

LEARNABLE FRACTIONAL FOURIER AND GRAPH FRACTIONAL OPERATORS FOR NONSTATIONARY GRAPH SIGNALS VALIDATED WITH EEG SEIZURE DETECTION

Anonymous authors

Paper under double-blind review

ABSTRACT

Nonstationary graph signals with time-varying spectral properties and evolving network topologies present fundamental challenges for existing deep learning architectures. We introduce learnable fractional operators that bridge time-frequency analysis and functional connectivity through trainable fractional orders. We propose EEG-GraphFrFT, a unified dual-path framework implementing this approach. The first path employs a fractional Fourier transform with a trainable order to adaptively capture nonstationary, transient patterns. The second path constructs functional networks derived from wPLI and Spectral Granger causality and applies graph fractional operators to model complex network interactions. We establish the minimal theoretical properties required for stable training, namely well-posedness and Hölder-type stability, under mild spectral assumptions. A parameter-efficient low-rank cross interaction integrates the two paths. As a challenging validation, we evaluate on epileptic seizure detection across three public datasets (FMCE, HUP, and Helsinki neonatal EEG) under strict subject-disjoint conditions (no leakage). EEG-GraphFrFT consistently outperforms strong baselines (e.g., EEG-Conformer, Mamba) by approximately 2–8 % in accuracy, while demonstrating robust performance under colored Gaussian noise and channel dropouts, with corresponding improvements in F1 and AUROC. Beyond EEG, the graph fractional operators are task-agnostic and apply broadly to nonstationary graph signals, e.g., traffic sensor networks, climate teleconnections, and multi-asset financial series.

1 INTRODUCTION

Epilepsy affects over 70 million people worldwide (Thijs et al., 2019). Accurate EEG-based diagnosis is clinically critical yet difficult (Stamoulis et al., 2012): EEG encodes rich spatiotemporal dynamics (Noorlag et al., 2022) and seizure-related state transitions (Li et al., 2019), but expert review of long multi-channel recordings is time-consuming and subjective. Reliable computer-aided diagnosis (CAD) is therefore a central goal (Pontes et al., 2024). Seizure EEG also exemplifies a broader challenge: nonstationary graph signals, where node features and graph topology evolve jointly. Standard deep models often assume stationary features or static graphs; performance degrades when both fail. Similar issues arise in stressed financial markets, disrupted traffic networks, and viral information spread.

Classical Fourier analysis cannot capture time-varying spectra (Yi et al., 2023). Graph Fourier transforms (Shuman et al., 2013) address static networks but not temporal evolution of graph spectra. Advances in fractional calculus and graph signal processing (Guibas et al., 2021a; Ortega et al., 2018) suggest that fractional-order operators are well-suited for such dynamics. AFNO leverages fractional transforms in vision (Guibas et al., 2021a); graph fractional Fourier transforms extend to irregular domains (Alikasıfoğlu et al., 2024). Yet existing approaches typically lack: (i) learnable fractional parameters, (ii) stability guarantees under graph perturbations and noise, and (iii) unified architectures that couple temporal nonstationarity with graph dynamics.

Related work. The Fractional Fourier Transform (FrFT) has emerged as a key operator in deep learning due to its unique properties. (Guibas et al., 2021b) developed the Adaptive Fourier Neural Operator (AFNO), which leverages FrFT’s rotational characteristics to enable quasi-linear token mixing in vision transformers, reducing computational complexity by 30% compared to standard self-attention. (Liu et al., 2019) combined FrFT with multi-scale wavelet decompositions, creating Fractional Wavelet Scattering Networks for pathological image analysis. (Chen et al., 2025) designed Multi-scale Fractional Fourier Convolutional Neural Networks (MFFCNN) with fractional orders α , achieving strong results on segmentation benchmarks. Concurrently, the Graph Fractional Fourier Transform (GFRFT) has extended these concepts to graph structures. (Alikasıfoğlu et al., 2024) established GFRFT as a continuous rotational operator in the vertex-frequency plane. Methodological innovations include fractional translation operators for vertex-localized filtering (Yan et al., 2020) and gradient-based order optimization (Zhang and Li, 2025), with applications demonstrating significant improvements in graph signal denoising (Gan et al., 2023) and classification accuracy (Wei and Yan, 2024).

Contributions. We address these gaps with a dual-path framework, instantiated for seizure EEG while remaining applicable to generic nonstationary graph signals. Path 1 performs adaptive time–frequency analysis via a fractional Fourier (FrFT) layer with a learnable order. Path 2 constructs dynamic functional networks using wPLI and Spectral Granger causality, then applies graph fractional operators to model evolving interactions. Under mild spectral assumptions we provide theoretical guarantees—well-posedness and Hölder-type stability—supporting stable training. A parameter-efficient low-rank bilinear fusion integrates the two paths.

On three public datasets (FMCE, HUP, Helsinki-neonatal) with strict subject-disjoint splits, our model consistently improves over strong baselines (EEG-Conformer, FreTS, iTransformer, Brain-JEPA, Mamba, FAPEX) by approximately 2–8 % in accuracy, with corresponding gains in F1 and AUROC, and remains robust under colored noise and channel dropouts. Although we focus on EEG, the proposed learnable fractional operators are task-agnostic and naturally extend to traffic sensor networks, climate teleconnections, and multi-asset financial series. Our contributions can be summarized as follows:

Learnable fractional operator framework. A trainable GraphFrFT and fractional operator layer with adaptive exponential modulation for nonstationary time–frequency analysis.

Graph fractional stability theory. Hölder-type stability bounds for fractional graph filters under bounded structural perturbations; training is well-posed under mild spectral conditions.

Unified nonstationary graph modeling. A dual-path design that fuses signal-space FrFT and graph-space fractional filtering via low-rank fusion, yielding state-of-the-art seizure detection while retaining cross-domain applicability.

2 METHOD

The EEG-GraphFrFT framework we proposed is a dual-path architecture designed to comprehensively capture the complex and non-stationary features of electroencephalogram (EEG) signals for seizure detection, as shown in Figure 1. This model processes the original multi-channel EEG input through two complementary and domain-specific channels, each of which is customized to extract different but complementary feature representations. CHAN1 directly operates on time series signals and uses Generalized Fractional Neural Operator (GFNO) with learnable fractional parameters for adaptive time-frequency analysis. CHAN2 builds a dynamic functional connectivity network from the same EEG input, quantifies undirected synchronization (via wPLI) and directed causal effects (via Spectral Granger causality), and then processes it through the graph fractional Fourier Transform (GFRFT) layer to extract the spectral features of the graph. A dedicated feature fusion module integrates the feature mappings of the two paths, and then generates the final prediction through a series of fusion post-processing modules and task-specific classifiers. This design enables the model to synergistically utilize signal space and graph space representations, providing a solid foundation for detecting pathological brain dynamics related to epilepsy.

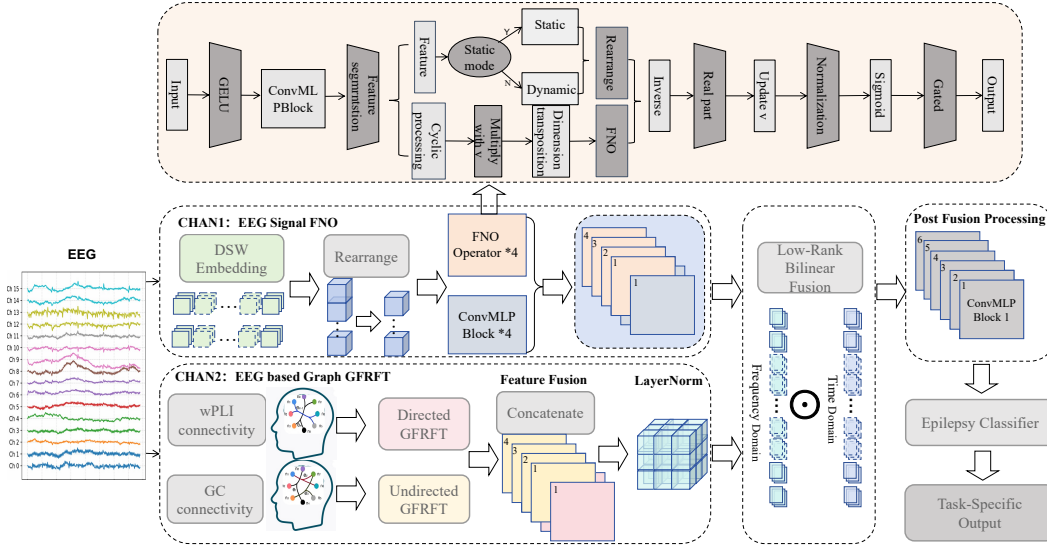


Figure 1: **The EEG-GraphFrFT framework employs a dual-path architecture:** CHAN1 applies trainable fractional neural operators for adaptive time–frequency analysis, while CHAN2 constructs functional networks (wPLI/GC) processed via graph fractional operators. Features are fused via parameter-efficient low-rank bilinear pooling for seizure detection.

2.1 GENERALIZED FRACTIONAL NEURAL OPERATOR

Real-world EEG signals are inherently non-stationary, exhibiting complex temporal dynamics such as drifting instantaneous frequencies, intermittent oscillatory patterns, and amplitude-phase modulations, particularly during epileptic seizures. Traditional signal processing techniques, including standard Fourier and wavelet transforms, often fail to adequately capture these behaviors due to their fixed basis functions and inherent assumption of stationarity. To address this limitation, we introduce a *Generalized Fractional Neural Operator (GFNO)*, a novel learnable operator that extends fractional pseudo-differential calculus within a deep learning framework, enabling adaptive time-frequency analysis tailored to non-stationary EEG characteristics. The mathematical foundation of our approach rests on the generalization of classical convolution and differentiation through fractional pseudo-differential operators. These operators enable adaptive joint time-frequency analysis by incorporating a rotational parameter that extends traditional Fourier analysis, enabling more flexible representation of non-stationary, highly variable signals such as EEG.

Definition 1 (Fractional Pseudo-Differential Operator). Let $a(t, \xi)$ be a sufficiently smooth symbol function defined on $\mathbb{R}_t \times \mathbb{R}_\xi$. For any rotation angle $\theta \notin \pi\mathbb{Z}$ and for any $\phi \in L^2(\mathbb{R})$, the associated *fractional pseudo-differential operator* $T_a^\theta : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ is defined as:

$$(T_a^\theta \phi)(t) = \int_{\mathbb{R}} K_{-\theta}(t, \xi) a(t, \xi) \hat{\phi}^\theta(\xi) d\xi, \quad (1)$$

where $\hat{\phi}^\theta(\xi) \equiv \mathcal{F}^\theta[\phi](\xi)$ denotes the Fractional Fourier Transform (FrFT) of ϕ of order θ , and $K_{-\theta}(t, \xi)$ represents the kernel associated with the fractional transformation.

To operationalize this mathematical framework within a deep learning architecture, we introduce the *Neural Fractional Operator (NFO)*, which parameterizes the symbol function $a(t, \xi)$ using learnable components while maintaining the theoretical properties of fractional operators. Let $X \in \mathbb{R}^{L \times C}$ represent a multivariate EEG signal with L time samples and C channels. The symbol function $a(t, \xi)$ is factorized into separable time-dependent and frequency-dependent components $a(t, \xi) = u(t)v(\xi)$, where $u : \mathbb{R} \rightarrow \mathbb{R}^C$ and $v : \mathbb{R} \rightarrow \mathbb{R}^C$ are learned functions capturing temporal and spectral profiles respectively. The Neural Fractional Operator $\mathcal{T}_a^\theta : L^2(\mathbb{R})^C \rightarrow L^2(\mathbb{R})^C$ of order $\theta \notin \pi\mathbb{Z}$ operates on signal $\varphi \in L^2(\mathbb{R})^C$ as:

$$(\mathcal{T}_a^\theta \varphi)(t) = u(t) \mathcal{F}^{-\theta} [v(\xi) \odot \mathcal{F}^\theta[\varphi](\xi)](t) \quad (2)$$

where $F^\theta[\cdot]$ denotes the FrFT of order θ , and \odot is channel-wise multiplication. The NFO is parameterized by two hypernets HyperNet^t and HyperNet^s producing $u(t)$ and $v(\xi)$. We set the trainable order as $\theta = \frac{\pi}{2} \tanh(\beta_\theta)$ with learnable β_θ (compact domain, stable gradients). Both hypernets are single-layer MLPs with Snake activation $A_\beta(x) = x + \beta^{-1} \sin^2(\beta x)$ and learnable $\beta > 0$ to mitigate spectral bias. To stabilize high-frequency behavior we apply a nonnegative frequency gate $w(\xi) = \text{SoftPlus}(\gamma_0 + \gamma_1|\xi|)$ and use $\tilde{v}(\xi) = \exp(-w(\xi))v(\xi)$ inside the operator; this yields monotone, learnable attenuation without numerical instability.

Positional encoding. Positional information is encoded via Fourier embeddings:

$$z(t) = [t_{\text{rescaled}}, \{\cos(f_k \cdot w)\}_{k=1}^B, \{\sin(f_k \cdot w)\}_{k=1}^B],$$

where $t_{\text{rescaled}} = t/T$, $w = 2\pi t/T$, and $f_k = 2^{k-1}$ for $k = 1, \dots, B$. Here T is the sequence length and $\{f_k\}$ form geometric frequency bands, yielding $z(t) \in \mathbb{R}^{1+2B}$.

2.2 GRAPH-BASED FUNCTIONAL CONNECTIVITY ANALYSIS WITH FRACTIONAL FOURIER TRANSFORM

Functional Connectivity Networks with wPLI and GC We have established a functional connection network. It captures the dynamic changes among complex regions in the brain activity of epilepsy patients. We use two complementary metrics. The Weighted Phase Lag Index (wPLI) measures phase-based synchronization (see Appendix B.2.1). Granger causality (GC) captures the directed causal information flow in the frequency domain (see Appendix B.2.2). The combination of this dual-function network construction method provides a complete capture of brain network information. It integrates undirected functional coupling with causal effects during epileptic seizures. See Appendix B.2.3 for details on construction of wPLI and GC graphs from raw signals. *Notation:* we use $(\cdot)^\top$ for transpose and $(\cdot)^H$ for conjugate transpose; for real-symmetric L_{sym} we write $L_{\text{sym}} = U\Lambda U^\top$, while for Hermitian L_H we write $L_H = U\Lambda U^H$.

Graph Fractional Operators via Functional Calculus (Unified) We operate *exclusively* on Hermitian (PSD) Laplacians to ensure real spectra and stable gradients. Let \mathbf{L}_\star denote the graph operator used for the spectral calculus: $\mathbf{L}_\star = \mathbf{L}_{\text{sym}}$ for undirected wPLI graphs, and $\mathbf{L}_\star = \mathbf{L}_H$ for GC graphs (Hermitianized; Appendix B.2.5). With $\mathbf{L}_\star = \mathbf{U}\mathbf{A}\mathbf{U}^H$ and $\alpha \in (0, 1]$, we define the fractional family

$$\mathbf{U}_{\theta, \alpha}(\mathbf{L}_\star) = e^{-i\theta \mathbf{L}_\star^\alpha}, \quad \mathbf{H}_{\tau, \alpha}(\mathbf{L}_\star) = e^{-\tau \mathbf{L}_\star^\alpha},$$

and use their real/imaginary parts as learnable graph-spectral filters. Gradients are analytic: $\partial_\alpha \mathbf{L}^\alpha = \mathbf{L}^\alpha \circ \log \mathbf{L}$ and $\partial_\theta e^{-i\theta \mathbf{L}^\alpha} = -i \mathbf{L}^\alpha e^{-i\theta \mathbf{L}^\alpha}$.

Implementation & complexity. During training we approximate \mathbf{L}_\star^α and the matrix exponentials by Chebyshev/Lanczos expansions of order K (typically $K = 8\text{--}16$), giving $\mathcal{O}(K|E|)$ time and $\mathcal{O}(K|V|)$ memory per layer.

Directed connectivity. For GC-derived graphs we build $\mathbf{L}_H = \mathbf{D}_s - \Gamma \odot \mathbf{W}_s$ (unit-modulus phase Γ ; Appendix B.2.2–B.2.3), which is Hermitian PSD. All fractional operators are applied to \mathbf{L}_H to guarantee a real spectrum and numerical stability.

Extended Fractional Graph Calculus The fractional graph derivative of order β is defined as $\mathcal{D}^\beta = F_G^{-\beta} \Lambda_\beta F_G^\beta$, where $\Lambda_\beta = \text{diag}((\lambda_k)^\beta)$ and λ_k are graph frequencies, extending classical fractional calculus to irregular domains. For a symmetrized adjacency matrix \mathbf{W}_s and unit-modulus phase matrix $\Gamma_q = (\Upsilon_q(i, j))$ satisfying $\Upsilon_q(i, j) = \overline{\Upsilon_q(j, i)}$ for all i, j , define $L_H = \mathbf{D}_s - \Gamma_q \odot \mathbf{W}_s$, where $\mathbf{D}_s = \text{diag}(\mathbf{W}_s \mathbf{1})$. Then $L_H = L_H^H$ and $x^H L_H x = \frac{1}{2} \sum_{i,j} W_{s,ij} |x_i - \Upsilon_q(i, j)x_j|^2 \geq 0$. Then, L_H is Hermitian ($L_H = L_H^*$), positive semidefinite ($x^* L_H x \geq 0$ for all $x \in \mathbb{C}^n$), recovers the standard symmetric Laplacian L_{sym} in the undirected case ($w_{ij} = w_{ji}$, $\Gamma_q \equiv 1$), and has eigenvalues on the nonnegative real axis by Gershgorin’s theorem (proof in Appendix B.2.6). For any diagonalization $L_H = UVU^*$ with $V = \text{diag}(\nu_k)$, define $\log V = \text{diag}(\log |\nu_k| + j \arg(\nu_k))$ and $\nu_k^\alpha = \exp(\alpha \log \nu_k)$, using the principal branch $\arg(\nu_k) \in (-\pi, \pi]$. Zero eigenvalues are shifted by a small ε , and tiny negative parts from numerical roundoff are clamped to zero, ensuring $\log V$ and V^α are well-defined and numerically stable.

GFRFT-Based Feature Extraction To extract discriminative features from functional connectivity networks, a dual-mode Graph Fourier Transform (GFT) framework is implemented, adapting to undirected (wPLI-based) and directed (GC-based) graph structures. The core innovation is an adaptive eigendecomposition strategy, based on graph directionality and a learnable fractional order parameter α .

Undirected Graph Processing (wPLI): For undirected networks derived from wPLI, the symmetric normalized Laplacian is computed as $\mathbf{L}_{\text{sym}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$, where \mathbf{D} is the degree matrix and \mathbf{A} is the symmetric adjacency matrix. Eigendecomposition preserves the spectral gap: $\mathbf{L}_{\text{sym}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top = \sum_{k=0}^{N-1} \lambda_k \mathbf{u}_k \mathbf{u}_k^\top$, with eigenvalues ordered $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1} = \lambda_{\text{max}}$. The learnable GFRFT is applied as $\mathbf{Z}_{\text{undir}} = \mathbf{U} \mathbf{\Lambda}^\alpha \mathbf{U}^\top \mathbf{X} = \sum_{k=0}^{N-1} \lambda_k^\alpha (\mathbf{u}_k^\top \mathbf{X}) \mathbf{u}_k$, where α is initialized at 0.5 and optimized via $\nabla_\alpha \mathcal{L} = \text{Tr} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{Z}_{\text{undir}}}^\top \mathbf{U} (\mathbf{\Lambda}^\alpha \circ \log \mathbf{\Lambda}) \mathbf{U}^\top \mathbf{X} \right)$.

Directed graphs (GC) via Hermitianization. For GC-derived directed connectivity we use the Hermitian Laplacian $L_H = D_s - \Gamma \odot W_s$ (unit-modulus phase Γ ; Appendix B.2.2–B.2.3). All fractional operators are applied to L_H , ensuring a real PSD spectrum. For GC-derived directed connectivity we apply fractional operators to L_H *per frequency band* (each band is Hermitianized and diagonalized independently); adjacent-band orders $\{\alpha_b\}$ are tied via a total-variation penalty to reduce overfitting.

Unified processing. We *always* apply fractional operators on a Hermitian PSD Laplacian:

$$\mathbf{Z} = \mathbf{U} \mathbf{\Lambda}^\alpha \mathbf{U}^H \mathbf{X}, \quad \mathbf{L}_* = \begin{cases} \mathbf{L}_{\text{sym}} & (\text{wPLI, undirected}), \\ \mathbf{L}_H & (\text{GC, directed; Hermitianized}). \end{cases}$$

This removes the unstable $\mathbf{A} = \mathbf{U} \mathbf{V} \mathbf{U}^{-1}$ branch and keeps gradients real-valued.

Multi-band Processing for GC: For Granger Causality networks, each frequency band is processed separately. GC matrices are computed for bands $\mathbf{A}_\delta, \mathbf{A}_\theta, \mathbf{A}_\alpha, \mathbf{A}_\beta, \mathbf{A}_\gamma$, and directed GFRFT is applied with independent α parameters. Band-specific features $\mathbf{h}_\delta, \mathbf{h}_\theta, \mathbf{h}_\alpha, \mathbf{h}_\beta, \mathbf{h}_\gamma$ are extracted and concatenated: $\mathbf{h}_{\text{gc}} = [\mathbf{h}_\delta \oplus \mathbf{h}_\theta \oplus \mathbf{h}_\alpha \oplus \mathbf{h}_\beta \oplus \mathbf{h}_\gamma]$. We share α across adjacent bands with a total-variation penalty on $\{\alpha_b\}$ to reduce overfitting.

Feature Fusion: The final representation combines modalities: $\mathbf{h}_{\text{final}} = \mathbf{h}_{\text{eeg}} \oplus \mathbf{h}_{\text{wpli}} \oplus \mathbf{h}_{\text{gc}}$, where \mathbf{h}_{eeg} are features from raw EEG, \mathbf{h}_{wpli} from wPLI graphs, and \mathbf{h}_{gc} from GC graphs.

Feature Fusion via Low-Rank Bilinear Interaction To integrate time-frequency (CHAN1) and graph (CHAN2) features, a low-rank bilinear fusion mechanism captures multiplicative interactions between feature vectors $\mathbf{s} \in \mathbb{R}^{d_s}$ and $\mathbf{g} \in \mathbb{R}^{d_g}$. This approach is more expressive than concatenation or addition and is parameter-efficient due to low-rank constraints. The output $\mathbf{y} \in \mathbb{R}^{d_o}$ is computed as:

$$\mathbf{y} = \mathbf{U}^\top [(\mathbf{V}_s^\top \mathbf{s}) \odot (\mathbf{V}_g^\top \mathbf{g})] + \mathbf{b}, \quad (3)$$

where $\mathbf{V}_s \in \mathbb{R}^{d_s \times r}$, $\mathbf{V}_g \in \mathbb{R}^{d_g \times r}$ project features to a shared r -dimensional space ($r \ll \min(d_s, d_g)$), $\mathbf{U} \in \mathbb{R}^{r \times d_o}$ projects to the output space, and \odot denotes the Hadamard product. This models second-order interactions efficiently, with details and derivations provided in Appendix B.2.9.

Stability of Graph Fractional Filters under Perturbations We study the stability of fractional graph filters under structural perturbations of the Laplacian. Throughout this subsection we assume that $L, \tilde{L} \in \mathbb{R}^{n \times n}$ are real symmetric positive semidefinite (PSD) Laplacians with spectra in $[0, \Lambda_{\text{max}}]$, and that $\alpha \in (0, 1], \tau > 0$. We write $\Delta = \tilde{L} - L$ and use $\|\cdot\|_2$ for the operator norm. Fractional powers are defined via functional calculus: $L^\alpha = U \mathbf{\Lambda}^\alpha U^\top$ for $L = U \mathbf{\Lambda} U^\top$. All proofs are deferred to Appendix B.2.7.

Theorem 1 (Hölder stability of fractional graph filters). *Let $L, \tilde{L} \succeq 0$ be PSD Laplacians, $\alpha \in (0, 1]$, and $\tau > 0$. Then there exists a constant $C_\alpha > 0$ depending only on α and a priori bounds on $\|L\|_2, \|\tilde{L}\|_2$ such that*

$$\|e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha}\|_2 \leq \tau C_\alpha \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (4)$$

Proof sketch. First, fractional powers are Hölder continuous on PSD cones: $\|L^\alpha - \tilde{L}^\alpha\|_2 \leq C_\alpha \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}$ for $\alpha \in (0, 1]$. Second, use the integral representation $e^{-A} - e^{-B} = \int_0^1 e^{-(1-s)A} (B - A) e^{-sB} ds$ with $A = \tau L^\alpha$, $B = \tau \tilde{L}^\alpha$; contractions $\|e^{-X}\|_2 \leq 1$ for $X \succeq 0$ give $\|e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha}\|_2 \leq \tau \|L^\alpha - \tilde{L}^\alpha\|_2$. Combine the two bounds. \square

Theorem 2 (High-frequency attenuation on spectral subspaces). *Let $P_E = \mathbf{1}_{[\lambda_0, \infty)}(L)$ be the spectral projector of L onto eigenvalues $\geq \lambda_0 > 0$. If $\|\tilde{L} - L\|_2 \leq \varepsilon$ with $\varepsilon \in [0, \lambda_0)$, then*

$$\left\| P_E \left(e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha} \right) P_E \right\|_2 \leq \tau e^{-\tau((\lambda_0 - \varepsilon)_+)^{\alpha}} \|L^\alpha - \tilde{L}^\alpha\|_2 \leq \tau C_\alpha e^{-\tau((\lambda_0 - \varepsilon)_+)^{\alpha}} \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (5)$$

where $(\cdot)_+ := \max(0, x)$. *Proof sketch. On the subspace E , the spectrum of L^α lies in $[\lambda_0^\alpha, \infty)$ while Weyl’s inequality yields a shift at most ε for \tilde{L} . Hence $\|e^{-\tau L^\alpha}|_E\|_2 \leq e^{-\tau \lambda_0^\alpha}$ and $\|e^{-\tau \tilde{L}^\alpha}|_E\|_2 \leq e^{-\tau((\lambda_0 - \varepsilon)_+)^{\alpha}}$. Apply the integral identity as in Theorem 1 restricted to E to obtain the exponential factor. \square*

Implications. The above bounds imply sublinear sensitivity to structural noise for $\alpha < 1$, and nonexpansiveness of heat-type filters ($\|e^{-\tau L^\alpha}\|_2 \leq 1$). Combined with our empirical noise and dropout robustness, they provide a theory-to-practice explanation for the gains of fractional operators in seizure detection. Proofs are in Appendix B.2.7.

3 EXPERIMENT

3.1 DATA PREPROCESSING

Neuroelectrical signals are prone to noise and artifacts that obscure neurodynamic features, particularly in high-frequency bands critical for epilepsy research. To preserve data integrity and physiological significance, we applied a strict preprocessing pipeline using MNE v1.9.0 and EEGLAB v2025.0. As summarized in Table 1, all EEG records were resampled to 512 Hz. A 1 Hz bandwidth notch filter removed 50/60 Hz line noise and harmonics, followed by a 0.1–200 Hz band-pass filter to capture relevant subbands (δ : 0.5–4 Hz, θ : 4–8 Hz, α : 8–12 Hz, β : 13–25 Hz, low γ : 26–80 Hz, HFOs: >80 Hz). Artifact subspace reconstruction (ASR; burst criterion: 10 SD) mitigated high-pass filtering artifacts. Independent component analysis (ICA) using the Picard algorithm identified and removed ocular, muscular, and cardiac artifacts through combined manual inspection and ICLabel automation. Common average referencing (CAR) was applied unless pre-referenced. Training parameters are provided in Table A3.

Table 1: All datasets were standardized to 64 channels based on a common electrode layout. When the original number of channels was less than 64, spatial interpolation was applied for upsampling. For recordings with more than 64 channels, electrodes were first selected according to the standard 10-20 system; if the number of standard channels still exceeded the target maximum, the top 64 channels were retained based on energy density. Then all data were sectioned into 4 seconds long sliding window slices of 0.5 seconds.

Dataset	Confidentiality	Species	# Subj.	Modality	# Ch.	# Samples	Duration	Stride
FMCE (Li et al., 2021)	Public	Human	65	ECoG/SEEG	52-232	34,320	4 s	0.5 s
HUP (Kini et al., 2019)	Public	Human	73	ECoG/SEEG	47-216	182,937	4 s	0.5 s
Helsinki Neonatal EEG (Stevenson et al., 2019)	Public	Human	79	EEG	21	805,097	4 s	0.5 s

3.2 RESULTS AND ANALYSIS

Our proposed EEG-GraphFrFT framework demonstrates superior and robust performance across all evaluation metrics and datasets, achieves the best performance among evaluated baselines for automated seizure detection. The comprehensive analysis below details its effectiveness, robustness, and the contribution of its core components.

Overall Performance EEG-GraphFrFT achieves strong performance across all datasets (Table 2), with accuracy scores of **94.12%** (HUP), **94.54%** (FMCE), and **97.44%** (Helsinki), substantially

outperforming second-best models—FAPEX (93.39% on HUP) and Mamba (91.82% on FMCE, 95.25% on Helsinki). The model demonstrates consistent superiority across all metrics, achieving the highest **F1 scores** and **AUC** values (0.9756 on FMCE vs. 0.9578 for EEG-Conformer and 0.9563 for Mamba), indicating exceptional precision-recall balance and ranking capability. Crucially, EEG-GraphFrFT exhibits outstanding sensitivity (**97.21%** on HUP) and specificity (**98.99%** on Helsinki), demonstrating equal proficiency in seizure detection and non-seizure rejection—critical for clinical deployment to minimize false alarms and missed detections. The performance gap highlights the necessity of domain-specific designs: the iTransformer lags considerably (85.87% vs. 94.12% on HUP), while frequency-based FrTS is surpassed by our integration of *learnable* time-frequency representations with graph-based connectivity analysis, outperforming fixed Fourier transforms.

Table 2: Performance comparison of different models on epileptic EEG classification tasks. The **best** results are highlighted in **red** and the **second-best** results are highlighted in **blue** for each metric within each dataset.

Models Metrics	HUP					FMCE					Helsinki Neonatal EEG				
	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec
EEG-Conformer (Song et al., 2023)	0.9026	0.9012	0.9259	0.9055	0.9102	0.8990	0.8990	0.9578	0.8807	0.9100	0.8843	0.8935	0.9095	0.9177	0.9134
Jepa (Dong et al., 2024)	0.7917	0.7655	0.8225	0.7884	0.8139	0.8691	0.8376	0.8578	0.8626	0.8898	0.8431	0.8628	0.8728	0.8333	0.8571
Mamba (Gui et al., 2024)	0.9107	0.9115	0.9485	0.9347	0.9308	0.9182	0.9183	0.9563	0.9533	0.9246	0.9525	0.9549	0.9839	0.9568	0.9644
iTransformer (Liu et al., 2023)	0.8587	0.8580	0.8717	0.8617	0.9033	0.8578	0.8574	0.9235	0.7896	0.8835	0.8725	0.8903	0.9357	0.9115	0.8639
FrTS (Yi et al., 2023)	0.8151	0.8092	0.8003	0.8367	0.8865	0.8460	0.8455	0.9104	0.8367	0.8865	0.9076	0.9087	0.9054	0.9522	0.9451
FAPEX (Zheng et al., 2025)	0.9339	0.9339	0.9692	0.9364	0.9728	0.9005	0.9003	0.9567	0.9341	0.9588	0.9404	0.9449	0.9816	0.9406	0.9611
EEG-GraphFrFT	0.9412	0.9406	0.9750	0.9721	0.9683	0.9454	0.9412	0.9756	0.9528	0.9746	0.9744	0.9739	0.9831	0.9608	0.9899

Protocol fairness. All baselines are tuned under the same subject-disjoint protocol and early-stopping rule; when authors recommend default configs (e.g., Brain-JEPA), we adopt them verbatim. To ensure the fairness of the experimental setup, a comprehensive comparison with additional baseline models is provided in Table A2.

Exceptional Noise Robustness EEG-GraphFrFT exhibits exceptional noise robustness in additive pink noise tests (Table 3), maintaining **93.76%** accuracy under maximum noise (SD=1)—a mere **0.8%** decrease from clean data versus **6.8%** (Mamba) and **4.0%** (EEG-Conformer) drops. F1 score decreases only **0.2%** compared to **7.1%** (Mamba) and **3.9%** (EEG-Conformer), highlighting superior stability. For complete noise robustness experimental data, refer to Table A1; for the visualization of EEG data with superimposed noise, see FigureA1; and for the robustness results of various models under different noise intensities for each metric, see FigureA2.

This robustness originates from our fractional operators: CHAN1’s learnable FrFT adaptively filters noise via discriminative fractional orders, while CHAN2’s GFRFT processes inherently stable functional connections. The gated fusion mechanism leverages the Hölder stability and exponential high-frequency attenuation properties of fractional graph operators to dynamically integrate noise-robust features from the fractional channels (Section 2.2), in stark contrast to EEG-Conformer’s self-attention that tends to overfit to noise patterns and Mamba’s state-space model that exhibits heightened sensitivity to perturbations due to the lack of such structural stability guarantees.

Table 3: Performance comparison of different models on epileptic EEG classification tasks under noise conditions for the FMCE dataset. For Pink Noise conditions, the **best** and **second-best** percentage changes (indicating noise robustness) are highlighted for each metric.

Models Metrics	Original					Pink Noise (SD=1)				
	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec
EEG-Conformer (Song et al., 2023)	0.8990	0.8990	0.9578	0.8807	0.9100	0.8630 (-4.0%)	0.8636 (-3.9%)	0.8662 (-9.6%)	0.8240 (-6.4%)	0.8910 (-2.1%)
Mamba (Gui et al., 2024)	0.9182	0.9183	0.9568	0.9533	0.9246	0.8559 (-6.8%)	0.8533 (-7.1%)	0.9337 (-2.4%)	0.8400 (-11.9%)	0.8700 (-5.9%)
FAPEX (Zheng et al., 2025)	0.9005	0.9003	0.9567	0.9341	0.9588	0.8899 (-1.2%)	0.8974 (-0.3%)	0.8976 (-6.2%)	0.8841 (-5.4%)	0.9388 (-2.1%)
EEG-GraphFrFT	0.9454	0.9412	0.9756	0.9528	0.9746	0.9376 (-0.8%)	0.9393 (-0.2%)	0.9706 (-0.5%)	0.9378 (-1.6%)	0.9696 (-0.5%)

Ablation Study and Component Analysis Ablation results (Table 4) reveal critical insights: 1) **Learnable α is fundamental:** Optimal configuration (both channels with trained α) achieves 94.5% accuracy, while fixed $\alpha = 1$ (standard FT) causes significant 5.0% drop (89.9%), confirming adaptivity’s importance for capturing non-stationary dynamics; 2) **Graph pathway synergy:** Removing CHAN2 causes 3.3% performance drop (91.4%), proving functional connectivity provides unique complementary information. Combined removal of CHAN2 and CHAN1 adaptivity collapses performance to 69.1%; 3) **Adaptive spectral filtering:** Learned α shows domain-specific optimiza-

tion—CHAN1 averages ≈ 0.5 (complete distribution of alpha values in 4 Layers is provided in Figure A3), while CHAN2 tailors the fractional order parameter to different graph types, with α values of 0.4926 for wPLI and 1.0065 for GC (complete distribution of alpha values in 3 Layers is provided in Figure A4). The concentration of CHAN1’s FNO average α near 0.5 achieves an optimal time–frequency balance for non-stationary EEG transients. In CHAN2, GFRFT’s α is 0.4926 for wPLI (undirected), maintaining a similar balance for synchronization analysis, while its α of 1.0065 for GC (directed) favors a pure spectral approach to model causal flow. This demonstrates the model’s adaptive learning of domain-specific fractional orders. Meanwhile, Results in Table 4 (row 4) reveal a critical divergence: the fixed-order transform ($\alpha=1$) increases AUC by 0.4% yet degrades essential clinical classification metrics. This shows the advantages of our learnable fractional order method in diagnostic indicators, thereby ensuring its application value in clinical practice.

Table 4: Performance of EEG-GraphFrFT model on FMCE dataset under different channel settings and fractional order (α) values. Percentage changes relative to the first row are shown in parentheses.

Channel Settings	Acc	F1	AUC	Sens	Spec
CHAN1: trained α , CHAN2: trained α	0.9454	0.9412	0.9756	0.9528	0.9746
CHAN1: trained α , CHAN2: $\alpha = 1$	0.9378 (-0.8%)	0.9377 (-0.4%)	0.9759 (+0.0%)	0.9368 (-1.7%)	0.9698 (-0.5%)
CHAN1: $\alpha = 1$, CHAN2: $\alpha = 1$	0.8986 (-5.0%)	0.8991 (-4.5%)	0.9795 (+0.4%)	0.8913 (-6.5%)	0.9120 (-6.4%)
CHAN1: $\alpha = 1.5$, CHAN2: $\alpha = 1.5$	0.9213 (-2.5%)	0.9210 (-2.1%)	0.9595 (-1.7%)	0.8902 (-6.6%)	0.9426 (-3.3%)
Remove CHAN2 (only CHAN1: trained α)	0.9138 (-3.3%)	0.9199 (-2.3%)	0.9558 (-2.0%)	0.9062 (-4.9%)	0.9494 (-2.6%)
Remove CHAN2 (only CHAN1: $\alpha = 1$)	0.6910 (-26.9%)	0.6984 (-25.8%)	0.7223 (-26.0%)	0.6371 (-33.1%)	0.7367 (-24.4%)

The exceptional performance and robustness of our EEG-GraphFrFT model are fundamentally attributed to its capacity to extract such highly discriminative features from the input signals. As demonstrated in Figure 2, the functional connectivity networks constructed by our framework successfully capture the stark contrast between interictal and ictal brain states. The wPLI analysis reveals a dramatic emergence of strong, widespread phase-synchronization across the brain network during a seizure—a hallmark of epileptic activity. It is precisely this characteristic *hypersynchronization* that our dual-path architecture, particularly the graph fractional Fourier transform in CHAN2, is designed to detect and characterize in an adaptive manner. The quantitative results from the ablation study (Table 4), which show a significant performance drop when the graph pathway (CHAN2) is removed, provide direct empirical evidence that leveraging this graph-structured information is crucial for the model’s success. Furthermore, analysis of specific channels (e.g., O1 and O2 in occipital lobe, T5 and T6 in posterior temporal lobe, Pz, P3, and P4 in parietal lobe) during seizures reveals enhanced synchronization and pathological activities in these regions, which our model effectively captures through GFRFT, demonstrating its sensitivity to localized epileptic networks.

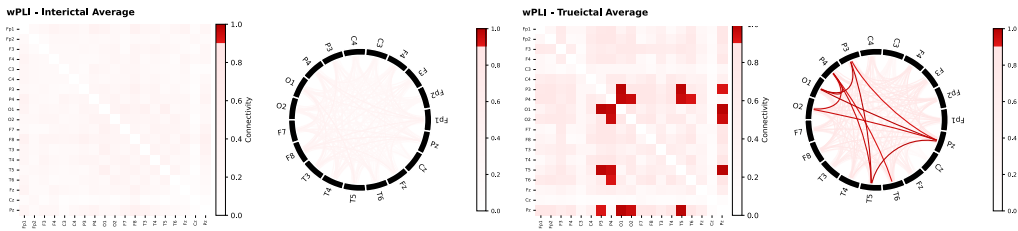


Figure 2: **Group-averaged functional connectivity using wPLI.** Left: Interictal state shows sparse and weak synchronization. Right: Ictal state exhibits dense hypersynchronization. Insets report two graph-spectral summaries *after* the fractional graph operator: the cumulative energy curve $E_{\text{cum}}(k) = \sum_{i=0}^k \|\hat{x}_i\|^2$ over Laplacian eigenmodes (ordered by λ), and the low-frequency energy ratio $R_{\text{low}} = (\sum_{\lambda_i \leq \lambda_0} \|\hat{x}_i\|^2) / (\sum_i \|\hat{x}_i\|^2)$. Ictal curves rise faster and yield higher R_{low} together with lower spectral entropy $H_g = -\sum_i p_i \log p_i$ (with $p_i \propto \|\hat{x}_i\|^2$), indicating that the fractional operator concentrates energy on large-scale coherent modes consistent with seizure dynamics. Additional wPLI and GC connectivity patterns for 10 samples are shown in Figure A5 and Figure A6, respectively. Full group-averaged graphs including GC analysis are provided in Figure A7.

Our designed GFRFT demonstrates exceptional feature discrimination capabilities when processing both directed Granger Causality (GC) and undirected wPLI networks. Taking GC as an example,

this advantage is further validated in Figure 3 (the complete results including wPLI are shown in Figure A8). The model transforms the raw, asymmetric GC input matrix which captures directed causal influences between brain regions—through three adaptive GFRFT layers. The key innovation lies in a mathematical transformation that shifts the analytical perspective from simple connection strength (the input view) to the extraction of significant dynamic network patterns, which are encoded in the magnitude and phase of complex-valued features. The ictal (seizure) state, characterized by a strong pathological causal network, activates the GFRFT filters to produce a sparse output of high-magnitude components (see Figure 3 b-d), effectively concentrating the signature of the epileptic seizure. In contrast, the interictal state yields a lower-energy, more diffuse response (see Figure 3 g-i). This demonstrates that our GFRFT not only preserves the input structure but, more importantly, actively refines and distills the complex directed brain dynamics into a potent and highly separable feature representation, which is ultimately projected by the linear layer into a clear and classifiable signature for seizure detection (see Figure 3 e,j). The increased concentration of energy in low-order graph modes, together with stronger phase consistency, is consistent with clinical observations of hypersynchronization and network densification during ictal periods.

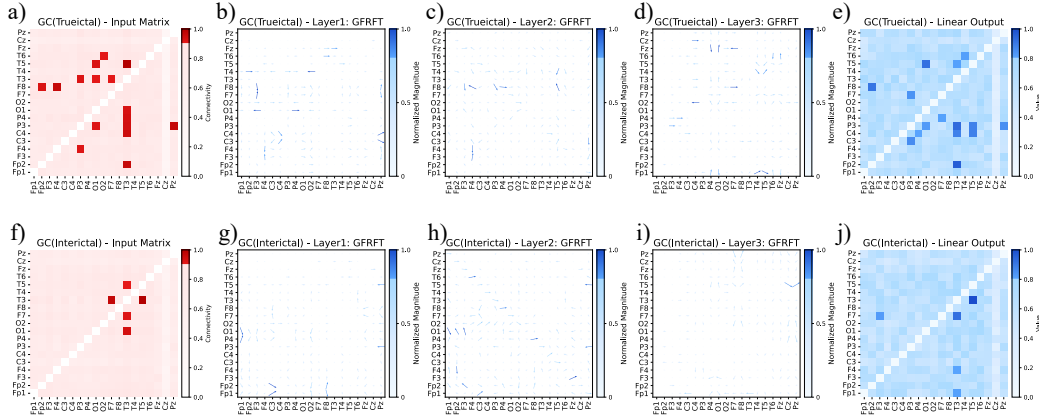


Figure 3: Processing pipeline for Granger-causality (GC) connectivity through three GFRFT layers. Panels (a–e) show an ictal trial and panels (f–j) an interictal trial. Arrow length encodes magnitude and arrow angle encodes phase of complex graph-spectral features. Across Layers 1–3 the ictal dynamics are progressively concentrated into a sparse set of high-magnitude, phase-aligned components, whereas interictal responses remain diffuse. The final linear projection separates classes. We also report layer-wise sparsity $S_k = 1 - \|\mathbf{z}\|_1 / (\sqrt{n} \|\mathbf{z}\|_2)$ and directional phase concentration $C_{\text{dir}} = \left| \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} e^{i\phi_e} \right|$. Ictal trials exhibit higher S_k and C_{dir} , consistent with a directed pathological flow enhanced by fractional orders.

4 CONCLUSION

EEG-GraphFrFT is a dual-path framework that learns fractional operators in the signal and graph domains to tackle nonstationary EEG. A trainable FrFT adapts its order to track transient, time-varying spectra, while graph fractional operators $e^{-i\theta L^\alpha}$ and $e^{-\tau L^\alpha}$ model evolving functional connectivity with controlled spectral selectivity. Across three datasets the model shows strong accuracy under colored noise and channel dropouts; ablations verify that learnable orders and the graph path are necessary. Spectral analyses tie seizures to energy concentration on low-order graph modes and increased phase alignment, yielding interpretable biomarkers consistent with clinical hypersynchronization. Hölder stability for fractional graph filters and scalable Chebyshev/Lanczos approximations for L^α link the observed robustness to theory with $O(K|E|)$ time and $O(K|V|)$ memory.

Limitations: accuracy relies on wPLI/GC and frequency bands; cross-center and patient metrics await larger tests; graph branch impedes real time.

Outlook: operators generalize to broader complex networks and graph analyses; next steps include self-supervised pretrain, adaptive connectivity, uncertainty calibration, on-device inference.

REFERENCES

- Timofey Adamovich, Ilya Zakharov, Anna Tabueva, and Sergey Malykh. The thresholding problem and variability in the eeg graph network parameters. *Scientific Reports*, 12(1):18659, 2022.
- Tuna Alikasıfoğlu, Bünyamin Kartal, and Aykut Koç. Graph fractional fourier transform: A unified theory. *IEEE Transactions on Signal Processing*, 72:3834–3850, 2024.
- John M Bernabei. *Quantitative Methods For Guiding Epilepsy Surgery From Intracranial Eeg*. PhD thesis, University of Pennsylvania, 2021.
- John M Bernabei, Nishant Sinha, T Campbell Arnold, Erin Conrad, Ian Ong, Akash R Pattnaik, Joel M Stein, Russell T Shinohara, Timothy H Lucas, Dani S Bassett, et al. Normative intracranial eeg maps epileptogenic tissues in focal epilepsy. *Brain*, 145(6):1949–1961, 2022.
- Boris C Bernhardt, SeokJun Hong, Andrea Bernasconi, and Neda Bernasconi. Imaging structural and functional brain networks in temporal lobe epilepsy. *Frontiers in Human Neuroscience*, 7: 624, 2013.
- Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5833–5848, 2023. doi: 10.1109/TPAMI.2022.3209686.
- Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, New York, 1997. ISBN 978-0-387-94846-1. doi: 10.1007/978-1-4612-0653-8.
- M. Sh. Birman and M. Z. Solomyak. Operator integration, lipschitz functions, and operator ideals. *Uspekhi Mat. Nauk*, 22(136):27–66, 1967. In Russian; English translation in Russian Mathematical Surveys.
- Susan M Bowyer. Coherence a measure of the brain networks: past and present. *Neuropsychiatric Electrophysiology*, 2(1):1, 2016.
- Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.
- Wuqi Chen, Junjie Ye, Chunna Zhao, and Yaquun Huang. Mffcnn: multi-scale fractional fourier transform convolutional neural network for multivariate time series forecasting. *The Journal of Supercomputing*, 81(2):416, 2025.
- Fan Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9: 1–19, 2005.
- Erin C Conrad, Alfredo Lucas, William KS Ojemann, Carlos A Aguila, Marissa Mojena, Joshua J LaRocque, Akash R Pattnaik, Ryan Gallagher, Adam Greenblatt, Ashley Tranquille, et al. Interictal intracranial eeg asymmetry lateralizes temporal lobe epilepsy. *Brain Communications*, 6(5): fcae284, 2024.
- Mukeshwar Dhamala, Govindan Rangarajan, and Mingzhou Ding. Estimating granger causality from fourier and wavelet transforms of time series data. *Physical Review Letters*, 100(1):018701, 2008.
- Stavros I Dimitriadis, Christos Salis, Ioannis Tarnanas, and David E Linden. Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel data-driven thresholding scheme based on orthogonal minimal spanning trees (omsts). *Frontiers in neuroinformatics*, 11:28, 2017.
- Mohammed Diykh, Firas Sabar Miften, Shahab Abdulla, Ravinesh C Deo, Siuly Siuly, Jonathan H Green, and Atheer Y Oudahb. Texture analysis based graph approach for automatic detection of neonatal seizure from multi-channel eeg signals. *Measurement*, 190:110731, 2022.
- Zijian Dong, Ruilin Li, Yilei Wu, Thuan Tinh Nguyen, Joanna Su Xian Chong, Fang Ji, Nathanael Ren Jie Tong, Christopher Li Hsian Chen, and Juan Helen Zhou. Brain-jepa: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. *Advances in Neural Information Processing Systems*, 37:86048–86073, 2024.

- Yu-Chen Gan, Jian-Yi Chen, and Bing-Zhao Li. The windowed two-dimensional graph fractional fourier transform. *Digital Signal Processing*, 162:105191, 2023.
- Sergi Gomez-Quintana, Alison O’Shea, Andreea Factor, Emanuel Popovici, and Andriy Temko. A method for ai assisted human interpretation of neonatal eeg. *Scientific Reports*, 12(1):10932, 2022.
- A Gramacki and J Gramacki. A deep learning framework for epileptic seizure detection based on neonatal eeg signals. *sci. rep.* 12, 13010 (2022).
- Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- Yiyu Gui, Mingzhi Chen, Yuqi Su, Guibo Luo, and Yuchao Yang. Eegmamba: Bidirectional state space models with mixture of experts for eeg classification. *CoRR*, abs/2407.20254, 2024. URL <https://doi.org/10.48550/arXiv.2407.20254>.
- John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers. *arXiv preprint arXiv:2111.13587*, nov 2021a. URL <https://arxiv.org/abs/2111.13587>.
- John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Efficient token mixing for transformers via adaptive fourier neural operators. In *International Conference on Learning Representations*, 2021b.
- Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Springer, New York, 2008. ISBN 978-0-387-78325-4.
- Robert Hogan, Sean R Mathieson, Aurel Luca, Soraia Ventura, Sean Griffin, Geraldine B Boylan, and John M O’Toole. Scaling convolutional neural networks achieves expert level seizure detection in neonatal eeg. *npj Digital Medicine*, 8(1):17, 2025.
- Radek Janca, Alena Jahodova, Jaroslav Hlinka, Petr Jezdik, Lenka Svobodova, Martin Kudr, Adam Kalina, Petr Marusic, Pavel Krsek, and Premysl Jiruska. Ictal gamma-band interactions localize ictogenic nodes of the epileptic network in focal cortical dysplasia. *Clinical Neurophysiology*, 132(8):1927–1936, 2021.
- Lurong Jiang, Jiawang He, Hangyi Pan, et al. Seizure detection algorithm based on improved functional brain network structure. *Biomedical Signal Processing and Control*, 79:104053, 2023.
- Lohith G Kini, John M Bernabei, Fadi Mikhail, Pnina Hadar, Pooja Shah, Ankit N Khambhati, Kristina Oechsel, Reese Archer, James Boccanfuso, Erin Conrad, Russell T Shinohara, Joel M Stein, Sandhitsu Das, Ammar Kheder, Timothy H Lucas, Kathryn A Davis, Danielle S Bassett, and Brian Litt. Virtual resection predicts surgical outcome for drug-resistant epilepsy. *Brain*, 142(12):3892–3905, 12 2019. doi: 10.1093/brain/awz303. URL <https://doi.org/10.1093/brain/awz303>.
- Adam Li, Chester Huynh, Zachary Fitzgerald, Iahn Cajigas, Damian Brusko, Jonathan Jagid, Angel O. Claudio, Andres M. Kanner, Jennifer Hopp, Stephanie Chen, Jennifer Haagensen, Emily Johnson, William Anderson, Nathan Crone, Sara Inati, Kareem A. Zaghloul, Juan Bulacio, Jorge Gonzalez-Martinez, and Sridevi V. Sarma. Neural fragility as an eeg marker of the seizure onset zone. *Nature Neuroscience*, 24(10):1465–1474, 10 2021. doi: 10.1038/s41593-021-00901-w. URL <https://www.nature.com/articles/s41593-021-00901-w>.
- Fali Li, Yi Liang, Luyan Zhang, Chanlin Yi, Yuanyuan Liao, Yuanling Jiang, Yajing Si, Yangsong Zhang, Dezhong Yao, Liang Yu, and Peng Xu. Transition of brain networks from an interictal to a preictal state preceding a seizure revealed by scalp eeg network analysis. *Cognitive Neurodynamics*, 13(2):175–181, 2019.
- Li Liu, Jiasong Wu, Dengwang Li, Lotfi Senhadji, and Huazhong Shu. Fractional wavelet scattering network and applications. *IEEE Transactions on Biomedical Engineering*, 66(2):553–563, 2019. doi: 10.1109/TBME.2018.2850356.

- Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023. doi: 10.48550/arXiv.2310.06625.
- Lingyan Mao, Gaoxing Zheng, Yang Cai, et al. Frontotemporal phase lag index correlates with seizure severity in patients with temporal lobe epilepsy. *Frontiers in Neurology*, 13:855842, 2022.
- Ana P Millán Vidal, Elisabeth CW van Straaten, Cornelis J Stam, Ida A Nissen, Ida Idema, Piet Van Mieghem, Arjan Hillebrand, et al. Individualized epidemic spreading models predict epilepsy surgery outcomes: A pseudo-prospective study. 2024.
- Madhusundar Nelson, Surendran Rajendran, Osamah Ibrahim Khalaf, and Habib Hamam. Deep-learning-based intelligent neonatal seizure identification using spatial and spectral gnn optimized with the aquila algorithm. *AIMS Mathematics*, 9(7):19645–19669, 2024.
- Lotte Noorlag, Nicole EC van Klink, Katsuhiro Kobayashi, Jean Gotman, Kees PJ Braun, and Maeike Zijlmans. High-frequency oscillations in scalp eeg: A systematic review of methodological choices and clinical findings. *Clinical Neurophysiology*, 137:46–58, 2022.
- Antonio Ortega, Pascal Frossard, Jelena Kovačević, José M. F. Moura, and Pierre Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828, may 2018. doi: 10.1109/JPROC.2018.2820126.
- Haldun M Ozaktas, Orhan Arikan, M Alper Kutay, and Gozde Bozdagt. Digital computation of the fractional fourier transform. *IEEE Transactions on Signal Processing*, 44(9):2141–2150, 1996.
- Sean O’Leary, Anne-Cecile Lesage, Liliana Camarillo-Rodriguez, Oliver Zhou, Diosely Silveira, Jiefei Wang, Sameer A Sheth, Michael S Beauchamp, Zhengjia Wang, John F Magnotti, et al. Integrating data across oscillatory power bands predicts the epileptogenic zone: the frequency range explorer epileptogenic zone (freez) identification algorithm. *bioRxiv*, pages 2024–05, 2024.
- Alison O’Shea, Gordon Lightbody, Geraldine Boylan, and Andriy Temko. Neonatal seizure detection from raw multi-channel eeg using a fully convolutional architecture. *Neural Networks*, 123:12–25, 2020.
- Edson David Pontes, Mauro Pinto, Fábio Lopes, and César Teixeira. Concept-drifts adaptation for machine learning eeg epilepsy seizure prediction. *Scientific Reports*, 14(1):8204, 2024.
- Brittany H Scheid, John M Bernabei, Ankit N Khambhati, Jay Jeschke, Danielle S Bassett, Danielle Becker, Kathryn A Davis, Tim Lucas, Werner Doyle, Edward F Chang, et al. Synchronizability predicts effective responsive neurostimulation for epilepsy prior to treatment. *medRxiv*, pages 2021–02, 2021.
- Brittany H Scheid, John M Bernabei, Ankit N Khambhati, Sofia Mouchtaris, Jay Jeschke, Dani S Bassett, Danielle Becker, Kathryn A Davis, Timothy Lucas, Werner Doyle, et al. Intracranial electroencephalographic biomarker predicts effective responsive neurostimulation for epilepsy prior to treatment. *Epilepsia*, 63(3):652–662, 2022.
- Yuanzhong Shen, Zijian Shen, Yimin Huang, Zhuojin Wu, Yixuan Ma, Feng Hu, and Kai Shu. Multimodal quantitative analysis guides precise preoperative localization of epilepsy. *Journal of Neurology*, 272(9):1–15, 2025.
- Laurent Sheybani, Umesh Vivekananda, Nishant Sinha, Erin Conrad, Brian Litt, Neil Burgess, Daniel Bush, and Matthew C Walker. Dynamic interplay between wake slow waves and epileptiform discharges in the epileptogenic zone. *Neurology*, 105(5):e214001, 2025.
- David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013. doi: 10.1109/MSP.2012.2235192.
- Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg-conformer: Convolutional transformer for eeg decoding and visualization. In *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, volume 31, pages 710–719, 2023. doi: 10.1109/TNSRE.2022.3230250.

- Catherine Stamoulis, Lawrence J Gruber, Donald L Schomer, and Bernard S Chang. High-frequency neuronal network modulations encoded in scalp eeg precede the onset of focal seizures. *Epilepsy Behavior*, 23(4):471–480, 2012.
- N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo. A dataset of neonatal eeg recordings with seizure annotations. *Scientific Data*, 6(1):190039, 3 2019. doi: 10.1038/sdata.2019.39. URL <https://www.nature.com/articles/sdata201939>.
- Lihan Tang and Menglian Zhao. Epileptic seizure detection in neonatal eeg using a multi-band graph neural network model. *Applied Sciences*, 14(21):9712, 2024.
- Karoliina T Tapani, Sampsa Vanhatalo, and Nathan J Stevenson. Time-varying eeg correlations improve automated neonatal seizure detection. *International journal of neural systems*, 29(04):1850030, 2019.
- Roland D Thijs, Rainer Surges, Terence J O’Brien, and Josemir W Sander. Epilepsy in adults. *The Lancet*, 393(10172):689–701, 2019. doi: 10.1016/S0140-6736(18)32596-0.
- John Thomas, Chifaou Abdallah, Kassem Jaber, Mays Khweileh, Olivier Aron, Irena Doležalová, Vadym Gnatkovsky, Daniel Mansilla, Päivi Nevalainen, Raluca Pana, et al. Development of a stereo-eeg based seizure matching system for clinical decision making in epilepsy surgery. *Journal of neural engineering*, 21(5):056025, 2024.
- Kevin Tyner. *Multimodal Investigations for the Identification of Surgically Relevant Brain Areas*. Ph.d. dissertation, University of Nebraska Medical Center, Omaha, Nebraska, May 2024. URL <https://digitalcommons.unmc.edu/etd/819>.
- Martin Vinck, Robert Oostenveld, Marijn van Wingerden, Francesco Battaglia, and Cyriel MA Pennartz. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage*, 55(4):1548–1565, 2011.
- Yuxia Wang, Shasha Yuan, Jin-Xing Liu, Wenrong Hu, Qingwei Jia, and Fangzhou Xu. Combining eeg features and convolutional autoencoder for neonatal seizure detection. *International Journal of Neural Systems*, 34(08):2450040, 2024.
- Lachlan Webb, Minna Kauppila, James A Roberts, Sampsa Vanhatalo, and Nathan J Stevenson. Automated detection of artefacts in neonatal eeg with residual neural networks. *Computer Methods and Programs in Biomedicine*, 208:106194, 2021.
- Deyun Wei and Zhenyang Yan. Generalized sampling of graph signals with prior information based on graph fractional fourier transform. *Signal Processing*, 214:109263, 2024.
- Deyun Wei and Shuangxiao Yun. Hermitian random walk graph fourier transform for directed graphs and its applications. *Digital Signal Processing*, 155:104751, 2024.
- Fang-Jia Yan and Bing-Zhao Li. Spectral graph fractional fourier transform for directed graphs and its application. *Signal Processing*, 210:109099, 2023.
- Fang-Jia Yan, Wen-Biao Gao, and Bing-Zhao Li. Windowed fractional fourier transform on graphs: Fractional translation operator and hausdorff-young inequality. In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 255–259, 2020.
- Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Defu Lian, Ning An, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.
- Haitao Yu, Jing Liu, Lihui Cai, et al. Functional brain networks in healthy subjects under acupuncture stimulation. *Physica A: Statistical Mechanics and its Applications*, 468:566–577, 2017. ISSN 0378-4371. doi: <https://doi.org/10.1016/j.physa.2016.10.068>.
- Yuan Yu, Zhiwei Zeng, Danlin Xie, Renliang Chen, Yongqiang Sha, Shiyang Huang, Wenjie Cai, Wanhua Chen, Wenjun Li, Rongqin Ke, et al. Interneuron origin and molecular diversity in the human fetal brain. *Nature neuroscience*, 24(12):1745–1756, 2021.

- Pengfei Zhang, Tianrui Li, Guoqiang Wang, et al. Multi-source information fusion based on rough set theory. *Information Fusion*, 68:85–117, 2021. doi: <https://doi.org/10.1016/j.inffus.2020.11.004>.
- Yu Zhang and Bing-Zhao Li. The graph fractional fourier transform in hilbert space. *IEEE Transactions on Signal and Information Processing over Networks*, 2025.
- Ruizhe Zheng, Jun Li, Yi Wang, Tian Luo, and Yuguo Yu. Scatterformer: locally-invariant scattering transformer for patient-independent multispectral detection of epileptiform discharges. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 148–158, 2023.
- Ruizhe Zheng, Lingyan Mao, Dingding Han, Tian Luo, Yi Wang, Jing Ding, and Yuguo Yu. Fapex: Fractional amplitude-phase expessor for robust cross-subject seizure prediction. In *Advances in Neural Information Processing Systems*, 2025.
- Weicheng Zhou, Wei Zheng, Youbing Feng, and Xiaolong Li. Lma-ecgnet: A lightweight multi-attention network for neonatal seizure detection using eeg signals. *Electronics*, 13(12):2354, 2024.

A THE USE OF LARGE LANGUAGE MODELS (LLMS)

We used large language models (LLMS) to assist in polishing the language and enhancing the clarity of the manuscript.

B APPENDIX

B.1 MORE RELATED WORK

Recent advances in functional brain network analysis focus on connectivity quantification, graph theory applications, and dynamic network modeling. Construction of functional networks using synchronization indicators like phase lag index (wPLI) (Mao et al., 2022) and coherence analysis (Bowyer, 2016) remains standard, with graph theory providing frameworks for topological analysis (Bullmore and Sporns, 2009). Current research addresses the *threshold problem* in network construction, where threshold selection introduces significant variability in network metrics (Adamovich et al., 2022), prompting data-driven approaches like orthogonal minimum Spanning Tree (OMST) (Dimitriadis et al., 2017). Implicit Neural Representations (INRs) have emerged for modeling dynamic functional connectivity, with (Yu et al., 2017) applying synchronization likelihood in INR frameworks and (Bessadok et al., 2023) developing Graph-Generative Networks (GGNs) for seizure propagation dynamics. Multi-modal integration techniques (Zhang et al., 2021) and challenges in spatial resolution and dynamic modeling persist, as most analyses assume stationarity despite millisecond-scale network reconfigurations being critical for understanding epileptogenesis (Jiang et al., 2023; Bernhardt et al., 2013).

B.2 MATHEMATICAL FOUNDATIONS

B.2.1 WEIGHTED PHASE LAG INDEX (wPLI)

The theoretical motivation for wPLI is that PLI can be unstable under small perturbations, as small phase fluctuations can flip lag into lead, reducing its sensitivity to weak interactions (Vinck et al., 2011). Similarly, Imaginary Coherency (ImC) normalizes by signal amplitude, making it more sensitive to added uncorrelated noise sources. In contrast, wPLI preserves the numerator of PLI while normalizing by $E[|\Im(S_{ij})|]$, ensuring that $0 \leq \text{wPLI} \leq 1$ and increasing robustness against both noise and volume conduction.

wPLI quantifies non-zero lag phase synchronization between electroencephalogram (EEG) signal channels, effectively suppressing volume conduction effects. For a windowed EEG segment $\mathbf{X} \in \mathbb{R}^{T \times C}$ with T time samples and C channels:

1. Compute the Fourier transforms for each channel at frequency f :

$$\mathbf{X}_f = \mathcal{F}(\mathbf{X}) \quad (6)$$

2. Calculate the cross-spectral density between channels i and j :

$$S_{ij} = X_i \cdot X_j^* \quad (7)$$

where $*$ denotes complex conjugation.

3. Extract the imaginary part of the cross-spectrum:

$$\Im(S_{ij}) = \text{imaginary component of } S_{ij} \quad (8)$$

The wPLI between channels i and j at frequency f is defined as:

$$\text{wPLI}_{ij}(f) = \frac{|\mathbb{E}[\Im(S_{ij})]|}{\mathbb{E}[|\Im(S_{ij})|]} \quad (9)$$

where \mathbb{E} represents expectation over trials or time windows. This formulation weights phase leads/lags by the magnitude of the cross-spectral imaginary component, emphasizing consistent non-zero lag interactions while suppressing volume-conducted zero-lag synchronization.

B.2.2 SPECTRAL GRANGER CAUSALITY (SGC)

Spectral GC measures directed frequency-specific information flow between EEG channels (Granger, 1969; Dhamala et al., 2008).

We model the EEG channel pair $x_i(t)$ and $x_j(t)$ with a stable bivariate VAR(p):

$$\begin{bmatrix} x_i(t) \\ x_j(t) \end{bmatrix} = \sum_{k=1}^p A_k \begin{bmatrix} x_i(t-k) \\ x_j(t-k) \end{bmatrix} + \begin{bmatrix} \epsilon_i(t) \\ \epsilon_j(t) \end{bmatrix}, \quad \Sigma = \text{cov} \begin{bmatrix} \epsilon_i(t) \\ \epsilon_j(t) \end{bmatrix}, \quad (10)$$

where $\mathbf{A}_k = \begin{bmatrix} a_{11}^{(k)} & a_{12}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} \end{bmatrix}$ are coefficient matrices, and ϵ are white noise residuals with covariance matrix Σ .

Define the frequency response

$$H(f) = \left(I - \sum_{k=1}^p A_k e^{-i2\pi f k} \right)^{-1} = \begin{bmatrix} H_{ii}(f) & H_{ij}(f) \\ H_{ji}(f) & H_{jj}(f) \end{bmatrix}, \quad (11)$$

so the spectral density is

$$S(f) = H(f) \Sigma H(f)^* = \begin{bmatrix} S_{ii}(f) & S_{ij}(f) \\ S_{ji}(f) & S_{jj}(f) \end{bmatrix}. \quad (12)$$

To obtain pairwise GC $i \rightarrow j$, we orthogonalize the innovations so that Σ is diagonal, which yields $\Sigma = \text{diag}(\Sigma_{ii}, \Sigma_{jj})$. Then the total spectrum of j decomposes as

$$S_{jj}(f) = |H_{ji}(f)|^2 \Sigma_{ii} + |H_{jj}(f)|^2 \Sigma_{jj}, \quad (13)$$

while the restricted (no- i) spectrum is

$$S_{jj|i}(f) = |H_{jj}(f)|^2 \Sigma_{jj}. \quad (14)$$

Thus the frequency-domain Granger causality reduces to the standard log-ratio:

$$F_{i \rightarrow j}(f) = \ln \frac{S_{jj}(f)}{S_{jj|i}(f)} = \ln \left(1 + \frac{|H_{ji}(f)|^2 \Sigma_{ii}}{|H_{jj}(f)|^2 \Sigma_{jj}} \right). \quad (15)$$

which matches the widely used formulation of (Dhamala et al., 2008). This is the expression we implement in our pipeline.

For statistical inference, we estimate significance by permutation tests with trial-label shuffling to generate the null distribution of $F_{i \rightarrow j}(f)$. To control for multiple comparisons across frequencies, we apply the Benjamini-Hochberg false discovery rate (FDR).

The computation involves three stages:

Stage 1: Time-domain Vector Autoregressive (VAR) Modeling For EEG signals $x_i(t)$ and $x_j(t)$, fit a bivariate VAR model:

$$\begin{bmatrix} x_i(t) \\ x_j(t) \end{bmatrix} = \sum_{k=1}^p \mathbf{A}_k \begin{bmatrix} x_i(t-k) \\ x_j(t-k) \end{bmatrix} + \begin{bmatrix} \epsilon_i(t) \\ \epsilon_j(t) \end{bmatrix} \quad (16)$$

where p is the model order (determined by AIC/BIC), \mathbf{A}_k are coefficient matrices, and ϵ are residuals with covariance Σ .

Stage 2: Spectral Transformation Compute the spectral transfer matrix:

$$\mathbf{H}(f) = \left(\mathbf{I} - \sum_{k=1}^p \mathbf{A}_k e^{-i2\pi f k} \right)^{-1} \quad (17)$$

Stage 3: Frequency-domain GC Calculation The spectral GC from i to j at frequency f is:

$$\text{sGC}_{i \rightarrow j}(f) = \ln \left(1 + \frac{|H_{ji}(f)|^2 \Sigma_{ii}}{|H_{jj}(f)|^2 \Sigma_{jj}} \right) \quad (18)$$

where $\Sigma = \text{cov}(\epsilon)$ is the residual covariance matrix.

The final GC metric integrates directional influence across frequency band F :

$$\text{GC}_{i \rightarrow j} = \frac{1}{|F|} \sum_{f \in F} \text{sGC}_{i \rightarrow j}(f) \quad (19)$$

Higher values indicate stronger directed causal influence from channel i to j .

B.2.3 CONNECTIVITY GRAPH CONSTRUCTION

Dynamic functional connectivity networks are constructed using a sliding window approach with a window size of $W = 1024$ samples (2 seconds at 512 Hz) and stride $S = 512$ samples (50% overlap). For each window, the process involves computing the wPLI matrix $\mathbf{W} \in \mathbb{R}^{C \times C}$ to capture undirected phase synchronization, followed by the GC tensor $\mathbf{G} \in \mathbb{R}^{F \times C \times C}$ for F frequency bands (delta: 1–4 Hz, theta: 4–8 Hz, alpha: 8–13 Hz, beta: 13–25 Hz, gamma: 26–80 Hz). Discriminative connections are enhanced through element-wise exponential modulation: $\mathbf{W}_{ij} \leftarrow \mathbf{W}_{ij} \cdot \exp(-\gamma_w \cdot \mathbb{I}(\mathbf{W}_{ij} < \tau_w))$ and $\mathbf{G}_{fij} \leftarrow \mathbf{G}_{fij} \cdot \exp(-\gamma_g \cdot \mathbb{I}(\mathbf{G}_{fij} < \tau_g))$, where γ_w, γ_g are learnable decay parameters and τ_w, τ_g are adaptive thresholds. The resulting network representations, capturing spatiotemporal dynamics of epileptic propagation, are given by $\mathcal{G} = \{\mathbf{W}^{(t)}, \mathbf{G}^{(t)}\}_{t=1}^N$, where N is the number of windows, $\mathbf{W}^{(t)}$ is the wPLI network at window t , and $\mathbf{G}^{(t)}$ is the multi-band GC tensor.

B.2.4 FAST FRFT COMPUTATION

The standard Ozaktas-style (Ozaktas et al., 1996) factorization for the fast FRFT:

$$\text{frft_ozaktas}(x, \alpha) = \text{scale} \cdot \text{post_chirp} \cdot \mathcal{F}^{-1} [\mathcal{F}[\text{pre_chirp} \cdot x] \odot \mathcal{F}[g]] \cdot \text{post_chirp} \quad (20)$$

The pre-chirp and post-chirp terms are defined as:

$$\text{pre_chirp}(t) = \exp(-j\pi t \tan \frac{\phi}{2} \cdot t^2), \quad \text{post_chirp}(u) = \exp(-j\pi u \tan \frac{\phi}{2} \cdot u^2) \quad (21)$$

Here, $\phi = \alpha(\pi/2)$. The convolution kernel is

$$h[k] = \frac{1}{\sqrt{|\sin \varphi|}} \exp\left(j\pi \frac{k^2}{\sin \varphi}\right),$$

implemented as a linear convolution via FFT with appropriate zero-padding.

Special cases are handled explicitly, without invoking the general FRFT kernel:

$$\alpha = 0 : \quad \text{identity transform} \quad (22)$$

$$\alpha = 1 : \quad \text{standard Fourier transform} \quad (23)$$

$$\alpha = -1 : \quad \text{inverse Fourier transform} \quad (24)$$

$$\alpha = 2 : \quad \text{time reversal} \quad (25)$$

The use of $\tan(\varphi/2)$ in the chirp factors avoids divergence near $\sin \varphi \rightarrow 0$, where $\cot \varphi$ is unstable. The kernel scaling $1/\sqrt{|\sin \varphi|}$ ensures unitarity up to a global phase. Near singular angles ($\varphi \approx 0, \pm\pi/2, \pi$), the algorithm snaps to the corresponding explicit branches.

B.2.5 THEORETICAL PROPERTIES OF GFRFT

Theorem 3 (Unitarity Preservation). When F_G is unitary ($F_G^H F_G = I$), its eigenvalues satisfy $|\lambda_i| = 1$. The fractional power becomes:

$$\lambda_i^a = e^{ja\theta_i}, \quad \lambda_i = e^{j\theta_i} \quad (26)$$

Thus:

$$(F_G^a)^H F_G^a = V(\Lambda^a)^H \Lambda^a V^{-1} = V \text{diag}(|\lambda_i^a|^2) V^{-1} = I. \quad (27)$$

For the undirected path, the construction is based on the symmetric Laplacian $L_{sym} = U\Lambda U^T$, where U is orthogonal and $\Lambda \geq 0$. The corresponding fractional operator is then defined spectrally as

$$L_{sym}^\alpha = U\Lambda^\alpha U^T, \quad (28)$$

which is guaranteed to be symmetric positive semidefinite but not necessarily unitary.

We have added the difference by distinguishing:

- The unitary case, where F_G itself is unitary and fractional powers preserve unitarity;
- The Laplacian-based case, where L_{sym} or L_H is diagonalized and fractional powers preserve Hermiticity and positive semidefiniteness, but not unitarity.

Theorem 4 (Index Additivity). Let $F_G^a = \exp(aM)$ and $F_G^b = \exp(bM)$ where $M = \log(F_G)$. Since M commutes with itself:

$$F_G^a F_G^b = \exp(aM) \exp(bM) = \exp((a+b)M) = F_G^{a+b} \quad (29)$$

B.2.6 EXTENDED FRACTIONAL CALCULUS ON GRAPHS

Definition 2 (Fractional Graph Derivative). The fractional derivative operator of order β is defined as:

$$\mathcal{D}^\beta = F_G^{-\beta} \Lambda_\beta F_G^\beta, \quad \Lambda_\beta = \text{diag}((\lambda_k)^\beta), \quad (30)$$

where λ_k are graph frequencies. This generalizes classical fractional calculus to irregular domains.

Proposition 1. Let W_s be the symmetrized adjacency matrix, and let $\Gamma_q = (\Upsilon_q(i, j))$ be a modulation matrix satisfying the conjugate-symmetry condition

$$\Upsilon_q(i, j) = \overline{\Upsilon_q(j, i)} \quad \forall i, j. \quad (31)$$

Define the directed Hermitian Laplacian

$$L_H = D_s - \Gamma_q \odot W_s, \quad D_s(i, i) = \sum_j W_s(i, j). \quad (32)$$

Then:

1. L_H is Hermitian, i.e. $L_H = L_H^*$.
2. L_H is positive semidefinite, i.e. $x^* L_H x \geq 0$ for all $x \in \mathbb{C}^n$.

3. In the undirected case ($w_{ij} = w_{ji}$, $\Gamma_q \equiv 1$), one recovers $L_H = L_{\text{sym}}$, the standard symmetric Laplacian.
4. By Gershgorin's theorem, all eigenvalues of L_H lie on the nonnegative real axis; hence the spectrum is bounded below by zero.

Proof. Since W_s is symmetric with nonnegative entries and $\Upsilon_q(i, j) = \overline{\Upsilon_q(j, i)}$, the off-diagonal entries satisfy

$$L_H(i, j) = -\Gamma_q(i, j) W_s(i, j), \quad L_H(j, i) = -\overline{\Gamma_q(i, j)} W_s(i, j), \quad (33)$$

so $L_H(j, i) = \overline{L_H(i, j)}$. Diagonal entries are real, thus $L_H = L_H^*$ (Yan and li, 2023; Wei and Yun, 2024).

For any $x \in \mathbb{C}^n$,

$$x^* L_H x = x^* D_s x - x^* (\Gamma_q \odot W_s) x = \sum_i \left(\sum_j W_s(i, j) \right) |x(i)|^2 - \sum_{i, j} W_s(i, j) \Upsilon_q(i, j) \overline{x(i)} x(j). \quad (34)$$

Using $W_s(i, j) = W_s(j, i)$ and $\Upsilon_q(j, i) = \overline{\Upsilon_q(i, j)}$, group (i, j) and (j, i) :

$$\begin{aligned} x^* L_H x &= \frac{1}{2} \sum_{i, j} W_s(i, j) \left(|x(i)|^2 + |x(j)|^2 - \Upsilon_q(i, j) \overline{x(i)} x(j) - \Upsilon_q(j, i) \overline{x(j)} x(i) \right) \\ &= \frac{1}{2} \sum_{i, j} W_s(i, j) \left(|x(i)|^2 + |x(j)|^2 - 2 \Re \{ \Upsilon_q(i, j) \overline{x(i)} x(j) \} \right) \\ &= \frac{1}{2} \sum_{i, j} W_s(i, j) \left| x(i) - \Upsilon_q(i, j) x(j) \right|^2 \geq 0. \end{aligned} \quad (35)$$

Hence $L_H \succeq 0$; see also Proposition 2 in (Yan and li, 2023). In the undirected case ($\Gamma_q \equiv 1$), equation 35 reduces to the standard proof for L_{sym} (Chung, 2005). Finally, since L_H is Hermitian, Gershgorin's theorem applies and all eigenvalues are real and nonnegative.

By Gershgorin's circle theorem, each eigenvalue λ of L_H lies in a disk centered at $L_H(i, i) \geq 0$ with radius $\sum_{j \neq i} |L_H(i, j)| = L_H(i, i)$. Thus $\Re(\lambda) \geq 0$, and since L_H is Hermitian, $\lambda \geq 0$. \square

For any diagonalization $L_H = UVU^*$ with eigenvalue matrix $V = \text{diag}(\nu_k)$, we define

$$\log V = \text{diag}(\log |\nu_k| + j \arg(\nu_k)), \quad \nu_k^\alpha = \exp(\alpha \log \nu_k), \quad (36)$$

using the principal branch $\arg(\nu_k) \in (-\pi, \pi]$. Zero eigenvalues are shifted by a small ε before applying log or powers, and tiny negative parts due to numerical roundoff are clamped to zero. Thus $\log V$ and V^α are well defined and numerically stable.

B.2.7 STABILITY OF GRAPH FRACTIONAL FILTERS UNDER STRUCTURAL PERTURBATIONS

Setting and notation. Let $L \in \mathbb{R}^{n \times n}$ be symmetric positive semidefinite (PSD) with spectral decomposition $L = U \Lambda U^\top$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $0 \leq \lambda_i \leq \Lambda_{\max}$. For $\alpha \in (0, 1]$ and $\tau > 0$ we consider the *fractional heat kernel* (graph fractional filter)

$$\mathcal{H}_{\tau, \alpha}(L) = U e^{-\tau \Lambda^\alpha} U^\top = e^{-\tau L^{\alpha p h a}}. \quad (37)$$

For a perturbed Laplacian $\tilde{L} = L + \Delta \succeq 0$ assume $\|\Delta\|_2 \leq \varepsilon$. Throughout, $\|\cdot\|_2$ denotes the spectral (operator) norm. For a Borel set $I \subseteq [0, \infty)$ we write $P_I(L)$ for the spectral projector of L onto eigenvalues in I .

Preliminaries

We collect two standard facts used repeatedly.

Lemma 1 (Operator Hölder continuity of fractional powers). Let $A, B \succeq 0$ and $\alpha \in (0, 1)$. There exists a constant C_α depending only on α (and, if desired, on a priori bounds for $\|A\|_2$ and $\|B\|_2$) such that

$$\|A^\alpha - B^\alpha\|_2 \leq C_\alpha \|A - B\|_2^\alpha. \quad (38)$$

Proof. This is classical; see, e.g., the (Birman and Solomyak, 1967) inequalities via double operator integrals or integral representations of fractional powers. Standard references include (Bhatia, 1997) and (Higham, 2008). \square

Lemma 2 (Difference of exponentials). If $X, Y \succeq 0$ are Hermitian PSD and $\tau > 0$, then

$$\|e^{-\tau X} - e^{-\tau Y}\|_2 \leq \tau \|X - Y\|_2. \quad (39)$$

Proof. Consider $F(\theta) = e^{-\tau(1-\theta)X} e^{-\tau\theta Y}$ for $\theta \in [0, 1]$. Then

$$\frac{d}{d\theta} F(\theta) = \tau e^{-\tau(1-\theta)X} (X - Y) e^{-\tau\theta Y}. \quad (40)$$

Hence

$$e^{-\tau X} - e^{-\tau Y} = -\tau \int_0^1 e^{-\tau(1-\theta)X} (X - Y) e^{-\tau\theta Y} d\theta. \quad (41)$$

Taking norms and using $\|e^{-\tau(1-\theta)X}\|_2 \leq 1$ and $\|e^{-\tau\theta Y}\|_2 \leq 1$ for PSD X, Y yields equation 39. \square

Main stability bound

Theorem 5 (Hölder stability of fractional graph filters). Let $L, \tilde{L} \succeq 0$, $\alpha \in (0, 1]$, and $\tau > 0$. Then

$$\|\mathcal{H}_{\tau, \alpha}(L) - \mathcal{H}_{\tau, \alpha}(\tilde{L})\|_2 \leq \tau C_\alpha \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (42)$$

In particular, for $\alpha \in (0, 1)$ we have the Hölder- α bound $\|e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha}\|_2 \leq \tau C_\alpha \|L - \tilde{L}\|_2^\alpha$, while for $\alpha = 1$ we recover the Lipschitz estimate $\|e^{-\tau L} - e^{-\tau \tilde{L}}\|_2 \leq \tau \|L - \tilde{L}\|_2$.

Proof. For $\alpha = 1$, apply Lemma 2 with $X = L$, $Y = \tilde{L}$. For $\alpha \in (0, 1)$,

$$\|e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha}\|_2 \leq \tau \|L^\alpha - \tilde{L}^\alpha\|_2 \quad \text{by Lemma 2,} \quad (43)$$

then use Lemma 1 to obtain

$$\|L^\alpha - \tilde{L}^\alpha\|_2 \leq C_\alpha \|L - \tilde{L}\|_2^\alpha. \quad (44)$$

\square

Remark 1 (Directed graphs via Hermitianization). If a directed functional network is represented by a Hermitian PSD operator L_H (e.g., via a conjugate-symmetric modulation of a symmetrized adjacency), then the same analysis applies verbatim to L_H and its perturbations \tilde{L}_H , since the functional calculus and the contraction property of $e^{-\tau X}$ require only $X \succeq 0$.

High-frequency attenuation on spectral subspaces

We make the high-frequency attenuation precise on a spectral subspace E of L .

Lemma 3 (Quadratic-form lower bound on E). Let $\tilde{L} = L + \Delta \succeq 0$ with $\|\Delta\|_2 \leq \varepsilon$, and fix $\lambda_0 > 0$. Let $P_E := P_{[\lambda_0, \infty)}(L)$ be the spectral projector of L onto eigenvalues $\geq \lambda_0$. Then for any $v \in E = \text{Ran}(P_E)$ and $\alpha \in (0, 1]$,

$$v^\top \tilde{L} v \geq v^\top L v - \|\Delta\|_2 \|v\|_2^2 \geq (\lambda_0 - \varepsilon) \|v\|_2^2, \quad (45)$$

$$v^\top \tilde{L}^\alpha v \geq (\lambda_0 - \varepsilon)_+^\alpha \|v\|_2^2, \quad (46)$$

whence, for any $s \in [0, 1]$ and $\tau > 0$,

$$\|e^{-\tau s \tilde{L}^\alpha} P_E\|_2 \leq e^{-\tau s (\lambda_0 - \varepsilon)_+^\alpha}, \quad \|P_E e^{-\tau(1-s)L^\alpha}\|_2 \leq e^{-\tau(1-s)\lambda_0^\alpha}. \quad (47)$$

Proof. Inequality equation 45 follows from $v^\top \Delta v \geq -\|\Delta\|_2 \|v\|_2^2$ and $v^\top L v \geq \lambda_0 \|v\|_2^2$ on E . Operator monotonicity of $t \mapsto t^\alpha$ on $[0, \infty)$ gives equation 46. The bounds in equation 47 then follow by the spectral theorem, since $e^{-\tau s \tilde{L}^\alpha} \preceq e^{-\tau s (\lambda_0 - \varepsilon)_+^\alpha} I$ on E while $L^\alpha \succeq \lambda_0^\alpha$ on E . \square

Theorem 6 (Exponential attenuation on high-frequency subspaces). *Let $L, \tilde{L} \succeq 0$ with $\tilde{L} = L + \Delta$, $\|\Delta\|_2 \leq \varepsilon$, and let $P_E = P_{[\lambda_0, \infty)}(L)$ with $\lambda_0 > 0$. For $\alpha \in (0, 1]$ and $\tau > 0$,*

$$\|P_E(e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha})P_E\|_2 \leq \tau e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L^\alpha - \tilde{L}^\alpha\|_2. \quad (48)$$

Consequently, by Lemma 1,

$$\|P_E(e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha})P_E\|_2 \leq \tau C_\alpha e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (49)$$

Proof. Apply the integral identity of Lemma 2 with $X = L^\alpha$ and $Y = \tilde{L}^\alpha$ and insert projectors:

$$P_E(e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha})P_E = -\tau \int_0^1 P_E e^{-\tau(1-s)L^\alpha} (L^\alpha - \tilde{L}^\alpha) e^{-\tau s \tilde{L}^\alpha} P_E ds. \quad (50)$$

Taking norms and using equation 47 gives

$$\|P_E(e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha})P_E\|_2 \leq \tau \|L^\alpha - \tilde{L}^\alpha\|_2 \int_0^1 e^{-\tau(1-s)\lambda_0^\alpha} e^{-\tau s(\lambda_0 - \varepsilon)_+^\alpha} ds \quad (51)$$

$$\leq \tau \|L^\alpha - \tilde{L}^\alpha\|_2 e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha}, \quad (52)$$

where the final step bounds the integral by the maximum of the integrand (attained at $s = 1$ since $(\lambda_0 - \varepsilon)_+^\alpha \leq \lambda_0^\alpha$). \square

Remark 2 (Perturbations confined to E). If $\Delta = P_E \Delta P_E$ (no cross-coupling between E and E^\perp), then E reduces both L and \tilde{L} , and

$$\|P_E(e^{-\tau L^\alpha} - e^{-\tau \tilde{L}^\alpha})P_E\|_2 = \|e^{-\tau(L|_E)^\alpha} - e^{-\tau(L|_E + \Delta|_E)^\alpha}\|_2 \leq \tau C_\alpha \|P_E \Delta P_E\|_2^{\min\{1, \alpha\}} \quad (53)$$

by combining Lemmas 2 and 1. For general Δ , Theorem 6 shows an additional *exponential* attenuation $e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha}$ on E .

Corollary 1 (Global-to-local high-frequency control). With the assumptions of Theorem 6,

$$\|\mathcal{H}_{\tau, \alpha}(L) - \mathcal{H}_{\tau, \alpha}(\tilde{L})\|_2 \geq \|P_E(\mathcal{H}_{\tau, \alpha}(L) - \mathcal{H}_{\tau, \alpha}(\tilde{L}))P_E\|_2 \leq \tau C_\alpha e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (54)$$

Thus perturbations that act predominantly on high-eigenvalue subspaces are strongly suppressed by the factor $e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha}$.

Implications: Robustness of graph FrFT and derived operators

We explain how the preceding results provide precise robustness guarantees for the graph fractional Fourier transform (FrFT) $\mathcal{F}_{\theta, \alpha}(L) = e^{-i\theta L^\alpha}$ and the unified complex-time family $\mathcal{G}_{\tau, \theta, \alpha}(L) = e^{-(\tau + i\theta)L^\alpha}$, $\tau \geq 0$.

Proposition 2 (Structural robustness of FrFT). For $L, \tilde{L} \succeq 0$, $\alpha \in (0, 1]$, $\theta \in \mathbb{R}$,

$$\|\mathcal{F}_{\theta, \alpha}(L) - \mathcal{F}_{\theta, \alpha}(\tilde{L})\|_2 \leq |\theta| C_\alpha \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (55)$$

Corollary 2 (Complex-time family: stability and high-frequency damping). For $\mathcal{G}_{\tau, \theta, \alpha}$,

$$\|\mathcal{G}_{\tau, \theta, \alpha}(L) - \mathcal{G}_{\tau, \theta, \alpha}(\tilde{L})\|_2 \leq \sqrt{\tau^2 + \theta^2} C_\alpha \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (56)$$

If $\tau > 0$ and $P_E = P_{[\lambda_0, \infty)}(L)$, then

$$\|P_E(\mathcal{G}_{\tau, \theta, \alpha}(L) - \mathcal{G}_{\tau, \theta, \alpha}(\tilde{L}))P_E\|_2 \leq \sqrt{\tau^2 + \theta^2} C_\alpha e^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (57)$$

Derived operators. (i) Many graph time–frequency atoms have the form $g(L) = e^{-(\tau + i\theta)L^\alpha}$ and inherit the bounds above. (ii) For the fractional resolvent $\mathcal{R}_{\tau, \alpha}(L) = (I + \tau L^\alpha)^{-1}$,

$$\|\mathcal{R}_{\tau, \alpha}(L) - \mathcal{R}_{\tau, \alpha}(\tilde{L})\|_2 \leq \tau C_\alpha \|L - \tilde{L}\|_2^{\min\{1, \alpha\}}. \quad (58)$$

(iii) Input-noise robustness: $\|\mathcal{F}_{\theta, \alpha}(L)\|_2 = 1$ and $\|\mathcal{G}_{\tau, \theta, \alpha}(L)\|_2 \leq 1$ for $\tau \geq 0$, so these operators do not amplify measurement noise.

B.2.8 DIFFERENTIABILITY OF GFRFT

We work specifically with the Hermitian directed Laplacian $L_H = D_s - \Gamma_q \odot W_s$, under the conjugate-symmetry condition $\Upsilon_q(i, j) = \overline{\Upsilon_q(j, i)}$ and $W_s = W_s^\top \geq 0$. Under these assumptions, $L_H = L_H^*$ and $L_H \succeq 0$.¹

Definition 3 (Spectral definition of L_H^α). Let the eigendecomposition of L_H be

$$L_H = U \Lambda U^H, \quad U^H U = I, \quad \Lambda = \text{diag}(\nu_1, \dots, \nu_n), \quad \nu_k \geq 0. \quad (59)$$

For any $\alpha \in \mathbb{R}$, the fractional power of L_H is defined by spectral calculus:

$$L_H^\alpha = U \Lambda^\alpha U^H, \quad \Lambda^\alpha = \text{diag}(\nu_1^\alpha, \dots, \nu_n^\alpha), \quad (60)$$

with the convention $0^\alpha = 0$ for $\alpha > 0$.

Well-posedness. Since L_H is Hermitian PSD, equation 60 is well-defined for all $\alpha \geq 0$. When derivatives with respect to α are required and zero eigenvalues are present, we use the standard ε -shift $\Lambda_\varepsilon = \Lambda + \varepsilon I$ and set $L_{H,\varepsilon}^\alpha = U \Lambda_\varepsilon^\alpha U^H$, then take $\varepsilon \downarrow 0$ after differentiation. This matches the principal-branch matrix-function calculus used in the paper for logarithm and exponentiation.

Proposition 3 (Basic properties). Let $L_H = U \Lambda U^H$ and L_H^α be defined by equation 60. For $\alpha, \beta \geq 0$:

1. **PSD and Hermitianity:** $L_H^\alpha = (L_H^\alpha)^H \succeq 0$.
2. **Semigroup (index additivity):** $L_H^\alpha L_H^\beta = L_H^{\alpha+\beta}$.
3. **Spectral monotonicity:** if $0 \leq \alpha \leq \beta$ then $L_H^\alpha \preceq L_H^\beta$.
4. **Continuity in α :** $\alpha \mapsto L_{H,\varepsilon}^\alpha$ is continuous and differentiable for every $\varepsilon > 0$, with derivative

$$\frac{\partial}{\partial \alpha} L_{H,\varepsilon}^\alpha = U (\Lambda_\varepsilon^\alpha \odot \log \Lambda_\varepsilon) U^H, \quad (61)$$

where “ \odot ” denotes the Hadamard product and $\log \Lambda_\varepsilon = \text{diag}(\log(\nu_k + \varepsilon))$.

Proof. All items follow from spectral calculus on the real nonnegative spectrum.

(1) Hermitianity is immediate since $U \Lambda^\alpha U^H$ with Λ^α real diagonal is Hermitian; PSD holds because each eigenvalue $\nu_k^\alpha \geq 0$.

(2) Using equation 60, $L_H^\alpha L_H^\beta = U \Lambda^\alpha U^H U \Lambda^\beta U^H = U \Lambda^{\alpha+\beta} U^H = L_H^{\alpha+\beta}$.

(3) For $0 \leq \alpha \leq \beta$, the scalar function $t \mapsto t^\gamma$ on $t \geq 0$ is monotone increasing in γ . Hence $\nu_k^\alpha \leq \nu_k^\beta$ for all k , which implies $L_H^\alpha \preceq L_H^\beta$.

(4) For $\varepsilon > 0$, each scalar map $\alpha \mapsto (\nu_k + \varepsilon)^\alpha$ is C^∞ with derivative $(\nu_k + \varepsilon)^\alpha \log(\nu_k + \varepsilon)$. Stacking on the diagonal and conjugating by U gives equation 61. \square

Action on features and gradient. Given features $X \in \mathbb{C}^{n \times d}$, the directed-graph branch applies L_H^α spectrally:

$$Z_{\text{dir}} = L_H^\alpha X = U \Lambda^\alpha U^H X. \quad (62)$$

Let \mathcal{L} be any real-valued loss on Z_{dir} . Using equation 61 with an ε -shift, the gradient w.r.t. α is

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \text{Re Tr} \left(\left(\frac{\partial \mathcal{L}}{\partial Z_{\text{dir}}} \right)^H U (\Lambda_\varepsilon^\alpha \odot \log \Lambda_\varepsilon) U^H X \right), \quad (63)$$

where the real part ensures a real gradient for complex-valued intermediates. This matches the matrix-function identities already used in the manuscript for $F_G^a = \exp(a \log F_G)$ and $\log(F_G^a) = a \log(F_G)$ under principal-branch assumptions.

¹Hermitian and PSD properties follow from the quadratic form $\frac{1}{2} \sum_{i,j} W_s(i,j) |x_i - \Upsilon_q(i,j) x_j|^2 \geq 0$.

B.2.9 BILINEAR FUSION DETAILS

The general bilinear interaction between feature vectors $\mathbf{s} \in \mathbb{R}^{d_s}$ (CHAN1) and $\mathbf{g} \in \mathbb{R}^{d_g}$ (CHAN2) is defined as $y_k = \mathbf{s}^\top \mathbf{W}_k \mathbf{g} + b_k$ for $k = 1, \dots, d_o$, where $\mathbf{W}_k \in \mathbb{R}^{d_s \times d_g}$ is a weight matrix for each output dimension k , and b_k is a bias term. The output $\mathbf{y} \in \mathbb{R}^{d_o}$ comprises elements y_k . This requires $d_o \cdot d_s \cdot d_g$ parameters, which is costly for high-dimensional features.

To improve efficiency, a low-rank constraint is imposed on \mathbf{W}_k , assuming rank $r \ll \min(d_s, d_g)$. Each \mathbf{W}_k is factorized as $\mathbf{W}_k = \mathbf{V}_{s,k} \mathbf{V}_{g,k}^\top$, where $\mathbf{V}_{s,k} \in \mathbb{R}^{d_s \times r}$ and $\mathbf{V}_{g,k} \in \mathbb{R}^{d_g \times r}$. Substituting yields:

$$y_k = \mathbf{s}^\top \mathbf{V}_{s,k} \mathbf{V}_{g,k}^\top \mathbf{g} + b_k = \sum_{i=1}^r (\mathbf{V}_{s,k}[i]^\top \mathbf{s}) (\mathbf{V}_{g,k}[i]^\top \mathbf{g}) + b_k, \quad (64)$$

or, using the Hadamard product \odot :

$$y_k = \mathbf{1}^\top [(\mathbf{V}_{s,k}^\top \mathbf{s}) \odot (\mathbf{V}_{g,k}^\top \mathbf{g})] + b_k. \quad (65)$$

To further reduce parameters, shared projection matrices $\mathbf{V}_s \in \mathbb{R}^{d_s \times r}$ and $\mathbf{V}_g \in \mathbb{R}^{d_g \times r}$ are used, with an output projection matrix $\mathbf{U} \in \mathbb{R}^{r \times d_o}$. The final output is:

$$\mathbf{y} = \mathbf{U}^\top [(\mathbf{V}_s^\top \mathbf{s}) \odot (\mathbf{V}_g^\top \mathbf{g})] + \mathbf{b}. \quad (66)$$

This reduces parameter complexity from $\mathcal{O}(d_s \cdot d_g \cdot d_o)$ to $\mathcal{O}((d_s + d_g + d_o) \cdot r)$, ensuring computational efficiency and strong performance.

B.3 FULL RESULTS

Table A1: Performance comparison of different models on epileptic EEG classification tasks under varying noise conditions for the FMCE dataset. For Pink Noise conditions (SD=0.5-1.0), the **best** and **second-best** percentage changes, indicating noise robustness, are highlighted for each metric.

Models Metrics	Original					Pink Noise (SD=0.5)					Pink Noise (SD=1)				
	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec
EEG-Conformer (Song et al., 2023)	0.8990	0.8990	0.9578	0.8807	0.9100	0.8792 (-2.2%)	0.8732 (-2.9%)	0.8728 (-8.9%)	0.8667 (-1.6%)	0.8967 (-1.5%)	0.8630 (-4.0%)	0.8636 (-3.9%)	0.8662 (-9.6%)	0.8240 (-6.4%)	0.8910 (-2.1%)
Manba (Gui et al., 2024)	0.9182	0.9183	0.9568	0.9533	0.9246	0.8989 (-2.1%)	0.8890 (-3.2%)	0.9521 (-0.5%)	0.9411 (-1.3%)	0.9100 (-1.6%)	0.8559 (-6.8%)	0.8533 (-7.1%)	0.9337 (-2.4%)	0.8400 (-11.9%)	0.8700 (-5.9%)
FAPEX (Zheng et al., 2025)	0.9005	0.9003	0.9567	0.9341	0.9588	0.9070 (+0.7%)	0.9073 (+0.8%)	0.9733 (+1.7%)	0.8934 (-4.4%)	0.9487 (-1.1%)	0.8899 (-1.2%)	0.8974 (-0.3%)	0.8976 (-6.2%)	0.8841 (-5.4%)	0.9388 (-2.1%)
EEG-GraphFFT	0.9454	0.9412	0.9756	0.9528	0.9746	0.9404 (-0.5%)	0.9462 (+0.5%)	0.9736 (-0.2%)	0.9451 (-0.8%)	0.9686 (-0.6%)	0.9376 (-0.8%)	0.9393 (-0.2%)	0.9706 (-0.5%)	0.9378 (-1.6%)	0.9696 (-0.5%)

Table A2: Performance comparison of more different baselines on epileptic EEG classification tasks.

Models Metrics	HUP					FMCE					Helsinki Neonatal EEG				
	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec	Acc	F1	AUC	Sens	Spec
Scattering Transformer (Zheng et al., 2023)	-	-	-	-	-	-	-	-	-	-	0.9055	0.7990	0.9638	-	-
MBGNN (Tang and Zhao, 2024)	-	-	-	-	-	-	-	-	-	-	-	0.9911	-	-	-
Extra Large (XL) ConNeXt (Hogan et al., 2025)	-	-	-	-	-	-	-	-	-	-	-	0.9280	-	-	-
Deep Residual Neural Network (Webb et al., 2021)	-	-	-	-	-	-	-	-	-	-	0.9550	0.9000	0.9620	-	-
CNN (Gomez-Quintana et al., 2022)	-	-	-	-	-	-	-	-	-	-	-	0.8640	0.8900	0.7800	-
Fd-CAE (Wang et al., 2024)	-	-	-	-	-	-	-	-	-	-	0.9234	0.9577	-	-	-
CNN (Gramacki and Gramacki)	-	-	-	-	-	-	-	-	-	-	0.9700	-	-	-	-
SRFE (O'Leary et al., 2024)	-	-	-	-	-	0.9520	-	0.9100	1.0000	-	-	-	-	-	-
seizure matching system (Thomas et al., 2024)	-	0.6700	0.7600	-	-	-	-	-	-	-	-	-	-	-	-
Random Forest Classifier (Bernabei et al., 2022)	-	-	0.7700	-	-	-	-	-	-	-	-	-	-	-	-
LASSO Logistic Regression (Conrad et al., 2024)	0.7930	-	0.8300	-	-	-	-	-	-	-	-	-	-	-	-
SDA (Tapani et al., 2019)	-	-	-	-	-	-	-	-	-	-	-	0.9880	0.7600	-	-
1D FCN, 2D FCN (O'Shea et al., 2020)	-	-	-	-	-	-	-	-	-	-	-	0.9000	-	-	-
Morse Wavelets based LBP coupled with an ensemble classifier (Diykh et al., 2022)	-	-	-	-	-	-	-	-	-	-	0.9700	-	0.9400	-	-
LMA-EEGNet (Zhou et al., 2024)	-	-	-	-	-	-	-	-	-	-	0.9571	-	0.9862	0.9500	-
SSGNN + AO (Nelson et al., 2024)	-	-	-	-	-	-	-	-	-	-	0.9870	-	0.9879	-	-
Logistic Regression Model (Conrad et al., 2024)	0.7930	-	0.8300	-	-	-	-	-	-	-	-	-	-	-	-
Simple RFO (Leary et al., 2024)	-	-	-	-	-	0.7800	-	0.1000	0.9900	-	-	-	-	-	-
Averaged RFO (Leary et al., 2024)	-	-	-	-	-	0.7900	-	0.1000	0.9900	-	-	-	-	-	-
Manifold RF (Yu et al., 2021)	-	-	-	-	-	-	0.8900	-	-	-	-	-	-	-	-
(Scheid et al., 2022)	-	-	0.8300	-	-	-	-	-	-	-	-	-	-	-	-
Virtual Resection (Bernabei, 2021)	0.8600	-	0.8900	-	-	-	-	-	-	-	-	-	-	-	-
Network-based epilepsy model (Tyner, 2024)	0.6600	-	-	0.8900	0.6500	-	-	-	-	-	-	-	-	-	-
functional connectivity networks (Scheid et al., 2021)	-	-	0.7500	-	-	-	-	-	-	-	-	-	-	-	-
IED-to-LoWS delay-based classifier (Sheyham et al., 2025)	0.6300	0.6700	0.6800	0.6700	0.6000	-	-	-	-	-	-	-	-	-	-
VEP (Shen et al., 2025)	-	-	0.7500	-	-	-	-	-	-	-	-	-	-	-	-
Neural Fragility model (Millán Vidal et al., 2024)	-	-	-	-	-	0.7900	0.7400	0.8200	-	-	-	-	-	-	-
SeizureNet (Janca et al., 2021)	-	-	-	-	-	-	-	-	-	-	0.8700	0.8400	0.9000	-	-

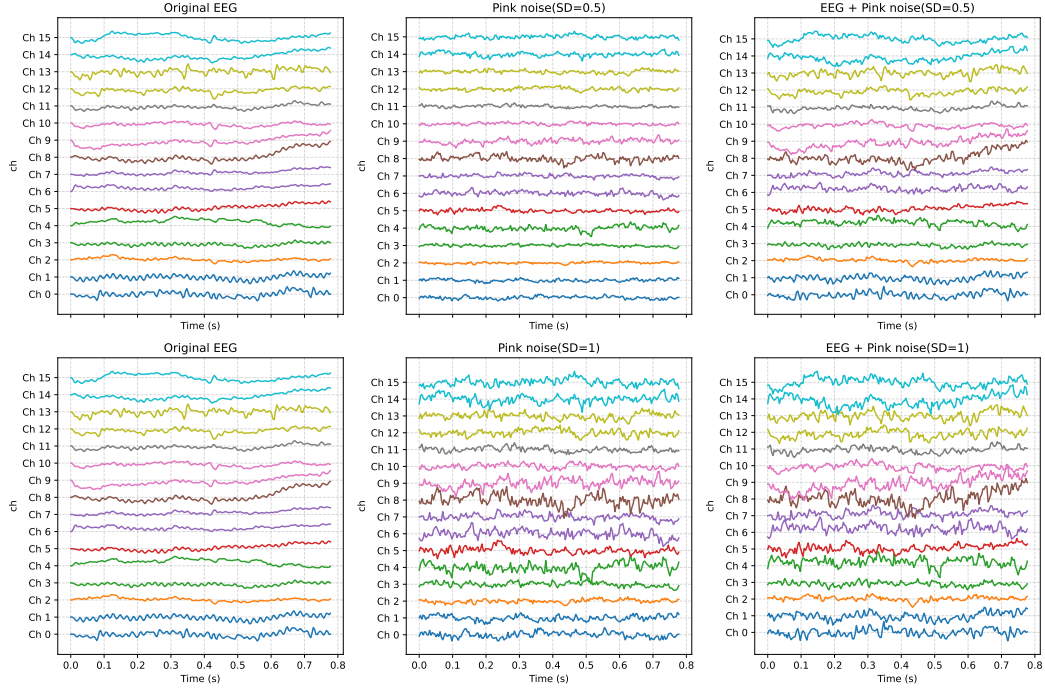


Figure A1: **Visualization comparison of original EEG signals and those superimposed with pink noise at different intensities.** (a) Original clean EEG signals; (b) Pink noise with standard deviation of 0.5; (c) Original EEG signals superimposed with SD=0.5 noise; (d) Pink noise with standard deviation of 1.0; (e) Original EEG signals superimposed with SD=1.0 noise. The figure displays signal waveforms of 16 electrode channels (Ch 0-Ch 15) within the 0-0.8 second time window, with different colors representing different channels. The comparison clearly demonstrates the contamination degree of EEG signals by noise at different intensities, providing visual evidence for evaluating model noise robustness.

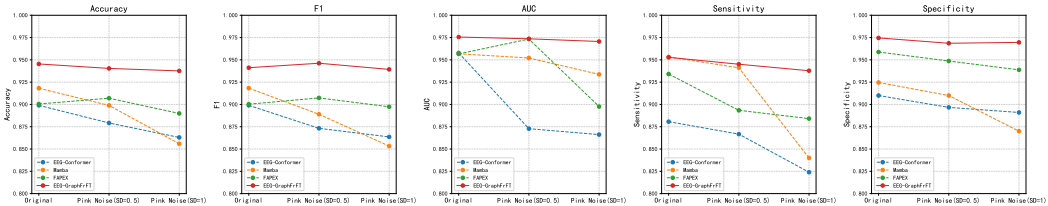


Figure A2: **Performance comparison of various models under different noise conditions.** The figure displays five metrics (Accuracy, F1 score, AUC, Sensitivity, and Specificity) for EEG-Conformer (blue dashed line), Mamba (orange dashed line), FAPEX (green solid line), and the proposed EEG-GraphFrFT (red solid line) across three scenarios: original EEG (no noise), pink noise with standard deviation 0.5, and pink noise with standard deviation 1.0. The results illustrate the robustness of each model against increasing noise intensity, with EEG-GraphFrFT demonstrating superior stability across all metrics.

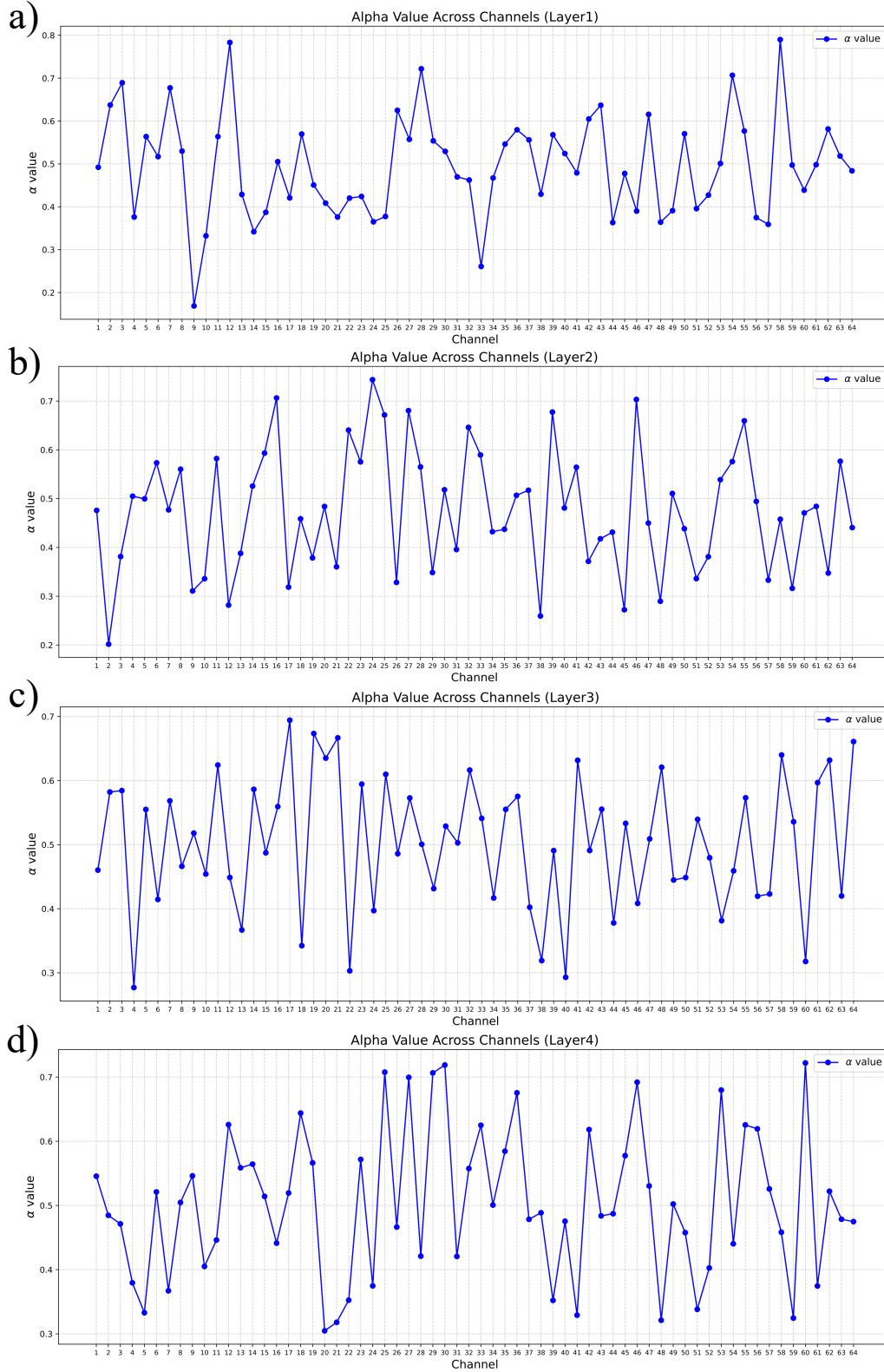


Figure A3: Distribution of learned fractional order parameter α across 64 different channels for four layers (Layer1 to Layer4) in CHAN1. The line plot illustrates the adaptive variation of α values (ranging 0.1-0.8) optimized for each EEG channel, demonstrating the model's capacity to tailor time-frequency resolution to channel-specific characteristics. The non-uniform distribution reflects the adaptive learning of fractional orders for capturing non-stationary patterns in epileptic EEG signals.

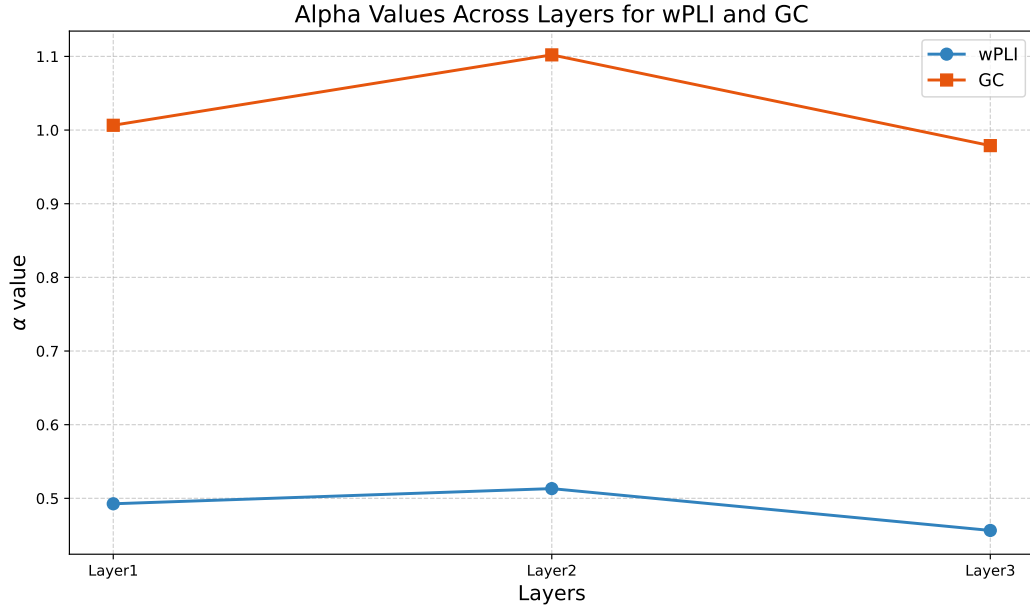


Figure A4: Distribution of the learned fractional order parameter α for the wPLI and GC algorithms across three layers (Layer 1 to Layer 3) in the CHAN2 model. The line plot illustrates the distinct distributions of the α values optimized for the functional networks constructed by the two algorithms. The α values for wPLI remain stable at approximately 0.5 across all layers, whereas the α values for GC are distributed around 1.05. This differential distribution demonstrates the capacity of our trainable GFRFT layer to tailor the time-frequency resolution specifically for the wPLI- and GC-based networks, reflecting the model's adaptive learning for capturing the complex non-stationary dynamics in epileptic EEG signals.

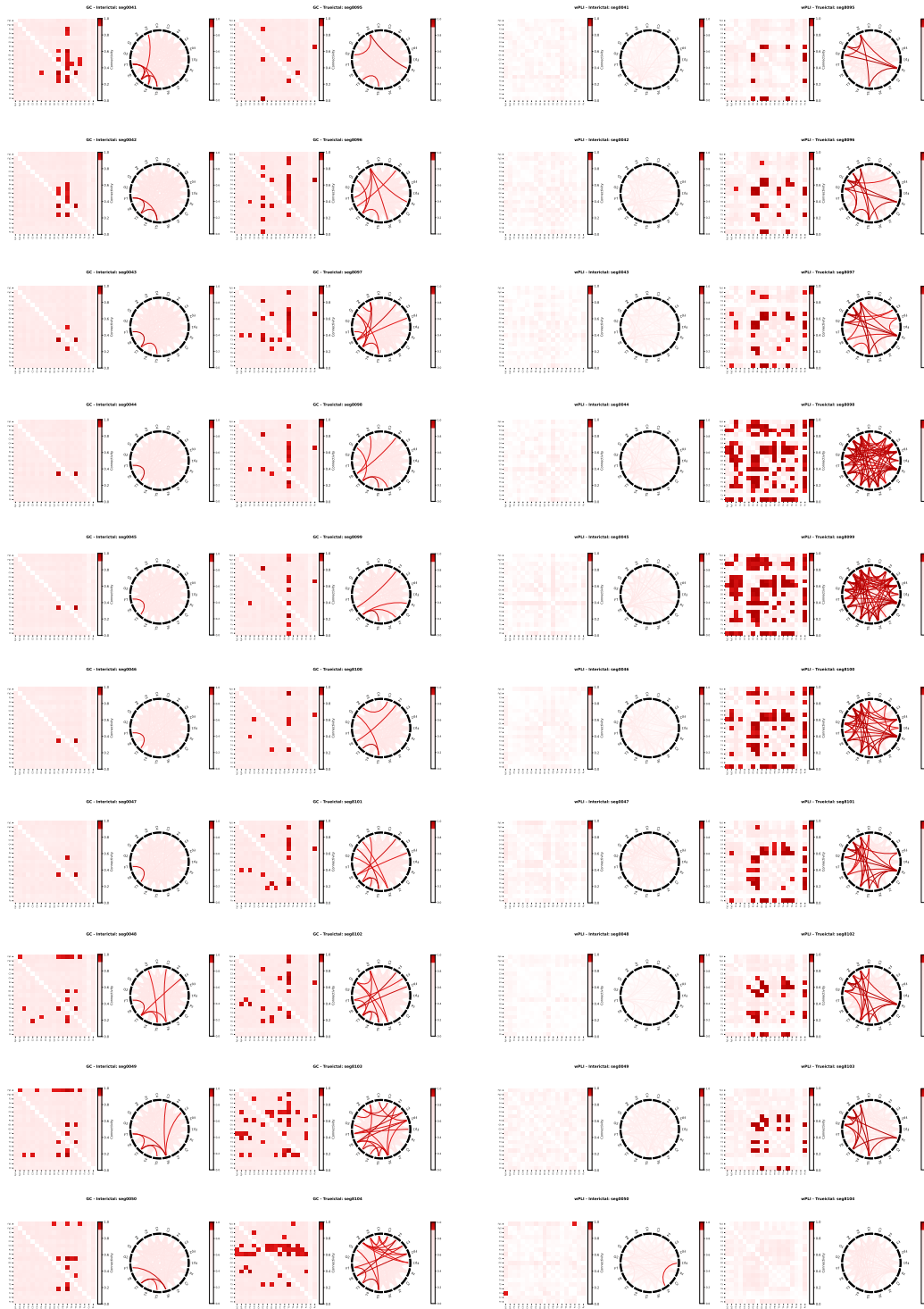


Figure A5: The figure shows functional connectivity matrices and brain network topology diagrams constructed using frequency-domain Granger Causality (GC) for 10 non-seizure samples (left two subfigures) and seizure samples (right two subfigures).

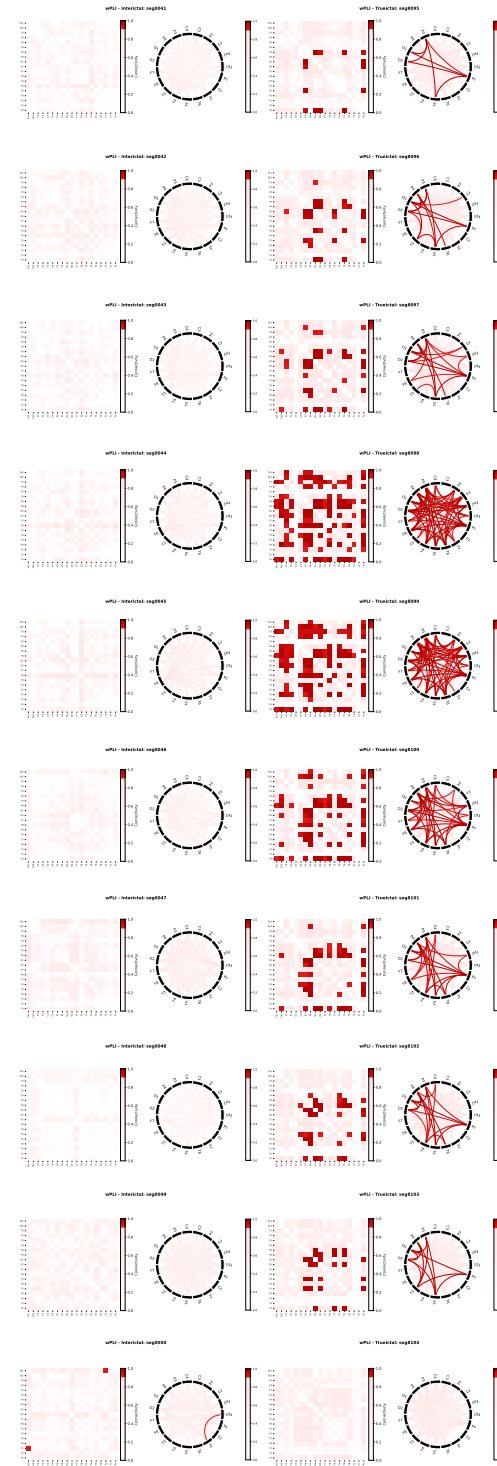


Figure A6: The figure shows functional connectivity matrices and brain network topology diagrams constructed using wPLI for 10 non-seizure samples (left two subfigures) and seizure samples (right two subfigures).

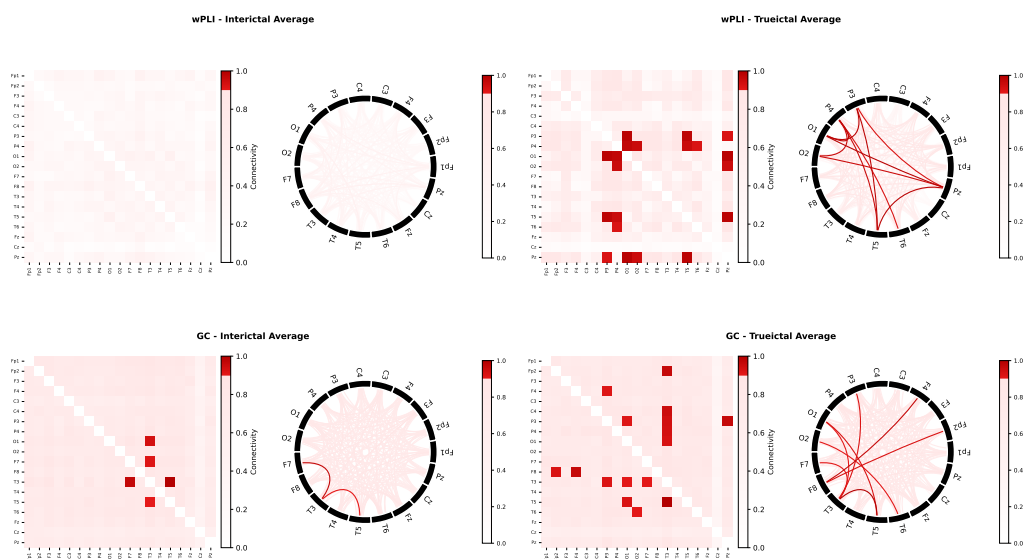


Figure A7: Average Functional connectivity analysis using wPLI and GC. **(Left)** Interictal period: sparse and weak synchronization. **(Right)** Ictal period: significantly enhanced phase-synchronization strength and density, demonstrating epileptic hypersynchronization.

REPRODUCIBILITY STATEMENT

This study was conducted under the following computational conditions to ensure reproducibility:

Table A3: Experimental Configuration Details

Parameter	Value
Hardware	
- GPUs	2 × NVIDIA A100 (80GB)
- CPU	Intel(R) Xeon(R) Platinum 8369B CPU @ 2.90GHz
- RAM	1TB DDR4
Software	
- OS	Ubuntu 22.04 LTS
- CUDA	11.8
- cuDNN	8.9
- Python	3.10
- PyTorch	2.1.0
Training Configuration	
- Batch size	16
- Epochs	100 (with early stopping)
- Random seed	42
- Optimizer	RAdam + Lookahead
• betas	(0.9, 0.999)
• eps	1e-8
• lookahead_k	5
• lookahead_alpha	0.5
- Learning rate	5e-4
- LR scheduler	ReduceLROnPlateau
• factor	0.5
• patience	3
• min_lr	1e-6
- Weight decay	1e-4
- Early stopping	
• patience	10
• min_delta	0.001
- Gradient clipping	Enabled
- Dropout rate	0.3
Model Architecture	
- Input channels	64
- Sequence length	2048 samples
- Embedding size	96