentiate these highly overlapping neurodegenerative conditions. Specifically, pure temporal networks, while effective at capturing localized rhythmic slowing, may not fully capture macroscopic structural degradation. Conversely, pure spatial graphs show higher prediction variance, likely due to the lack of dynamic contextual constraints. This observation implies that coupling local signal characteristics with global network connectivity could yield a more comprehensive representation of neurodegeneration, thereby stabilizing the diagnostic performance.

Furthermore, the direct integration of spatial and temporal features can introduce feature interference. As observed in the unrectified fusion variant (w/o STSBR), directly combining the two streams leads to a performance drop. Because short-window scalp EEG inherently contains temporal noise and non-stationarity, it not only affects the temporal representations but also introduces spurious causality in the estimated effective connectivity graphs. Direct integration may propagate both the temporal noise and these spurious causal edges to the classifier, exacerbating the interference between the spatial and temporal domains. The STSBR mechanism mitigates this issue by utilizing temporal dynamics to evaluate and prune unreliable spatial links prior to deep integration, indicating that dynamic cross-calibration can be a beneficial step when fusing spatial and temporal EEG features.

Despite these methodological observations, several constraints affect the framework’s immediate clinical deployment. First, the evaluation relies solely on resting-state EEG records from a single database; thus,

further validation on multi-center datasets is necessary to assess its generalization ability. Second, relying on isolated EEG segments (epochs) prevents the framework from tracking the long-term evolution of brain function as the disease progresses. Finally, the computational overhead of the dual-stream architecture restricts its real-time execution on resource-constrained edge devices. Future research will explore architectural compression techniques to facilitate portable screening and integrate longitudinal multimodal indices to further refine diagnostic reliability.

6 Conclusions

In conclusion, this paper presents a dual-stream framework that integrates temporal sequences and spatial topologies for the differential diagnosis of AD and FTD using resting-state EEG. Our evaluations indicate that this coupled representation compensates for the limitations of single-domain analyses, yielding more stable diagnostic performance. By employing the STSBR mechanism, the framework mitigates feature interference by filtering temporal noise and spurious causal edges across both streams prior to deep fusion. Furthermore, the mask-guided distillation strategy provides a practical regularization approach to address the inherent data imbalance in clinical cohorts. While further validation on multi-center datasets is necessary, this framework offers a promising and non-invasive computational tool to assist in the screening of neurodegenerative diseases.

Declarations

- **Funding:** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
- **Conflict of interest:** The authors declare no competing interests.
- **Data availability:** The ADFTD-RS dataset analyzed during the current study is available in the OpenNeuro repository (ds004504).

References

- [1] Hansson, O.: Biomarkers for neurodegenerative diseases. *Nature Medicine* **27**(6), 954–963 (2021)
- [2] Miltiadous, A., Tzamourta, K.D., Afrantou, T., *et al.*: A dataset of scalp eeg recordings of alzheimer’s disease, frontotemporal dementia and healthy subjects from routine eeg. *Data* **8**(6), 95 (2023)
- [3] Miltiadous, A., Tzamourta, K.D., Giannakeas, N., *et al.*: Alzheimer’s disease and frontotemporal dementia: A robust classification method of eeg signals and a comparison of validation methods. *Diagnostics* **11**(8), 1437 (2021)
- [4] Cohen, M.X.: Where does eeg come from and what does it mean? *Trends in Neurosciences* **40**(4), 208–218 (2017)
- [5] Kim, M., Youn, Y.C., Paik, J.: Deep learning-based eeg analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset. *NeuroImage* **272**, 120054 (2023)
- [6] Rossini, P.M., Miraglia, F., Vecchio, F.: Early dementia diagnosis, mci-to-dementia risk prediction, and the role of machine learning methods for feature extraction from integrated biomarkers, in particular for eeg signal analysis. *Alzheimer’s & Dementia* **18**(12), 2699–2706 (2022)
- [7] Chaddad, A., Wu, Y., Kateb, R., *et al.*: Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors* **23**(14), 6434 (2023)
- [8] Qiu, S., Miller, M.I., Joshi, P.S., *et al.*: Multimodal deep learning for alzheimer’s disease dementia assessment. *Nature Communications* **13**(1), 3404 (2022)
- [9] Cheng, J., Chen, M., Li, C., *et al.*: Emotion recognition from multi-channel eeg via deep forest. *IEEE Journal of Biomedical and Health Informatics* **25**(2), 453–464 (2020)
- [10] Yang, Y., Wu, Q., Qiu, M., *et al.*: Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–7 (2018). IEEE
- [11] Vaswani, A., Shazeer, N., Parmar, N., *et al.*: Attention is all you need. *Advances in Neural Information Processing Systems* **30**, 5998–6008 (2017)
- [12] Wang, Y., Huang, N., Li, T., *et al.*: Medformer: A multi-granularity patching transformer for medical time-series classification. *Advances in Neural Information Processing Systems* **37**, 36314–36341 (2024)
- [13] Miltiadous, A., Gionanidis, E., Tzamourta, K.D., *et al.*: Dice-net: a novel convolution-transformer architecture for alzheimer detection in eeg signals. *IEEE Access* **11**, 71840–71858 (2023)
- [14] Mlinarič, T., Van Den Kerchove, A., Barinaga, Z.I., *et al.*: Eeg-based classification of alzheimer’s disease and frontotemporal dementia using functional connectivity. *Scientific Reports* **16**, 35316 (2026)
- [15] Bessadok, A., Mahjoub, M.A., Rekik, I.: Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(5), 5833–5848 (2022)

- [16] Yang, H., Huang, R., Ye, S., *et al.*: A review of graph neural networks for brain diseases analysis. *Neurocomputing* **589**, 133174 (2026)
- [17] Bagherzadeh, S., Shalbaf, A., Shoeibi, A., *et al.*: Developing an eeg-based emotion recognition using ensemble deep learning methods and fusion of brain effective connectivity maps. *IEEE Access* **12**, 50949–50965 (2024)
- [18] Zhao, H., Wang, Z., Peng, D., *et al.*: St-decnet: A spatio-temporal graph learning framework for epileptic seizure prediction via dynamic effective connectivity in eegs. *IEEE Transactions on Instrumentation and Measurement* **74**, 9512812 (2025)
- [19] Yuan, N., Guan, D., Li, S., *et al.*: Enhancing neurodegenerative disease diagnosis through confidence-driven dynamic spatio-temporal convolutional network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **33**, 1245–1256 (2025)
- [20] Liu, Y., Deng, J., Xu, R., *et al.*: Stcwtntnet: Spatio-temporal channel-wise transformer network for interpretable cross-subject mild cognitive impairment detection using eeg signals. *Biomedical Signal Processing and Control* **118**, 109564 (2026)
- [21] Khalfallah, S., Puech, W., Tlija, M., *et al.*: Exploring the effectiveness of machine learning and deep learning techniques for eeg signal classification in neurological disorders. *IEEE Access* **13**, 17002–17015 (2025)
- [22] Kowshiga, A., Pavithra, T., Priyanka, V.: Deep learning into the future: Hybrid cnn-rnn for early detection of alzheimer’s disease. In: 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 940–946 (2024). IEEE
- [23] Zini, S., Barbera, T., Bianco, S., *et al.*: Alzheimer’s disease classification from eeg using a multiscale temporal deep network. *Biomedical Signal Processing and Control* **114**, 109321 (2026)
- [24] Khan, W., Khan, M.S., Qasem, S.N., *et al.*: An explainable and efficient deep learning framework for eeg-based diagnosis of alzheimer’s disease and frontotemporal dementia. *Frontiers in Medicine* **12**, 1590201 (2025)
- [25] Acharya, M., Deo, R.C., Barua, P.D., *et al.*: Eegconvnext: a novel convolutional neural network model for automated detection of alzheimer’s disease and frontotemporal dementia using eeg signals. *Computer Methods and Programs in Biomedicine* **262**, 108652 (2025)
- [26] Xu, Q., An, L., Yang, H., *et al.*: A multi-graph convolutional network method for alzheimer’s disease diagnosis based on multi-frequency eeg data with dual-mode connectivity. *Frontiers in Neuroscience* **19**, 1555657 (2025)
- [27] Cao, J., Li, B., Li, X.: Identification of alzheimer’s disease brain networks based on eeg phase synchronization. *Biomedical Engineering Online* **24**(1), 32 (2025)
- [28] Cao, J., Yang, L., Sarrigiannis, P.G., *et al.*: Dementia classification using a graph neural network on imaging of effective brain connectivity. *Computers in Biology and Medicine* **168**, 107701 (2024)
- [29] Dong, A., Cao, J., Sarrigiannis, P.G., *et al.*: A causal validation augmented temporal convolutional framework for brain effective connectivity networks estimation. *Neural Networks* **178**, 108405 (2025)
- [30] Zhang, J., Liu, H., Yang, K., *et al.*: Cmx: Cross-modal fusion for rgb-x semantic segmentation with transformers. *IEEE Transactions on Intelligent Transportation Systems* **24**(12), 14679–14694 (2023)
- [31] Yan, N., Chen, G., Zhang, X.: Emotion recognition using multi-view eeg-fnirs and cross-attention feature fusion. *Biosensors* **16**(3), 145 (2026)
- [32] Shi, K., Liu, A., Zhang, J., *et al.*: Medical image fusion based on multilevel bidirectional feature interaction network. *IEEE Sensors Journal* **24**(12), 19428–19441 (2024)