MACHINE LEARNING
Science and Technology

PURPOSE-LED PUBLISHING™

Part of the MACHINE LEARNING Series

PAPER • OPEN ACCESS

Learning to rotate and diffuse: unified fractional operators for nonstationary graph dynamics

To cite this article: Zhiwen Ren *et al* 2026 *Mach. Learn.: Sci. Technol.* **7** 045068

MACHINE
LEARNING
Science and Technology

Check for updates

OPEN ACCESS

RECEIVED
20 January 2026

REVISED
4 July 2026

ACCEPTED FOR PUBLICATION
10 July 2026

PUBLISHED
24 August 2026

PAPER

Learning to rotate and diffuse: unified fractional operators for nonstationary graph dynamics

Zhiwen Ren[1,3], Ruizhe Zheng[2,3], Chendrayan Dineshkumar[2], Yansong Li[1], Yang Du[1], Yuguo Yu[2,*] and Dingding Han[1,*]

[1] College of Future Information Technology, Fudan University, Shanghai 200433, People's Republic of China
[2] MOE Frontiers Center for Brain Science, and State Key Laboratory of Brain Function and Disorders, Research Institute of Intelligent and Complex Systems, Fudan University, Shanghai 200433, People's Republic of China
[3] These authors contributed equally to this work.
[*] Authors to whom any correspondence should be addressed.

E-mail: yuyuguo@fudan.edu.cn and ddhan@fudan.edu.cn

Keywords: AI for health, graph fractional operator, epilepsy, fractional Fourier transform, graph neural network

Abstract

Modeling nonstationary graph signals—where node features exhibit time-varying spectral properties and the underlying network topology evolves dynamically—presents a fundamental challenge in geometric deep learning. Standard approaches, constrained by fixed integer-order operators and isotropic diffusion, often fail to capture transient spectral shifts and preserve high-frequency structures in deep layers. To address this, we introduce a unified mathematical framework based on learnable fractional operators that generalizes time–frequency analysis and graph signal processing into a continuous parameter space. We propose GraphFrFT, a dual-path architecture that employs: (1) a fractional Fourier transform (FrFT) with a trainable order to adaptively rotate the time–frequency plane for capturing transient, chirp-like dynamics; and (2) graph fractional operators that induce tunable anomalous diffusion on dynamic graphs to mitigate over-smoothing. We establish theoretical guarantees for the framework, including well-posedness and Hölder-type stability under spectral perturbations. To validate the limits of this theoretical framework, we evaluate it on epileptic electroencephalography (EEG) seizure detection—a domain representing the 'extreme case' of dual nonstationarity due to its chaotic, fractal, and rapidly reconfiguring network dynamics. On three public datasets (FMCE, HUP, and Helsinki), our model consistently outperforms strong baselines (e.g. Mamba, Conformer) by 2%–8% in accuracy and demonstrates exceptional robustness to noise, effectively supporting the framework's ability to disentangle complex spatiotemporal dynamics. While validated on EEG, the proposed learnable fractional operators are task-agnostic and demonstrate potential for extension to other nonstationary systems such as traffic flow and climate teleconnections.

1. Introduction

Nonstationary graph signals, in which both node features and graph topology evolve over time, remain a central challenge in graph signal processing (GSP) and machine learning. This difficulty is particularly acute in epileptic electroencephalography (EEG): epilepsy affects over 70 million people worldwide [1], and accurate EEG-based diagnosis is clinically critical yet notoriously difficult [2]. EEG encodes rich spatiotemporal dynamics [3] and seizure-related state transitions [4], but expert inspection of long, multi-channel recordings is time-consuming and subjective. Reliable computer-aided diagnosis (CAD) has therefore become a central goal in epilepsy research and clinical practice [5–7]. From a modeling standpoint, seizure EEG exemplifies nonstationary graph signals: temporal statistics drift while functional brain networks reorganize, and similar joint nonstationarities arise in stressed financial markets, disrupted traffic networks, and viral information spread. Standard deep models typically assume approximately stationary features or static graphs, so their performance degrades when both assumptions break down. Recent deep learning advances have substantially improved EEG-based CAD on nonstationary data, yet important limitations remain. FreTS uses a frequency-domain MLP for a global view and energy compaction to capture long-range dependencies in epileptic EEG [8], but its reliance on a fixed Fourier basis restricts adaptation to transient, non-stationary components. Brain JEPA employs a joint embedded prediction architecture with brain gradient localization for transferable representations across tasks [9], potentially overlooking fine-grained, evolving spatial dependencies within a single seizure event. EEG-Conformer integrates convolution and self-attention with class activation mapping for interpretable visualization [10], yet its computational complexity and sensitivity to noise pose challenges for long-sequence EEG analysis. EEG-Mamba and iTransformer leverage bidirectional state-space models for efficient long-sequence processing and multi-task learning [11, 12], but their primary focus on temporal modeling often comes at the expense of explicitly capturing dynamic functional connectivity patterns. Altogether, these observations highlight the need for models that jointly handle temporal nonstationarity and evolving graph structure in a principled way.

To bridge this gap, we advocate a paradigm shift from rigid integer-order models and fixed time–frequency bases to learnable fractional-order operators. Fractional calculus generalizes differentiation and integration to non-integer orders, providing a powerful toolkit for systems with memory, non-locality, and fractal characteristics. Classical time–frequency techniques struggle to accommodate the fractional-order dynamics that characterize epileptic EEG: wavelet transforms depend on fixed mother wavelets, and integer-order differential frames make insufficient use of phase information, which degrades instantaneous frequency estimation and causes time–frequency energy dispersion and pronounced aliasing between distinct components [13–16]. Short-time Fourier transforms (STFT) use fixed-length windows, inducing a rigid trade-off between time and frequency resolution and exhibiting sensitivity to noise and nonstationary interference, which limits performance on abrupt, nonstationary signals such as EEG [17–20]. Although adaptive time–frequency representations can optimize local temporal-spectral adaptation, they still neglect the dynamic spatial couplings among brain regions during seizure propagation [21–23]. In contrast, fractional-order operators, such as the fractional Fourier transform (FrFT), implement a continuous rotation of the time–frequency plane via adjustable order parameters, enabling energy concentration on relevant structures and accurate capture of the long-range correlations and fractal characteristics of epileptic EEG [24–26]. At the same time, classical Fourier analysis cannot fully capture time-varying spectra [8], and graph Fourier transforms extend spectral analysis to static networks but not to the temporal evolution of graph spectra [27]. Advances in fractional calculus and GSP suggest that fractional-order operators are particularly well-suited for such dynamics: adaptive Fourier neural operator (AFNO) performs adaptive token mixing in the Fourier domain in vision [28], and graph FrFTs (GFRFTs) extend these ideas to irregular graph domains [29], building on the broader theory of GSP [30]. Concurrently, the broader emerging field of Fractional Deep Learning provides robust evidence for the efficacy of trainable fractional orders; for instance, recent advancements in neural fractional differential equations (FDEs) integrate variable-order fractional derivatives into neural networks to dynamically capture complex, memory-dependent dynamics [31, 32]. However, 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. In the graph domain, fractional operators naturally address oversmoothing in deep graph neural networks (GNN), where repeated message passing with integer-order Laplacians drives node representations toward indiscriminate convergence. Fractional graph Laplacians induce anomalous diffusion that significantly slows spectral convergence and preserves local high-frequency details [33], while fractional-order dynamics have been analytically shown to exhibit slower algebraic convergence than their integer-order counterparts, thereby retaining richer discriminative information across layers [34]. These properties are especially valuable for epileptic EEG, where oversmoothing would otherwise blur critical hypersynchronization patterns across brain regions.

Among nonstationary graph signals, epileptic EEG provides an ideal 'worst-case' benchmark for testing fractional-operator-based models. Epileptic seizures originate from an imbalance in the neuronal microenvironment, giving rise to EEG with pronounced fractional-order dynamic characteristics, high fractal dimension, and long-term temporal dependence—core signatures that distinguish epileptic signals from healthy EEG. Seizure events feature abrupt, chirp-like frequency drifts and rapid reorganization of functional brain networks, precisely capturing the dual nonstationarity where node features and graph topology evolve jointly. If a model built on generalized fractional operators can robustly disentangle such chaotic seizure dynamics and support reliable CAD, it offers a strong proof-of-concept for a broader class of nonstationary graph signals, including those arising in stressed financial markets, disrupted traffic networks, and contagion-like information spread.

Motivated by these observations, we introduce *EEG-GraphFrFT*, a unified dual-path framework grounded in fractional signal processing that directly addresses dual nonstationarity in seizure EEG. The temporal path implements a generalized fractional neural operator (GFNO) with a trainable order θ, which learns to rotate the time–frequency basis so as to optimally concentrate energy on transient seizure patterns and fractal components, thereby overcoming the rigidity of fixed Fourier bases used in conventional spectral methods and models such as FreTS [8]. The graph path constructs dynamic functional brain networks from multichannel EEG—e.g. via weighted phase lag index (wPLI) and spectral Granger causality (GC)—and applies graph fractional operators L^{α} as adaptive spectral filters on the evolving topology. By tuning the fractional order, these operators interpolate between sub- and super-diffusive regimes, providing a data-driven balance between global integration and the preservation of local hypersynchronization patterns that carry diagnostic value. In combination, the two paths form a collaborative 'fractional time series—dynamic space' representation that jointly models fractional temporal dynamics and dynamic spatial connectivity, achieving a level of coupled temporal-graph adaptivity that traditional time–frequency methods and standard GNNs cannot match. Unlike prior fractional-transform and graph-fractional approaches [28–30], EEG-GraphFrFT endows the fractional orders in both temporal and graph domains with learnable parameters. Furthermore, while recent breakthroughs in differentiable graph learning—such as iterative deep graph learning (IDGL) and neural relational inference (NRI)—have made significant strides in learning graph topologies directly from data rather than relying on predefined heuristics [35–38], we explicitly position our approach as a hybrid framework. It leverages domain-knowledge-based deterministic metrics as structural priors to balance physiological interpretability and computational efficiency. Crucially, to process directed Granger-predictive influence (e.g. GC) without losing unidirectionality, we utilize a Hermitianization technique. This approach relates closely to the established framework of Magnetic Laplacians, which natively preserves asymmetric edge directionality in the complex domain for directed GNNs [39, 40]. These hybrid, direction-aware spatial priors are then adaptively filtered through our learnable fractional operators.

We further provide rigorous theoretical analysis of the proposed graph fractional operators, proving the well-posedness of the resulting filters and establishing Hölder-type stability under structural perturbations of the underlying graph—an essential property for processing noisy, artifact-contaminated EEG recordings. Extensive experiments on three large-scale benchmarks (FMCE, HUP, and Helsinki-neonatal) show that EEG-GraphFrFT significantly outperforms state-of-the-art baselines, including recent Mamba- and Transformer-based models such as EEG-Mamba and iTransformer [11, 12], as well as frequency-domain and attention-based architectures like FreTS, EEG-Conformer, and Brain JEPA [8–10]. Crucially, the model exhibits superior robustness to noise, channel dropouts, and graph estimation errors, empirically validating the theoretical stability conferred by fractional operators in both temporal and graph domains. While our empirical focus is on epileptic EEG, the proposed learnable fractional operators are domain-agnostic and extend naturally to other complex systems—such as traffic sensor networks and multi-asset financial series—where fractal temporal dynamics and evolving interaction graphs jointly shape the observed data.

Related work. The FrFT has emerged as a key operator in deep learning due to its unique properties. Guibas *et al* [41] developed the AFNO, which performs adaptive token mixing in the Fourier domain. Li *et al* [42] combined FrFT with multi-scale wavelet decompositions, while [43] designed multi-scale fractional Fourier convolutional neural networks with fractional orders α. Beyond specific transforms, the broader field of Fractional Deep Learning is rapidly expanding to model complex, memory-dependent dynamics. Recent breakthroughs in FDEs have successfully integrated variable-order fractional derivatives into neural networks [31, 32]. These foundational advancements in Neural FDEs strongly motivate our design of the 'Learnable α', grounding our adaptive fractional operators in a mathematically rigorous framework rather than empirical heuristics.

Concurrently, the GFRFT has extended these concepts to irregular graph domains [29]. Yan *et al* [44] introduced windowed GFRFT for localized filtering, while [45] applied spectral GFRFT to directed graphs with notable success in causal analysis. Processing directed graph topologies remains a fundamental challenge, as standard GNNs often fail to preserve asymmetric edge directionality. In this context, Magnetic Laplacians have been established as a robust mathematical framework for encoding directionality in the complex domain [39, 40]. Our Hermitianization technique for processing directed GC networks shares this core philosophy: by mapping asymmetric relations into a complex Hermitian space, we ensure that the unidirectionality of directed Granger-predictive influence is strictly maintained during fractional spectral filtering.

Recent studies have advanced dynamic graph modeling for EEG signals, particularly in epilepsy applications. Gao and Ribeiro [46] established a theoretical foundation for temporal graph representations, proving that time-then-graph frameworks are strictly more expressive than time-and-graph approaches. Building on this, EvoBrain explicitly captures evolving brain networks through a two-stream Mamba architecture [47], and GRAPHS4MER combines structured state space models (S4) with GNNs for spatiotemporal modeling [48]. Although focused on SAR images, Chen *et al* [49] utilized a shallow broad wavelet scattering network with fractional operators for noise-robust feature extraction, aligning with our approach of integrating spatial-frequency analysis with learnable fractional orders.

In functional brain network analysis, capturing dynamic topology is critical. While traditional approaches heavily rely on handcrafted synchronization indicators like wPLI [50] and coherence analysis [51], they face the threshold problem [52]. To overcome reliance on predefined heuristics, Differentiable Graph Learning has emerged as a powerful paradigm. Prominent frameworks such as NRI [37, 38] and IDGL [35, 36] jointly learn both the graph topology and node representations purely from observational data. While fully end-to-end approaches offer maximum flexibility, they incur massive computational overhead and may generate topologies lacking physiological interpretability. Therefore, our EEG-GraphFrFT adopts a principled hybrid strategy: we utilize deterministic connectivity metrics (wPLI and GC) to establish reliable, domain-grounded structural priors, which are subsequently refined and optimized through our learnable graph fractional operators.

Recent surveys and empirical studies further highlight the momentum of deep learning in epileptic EEG classification. [53] comprehensively reviewed convolutional, recurrent, and hybrid models for automatic seizure detection. Sharma *et al* [54] demonstrated the effectiveness of FrFT-based feature extraction in EEG-based neurological disorder identification, providing direct evidence for extending fractional methods to epilepsy.

Contributions. Analyzing nonstationary graph signals is a core challenge in GSP. These signals exhibit dual nonstationarity: the statistical properties of node features drift over time, while the underlying graph topology evolves simultaneously. Standard deep learning models struggle with this complexity. Time-frequency tools (e.g. STFT, FreTS) use fixed basis functions, limiting adaptability to transient components. Spatially, standard GNNs with integer-order Laplacians suffer from over-smoothing, erasing critical high-frequency details as network depth increases.

We propose a shift to learnable fractional-order operators. Fractional calculus models systems with memory and non-locality. Temporally, the FrFT rotates the time–frequency plane via a fractional order, optimizing energy concentration for non-stationary signals. On graphs, fractional Laplacian powers (L^{α}) induce anomalous diffusion, slowing spectral convergence and alleviating over-smoothing. Making these orders learnable allows the model to adapt its spectral resolution and diffusion dynamics to the data's intrinsic complexity.

We validate our framework on epileptic EEG, an ideal 'worst-case' benchmark exhibiting pronounced dual nonstationarity: abrupt, chirp-like frequency drifts and rapidly reorganizing functional brain networks. Its fractal properties and long-range dependencies challenge integer-order models. Success here provides a rigorous proof-of-concept for broader nonstationary graph signals.

We introduce EEG-GraphFrFT, a unified dual-path framework. Path 1 (Temporal) uses a GFNO with trainable order θ to learn an optimal time–frequency basis for transient patterns. Path 2 (Graph) constructs dynamic functional networks and applies Graph Fractional Operators for adaptive spectral filtering on the evolving topology, balancing global integration with local pattern preservation.

We provide theoretical analysis proving the well-posedness and stability of our fractional graph filters. Extensive experiments on three benchmarks show EEG-GraphFrFT outperforms state-of-the-art models (e.g. Mamba, Transformers), with superior robustness to noise and channel dropouts. While empirically focused on epilepsy, the learnable operators are domain-agnostic, applicable to traffic networks and financial series. Our contributions are:

Learnable fractional operator framework: bridging memory and locality. A trainable GraphFrFT and fractional operator layer with adaptive exponential modulation for nonstationary time–frequency analysis. Fractional operators inherently capture long-range dependencies and fractal structures, overcoming the locality of integer-order calculus.

Graph fractional stability theory: tunable anomalous diffusion. Hölder-type stability bounds for fractional graph filters under bounded structural perturbations; the graph fractional operator induces a tunable anomalous diffusion process, compressing the Laplacian spectrum to suppress noise while preserving mid-high frequency structures, thus theoretically addressing the over-smoothing bottleneck in deep GNNs.

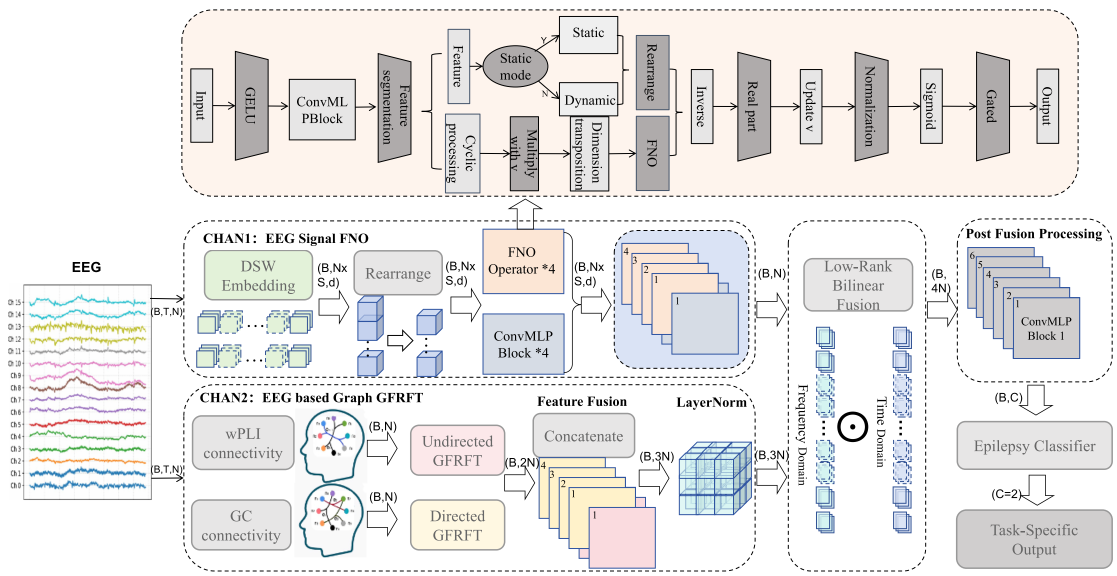

Figure 1. The EEG-GraphFrFT framework employs a dual-path architecture: CHAN1 applies trainable fractional neural operators for adaptive time–frequency analysis, while CHAN2 constructs functional networks (wPLI/GC) processed via graph fractional operators. Features are fused via parameter-efficient low-rank bilinear pooling for seizure detection. Here, B, T, N, S, d, and C denote the number of frequency bands, raw temporal length, number of EEG channels (or graph nodes), temporal segments, hidden embedding dimension, and number of output classes ($C = 2$), respectively.

Hybrid nonstationary graph modeling via operator-level fusion. A dual-path design that fuses signal-space FrFT and graph-space fractional filtering via low-rank fusion. By embedding deterministic domain-knowledge connectivity priors into learnable fractional filters, it offers a mathematically unified paradigm through shared fractional transformations, rather than simple feature concatenation, for consistent handling of dual nonstationarities.

2. Method

The EEG-GraphFrFT framework we propose 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 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 predictive dependence (via spectral GC), and then processes it through the 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.

2.1. GFNO

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 *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 dimensionless FrFT order $\sin\left(\frac{\pi\theta}{2}\right) \neq 0$ and for any $\phi \in L^2(\mathbb{R})$, the associated *fractional pseudo-differential operator* $T_a^\theta : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is defined as:

$$\left(T_a^\theta \phi\right)(t) = \int_{\mathbb{R}} K_{-\theta}(t,\xi)\, a(t,\xi)\, \widehat{\phi}^\theta(\xi)\, \mathrm{d}\xi, \tag{1}$$

where $\widehat{\phi}^\theta(\xi) \equiv \mathcal{F}^\theta[\phi](\xi)$ denotes the FrFT of ϕ of order θ, and $K_{-\theta}(t,\xi)$ represents the kernel associated with the fractional transformation. Let θ denote the dimensionless FrFT order, with the corresponding rotation angle $\varphi = \pi\theta/2$.

A critical aspect of this definition is that the dimensionless FrFT order θ serves as a fractional order parameter, which is not predetermined but learned during training. This learnability allows the operator to dynamically adjust the time–frequency resolution based on the input signal characteristics, overcoming the limitations of fixed-order transforms. By optimizing θ through backpropagation, the GFNO can capture non-stationary features such as abrupt frequency changes in epileptic seizures more effectively than traditional methods.

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 $u : \mathbb{R} \to \mathbb{R}^C$ and $v : \mathbb{R} \to \mathbb{R}^C$ are learned functions capturing temporal and spectral profiles respectively. The NFO $\mathcal{T}_a^\theta : L^2(\mathbb{R})^C \to L^2(\mathbb{R})^C$ of order $\theta \notin \pi\mathbb{Z}$ operates on signal $\varphi \in L^2(\mathbb{R})^C$ as:

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

where $F^\theta[\cdot]$ denotes the FrFT of order θ, and $\odot$ is channel-wise multiplication. The NFO is parameterized by two hypernets $\mathrm{HyperNet}^t$ and $\mathrm{HyperNet}^\xi$ producing $u(t)$ and $v(\xi)$. For each latent feature dimension $q = 1,\ldots,d$, we introduce an unconstrained trainable parameter $\beta_{\theta,q} \in \mathbb{R}$. The normalized FrFT order is parameterized as $\theta_q = \sigma(\beta_{\theta,q}) \in (0,1)$, where $\sigma(\cdot)$ denotes the sigmoid function. The corresponding rotation angle is $\phi_q = \frac{\pi}{2}\theta_q$. The inverse fractional transform is evaluated using order $-\theta_q$. We initialize $\beta_{\theta,q} \sim \mathcal{N}(0, 0.5^2)$, which produces initial fractional orders concentrated around 0.5. The parameter θ_q is associated with the q-th latent feature dimension after temporal embedding, rather than directly with an original EEG electrode channel. 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 $w(\xi) = \mathrm{SoftPlus}(\gamma_0 + \gamma_1|\xi|)$ 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:

$$z(t) = \left[t_{\text{rescaled}}, \{\cos(f_k \cdot w)\}_{k=1}^B, \{\sin(f_k \cdot w)\}_{k=1}^B\right] \tag{3}$$

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

Fast FRFT computation The standard Ozaktas-style [55] factorization for the fast FRFT:

$$\mathrm{frft\_ozaktas}(x,\theta) = \mathrm{scale} \cdot \mathrm{post\_chirp} \cdot \mathcal{F}^{-1}\left[\mathcal{F}[\mathrm{pre\_chirp} \cdot x] \odot \mathcal{F}[g]\right] \cdot \mathrm{post\_chirp} \tag{4}$$

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

$$\mathrm{pre\_}chirp(t) = \exp\left(-j\pi\, tan\frac{\phi}{2} \cdot t^2\right), \quad \mathrm{post\_}chirp(u) = \exp\left(-j\pi\, tan\frac{\phi}{2} \cdot u^2\right) \tag{5}$$

Here, θ denotes the dimensionless FrFT order, whereas $\phi = \pi\theta/2$ denotes the corresponding rotation angle.. The convolution kernel is

$$h[k] = \frac{1}{\sqrt{|\sin\phi|}} \exp\left(j\pi \frac{k^2}{\sin\phi}\right) \tag{6}$$

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

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

$$\theta = 0: \quad \text{identity transform} \tag{7}$$

$$\theta = 1: \quad \text{standard Fourier transform} \tag{8}$$

$$\theta = -1: \quad \text{inverse Fourier transform} \tag{9}$$

$$\theta = 2: \quad \text{time reversal.} \tag{10}$$

The use of $\tan(\phi/2)$ in the chirp factors avoids divergence near $\sin\phi \to 0$, where $\cot\phi$ is unstable. The kernel scaling $1/\sqrt{|\sin\phi|}$ ensures unitarity up to a global phase. Near singular angles ($\varphi \approx k\pi$), the implementation uses explicit limiting branches. The $\varphi = \pm\pi/2$ Fourier-transform cases are also handled explicitly for numerical accuracy and efficiency.

2.2. Graph-based functional connectivity analysis with graph fractional operators

Functional connectivity networks with wPLI and GC We have established a functional connection network that integrates domain expertise with learnable spectral filtering. It captures the dynamic changes among complex regions in the brain activity of epilepsy patients. We use two complementary deterministic metrics calculated across predefined frequency bands. The wPLI measures phase-based synchronization. GC captures the directed Granger-predictive influence in the frequency domain. While our graph fractional operators are fully differentiable and theoretically capable of purely end-to-end topological learning from scratch, we explicitly incorporate these deterministic connectivity metrics as structural priors. This hybrid design is a deliberate trade-off: it significantly accelerates model convergence, mitigates the prohibitive training time overhead associated with learning dense dynamic graphs purely from data, and ensures that the initial topology remains physiologically interpretable. The combination of this dual-function network construction method provides a more comprehensive characterization of brain-network interactions. It integrates undirected functional coupling with directed predictive dependence during epileptic seizures, which is subsequently processed and refined by our learnable fractional operators. *Notation:* we use $(\cdot)^\top$ for transpose and $(\cdot)^{\mathrm{H}}$ for conjugate transpose; for real-symmetric L_{sym} we write $L_{\mathrm{sym}} = U\Lambda U^\top$, while for Hermitian L_{H} we write $L_{\mathrm{H}} = U\Lambda U^{\mathrm{H}}$.

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 [56]. Similarly, imaginary coherency normalizes by signal amplitude, making it more sensitive to added uncorrelated noise sources. wPLI weights the sign consistency of non-zero phase lag by the magnitude of the imaginary cross-spectrum, improving robustness to small perturbations and reducing the contribution of near-zero-lag interactions. wPLI quantifies non-zero lag phase synchronization between 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:

$$\mathbf{X}_f = \mathcal{F}(\mathbf{X}) \tag{11}$$

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

$$S_{ij} = X_i \cdot X_j^* \tag{12}$$

where * denotes complex conjugation.

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

$$\Im(S_{ij}) = \text{imaginary component of } S_{ij} \tag{13}$$

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

$$\mathrm{wPLI}_{ij}(f) = \frac{|\mathbb{E}[\Im(S_{ij})]|}{\mathbb{E}[|\Im(S_{ij})|]} \tag{14}$$

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.

Spectral GC measures directed frequency-specific information flow between EEG channels [57, 58]. We model the EEG channel pair $x_i(t)$ and $x_j(t)$ with a stable bivariate VAR(p):

$$\begin{bmatrix} x_i(t) \\ x_j(t) \end{bmatrix} = \sum_{k=1}^{p} A_k \begin{bmatrix} x_i(t-k) \\ x_j(t-k) \end{bmatrix} + \begin{bmatrix} \epsilon_i(t) \\ \epsilon_j(t) \end{bmatrix}, \qquad \Sigma = \mathrm{cov}\begin{bmatrix} \epsilon_i(t) \\ \epsilon_j(t) \end{bmatrix}, \tag{15}$$

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 $\boldsymbol{\Sigma}$.

Define the frequency response

$$H(f) = \left(I - \sum_{k=1}^{p} A_k \mathrm{e}^{-\mathrm{i}2\pi fk}\right)^{-1} = \begin{bmatrix} H_{ii}(f) & H_{ij}(f) \\ H_{ji}(f) & H_{jj}(f) \end{bmatrix}, \tag{16}$$

so the spectral density is

$$S(f) = H(f)\,\Sigma H(f)^* = \begin{bmatrix} S_{ii}(f) & S_{ij}(f) \\ S_{ji}(f) & S_{jj}(f) \end{bmatrix}. \tag{17}$$

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

$$S_{jj}(f) = |H_{ji}(f)|^2\,\Sigma_{ii} + |H_{jj}(f)|^2\,\Sigma_{jj}, \tag{18}$$

while the restricted (no-i) spectrum is

$$S_{jj|i}(f) = |H_{jj}(f)|^2\,\Sigma_{jj}. \tag{19}$$

Thus the frequency-domain GC reduces to the standard log-ratio:

$$F_{i\to j}(f) = \ln \frac{S_{jj}(f)}{S_{jj|i}(f)} = \ln\left(1 + \frac{|H_{ji}(f)|^2\,\Sigma_{ii}}{|H_{jj}(f)|^2\,\Sigma_{jj}}\right). \tag{20}$$

which matches the widely used formulation of [58]. This is the expression we implement in our pipeline.

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:

$$\begin{bmatrix} x_i(t) \\ x_j(t) \end{bmatrix} = \sum_{k=1}^{p} \mathbf{A}_k \begin{bmatrix} x_i(t-k) \\ x_j(t-k) \end{bmatrix} + \begin{bmatrix} \epsilon_i(t) \\ \epsilon_j(t) \end{bmatrix} \tag{21}$$

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

Stage 2: Spectral transformation Compute the spectral transfer matrix:

$$\mathbf{H}(f) = \left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_k \mathrm{e}^{-\mathrm{i}2\pi fk}\right)^{-1} \tag{22}$$

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

$$\mathrm{sGC}_{i\to j}(f) = \ln\left(1 + \frac{|H_{ji}(f)|^2 \Sigma_{ii}}{|H_{jj}(f)|^2 \Sigma_{jj}}\right) \tag{23}$$

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

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

$$\mathrm{GC}_{i\to j} = \frac{1}{|F|} \sum_{f\in F} \mathrm{sGC}_{i\to j}(f) \tag{24}$$

Higher values indicate stronger directed causal influence from channel ii to jj.

Dynamic functional connectivity networks are constructed using a sliding window approach. It is important to note a fundamental trade-off in our methodology: while we emphasize the severe non-stationarity and abrupt frequency drifts of epileptic EEG, estimating reliable functional connectivity (wPLI and GC) inherently requires a temporal window of sufficient length to compute stable cross-spectral densities and autoregressive coefficients. This presents a 'stationary window paradox'—a window too short yields noisy, statistically unreliable connectivity estimates, whereas a window too long oversmooths transient seizure dynamics and violates local stationarity. To balance temporal resolution with estimation reliability, we systematically evaluated varying window sizes. As detailed in our ablation studies (section 3.2), dynamic functional connectivity networks are constructed from overlapping temporal windows. Each window contains $W = 2048$ samples, corresponding to 4~s at a sampling rate of 512~Hz. Consecutive windows are extracted using a stride of $R = 256$ samples, corresponding to 0.5~s and an overlap of 1792 samples (3.5~s). For a recording containing T_{rec} samples, the number of extracted windows is $S = \left\lfloor \frac{T_{\mathrm{rec}} - W}{R} \right\rfloor + 1$.For each temporal segment, the process involves computing the wPLI matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ to capture undirected phase synchronization, followed by the GC tensor $\mathbf{G} \in \mathbb{R}^{B\times N\times N}$ for B frequency bands (delta: 1–4 Hz, theta: 4–8 Hz, alpha: 8–13 Hz, beta: 13–25 Hz, gamma: 26–80 Hz). To enable end-to-end threshold optimization, we replace the non-differentiable hard indicator with a temperature-controlled sigmoid approximation. For the wPLI and GC branches, respectively, the soft edge masks are defined as $M_{ij}^{(w)} = \sigma\left(\kappa(\tau_w - W_{ij})\right)$, $M_{fij}^{(g)} = \sigma\left(\kappa(\tau_g - G_{fij})\right)$, where $\sigma(z) = 1/(1+\exp(-z))$ and $\kappa = 20$ is a fixed temperature coefficient. Connections below the learned threshold produce mask values close to one, whereas connections above the threshold produce mask values close to zero.

The modulated connectivity matrices are computed as $\widetilde{W}_{ij} = W_{ij}\exp\left(-\gamma_w M_{ij}^{(w)}\right)$, $\widetilde{G}_{fij} = G_{fij}\exp\left(-\gamma_g M_{fij}^{(g)}\right)$. The parameter pairs (τ_w, γ_w) and (τ_g, γ_g) are independently instantiated for the wPLI and GC branches and jointly optimized with the remaining model parameters. The thresholds are differentiable because $\frac{\partial M}{\partial \tau} = \kappa M(1-M)$, which provides valid gradients around the threshold transition. The modulated matrices $\widetilde{W}$ and $\widetilde{G}$ are subsequently used to construct the corresponding graph spectral operators. After differentiable graph modulation, the resulting sequence of time-varying functional-connectivity representations is defined as $\mathcal{G} = \left\{\left(\widetilde{\mathbf{W}}^{(t)}, \widetilde{\mathbf{G}}^{(t)}\right)\right\}_{t=1}^{S}$, where S denotes the number of temporal windows, $\widetilde{\mathbf{W}}^{(t)} \in \mathbb{R}^{N\times N}$, is the modulated wPLI connectivity matrix at window t, and$\widetilde{\mathbf{G}}^{(t)} \in \mathbb{R}^{B\times N\times N}$ is the modulated multi-band GC tensor. In the present implementation, $B = 5$ corresponds to the δ, θ, α, β, and γ frequency bands. These modulated connectivity representations are subsequently passed to the graph fractional feature-extraction layers.

Graph fractional operators via functional calculus We operate *exclusively* on Hermitian (PSD) Laplacians to ensure real spectra and stable gradients. Let $\mathbf{L}_*$ denote the graph operator used for the spectral calculus: $\mathbf{L}_* = \mathbf{L}_{\mathrm{sym}}$ for undirected wPLI graphs, and $\mathbf{L}_* = \mathbf{L}_{\mathrm{H}}$ for GC graphs (Hermitianized; appendix A.1.1). With $\mathbf{L}_* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and $\alpha \in (0,1]$, we define the fractional family

$$\mathbf{U}_{\theta,\alpha}(\mathbf{L}_*) = \mathrm{e}^{-\mathrm{i}\theta\mathbf{L}_*^{\alpha}}, \qquad \mathbf{H}_{\tau,\alpha}(\mathbf{L}_*) = \mathrm{e}^{-\tau\mathbf{L}_*^{\alpha}}, \tag{25}$$

and use their real/imaginary parts as learnable graph-spectral filters. Gradients are analytic: $\partial_\alpha \mathbf{\Lambda}^\alpha = \mathbf{\Lambda}^\alpha \circ \log\mathbf{\Lambda}$ and $\partial_\theta \mathrm{e}^{-\mathrm{i}\theta\mathbf{\Lambda}^\alpha} = -\mathrm{i}\,\mathbf{\Lambda}^\alpha \mathrm{e}^{-\mathrm{i}\theta\mathbf{\Lambda}^\alpha}$. In the reported implementation, the GFT matrix is obtained by exact eigendecomposition of each graph-shift matrix, and the fractional power of the GFT matrix is also evaluated by eigendecomposition. Therefore, the dense spectral computation requires O(N^3) time and O(N^2) memory per graph. In the optimized data pipeline, the GFT matrices are precomputed and reused, reducing repeated forward-pass overhead without changing their asymptotic construction cost. For GC-derived graphs we build $\mathbf{L}_{\mathrm{H}} = \mathbf{D}_s - \mathbf{\Gamma} \odot \mathbf{W}_s$ (unit-modulus phase Γ), which is Hermitian PSD. All fractional operators are applied to $\mathbf{L}_{\mathrm{H}}$ to guarantee a real spectrum and numerical stability. The exponential families in equation (25) are theoretical auxiliary operators and are not used in the implemented network forward pass.

Extended fractional graph calculus The fractional graph derivative of order β is defined as $\mathcal{D}^\beta = F_{\mathrm{G}}^{-\beta}\Lambda_\beta F_{\mathrm{G}}^{\beta}$, where $\Lambda_\beta = \mathrm{diag}((\lambda_k)^\beta)$ and λ_k are graph frequencies, extending classical fractional calculus to irregular domains. For a symmetrized adjacency matrix 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 $L_{\mathrm{H}} = D_s - \Gamma_q \odot W_s$, where $D_s = \mathrm{diag}(W_s\mathbf{1})$. Then $L_{\mathrm{H}} = L_{\mathrm{H}}^*$ and $x^H L_{\mathrm{H}} x = \frac{1}{2}\sum_{i,j} W_{s,ij}|x_i - \Upsilon_q(i,j)x_j|^2 \geqslant 0$. Then, L_{H} is Hermitian ($L_{\mathrm{H}} = L_{\mathrm{H}}^*$), positive semidefinite (PSD) ($x^* L_{\mathrm{H}} x \geqslant 0$ 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 A.1.2). For any diagonalization $L_{\mathrm{H}} = UVU^*$ with $V = \mathrm{diag}(\nu_k)$, define $\log V = \mathrm{diag}(\log|\nu_k| + j\arg(\nu_k))$ 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. Specifically, we use $\varepsilon = 10^{-8}$ in all experiments, which is sufficiently small to have negligible impact on the spectrum.

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 $\mathbf{L}_{\mathrm{sym}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the degree matrix and $\mathbf{A}$ is the symmetric adjacency matrix. Eigendecomposition preserves the spectral gap: $\mathbf{L}_{\mathrm{sym}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathsf{T}} = \sum_{k=0}^{N-1} \lambda_k \mathbf{u}_k \mathbf{u}_k^{\mathsf{T}}$, with eigenvalues ordered $0 = \lambda_0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_{N-1} = \lambda_{\max}$. The learnable GFRFT is applied as $\mathbf{Z}_{\mathrm{undir}} = \mathbf{U}\boldsymbol{\Lambda}^\alpha \mathbf{U}^{\mathsf{T}}\mathbf{X} = \sum_{k=0}^{N-1} \lambda_k^\alpha (\mathbf{u}_k^{\mathsf{T}}\mathbf{X})\mathbf{u}_k$, where α is initialized at 0.5 and optimized via $\nabla_\alpha \mathcal{L} = \mathrm{Tr}\left(\frac{\partial \mathcal{L}}{\partial \mathbf{Z}_{\mathrm{undir}}}^{\mathsf{T}} \mathbf{U}(\boldsymbol{\Lambda}^\alpha \circ \log \boldsymbol{\Lambda})\mathbf{U}^{\mathsf{T}}\mathbf{X}\right)$.

Directed graphs (GC) via Hermitianization. For GC-derived directed connectivity we use the Hermitian Laplacian $L_{\mathrm{H}} = D_s - \Gamma \odot W_s$ (unit-modulus phase Γ). 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 supplied graph-shift matrix is processed independently using exact eigendecomposition; Hermitian inputs use a Hermitian eigensolver, whereas general inputs use a general eigensolver); 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:

$$\mathbf{Z} = \mathbf{U}\,\boldsymbol{\Lambda}^\alpha\,\mathbf{U}^H\mathbf{X}, \quad \mathbf{L}_* = \begin{cases} \mathbf{L}_{\mathrm{sym}} & \text{(wPLI, undirected)}, \\ \mathbf{L}_{\mathrm{H}} & \text{(GC, directed; Hermitianized)} \end{cases}. \tag{26}$$

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 GC 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: $\mathbf{h}_{\mathrm{gc}} = [\mathbf{h}_\delta \oplus \mathbf{h}_\theta \oplus \mathbf{h}_\alpha \oplus \mathbf{h}_\beta \oplus \mathbf{h}_\gamma]$. We share α across adjacent bands with a total-variation penalty on $\{\alpha_b\}$ to reduce overfitting.

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

$$\mathbf{y} = \mathbf{U}^\top \left[\left(\mathbf{V}_s^\top \mathbf{s}\right) \odot \left(\mathbf{V}_g^\top \mathbf{g}\right)\right] + \mathbf{b}, \tag{27}$$

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 A.1.5.

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 PSD Laplacians with spectra in $[0, \Lambda_{\max}]$, 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: $L^\alpha = U\Lambda^\alpha U^\top$ for $L = U\Lambda U^\top$. All proofs are deferred to appendix A.1.3.

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*

$$\left\|\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau \tilde{L}^\alpha}\right\|_2 \leqslant \tau\, C_\alpha\, \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{28}$$

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

Theorem 2 (High-frequency attenuation on spectral subspaces). *Let* $P_\mathrm{E} = \mathbf{1}_{[\lambda_0,\infty)}(L)$ *be the spectral projector of L onto eigenvalues* $\geqslant\lambda_0 > 0$. *Assume that E=Ran(*P_E*) is a reducing subspace for both L and* $\widetilde{L}$, *equivalently* $P_\mathrm{E}\widetilde{L}=\widetilde{L}P_\mathrm{E}$. *If* $\|\tilde{L} - L\|_2 \leqslant \varepsilon$ *with* $\varepsilon \in [0,\lambda_0)$, *then*

$$\left\|P_E\left(\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau\tilde{L}^\alpha}\right)P_E\right\|_2 \leqslant \tau\,\mathrm{e}^{-\tau\left((\lambda_0-\varepsilon)_+\right)^\alpha}\|L^\alpha - \tilde{L}^\alpha\|_2 \leqslant \tau\,C_\alpha\,\mathrm{e}^{-\tau\left((\lambda_0-\varepsilon)_+\right)^\alpha}\|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{29}$$

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* $\|\mathrm{e}^{-\tau L^\alpha}|_E\|_2 \leqslant \mathrm{e}^{-\tau\lambda_0^\alpha}$ *and* $\|\mathrm{e}^{-\tau\tilde{L}^\alpha}|_E\|_2 \leqslant \mathrm{e}^{-\tau((\lambda_0-\varepsilon)_+)^\alpha}$. *Apply the integral identity as in theorem 1 restricted to E to obtain the exponential factor.*

Remark 1. Theorems 1 and 2 are stated for $\alpha \in (0,1]$, where Hölder-type sensitivity holds. For $\alpha > 1$, the scalar spectral map $f(\lambda) = \lambda^\alpha$ is continuously differentiable with bounded derivative on $[0,\Lambda_{\max}]$. Hence the same perturbation argument yields a Lipschitz-type bound on bounded PSD spectra. Therefore, stability is preserved, although the sublinear Hölder sensitivity is replaced by linear sensitivity.

Corollary 1 (Practical stability under approximation). *Let* $L,\tilde{L} \succeq 0$ *be PSD Laplacians with spectra in* $[0,\Lambda_{\max}]$, *let* $\alpha \in (0,1]$, $\tau > 0$, *and define*

$$H_{\tau,\alpha}(L) = \mathrm{e}^{-\tau L^\alpha}.$$

Let $\widehat{H}_{\tau,\alpha,K}(L)$ *denote the order-K Chebyshev/Lanczos approximation of* $H_{\tau,\alpha}(L)$, *and assume that*

$$\|\widehat{H}_{\tau,\alpha,K}(L) - H_{\tau,\alpha}(L)\|_2 \leqslant \varepsilon_K, \quad \|\widehat{H}_{\tau,\alpha,K}(\tilde{L}) - H_{\tau,\alpha}(\tilde{L})\|_2 \leqslant \varepsilon_K.$$

Then

$$\|\widehat{H}_{\tau,\alpha,K}(\tilde{L}) - \widehat{H}_{\tau,\alpha,K}(L)\|_2 \leqslant \tau C_\alpha \|\tilde{L} - L\|_2^{\min(1,\alpha)} + 2\varepsilon_K.$$

Therefore, the practical operator inherits the Hölder-type stability of theorem 1 up to an additive truncation error that decreases with the approximation order K.

Proof. We decompose the error using the triangle inequality:

$$\|\widehat{H}_{\tau,\alpha,K}(\tilde{L}) - \widehat{H}_{\tau,\alpha,K}(L)\|_2 \leqslant \|\widehat{H}_{\tau,\alpha,K}(\tilde{L}) - H_{\tau,\alpha}(\tilde{L})\|_2 + \|H_{\tau,\alpha}(\tilde{L}) - H_{\tau,\alpha}(L)\|_2 + \|H_{\tau,\alpha}(L) - \widehat{H}_{\tau,\alpha,K}(L)\|_2.$$

The middle term is bounded by theorem 1:

$$\|H_{\tau,\alpha}(\tilde{L}) - H_{\tau,\alpha}(L)\|_2 \leqslant \tau C_\alpha \|\tilde{L} - L\|_2^{\min(1,\alpha)}.$$

The first and third terms are bounded by the truncation error:

$$\|\widehat{H}_{\tau,\alpha,K}(\tilde{L}) - H_{\tau,\alpha}(\tilde{L})\|_2 \leqslant \varepsilon_K, \quad \|H_{\tau,\alpha}(L) - \widehat{H}_{\tau,\alpha,K}(L)\|_2 \leqslant \varepsilon_K.$$

Combining the bounds gives

$$\|\widehat{H}_{\tau,\alpha,K}(\tilde{L}) - \widehat{H}_{\tau,\alpha,K}(L)\|_2 \leqslant \tau C_\alpha \|\tilde{L} - L\|_2^{\min(1,\alpha)} + 2\varepsilon_K.$$

□

In practice, we approximate L^α and the associated matrix exponentials using Chebyshev/Lanczos expansions of order $K = 8-16$. This introduces a truncation error ε_K, which is explicitly accounted for in corollary 1. For the chosen values of K, this error is small, and the overall Hölder-type stability is effectively preserved.

Implications. The above bounds imply sublinear sensitivity to structural noise for $\alpha < 1$, and nonexpansiveness of heat-type filters ($\|\mathrm{e}^{-\tau L^\alpha}\|_2 \leqslant 1$). 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 A.1.3.

Table 1. Summary of the three EEG datasets. Channel processing was performed separately for scalp EEG and intracranial ECoG/SEEG recordings. For FMCE and HUP, channels marked as bad, non-neural, or reference-only were excluded according to the available electrode metadata. The scalp 10–20 system and spatial interpolation were not used to generate artificial intracranial contacts. Recordings with fewer than 64 valid contacts were zero-padded and accompanied by a binary node mask, whereas recordings with more than 64 valid contacts were reduced using a fixed recording-level contact-selection procedure. For the Helsinki neonatal dataset, the original 21-electrode scalp montage was retained; the remaining input positions were zero-padded rather than spatially interpolated. After subject- or recording-level partitioning, all recordings were segmented into 4~s windows ($W = 2048$ samples at 512~Hz) using a stride of 0.5~s ($R = 256$ samples), corresponding to an overlap of 3.5~s.

Dataset	Confidentiality	Species	# Subj.	Modality	# Ch.	# Samples	Duration	Stride
FMCE [59]	Public	Human	65	ECoG/SEEG	52–232	34 320	4 s	0.5 s
HUP [60]	Public	Human	73	ECoG/SEEG	47–216	182 937	4 s	0.5 s
Helsinki Neonatal EEG [61]	Public	Human	79	EEG	21	805 097	4 s	0.5 s

3. Experiments

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 sub-bands (δ: 1–4 Hz, θ: 4–8 Hz, α: 8–13 Hz, β: 13–25 Hz, low γ: 26–80 Hz, HFOs: >80 Hz). Artifact subspace reconstruction (ASR; burst criterion: 10 SD) was used to suppress transient high-amplitude artifacts. Independent component analysis using the Picard algorithm identified and removed ocular, muscular, and cardiac artifacts through combined manual inspection and ICLabel automation. Common average referencing was applied unless pre-referenced. Training parameters are provided in table A8.

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). EEG-GraphFrFT achieves the best result on 12 of the 15 reported dataset–metric combinations, achieving the highest ***F*1 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 FreTS is surpassed by our integration of *learnable* time–frequency representations with graph-based connectivity analysis, outperforming fixed Fourier transforms.

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 A4.

Exceptional noise robustness

EEG-GraphFrFT exhibits exceptional noise robustness in additive pink noise tests (table 3), maintaining **93.76%** accuracy under moderate noise (SD = 1)—a mere **0.8%** decrease from clean data versus **6.8%** (Mamba) and **4.0%** (EEG-Conformer) drops. Notably, under mild noise (SD = 0.5), it achieves **94.04%** accuracy (**0.5%** decrease) and even shows a **0.5%** improvement in $F1$ score, while other models exhibit declines. The $F1$ score decreases only **0.2%** compared to **7.1%** (Mamba) and **3.9%** (EEG-Conformer), highlighting superior stability. Under strong noise (SD = 5), EEG-GraphFrFT demonstrates remarkable

Table 2. Performance comparison of different models on epileptic EEG classification tasks. The best results are shown in bold, and the second-best results are underlined for each metric within each dataset.

Models	HUP					FMCE					Helsinki Neonatal EEG				
Metrics	Acc	*F*1	AUC	Sens	Spec	Acc	*F*1	AUC	Sens	Spec	Acc	*F*1	AUC	Sens	Spec
EEG-Conformer [10]	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 [9]	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 [11]	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 [12]	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
FreTS [8]	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 [62]	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**

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 percentage changes, indicating noise robustness, are shown in bold, and the second-best percentage changes are underlined for each metric. Tied second-best results are underlined together.

Models	Original					Pink Noise (SD = 1)				
Metrics	Acc	$F1$	AUC	Sens	Spec	Acc	$F1$	AUC	Sens	Spec
EEG-Conformer [10]	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 [11]	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 [62]	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%**)

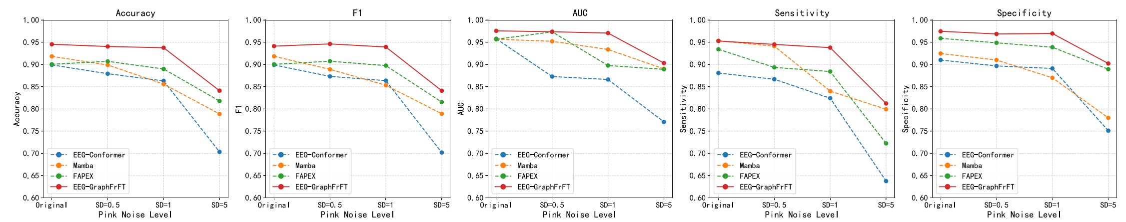

Figure 2. Performance comparison of various models under different noise conditions. The figure displays five metrics (Accuracy, $F1$ score, AUC, Sensitivity, and Specificity) for EEG-Conformer (blue dashed line), Mamba (orange dashed line), FAPEX (green solid line), and the proposed EEG-GraphFrFT (red solid line) across four noise conditions: original EEG (no noise), pink noise with standard deviation 0.5, pink noise with standard deviation 1.0 and pink noise with standard deviation 5.0. EEG-GraphFrFT maintains the strongest overall absolute performance under severe noise, although FAPEX exhibits smaller relative degradation on several metrics.

resilience with **84.10%** accuracy (**10.4%** drop), significantly outperforming Mamba (**13.0%** drop) and EEG-Conformer (**19.6%** drop). For complete noise robustness experimental data, refer to table A1; for the visualization of EEG data with superimposed noise, see figure A1; and for the robustness results of various models under different noise intensities for each metric, see figure 2.

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). This stands 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.

Ablation study and component analysis Ablation results (table 4) reveal critical insights: 1) **Learnable fractional orders are fundamental**: Optimal configuration (both channels with trained fractional orders: CHAN1 with θ and CHAN2 with α) achieves 94.5% accuracy, while fixed orders ($\theta = 1$ corresponds to the standard Fourier case in CHAN1, whereas $\alpha = 1$corresponds to the standard Fourier case in CHAN1, whereas $\alpha = 1$ corresponds to the integer-order graph spectral case in CHAN2, equivalent to 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 fractional orders show domain-specific optimization—CHAN1's parameter θ averages $\approx$0.5 (complete distribution of θ values in four Layers is provided in figure A3), while CHAN2's parameter α tailors to different graph types, with α values of 0.4926 for wPLI and 1.0065 for GC (complete distribution of α values in three Layers is provided in figure A2). The value $\alpha = 1.0065$ arises from unconstrained optimization and lies very close to the theoretical boundary $\alpha = 1$, indicating that the learned optimum is effectively near the standard spectral regime. The concentration of CHAN1's FNO average θ 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 directed predictive influence. 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 ($\theta = 1$ for CHAN1 and $\alpha = 1$ for CHAN2) 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. 4) **Sensitivity to initial values and gradient stability**: To deeply explore the robustness of the gradient updates and the model's sensitivity to initial conditions, we systematically evaluated performance across a grid search of initial parameters $\theta_0, \alpha_0 \in \{0.0, 0.5, 1.0\}$. As illustrated in the 5-panel 3D sensitivity surface plot (figure A9), the model demonstrates remarkable robustness regardless of the initialization states. Across the entire parameter grid, the performance metrics remain overwhelmingly stable, consistently maintaining test accuracies above 91.1% and AUC-ROC above 0.935. Crucially, the multi-metric visualization reveals a relatively smooth performance surface over the sampled initialization grid where the collective optimum converges precisely near the 'empirically favorable' regime ($\theta_0 = 0.5, \alpha_0 = 0.5$), yielding the highest Accuracy (94.54%) alongside optimally balanced Sensitivity and Specificity. This extensive empirical evidence effectively demonstrates that our

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

Channel Settings	Acc	$F1$	AUC	Sens	Spec
CHAN1: trained θ, CHAN2: trained α	**0.9454**	**0.9412**	**0.9756**	**0.9528**	**0.9746**
CHAN1: trained θ, CHAN2: $\alpha = 1$	0.9378 (−0.8%)	0.9377 (−0.4%)	0.9759 (+0.0%)	0.9368 (−1.7%)	0.9698 (−0.5%)
CHAN1: $\theta = 1$, CHAN2: trained α	0.9373 (−0.9%)	0.9373 (−0.4%)	0.9750 (−0.1%)	0.9365 (−1.7%)	0.9399 (−3.6%)
CHAN1: $\theta = 1$, CHAN2: $\alpha = 1$	0.8986 (−5.0%)	0.8991 (−4.5%)	0.9795 (+0.4%)	0.8913 (−6.5%)	0.9120 (−6.4%)
CHAN1: $\theta = 1.5$, 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 θ)	0.9138 (−3.3%)	0.9199 (−2.3%)	0.9558 (−2.0%)	0.9062 (−4.9%)	0.9494 (−2.6%)
Remove CHAN2 (only CHAN1: $\theta = 1$)	0.6910 (−26.9%)	0.6984 (−25.8%)	0.7223 (−26.0%)	0.6371 (−33.1%)	0.7367 (−24.4%)

gradient updates are highly stable, resisting divergence from arbitrary initializations and reliably navigating to the optimal functional connectivity representations. 5) **Hyperparameter sensitivity of exponential modulation**: To construct robust functional connectivity networks, our framework employs an element-wise exponential modulation mechanism with adaptive thresholds (τ_w, τ_g) and decay parameters (γ_w, γ_g). Crucially, in our EEG-GraphFrFT architecture, these are designed as *learnable parameters* updated via backpropagation rather than fixed hyperparameters. To evaluate the model's sensitivity to their initialization, we tested various configurations (detailed in appendix table A5). Extreme initializations, such as the absence of modulation (all 0 s) or aggressive pruning (all 1 s), degrade performance ($F1$ scores drop to 92.62% and 93.07%, respectively) by either retaining noisy spurious connections or destroying critical pathological topology. A uniform initialization (all 0.5 s) also performs suboptimally because it ignores the distinct scale differences between phase-based wPLI and directed GC metrics. In contrast, our optimal domain-aware initialization ($\tau_w = 0.5, \gamma_w = 1.0, \tau_g = 0.1, \gamma_g = 1.0$) yields the highest Accuracy (94.54%) and $F1$ score (94.12%). Assigning a lower initial threshold to GC (τg = 0.1) prevents the premature removal of sparse directed connections, while a higher initial threshold for wPLI (τw = 0.5) downweights weak or potentially spurious residual connections. The wPLI construction itself reduces the contribution of zero-lag volume conduction. Together, these settings provide an effective starting point for gradient descent to fine-tune the functional network topology.

To address the trade-off between capturing abrupt nonstationary transitions and maintaining reliable functional connectivity estimation, we evaluated varying sliding window sizes for graph construction (table A3). A short 0.5 s window results in lower performance (92.24% accuracy) due to insufficient temporal samples for stable wPLI and GC estimation, leading to noisy and statistically unreliable structural priors. Increasing the window to 2.0 s improves stability (93.08% accuracy), but the optimal balance is achieved at 4.0 s (94.54% accuracy), providing adequate data for robust spectral estimation without excessively blurring transient seizure dynamics. Further extending the window to 8.0 s degrades performance (92.72% accuracy), confirming that overly long windows violate the local stationarity assumption, oversmooth critical high-frequency temporal fluctuations, and fail to capture the abrupt structural reorganization of the brain network during seizures.

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 3, 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 GFRFT in CHAN2, is designed to detect and characterize in an adaptive manner. Notably, while the theoretical definition of fractional operators is typically constrained to the (0,1] interval, our implementation allows the model to explore fractional orders beyond this range ($\alpha > 1$), expanding the parameter space for optimal feature learning. Empirical results show that the learned parameters predominantly converge to physiologically interpretable regimes within the conventional range, validating the stability of our approach. The quantitative results from the ablation study

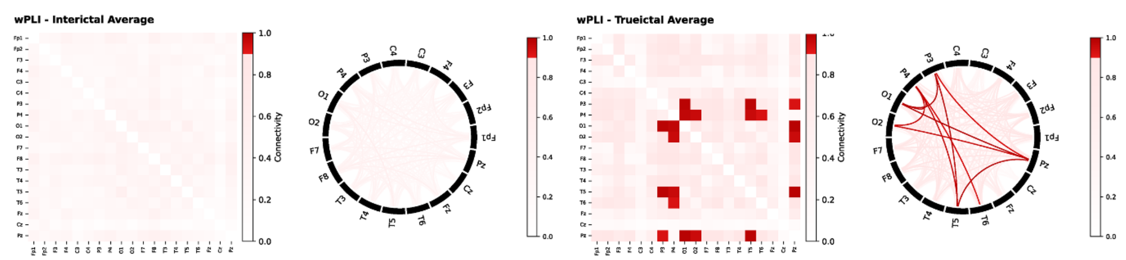

Figure 3. Group-averaged functional connectivity using wPLI. Left: Interictal state shows sparse and weak synchronization. Right: Ictal state exhibits dense hypersynchronization. Insets report two graph-spectral summaries *after* the fractional graph operator: the cumulative energy curve $E_{\mathrm{cum}}(k) = \sum_{i=0}^{k} \|\hat{x}_i\|^2$ over Laplacian eigenmodes (ordered by λ), and the low-frequency energy ratio $R_{\mathrm{low}} = \left(\sum_{\lambda_i \leqslant \lambda_0} \|\hat{x}_i\|^2\right) / \left(\sum_i \|\hat{x}_i\|^2\right)$. 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 figures A4 and A5, respectively. Full group-averaged graphs including GC analysis are provided in figure A6.

(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 O_2 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.

Our designed GFRFT demonstrates exceptional feature discrimination capabilities when processing both directed GC and undirected wPLI networks. Taking GC as an example, this advantage is further validated in figure 4 (the complete results including wPLI are shown in figure A7). 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 figures 4(b)–(d)), effectively concentrating the signature of the epileptic seizure. In contrast, the interictal state yields a lower-energy, more diffuse response (see figures 4(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 figures 4(e) and (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.

We evaluate the inference efficiency and performance trade-offs of various models based on comprehensive experiments conducted on an A100 GPU with 1000 inference samples. As illustrated in figure 5, the per-sample latency generally decreases with larger batch sizes due to parallelization, yet our EEG-GraphFrFT exhibits moderate latency compared to top-efficient models like Mamba. Table A2 confirms that EEG-GraphFrFT is not the fastest (achieving 20.6 samples/second throughput), but this computational cost is offset by its exceptional performance. Furthermore, to comprehensively address the computational overhead, we evaluate the training time and peak GPU memory consumption across models, as illustrated in figure A8. EEG-GraphFrFT requires a longer training duration, approximately 144 min per epoch on a single GPU, mainly due to dynamic graph construction and fractional operator optimization. However, its peak GPU memory consumption is only 950.89 MB, indicating that the main bottleneck is computation time rather than memory capacity. This represents a remarkable 98.4% memory reduction compared to large-scale models like EEG-Conformer, allowing our framework to be efficiently trained on standard consumer-grade GPUs without relying on expensive computing clusters. Crucially, table 2 shows that EEG-GraphFrFT achieves state-of-the-art results across all three datasets, leading in 12 out of 15 reported dataset–metric combinations. Thus, for accuracy-critical applications like epileptic EEG classification, our model offers a favorable balance, prioritizing diagnostic reliability and memory efficiency over minimal latency and training speed.

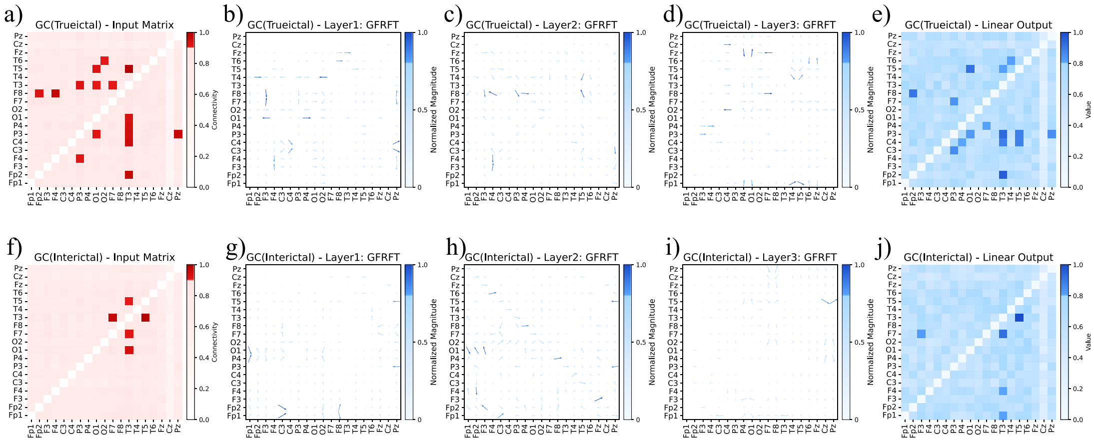

Figure 4. Processing pipeline for Granger-causality (GC) connectivity through three GFRFT layers. Panels (a)–(e) show an ictal trial and panels (f)–(j) an interictal trial. Arrow length encodes magnitude and arrow angle encodes phase of complex graph-spectral features. Across Layers 1–3 the ictal dynamics 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 $C_{\mathrm{dir}} = \left|\frac{1}{|\mathcal{E}|}\sum_{e\in\mathcal{E}} \mathrm{e}^{i\phi_e}\right|$. Ictal trials exhibit higher S_k and C_{dir}, consistent with a directed pathological flow enhanced by fractional orders.

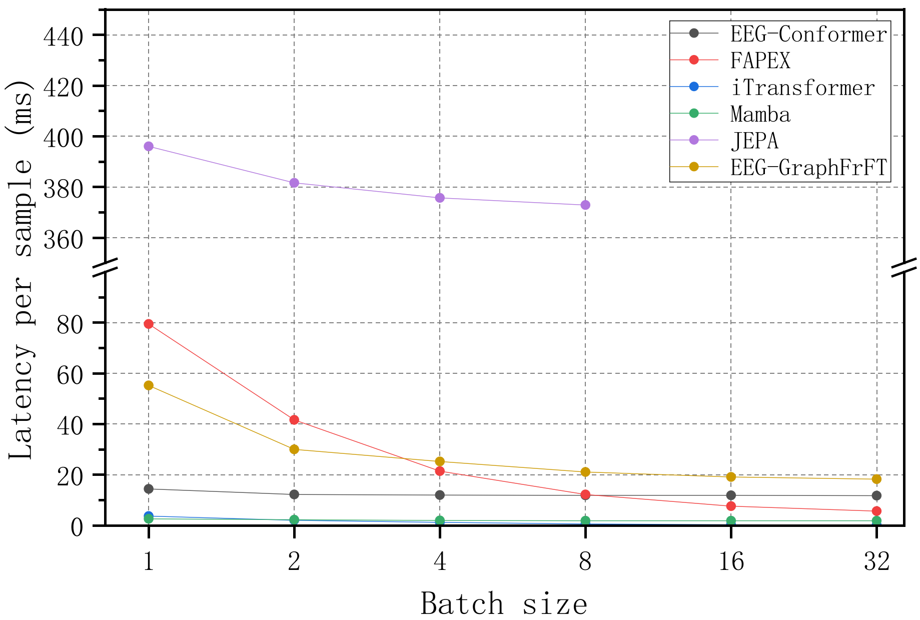

Figure 5. Per-sample inference latency comparison of different models across varying batch sizes. The plot demonstrates the computational efficiency of each model under different operational conditions. Notably, JEPA exhibits significantly higher latency and has no results for batch sizes larger than 8 due to cache overflow, making it less suitable for time-sensitive scenarios. As batch size increases, most models show reduced per-sample latency due to parallelization benefits.

Generalization to non-biological systems To empirically validate the task-agnostic nature of our framework and its applicability to broader nonstationary systems (e.g. traffic flow, climate teleconnections), we conducted experiments on a synthetic dynamic network dataset based on the Kuramoto model. The Kuramoto model of coupled oscillators is a fundamental paradigm in physics and network science for studying chaotic phase synchronization in complex macro-systems. We simulated highly nonstationary phase dynamics across 64 nodes for 2048 time steps, driven by varying coupling strengths and external forcing. To model distinct structural regimes, the underlying graph topologies were randomly generated as either Watts-Strogatz (WS) small-world networks or Barabási-Albert (BA) scale-free networks. The observation sequences were further corrupted by $1/f$ fractional noise to mimic real-world measurement uncertainties. The task was formulated as a binary classification (WS vs BA) using exclusively the observed chaotic time-series signals. The detailed parameters for the generative dynamic network, including graph topology configurations and non-linear oscillator dynamics, are summarized in table A7.

As demonstrated in table A6, conventional time-series and transformer-based architectures (EEG-Conformer, iTransformer, JEPA) severely struggle with this task, collapsing to near-random performance (Accuracy $\approx$ 51%, $F1 \leqslant$ 35%) due to the chaotic and nonstationary nature of the phase dynamics. While state-space models (Mamba) and physics-informed baselines (FAPEX) show improved resilience, our EEG-GraphFrFT framework achieves the highest Accuracy (84.11%) and $F1$ score (84.10%). This robust performance provides direct empirical evidence that the learnable fractional operators successfully disentangle complex graph-spectral dynamics and nonstationary temporal shifts without relying on biological priors, confirming the framework's broad applicability to non-biological complex systems.

Theoretical framework and generalization. This work introduces a unified mathematical framework of learnable fractional operators for analyzing nonstationary graph signals—a fundamental challenge spanning GSP and machine learning. We establish that fractional-order operators provide a principled generalization beyond integer-order calculus, naturally bridging temporal memory effects and spatial locality through continuous parameterization. The core theoretical contribution is a trainable dual-path architecture where FrFTs in the signal domain and fractional Laplacians in the graph domain share the same mathematical foundation of rotational and scaling operations, enabling consistent handling of dual nonstationarities.

Epileptic EEG serves as an ideal validation testbed—a 'worst-case' scenario where both signal statistics and network topology evolve rapidly and nonlinearly. Compared with the strongest baseline on each dataset, EEG-GraphFrFT improves accuracy by approximately 0.7–2.7 percentage points, while larger gains are observed relative to several other baselines.. More importantly, the consistent performance advantages under noise corruption and structural perturbations are consistent with the theoretical stability properties derived for fractional graph filters.

Mathematical foundations and mechanism insights. The framework's effectiveness stems from deep mathematical principles:

(i) **Fractional operators as continuity bridges.** By generalizing integer-order derivatives to continuous fractional orders, our framework creates a smooth interpolation between local and global, memoryless and historical dependencies. The learnable fractional parameter α enables adaptive control over the trade-off between feature preservation and noise suppression—a fundamental limitation of fixed-order methods. This continuous parameterization is particularly suited for epileptic dynamics, which exhibit fractal characteristics and long-range correlations that integer-order models fail to capture efficiently.

(ii) **Unified time-graph spectral processing.** The rotational interpretation of FrFT in the time–frequency plane and GraphFrFT in the graph spectral domain provides a coherent mathematical language for spatiotemporal analysis. This unification explains why our model outperforms approaches that process temporal and graph features separately: the shared fractional transformation paradigm ensures consistent treatment of nonstationarity across domains. The Hölder-type stability guarantees further provide rigorous foundations for real-world deployment under measurement uncertainties.

(iii) **Anomalous diffusion as implicit regularization.** The fractional graph Laplacian L^α induces a tunable anomalous diffusion process that fundamentally addresses the oversmoothing problem in deep GNNs. By compressing the spectral domain according to learned α values, our framework automatically balances global integration with local structure preservation—the learned parameters ($\alpha \approx$ 0.49 for synchronization networks, $\alpha \approx$ 1.0 for causal networks) empirically converge to biologically meaningful configurations that optimize this trade-off for seizure dynamics.

Generalization beyond epileptic EEG. While epileptic EEG provides a rigorous validation domain, the proposed mathematical framework is fundamentally task-agnostic. The core challenges of nonstationary graph signals, which is dual evolution of node features and network topology, appear across numerous domains. **Traffic sensor networks:** Road segment velocities (node features) and congestion propagation patterns (edge weights) change jointly during rush hours or incidents. **Climate teleconnections:** Regional climate indices and their long-distance correlation patterns evolve under climate change dynamics. **Multi-asset financial series:** Asset prices and their volatility spillover networks co-evolve during market regimes.

In each case, the learnable fractional operators can provide a principled approach to adaptively capture the intrinsic time-scale and spatial-scale characteristics without pre-specified basis functions or fixed diffusion rates. The theoretical guarantees of well-posedness and stability ensure reliable performance across different graph geometries and signal characteristics.

Limitations and future directions. Current performance still depends on the choice of connectivity metrics (wPLI/GC) and predefined frequency bands; generalization across recording centers and patient-specific calibration requires larger heterogeneous cohorts. Furthermore, while the accuracy gains are critical for medical diagnostics, the dual-path architecture—particularly the graph branch—introduces non-negligible computational overhead. As indicated by our inference analysis, the average latency of 48.606 ms per sample (approximately 20.57 samples/sec) is significantly slower than highly efficient baselines like Mamba and iTransformer. Consequently, the current framework is not suitable for real-time deployment on resource-constrained edge devices, such as portable or wearable medical monitoring equipment. The proposed fractional operators are task-agnostic and readily extend to other nonstationary graph signal domains (traffic networks, climate teleconnections, financial time series). To alleviate the current computational bottlenecks, promising next steps include exploring knowledge distillation to transfer learned representations into lightweight student models, and implementing dynamic graph sparsification to reduce message-passing complexity. Additional future directions involve self-supervised pretraining on unlabeled EEG, fully adaptive connectivity estimation, uncertainty quantification via fractional diffusion ensembles, and hardware-aware compression. Incorporating multi-dimensional GFRFT [63] or fractional Hilbert transforms [64] may further enhance phase-based seizure forecasting.

Conclusion. We have developed a general mathematical framework of learnable fractional operators for nonstationary graph signals, with rigorous theoretical foundations and empirical validation on challenging epileptic EEG data. The framework transcends domain-specific heuristics by providing a unified approach to dual nonstationarity through fractional calculus principles. By demonstrating both theoretical soundness and practical efficacy, this work establishes fractional operator learning as a promising direction for GSP, with potential applications spanning biomedical monitoring, infrastructure networks, and complex system modeling.

Acknowledgments

We acknowledge the support of National Natural Science Foundation of China (Grant Nos. 11875133, 11075057, 12547102 and 9257020), the National Key Research and Development Program of China under Contract No. 2024YFA1610902 and 2018YFB2101302, Science and Technology Innovation 2030 - Brain Science and Brain-Inspired Intelligence Project (2021ZD0201301), Shanghai Municipal Science and Technology (24JS2810400). The authors thank the Shanghai Institute for Mathematics and Interdisciplinary Sciences (SIMIS) for their financial support (SIMISID-2025-NC). The computations in this research were performed using the CFFF platform of Fudan University.

Data availability statement

The datasets used in this study are publicly available from the sources cited for the FMCE, HUP and Helsinki datasets [59–61]. The HUP iEEG Epilepsy Dataset is additionally available from OpenNeuro as dataset ds004100, version 1.1.3 [95].

Author contributions

Zhiwen Ren 0009-0006-8146-8132
Data curation (equal), Formal analysis (equal), Methodology (equal), Resources (equal), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (supporting)

Ruizhe Zheng 0009-0003-8531-3929
Data curation (equal), Formal analysis (equal), Methodology (equal), Resources (equal), Software (supporting), Validation (supporting), Visualization (supporting), Writing – original draft (equal), Writing – review & editing (supporting)

Chendrayan Dineshkumar
Formal analysis (equal), Validation (equal), Writing – original draft (supporting), Writing – review & editing (supporting)

Yansong Li
Data curation (supporting), Software (supporting), Visualization (supporting)

Yang Du
Data curation (supporting), Visualization (supporting)

Yuguo Yu 0000-0003-0603-2890
Conceptualization (equal), Investigation (equal), Project administration (equal), Supervision (equal), Writing – review & editing (equal)

Dingding Han 0000-0001-5340-9086
Conceptualization (equal), Funding acquisition (lead), Investigation (lead), Project administration (lead), Resources (equal), Supervision (equal), Writing – review & editing (equal)

Appendix

A.1. Mathematical foundations

A.1.1. 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:*

$$\lambda_i^a = e^{ja\theta_i}, \quad \lambda_i = e^{j\theta_i}. \tag{A30}$$

Thus:

$$(F_G^a)^H F_G^a = V(\Lambda^a)^H \Lambda^a V^{-1} = V\mathrm{diag}\left(|\lambda_i^a|^2\right) V^{-1} = I. \tag{A31}$$

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

$$L_{\mathrm{sym}}^{\alpha} = U\Lambda^{\alpha} U^{\mathrm{T}}, \tag{A32}$$

which is guaranteed to be symmetric PSD 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:*

$$F_G^a F_G^b = \exp(aM)\exp(bM) = \exp((a+b)M) = F_G^{a+b}. \tag{A33}$$

A.1.2. Extended fractional calculus on graphs

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

$$\mathcal{D}^{\beta} = F_G^{-\beta}\Lambda_{\beta}F_G^{\beta}, \qquad \Lambda_{\beta} = \mathrm{diag}\left(\lambda_k^{\beta}\right), \tag{A34}$$

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 unit-modulus modulation matrix encoding directional interactions in the graph, satisfying the conjugate symmetry condition*

$$\Upsilon_q(i,j) = \overline{\Upsilon_q(j,i)} \quad \forall i,j, \tag{A35}$$

where $\Gamma_q = (\Upsilon_q(i,j))$ denotes the matrix with entries $\Upsilon_q(i,j)$.

Define the directed Hermitian Laplacian

$$L_{\mathrm{H}} = D_s - \Gamma_q \odot W_s, \qquad D_s(i,i) = \sum_j W_s(i,j), \tag{A36}$$

where the complex phase of $\Upsilon_q(i,j)$ captures the directional asymmetry of the underlying causal interactions between nodes i and j.
Then:

1. *L_{H} is Hermitian, i.e. $L_{\mathrm{H}} = L_{\mathrm{H}}^*$.*
2. *L_{H} is PSD, i.e. $x^* L_{\mathrm{H}} x \geqslant 0$ for all $x \in \mathbb{C}^n$.*
3. *In the undirected case ($w_{ij} = w_{ji}$, $\Gamma_q \equiv 1$), one recovers $L_{\mathrm{H}} = L_{\mathrm{sym}}$, the standard symmetric Laplacian.*
4. *All eigenvalues of L_{H} are real and nonnegative.*

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

$$L_{\mathrm{H}}(i,j) = -\Upsilon_q(i,j)\, W_s(i,j), \qquad L_{\mathrm{H}}(j,i) = -\overline{\Upsilon_q(i,j)}\, W_s(i,j), \tag{A37}$$

so $L_{\mathrm{H}}(j,i) = \overline{L_{\mathrm{H}}(i,j)}$. Diagonal entries are real, thus $L_{\mathrm{H}} = L_{\mathrm{H}}^*$ [, 65].

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

$$x^* L_{\mathrm{H}} x = x^* D_s x - x^* \left(\Gamma_q \odot W_s\right) x = \sum_i \Big(\sum_j W_s(i,j)\Big)|x(i)|^2 - \sum_{i,j} W_s(i,j)\, \Upsilon_q(i,j)\, \overline{x(i)} x(j). \tag{A38}$$

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):

$$\begin{aligned} x^* L_{\mathrm{H}} x &= \tfrac{1}{2}\sum_{i,j} W_s(i,j)\left(|x(i)|^2 + |x(j)|^2 - \Upsilon_q(i,j)\,\overline{x(i)}x(j) - \Upsilon_q(j,i)\,\overline{x(j)}x(i)\right) \\ &= \tfrac{1}{2}\sum_{i,j} W_s(i,j)\left(|x(i)|^2 + |x(j)|^2 - 2\Re\left\{\Upsilon_q(i,j)\,\overline{x(i)}x(j)\right\}\right) \\ &= \tfrac{1}{2}\sum_{i,j} W_s(i,j)\left|x(i) - \Upsilon_q(i,j)\,x(j)\right|^2 \geqslant 0. \end{aligned} \tag{A39}$$

Hence $L_{\mathrm{H}} \succeq 0$; see also proposition 2 in []. In the undirected case ($\Gamma_q \equiv 1$), equation (A39) reduces to the standard proof for L_{sym} [66]. 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 $L_{\mathrm{H}}(i,i) \geqslant 0$ with radius $\sum_{j \neq i} |L_{\mathrm{H}}(i,j)| = L_{\mathrm{H}}(i,i)$. Thus $\Re(\lambda) \geqslant 0$, and since L_{H} is Hermitian, $\lambda \geqslant 0$.

□

Remark 2 (Preservation of directionality). Although the constructed Laplacian $L_{\mathrm{H}} = D_s - \Gamma_q \odot W_s$ is Hermitian, it does not discard the directional information inherent in the original GC graph.

Indeed, the modulation matrix $\Upsilon_q(i,j)$ encodes directional interactions through complex phases. Since $|\Upsilon_q(i,j)| = 1$, one can represent it as

$$\Upsilon_q(i,j) = \mathrm{e}^{\mathrm{i}\phi_{ij}}, \quad \Upsilon_q(j,i) = \overline{\Upsilon_q(i,j)} = \mathrm{e}^{-\mathrm{i}\phi_{ij}},$$

where $\phi_{ij} \in \mathbb{R}$ denotes the phase associated with the directed interaction between nodes i and j.

Thus, an interaction from node i to node j corresponds to a phase shift $+\phi_{ij}$, while the reverse direction corresponds to $-\phi_{ij}$. Therefore, directional (causal) asymmetry is preserved through phase differences, even though the resulting operator is Hermitian.

This construction is closely related to the magnetic Laplacian, where edge directionality is similarly encoded via complex phase factors while maintaining a Hermitian structure for stable spectral analysis.

Thus, Hermitianization does not symmetrize the graph in the real domain, but instead represents directional structure in the complex phase domain.

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

$$\log V = \mathrm{diag}(\log|\nu_k| + j\arg(\nu_k)), \quad \nu_k^\alpha = \exp(\alpha \log \nu_k), \tag{A40}$$

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.

A.1.3. Stability of graph fractional filters under structural perturbations

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

$$\mathcal{H}_{\tau,\alpha}(L) = U \mathrm{e}^{-\tau\Lambda^\alpha} U^\top = \mathrm{e}^{-\tau L^\alpha}. \tag{A41}$$

For a perturbed Laplacian $\tilde{L} = L + \Delta \succeq 0$ assume $\|\Delta\|_2 \leqslant \varepsilon$. 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*

$$\|A^\alpha - B^\alpha\|_2 \leqslant C_\alpha \|A - B\|_2^\alpha. \tag{A42}$$

Proof. This is classical; see, e.g. the [67] inequalities via double operator integrals or integral representations of fractional powers. Standard references include [68, 69]. □

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

$$\|\mathrm{e}^{-\tau X} - \mathrm{e}^{-\tau Y}\|_2 \leqslant \tau \|X - Y\|_2. \tag{A43}$$

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

$$\frac{d}{d\theta} F(\theta) = \tau\, \mathrm{e}^{-\tau(1-\theta)X} (X - Y)\, \mathrm{e}^{-\tau\theta Y}. \tag{A44}$$

Hence

$$\mathrm{e}^{-\tau X} - \mathrm{e}^{-\tau Y} = -\tau \int_0^1 \mathrm{e}^{-\tau(1-\theta)X} (X - Y)\, \mathrm{e}^{-\tau\theta Y}\, d\theta. \tag{A45}$$

Taking norms and using $\|\mathrm{e}^{-\tau(1-\theta)X}\|_2 \leqslant 1$ and $\|\mathrm{e}^{-\tau\theta Y}\|_2 \leqslant 1$ for PSD X, Y yields equation (A43). □

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*

$$\left\| \mathcal{H}_{\tau,\alpha}(L) - \mathcal{H}_{\tau,\alpha}(\tilde{L}) \right\|_2 \leqslant \tau C_\alpha \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A46}$$

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

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

$$\|\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau\tilde{L}^\alpha}\|_2 \leqslant \tau \|L^\alpha - \tilde{L}^\alpha\|_2 \quad \text{by lemma 2}, \tag{A47}$$

then use lemma 1 to obtain

$$\|L^\alpha - \tilde{L}^\alpha\|_2 \leqslant C_\alpha \|L - \tilde{L}\|_2^\alpha. \tag{A48}$$

□

Remark 3 (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 $\tilde{L}_{\mathrm{H}}$, since the functional calculus and the contraction property of $\mathrm{e}^{-\tau X}$ 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 $\|\Delta\|_2 \leqslant \varepsilon$, and fix $\lambda_0 > 0$. Let $P_E := P_{[\lambda_0,\infty)}(L)$ be the spectral projector of L onto eigenvalues $\geqslant \lambda_0$. Assume that $E = Ran(P_E)$ is a reducing subspace for both L and $\tilde{L}$. Then for any $v \in E = \mathrm{Ran}(P_E)$ and $\alpha \in (0,1]$,*

$$v^\top \tilde{L} v \geqslant v^\top L v - \|\Delta\|_2 \|v\|_2^2 \geqslant (\lambda_0 - \varepsilon) \|v\|_2^2, \tag{A49}$$

$$v^\top \tilde{L}^\alpha v \geqslant (\lambda_0 - \varepsilon)_+^\alpha \|v\|_2^2, \tag{A50}$$

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

$$\|\mathrm{e}^{-\tau s \tilde{L}^\alpha} P_E\|_2 \leqslant \mathrm{e}^{-\tau s (\lambda_0 - \varepsilon)_+^\alpha}, \qquad \|P_E \mathrm{e}^{-\tau(1-s)L^\alpha}\|_2 \leqslant \mathrm{e}^{-\tau(1-s)\lambda_0^\alpha}. \tag{A51}$$

Proof. Inequality equation (A49) follows from $v^\top \Delta v \geqslant -\|\Delta\|_2 \|v\|_2^2$ and $v^\top L v \geqslant \lambda_0 \|v\|_2^2$ on E. Operator monotonicity of $t \mapsto t^\alpha$ on $[0,\infty)$ gives equation (A50). The bounds in equation (A51) then follow by the spectral theorem, since $\mathrm{e}^{-\tau s \tilde{L}^\alpha} \preceq \mathrm{e}^{-\tau s(\lambda_0 - \varepsilon)_+^\alpha} I$ on E while $L^\alpha \succeq \lambda_0^\alpha$ on E. □

Theorem 6 (Exponential attenuation on high-frequency subspaces). *Let $L, \tilde{L} \succeq 0$ with $\tilde{L} = L + \Delta$, $\|\Delta\|_2 \leqslant \varepsilon$, and let $P_E = P_{[\lambda_0,\infty)}(L)$ with $\lambda_0 > 0$. Assume that $E = Ran(P_E)$ is a reducing subspace for both L and $\tilde{L}$. For $\alpha \in (0,1]$ and $\tau > 0$,*

$$\left\|P_E \left(\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau \tilde{L}^\alpha}\right) P_E\right\|_2 \leqslant \tau \, \mathrm{e}^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L^\alpha - \tilde{L}^\alpha\|_2. \tag{A52}$$

Consequently, by lemma 1,

$$\left\|P_E \left(\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau \tilde{L}^\alpha}\right) P_E\right\|_2 \leqslant \tau \, C_\alpha \, \mathrm{e}^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A53}$$

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

$$P_E \left(\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau \tilde{L}^\alpha}\right) P_E = -\tau \int_0^1 P_E \mathrm{e}^{-\tau(1-s)L^\alpha} \left(L^\alpha - \tilde{L}^\alpha\right) \mathrm{e}^{-\tau s \tilde{L}^\alpha} P_E \, ds. \tag{A54}$$

Taking norms and using equation (A51) gives

$$\|P_E \left(\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau \tilde{L}^\alpha}\right) P_E\|_2 \leqslant \tau \|L^\alpha - \tilde{L}^\alpha\|_2 \int_0^1 \mathrm{e}^{-\tau(1-s)\lambda_0^\alpha} \mathrm{e}^{-\tau s(\lambda_0 - \varepsilon)_+^\alpha} \, ds \tag{A55}$$

$$\leqslant \tau \|L^\alpha - \tilde{L}^\alpha\|_2 \, \mathrm{e}^{-\tau(\lambda_0 - \varepsilon)_+^\alpha}, \tag{A56}$$

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

Remark 4 (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

$$\|P_E \left(\mathrm{e}^{-\tau L^\alpha} - \mathrm{e}^{-\tau \tilde{L}^\alpha}\right) P_E\|_2 = \|\mathrm{e}^{-\tau (L|_E)^\alpha} - \mathrm{e}^{-\tau \left(\tilde{L}|_E\right)^\alpha}\|_2 \leqslant \tau \, C_\alpha \|P_E \Delta P_E\|_2^{\min\{1,\alpha\}}. \tag{A57}$$

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

Corollary 2 (Global-to-local high-frequency control). *With the assumptions of theorem 6,*

$$\|\mathcal{H}_{\tau,\alpha}(L) - \mathcal{H}_{\tau,\alpha}(\tilde{L})\|_2 \geqslant \|P_E \left(\mathcal{H}_{\tau,\alpha}(L) - \mathcal{H}_{\tau,\alpha}(\tilde{L})\right) P_E\|_2 \leqslant \tau \, C_\alpha \, \mathrm{e}^{-\tau(\lambda_0 - \varepsilon)_+^\alpha} \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A58}$$

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

Implications: Robustness of GFRFT and derived operators
We explain how the preceding results provide precise robustness guarantees for the GFRFT $\mathcal{F}_{t,\alpha}(L) = \mathrm{e}^{-itL^{\alpha}}$ and the unified complex-time family $\mathcal{G}_{\tau,t,\alpha}(L) = \mathrm{e}^{-(\tau+it)L^{\alpha}}$, $\tau \geqslant 0$.

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

$$\|\mathcal{F}_{\theta,\alpha}(L) - \mathcal{F}_{\theta,\alpha}(\tilde{L})\|_2 \leqslant |\theta|\, C_{\alpha}\, \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A59}$$

Corollary 3 (Complex-time family: stability and high-frequency damping). *For $\mathcal{G}_{\tau,\theta,\alpha}$,*

$$\|\mathcal{G}_{\tau,\theta,\alpha}(L) - \mathcal{G}_{\tau,\theta,\alpha}(\tilde{L})\|_2 \leqslant \sqrt{\tau^2+\theta^2}\, C_{\alpha}\, \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A60}$$

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

$$\|P_E\left(\mathcal{G}_{\tau,\theta,\alpha}(L) - \mathcal{G}_{\tau,\theta,\alpha}(\tilde{L})\right)P_E\|_2 \leqslant \sqrt{\tau^2+\theta^2}\, C_{\alpha}\, \mathrm{e}^{-\tau(\lambda_0-\varepsilon)_+^{\alpha}}\, \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A61}$$

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

$$\|\mathcal{R}_{\tau,\alpha}(L) - \mathcal{R}_{\tau,\alpha}(\tilde{L})\|_2 \leqslant \tau\, C_{\alpha}\, \|L - \tilde{L}\|_2^{\min\{1,\alpha\}}. \tag{A62}$$

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

A.1.4. Differentiability of GFRFT
We work specifically with the Hermitian directed Laplacian $L_{\mathrm{H}} = D_s - \Gamma_q \odot W_s$, 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, $L_{\mathrm{H}} = L_{\mathrm{H}}^{*}$ and $L_{\mathrm{H}} \succeq 0$[3].

Definition 3 (Spectral definition of L_{H}^{α}). Let the eigendecomposition of L_{H} be

$$L_{\mathrm{H}} = U\Lambda U^{\mathrm{H}}, \qquad U^{\mathrm{H}}U = I, \quad \Lambda = \mathrm{diag}(\nu_1,\ldots,\nu_n),\ \nu_k \geq 0, \tag{A63}$$

For $\alpha > 0$, the fractional power of L_{H} is defined by spectral calculus:

$$L_{\mathrm{H}}^{\alpha} \;=\; U\Lambda^{\alpha}U^{\mathrm{H}}, \qquad \Lambda^{\alpha} \;=\; \mathrm{diag}(\nu_1^{\alpha},\ldots,\nu_n^{\alpha}), \tag{A64}$$

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

Well-posedness. Since L_{H} is Hermitian PSD, equation (A64) is well-defined for $\alpha > 0$. 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 $L_{\mathrm{H},\varepsilon}^{\alpha} = U\Lambda_{\varepsilon}^{\alpha}U^{\mathrm{H}}$, 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 $L_{\mathrm{H}} = U\Lambda U^{\mathrm{H}}$ and L_{H}^{α} be defined by equation (A64). For $\alpha, \beta \geq 0$:*

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

$$\frac{\partial}{\partial\alpha} L_{\mathrm{H},\varepsilon}^{\alpha} \;=\; U\left(\Lambda_{\varepsilon}^{\alpha} \odot \log\Lambda_{\varepsilon}\right)U^{\mathrm{H}}, \tag{A65}$$

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

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

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

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

(2) Using equation (A64), $L_{\mathrm{H}}^{\alpha} L_{\mathrm{H}}^{\beta} = U\Lambda^{\alpha} U^{\mathrm{H}} U \Lambda^{\beta} U^{\mathrm{H}} = U\Lambda^{\alpha+\beta} U^{\mathrm{H}} = L_{\mathrm{H}}^{\alpha+\beta}$.

(3) 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 (A65).

□

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

$$Z_{\mathrm{dir}} \;=\; L_{\mathrm{H}}^{\alpha} X \;=\; U\Lambda^{\alpha} U^{\mathrm{H}} X. \tag{A66}$$

Let $\mathcal{L}$ be any real-valued loss on Z_{dir}. Using equation (A65) with an ε-shift, the gradient w.r.t. α is

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \operatorname{Re}\operatorname{Tr}\left(\left(\tfrac{\partial \mathcal{L}}{\partial Z_{\mathrm{dir}}}\right)^{H} U(\Lambda_{\varepsilon}^{\alpha} \circ \log \Lambda_{\varepsilon})\, U^{\mathrm{H}} X\right). \tag{A67}$$

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

A.1.5. 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 $k = 1,\ldots,d_o$, 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:

$$y_k = \mathbf{s}^{\top}\mathbf{V}_{s,k}\mathbf{V}_{g,k}^{\top}\mathbf{g} + b_k = \sum_{i=1}^{r}\left(\mathbf{V}_{s,k}[i]^{\top}\mathbf{s}\right)\left(\mathbf{V}_{g,k}[i]^{\top}\mathbf{g}\right) + b_k, \tag{A68}$$

or, using the Hadamard product $\odot$:

$$y_k = \mathbf{1}^{\top}\left[\left(\mathbf{V}_{s,k}^{\top}\mathbf{s}\right) \odot \left(\mathbf{V}_{g,k}^{\top}\mathbf{g}\right)\right] + b_k. \tag{A69}$$

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:

$$\mathbf{y} = \mathbf{U}^{\top}\left[\left(\mathbf{V}_s^{\top}\mathbf{s}\right) \odot \left(\mathbf{V}_g^{\top}\mathbf{g}\right)\right] + \mathbf{b}. \tag{A70}$$

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.

A.2. Full results

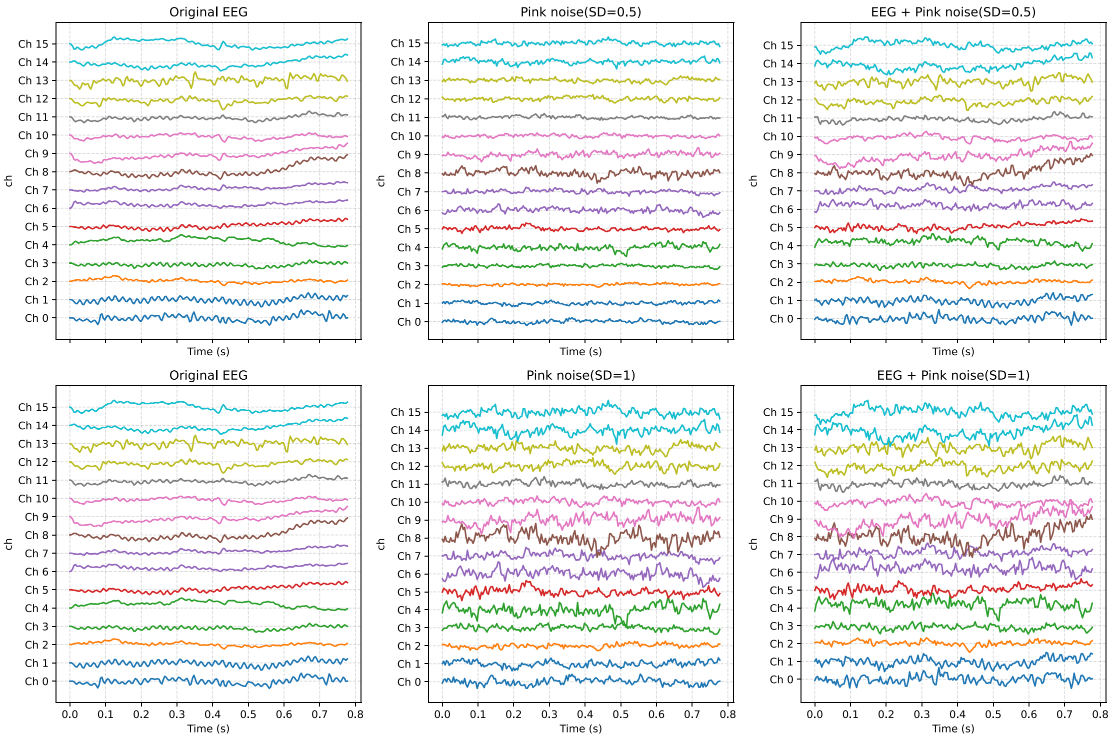

Figure A1. Visualization comparison of original EEG signals and those superimposed with pink noise at different intensities. (a) Original clean EEG signals; (b) Pink noise with standard deviation of 0.5; (c) Original EEG signals superimposed with SD = 0.5 noise; (d) Pink noise with standard deviation 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 s 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.

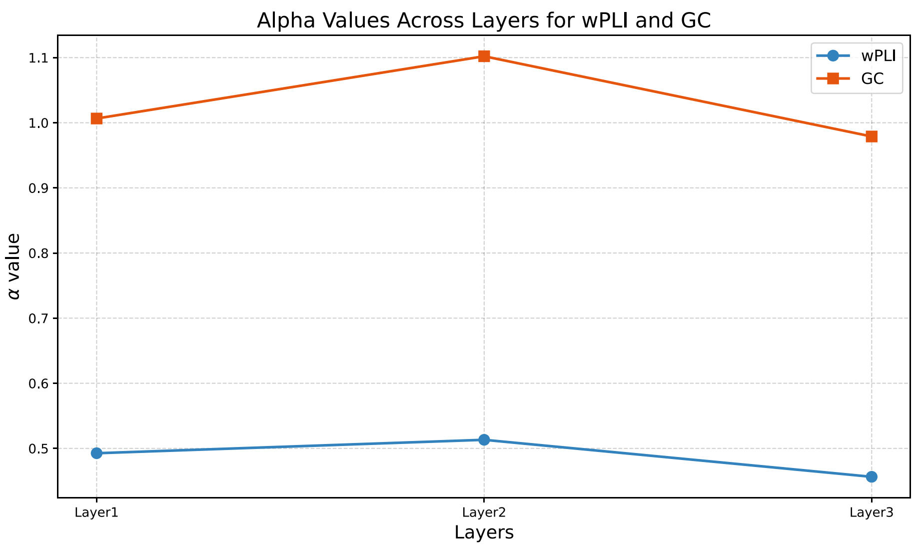

Figure A2. Distribution of the learned fractional order parameter α for the wPLI and GC algorithms across three layers (Layer 1 to Layer 3) in the CHAN2 model. The line plot illustrates the distinct distributions of the α values optimized for the functional networks constructed by the two algorithms. 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.

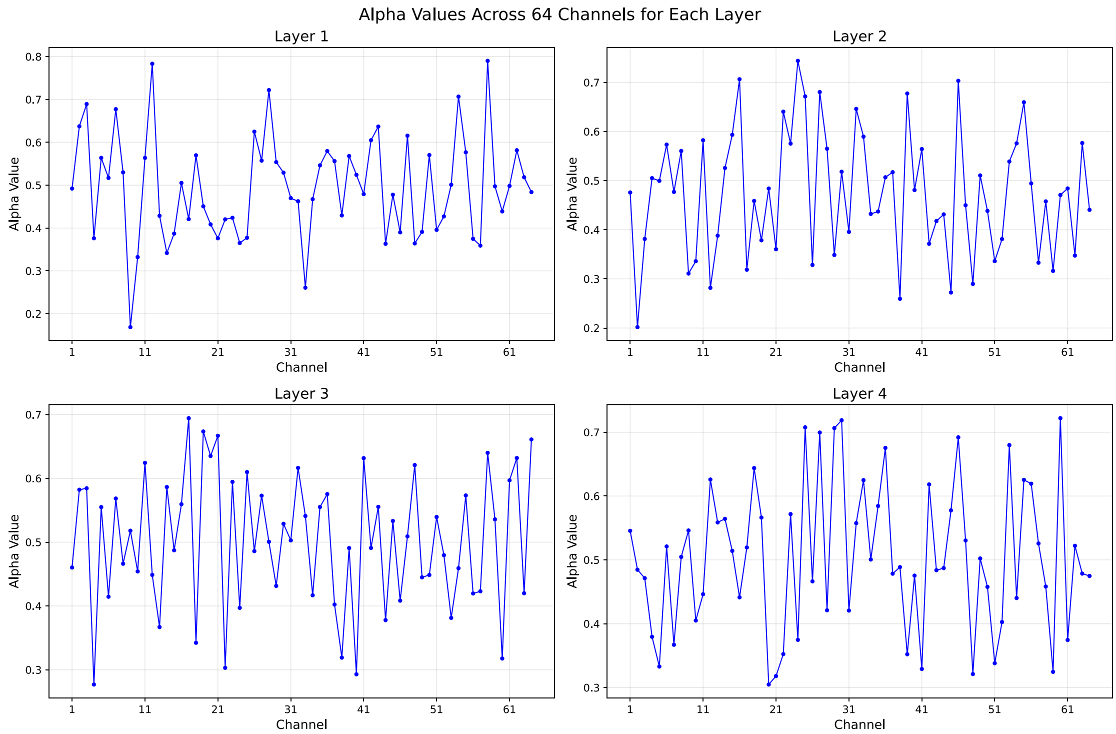

Figure A3. Distribution of learned fractional order parameter θ across 64 different channels for four layers (Layer1 to Layer4) in CHAN1. The line plot illustrates the adaptive variation of α values (ranging 0.1–0.8) optimized for each EEG channel, demonstrating the model's capacity to tailor time–frequency 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.

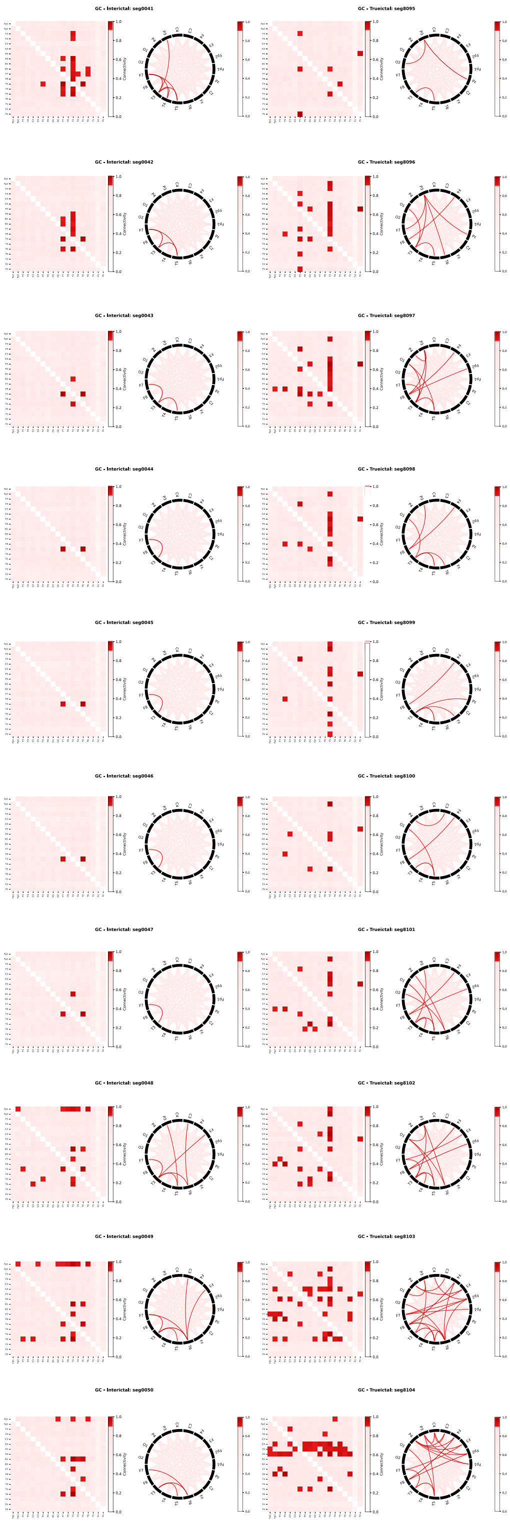

Figure A4. The figure shows functional connectivity matrices and brain network topology diagrams constructed using frequency-domain Granger Causality (GC) for 10 non-seizure samples (left two subfigures) and seizure samples (right two subfigures).

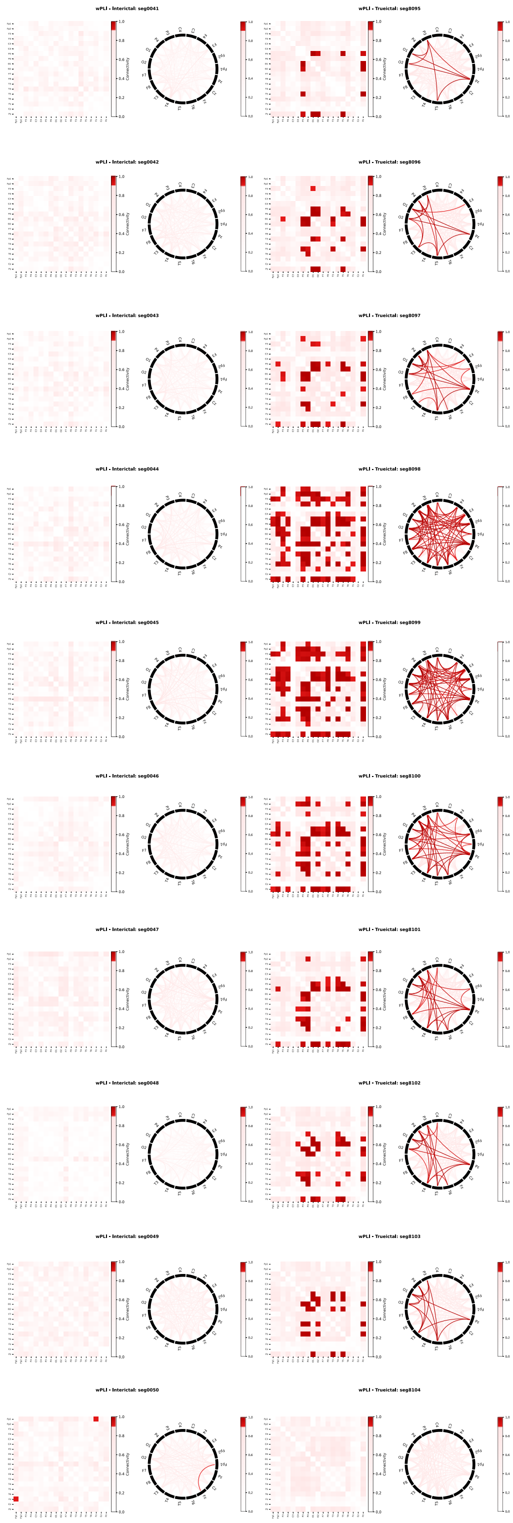

Figure A5. The figure shows functional connectivity matrices and brain network topology diagrams constructed using wPLI for 10 non-seizure samples (left two subfigures) and seizure samples (right two subfigures).

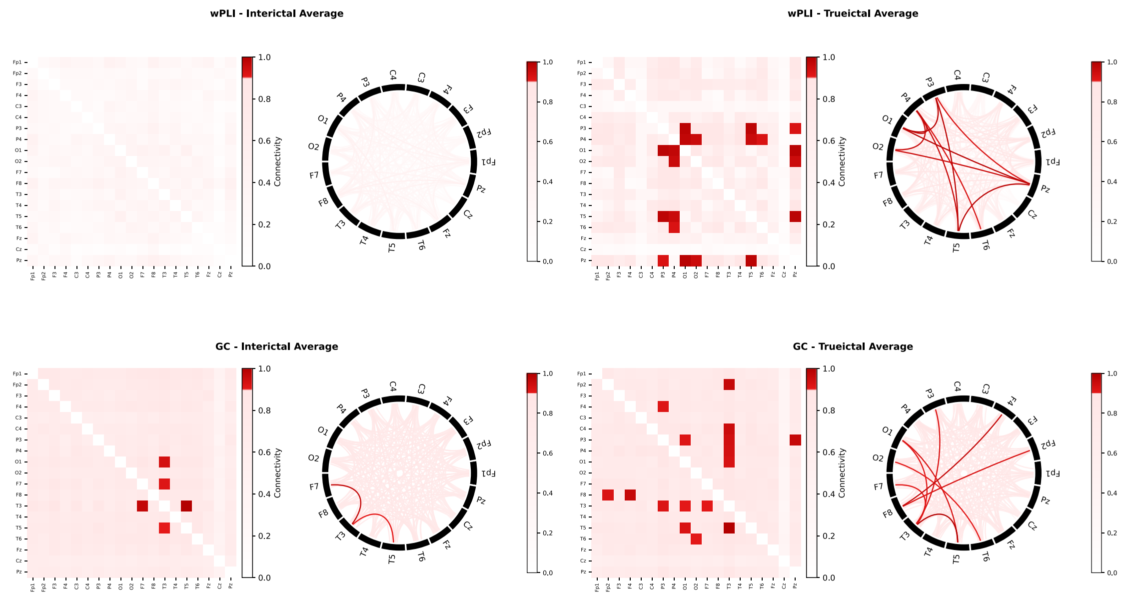

Figure A6. Average Functional connectivity analysis using wPLI and GC. (**Left**) Interictal period: sparse and weak synchronization. (**Right**) Ictal period: significantly enhanced phase-synchronization strength and density, demonstrating epileptic hypersynchronization.

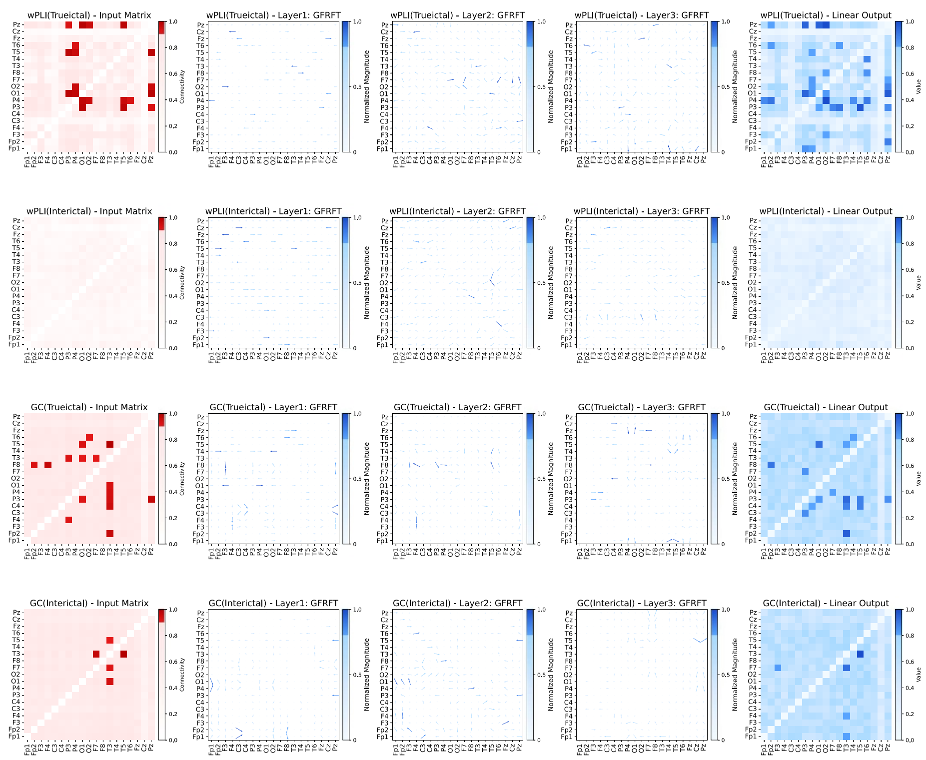

Figure A7. The transformation pipeline of wPLI and frequency-domain Granger Causality (GC) EEG functional connectivity data through the intermediate layers of the Graph Fractional Fourier Transform (GFRFT) is illustrated in the figure.

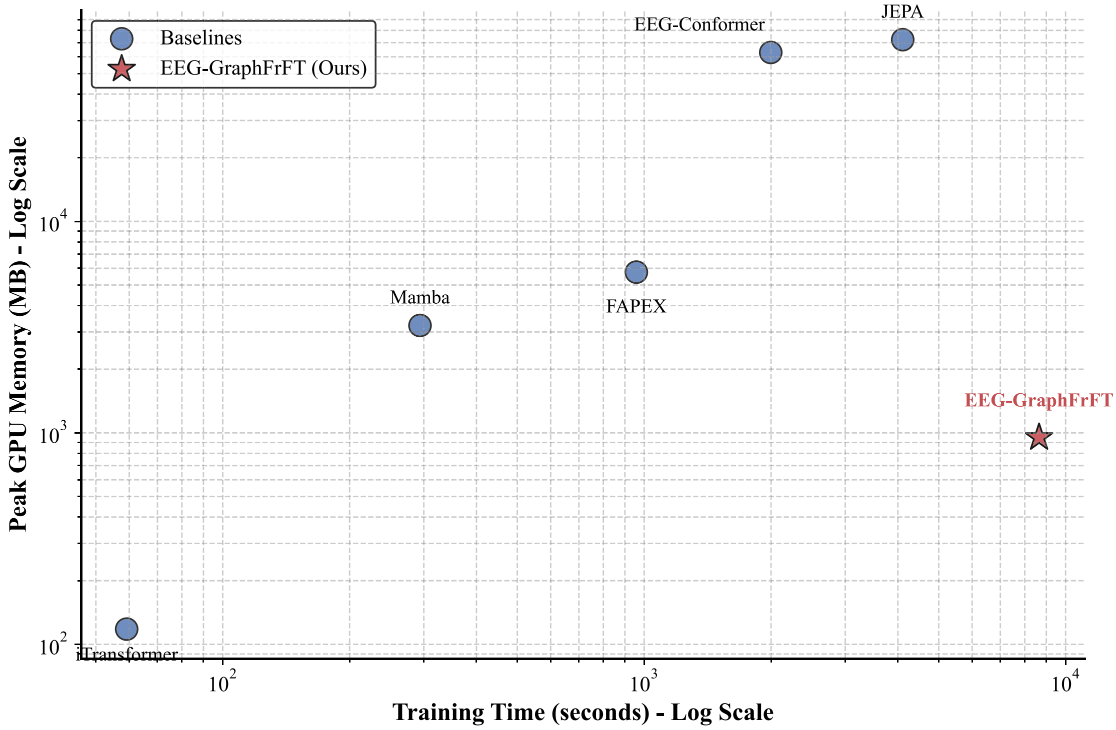

Figure A8. Comparison of training time and peak GPU memory consumption across different models for one training epoch on the FMCE dataset. Both axes are presented on a logarithmic scale to accommodate the wide range of computational requirements. While EEG-GraphFrFT requires a longer training duration due to its dynamic graph construction, it demonstrates exceptional memory efficiency, enabling training on standard consumer-grade GPUs.

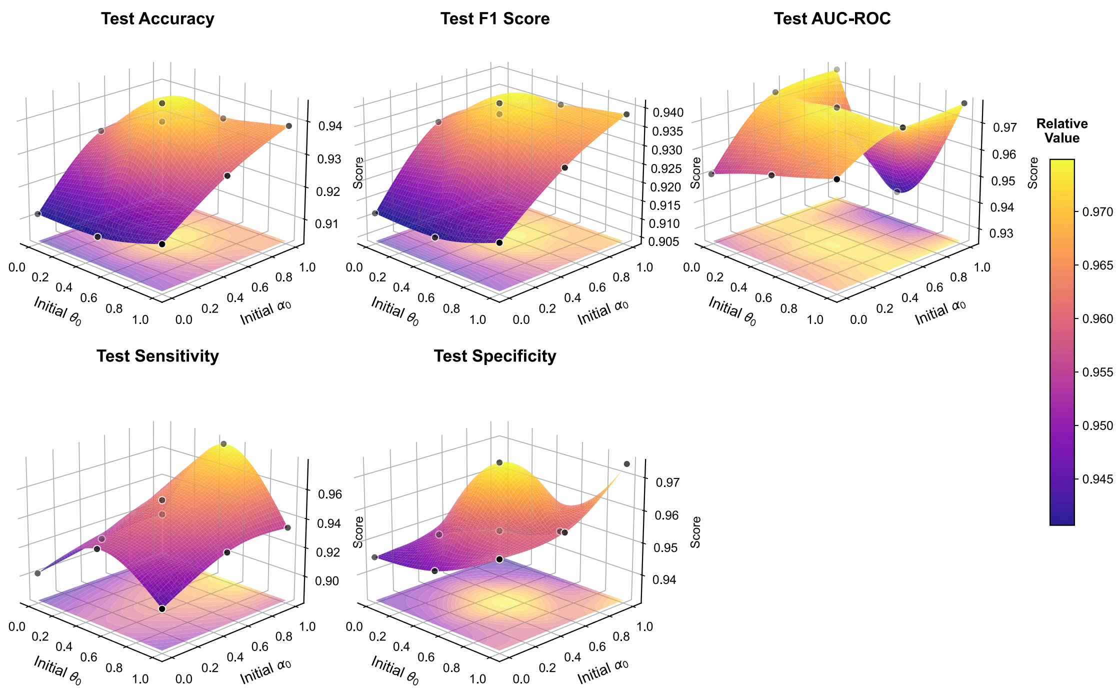

Figure A9. Sensitivity analysis of initial parameters θ_0 and α_0 across five critical performance metrics. The 3D surface plots illustrate the model's robustness to varying initializations ($\{0.0, 0.5, 1.0\}$) while demonstrating a clear and stable convergence toward an optimal performance peak near the ($\theta_0 = 0.5, \alpha_0 = 0.5$) regime.

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 = 0.5–5.0), the best percentage changes, indicating noise robustness, are shown in bold, and the second-best percentage changes are underlined for each metric. Tied second-best results are underlined together.

Noise condition	Metric	Models			
		EEG-conformer [10]	Mamba [11]	FAPEX [62]	**EEG-GraphFrFT**
Original	Acc	0.8990	0.9182	0.9005	0.9454
	$F1$	0.8990	0.9183	0.9003	0.9412
	AUC	0.9578	0.9563	0.9567	0.9756
	Sens	0.8807	0.9533	0.9341	0.9528
	Spec	0.9100	0.9246	0.9588	0.9746
Pink Noise (SD = 0.5)	Acc	0.8792 (−2.2%)	0.8989 (−2.1%)	0.9070 (**+0.7%**)	0.9404 (<u>−0.5%</u>)
	$F1$	0.8732 (−2.9%)	0.8890 (−3.2%)	0.9073 (**+0.8%**)	0.9462 (<u>+0.5%</u>)
	AUC	0.8728 (−8.9%)	0.9521 (−0.5%)	0.9733 (**+1.7%**)	0.9736 (<u>−0.2%</u>)
	Sens	0.8667 (−1.6%)	0.9411 (<u>−1.3%</u>)	0.8934 (−4.4%)	0.9451 (**−0.8%**)
	Spec	0.8967 (−1.5%)	0.9100 (−1.6%)	0.9487 (<u>−1.1%</u>)	0.9686 (**−0.6%**)
Pink Noise (SD = 1)	Acc	0.8630 (−4.0%)	0.8559 (−6.8%)	0.8899 (<u>−1.2%</u>)	0.9376 (**−0.8%**)
	$F1$	0.8636 (−3.9%)	0.8533 (−7.1%)	0.8974 (<u>−0.3%</u>)	0.9393 (**−0.2%**)
	AUC	0.8662 (−9.6%)	0.9337 (<u>−2.4%</u>)	0.8976 (−6.2%)	0.9706 (**−0.5%**)
	Sens	0.8240 (−6.4%)	0.8400 (−11.9%)	0.8841 (<u>−5.4%</u>)	0.9378 (**−1.6%**)
	Spec	0.8910 (<u>−2.1%</u>)	0.8700 (−5.9%)	0.9388 (<u>−2.1%</u>)	0.9696 (**−0.5%**)
Pink Noise (SD = 5)	Acc	0.7033 (−19.6%)	0.7887 (−13.0%)	0.8177 (**−8.3%**)	0.8410 (<u>−10.4%</u>)
	$F1$	0.7019 (−19.7%)	0.7892 (−12.9%)	0.8154 (**−8.5%**)	0.8409 (<u>−10.0%</u>)
	AUC	0.7708 (−18.7%)	0.8901 (**−6.7%**)	0.8890 (<u>−6.8%</u>)	0.9031 (−7.3%)
	Sens	0.6376 (−24.3%)	0.7992 (<u>−15.4%</u>)	0.7224 (−21.2%)	0.8122 (**−14.1%**)
	Spec	0.7510 (−15.9%)	0.7800 (−14.5%)	0.8893 (**−7.0%**)	0.9022 (<u>−7.2%</u>)

Table A2. Inference latency comparison of different models on epileptic EEG classification tasks. For latency metrics, lower values are better; for samples per second, higher values are better. The best results are shown in bold, and the second-best results are underlined for each metric.

Model	Mean latency	Std latency	Min latency	Max latency	p95 latency	p99 latency	Samples per sec
EEG-Conformer	14.391	<u>0.242</u>	14.278	19.056	14.410	14.873	69.487
FAPEX	82.638	4.528	75.834	136.352	86.154	99.307	12.101
JEPA	395.969	0.340	395.776	399.516	396.191	397.883	2.680
iTransformer	<u>3.753</u>	1.335	<u>3.558</u>	44.093	<u>3.793</u>	<u>6.366</u>	<u>266.454</u>
FreTS	7.854	**0.169**	7.787	<u>10.332</u>	7.867	8.431	127.318
Mamba	**2.621**	0.268	**2.506**	**5.972**	**3.040**	**3.183**	**381.535**
EEG-GraphFrFT	48.606	28.394	20.802	199.314	86.635	92.138	20.574

Table A3. Ablation study on the sliding window size for functional connectivity construction on the FMCE dataset.

Window size	Acc	$F1$	AUC	Sens	Spec
0.5 s	0.9224	0.9223	0.9718	0.8963	0.9422
2.0 s	0.9308	0.9311	0.9529	0.9130	0.9605
4.0 s (Proposed)	**0.9454**	**0.9412**	**0.9756**	**0.9528**	**0.9746**
8.0 s	0.9272	0.9369	0.9492	0.9310	0.9550

Table A4. Performance comparison with additional baseline methods on epileptic EEG classification tasks.

Models	HUP					FMCE					Helsinki Neonatal EEG				
Metrics	Acc	*F*1	AUC	Sens	Spec	Acc	*F*1	AUC	Sens	Spec	Acc	*F*1	AUC	Sens	Spec
Scattering transformer [70]	—	—	—	—	—	—	—	—	—	—	0.9055	0.7990	0.9638	—	—
MBGNN [71]	—	—	—	—	—	—	—	—	—	—	—	—	0.9911	—	—
Extra large (XL) ConvNeXt [72]	—	—	—	—	—	—	—	—	—	—	—	—	0.9280	—	—
Deep residual neural network [73]	—	—	—	—	—	—	—	—	—	—	0.9550	0.9000	0.9620	—	—
CNN [74]	—	—	—	—	—	—	—	—	—	—	—	—	0.8640	0.8900	0.7800
Fd-CAE [75]	—	—	—	—	—	—	—	—	—	—	0.9234	0.9577	—	—	—
CNN [76]	—	—	—	—	—	—	—	—	—	—	0.9700	—	—	—	—
SRFE [77]	—	—	—	—	—	0.9520	—	0.9100	1.0000	—	—	—	—	—	—
seizure matching system [78]	—	0.6700	0.7600	—	—	—	—	—	—	—	—	—	—	—	—
Random forest classifier [79]	—	—	0.7700	—	—	—	—	—	—	—	—	—	—	—	—
LASSO logistic regression [80]	0.7930	—	0.8300	—	—	—	—	—	—	—	—	—	—	—	—
SDA [81]	—	—	—	—	—	—	—	—	—	—	—	—	0.9880	0.7600	—
1D FCN,2D FCN [82]	—	—	—	—	—	—	—	—	—	—	—	—	0.9000	—	—
Morse Wavelets based LBP coupled with an ensemble classifier [83]	—	—	—	—	—	—	—	—	—	—	0.9700	—	0.9400	—	—
LMA-EEGNet [84]	—	—	—	—	—	—	—	—	—	—	0.9571	—	0.9862	0.9500	—
SSGNN + AO [85]	—	—	—	—	—	—	—	—	—	—	0.9870	—	0.9879	—	—
Logistic regression model [80]	0.7930	—	0.8300	—	—	—	—	—	—	—	—	—	—	—	—
Simple RF [77]	—	—	—	—	—	0.7800	—	—	0.1000	0.9900	—	—	—	—	—
Averaged RF [77]	—	—	—	—	—	0.7900	—	—	0.1000	0.9900	—	—	—	—	—
Manifold RF [86]	—	—	—	—	—	—	—	0.8900	—	—	—	—	—	—	—
[87]	—	—	0.8300	—	—	—	—	—	—	—	—	—	—	—	—
Virtual resection [88]	0.8600	—	0.8900	—	—	—	—	—	—	—	—	—	—	—	—
Network-based epilepsy model [89]	0.6600	—	—	0.8900	0.6500	—	—	—	—	—	—	—	—	—	—
functional connectivity networks [90]	—	—	0.7500	—	—	—	—	—	—	—	—	—	—	—	—
IED-to-LoWS delay-based classifier [91]	0.6300	0.6700	0.6800	0.6700	0.6000	—	—	—	—	—	—	—	—	—	—
VEP [92]	—	—	0.7500	—	—	—	—	—	—	—	—	—	—	—	—
Neural fragility model [93]	—	—	—	—	—	0.7900	0.7400	0.8200	—	—	—	—	—	—	—
SeizureNet [94]	—	—	—	—	—	—	—	—	—	—	0.8700	0.8400	0.9000	—	—

Table A5. Initialization sensitivity of the differentiable graph-modulation parameters. The first four columns report only the initial values of the trainable parameters τ_w, γ_w,τ_g, and γ_g; all four parameters are subsequently updated through backpropagation. The best results are highlighted in bold.

Parameter Initialization				Metrics				
τ_w	γ_w	τ_g	γ_g	**Acc**	***F*1**	**AUC**	**Sens**	**Spec**
0.0	0.0	0.0	0.0	0.9264	0.9262	**0.9771**	0.9011	0.9450
0.5	0.5	0.5	0.5	0.9303	0.9302	0.9697	0.9293	0.9453
1.0	1.0	1.0	1.0	0.9310	0.9307	0.9744	0.8857	0.9657
0.5	**1.0**	**0.1**	**1.0**	**0.9454**	**0.9412**	0.9756	**0.9528**	**0.9746**

Table A6. Performance comparison on the non-biological Kuramoto coupled oscillator dataset. The best results are highlighted in **bold**.

Models	Acc	*F*1	AUC	Sens	Spec
EEG-Conformer	0.5122	0.3559	0.4935	0.0091	0.9891
JEPA	0.5111	0.0000	0.5000	0.0000	**1.0000**
iTransformer	0.5111	0.3458	0.5344	0.0000	**1.0000**
Mamba	0.7944	0.7935	0.9034	0.8636	0.6826
FAPEX	0.8397	0.8269	**0.9679**	**0.8711**	0.8072
EEG-GraphFrFT	**0.8411**	**0.8410**	0.9151	0.8295	0.8891

Table A7. Parameters used for generating the synthetic Kuramoto dynamic network dataset. Variables defined by ranges indicate values uniformly sampled for each instantiation.

Category	Parameter	Value / Range
Dataset config	Number of variables/nodes (C)	64
	Samples per class	3000
	Retained time steps (T)	2048
	Integration step (dt)	0.03
	Burn-in steps	300
Graph topology	Average degree (k, m)	$[6, 10]$
	WS rewiring probability (p)	$[0.10, 0.30]$
	Edge perturbation fraction	$[0.00, 0.05]$
Oscillator dynamics	Damping (γ)	$[0.6, 1.4]$
	Mass / Inertia	$[0.6, 1.6]$
	Natural frequency std (σ_ω)	$[0.4, 1.2]$
	Coupling strengths (K_1, K_2)	$[0.8, 2.0]$, $[0.15, 0.8]$
	Phase shifts (α, β)	$[0.6, 1.2]$, $[0.2, 1.0]$
External forcing & noise	Forcing amplitude (F)	$[0.0, 0.35]$
	Forcing frequency (Ω)	$[0.4, 1.4]$
	Fractional measurement noise ($1/f^\beta$)	Level $= 1.5, \beta = 1.0$

Reproducibility statement

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

Table A8. 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	1×10^{-8}
• lookahead_k	5
• lookahead_alpha	0.5
- Learning rate	5×10^{-4}
- LR scheduler	ReduceLROnPlateau
• factor	0.5
• patience	3
• min_lr	1×10^{-6}
- Weight decay	1×10^{-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

References

[1] Thijs R D, Surges R, O'Brien T J and Sander J W 2019 Epilepsy in adults *The Lancet* **393** 689–701

[2] Stamoulis C, Gruber L J, Schomer D L and Chang B S 2012 High-frequency neuronal network modulations encoded in scalp EEG precede the onset of focal seizures *Epilepsy Behav.* **23** 471–80

[3] Noorlag L, van Klink N E C, Kobayashi K, Gotman J, Braun K P J and Zijlmans M 2022 High-frequency oscillations in scalp EEG: a systematic review of methodological choices and clinical findings *Clin. Neurophysiol.* **137** 46–58

[4] Li F *et al* 2019 Transition of brain networks from an interictal to a preictal state preceding a seizure revealed by scalp EEG network analysis *Cogn. Neurodyn.* **13** 175–81

[5] Pontes E D, Pinto M, Lopes F and Teixeira C 2024 Concept-drifts adaptation for machine learning EEG epilepsy seizure prediction *Sci. Rep.* **14** 8204

[6] Alshayeji M H 2023 Cad system for epileptic seizure detection from EEG through image processing and SURF-BOF technique *Mach. Learn.: Sci. Technol.* **4** 045029

[7] Alahmadi T J, Ur Rahman A, Alhababi Z A, Ali S and Alkahtani H K 2024 Prediction of mild cognitive impairment using EEG signal and BiLSTM network *Mach. Learn.: Sci. Technol.* **5** 025028

[8] Yi K, Zhang Q, Fan W, Wang S, Wang P, He H, Lian D, An N, Cao L and Niu Z 2023 Frequency-domain mlps are more effective learners in time series forecasting *Advances in Neural Information Processing Systems* vol 36 pp 76656–79

[9] Dong Z, Li R, Wu Y, Nguyen T T, Chong J S X, Ji F, Tong N R J, Chen C Li H and Zhou J H 2024 Brain-JEPA: brain dynamics foundation model with gradient positioning and spatiotemporal masking *Advances in Neural Information Processing Systems* vol 37 pp 86048–73

[10] Song Y, Zheng Q, Liu B and Gao X 2023 Eeg-conformer: Convolutional transformer for EEG decoding and visualization *IEEE Trans. Neural Syst. Rehabil. Eng.* **31** 710–9

[11] Gui Y, Chen M, Su Y, Luo G and Yang Y 2024 EEGMamba: bidirectional state space models with mixture of experts for EEG classification *CoRR* (arXiv:2407.20254)
[12] Liu Y, Hu T, Zhang H, Wu H, Wang S, Ma L and Long M 2023 iTransformer: inverted transformers are effective for time series forecasting (arXiv:2310.06625)
[13] Guariglia E, Guido R C and Dalalana G J P 2023 From wavelet analysis to fractional calculus: a review *Mathematics* **11** 1606
[14] Fei S, Chen J and Hu Y 2025 A novel time-frequency feature extraction method of EEG signals utilizing fractional synchrosqueezing wavelet transform *Phys. Eng. Sci. Med.* **48** 1237–47
[15] Zhang F, Yang C, You L, Wang X, Yuan Y, Le Bouquin Jeannès R, Shu H and Xiang W 2025 WS-BiLSTM-MA: wavelet scattering-based BiLSTM with mixed attention block for MDD recognition using multichannel EEG signals *IEEE Trans. Instrum. Meas.* **74** 1–13
[16] Wang X Y, Li C, Zhang R, Wang L, Tan J L and Wang H 2022 Intelligent extraction of salient feature from electroencephalogram using redundant discrete wavelet transform *Front. Neurosci.* **16** 921642
[17] Peng P, Song Y, Yang L and Wei H 2022 Seizure prediction in EEG signals using STFT and domain adaptation *Front. Neurosci.* **15** 825434
[18] Babu N R and Vadivel V 2025 MFENet: a multi-feature extraction network for enhanced emotion detection using EEG and STFT *IEEE Access* **13** 133338–50
[19] Ali O, Saif-ur-Rehman M, Dyck S, Glasmachers T, Iossifidis I and Klaes C 2022 Enhancing the decoding accuracy of EEG signals by the introduction of anchored-STFT and adversarial data augmentation method *Sci. Rep.* **12** 4245
[20] Beeraka S M, Kumar A, Sameer M, Ghosh S and Gupta B 2022 Accuracy enhancement of epileptic seizure detection: a deep learning approach with hardware realization of STFT *Circuits Syst. Signal Process.* **41** 461–84
[21] Xu Z, Philip Chen C L and Zhang T 2025 Tfagl: a novel agent graph learning method using time-frequency EEG for major depressive disorder detection *IEEE Trans. Affective Comput.* **16** 1592–605
[22] Huang Q, Wei S and Zhang L 2023 Robust adaptive time-frequency filtering for frequency shift jamming suppression *IEEE Trans. Aerosp. Electron. Syst.* **59** 7963–76
[23] Zheng N and Li Y 2024 Adaptive radial basis functions based time-varying model for EEG analysis in patients with cervical dystonia *Biomed. Signal Process. Control* **92** 106135
[24] Wang Y and Li B 2018 The fractional fourier transform on graphs: sampling and recovery *2018 14th IEEE Int. Conf. on Signal Processing (ICSP)* pp 1103–8
[25] Zhang Y and Li B-Z 2026 Sampling of graph signals based on joint time-vertex fractional Fourier transform *Signal Process.* **239** 110309
[26] Wei D and Yan Z 2024 Generalized sampling of graph signals with the prior information based on graph fractional Fourier transform *Signal Process.* **214** 109263
[27] Shuman D I, Narang S K, Frossard P, Ortega A and Vandergheynst P 2013 The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains *IEEE Signal Process. Mag.* **30** 83–98
[28] Guibas J, Mardani M, Li Z, Tao A, Anandkumar A and Catanzaro B 2021 Adaptive Fourier neural operators: efficient token mixers for transformers (arXiv:2111.13587)
[29] Alikaşifoğlu T, Kartal B and Koç A 2024 Graph fractional fourier transform: a unified theory *IEEE Trans. Signal Process.* **72** 3834–50
[30] Ortega A, Frossard P, Kovačević J, Moura J M F and Vandergheynst P 2018 Graph signal processing: overview, challenges and applications *Proc. IEEE* **106** 808–28
[31] Cui W, Kang Q, Li X, Zhao K, Tay W P, Deng W and Li Y 2025 Neural variable-order fractional differential equation networks *Proc. AAAI Conf. on Artificial Intelligence*
[32] Coelho C, Costa M F P and Ferrás L L 2025 Neural fractional differential equations *Appl. Math. Modelling* **144** 116060
[33] Paolino R, Maskey S, Delvenne J-C and Vandergheynst P 2023 A fractional graph Laplacian approach to oversmoothing *Advances in Neural Information Processing Systems* 36 (arXiv:2305.13084)
[34] Kang Q, Zhao K, Ding Q, Ji F, Li X, Liang W, Song Y and Tay W P 2024 Unleashing the potential of fractional calculus in graph neural networks with FROND Accepted at ICLR 2024.
[35] Chen Y, Wu L and Zaki M J 2020 Iterative deep graph learning for graph neural networks: better and robust node embeddings *Advances in Neural Information Processing Systems (NeurIPS)* vol 33 pp 19314–26
[36] Wu W, Song C, Zhao J and Xu Z 2023 Iterative deep graph learning and embedding for industrial fault diagnosis with limited labeled data *IEEE Trans. Instrum. Meas.* **73** 1–15
[37] Kipf T, Fetaya E, Wang K-C, Welling M and Zemel R 2018 Neural relational inference for interacting systems *Int. Conf. on Machine Learning (ICML)* (PMLR) pp 2688–97
[38] Tygesen M N, Pereira F C and Rodrigues F 2023 Unboxing the graph: towards interpretable graph neural networks for transport prediction through neural relational inference *Transp. Res. C: Emerg. Technol.* **146** 103946
[39] Zhang X, He Y, Brugnone N, Perlmutter M and Hirn M 2021 Magnet: a neural network for directed graphs *Advances in Neural Information Processing Systems (NeurIPS)* vol 34 pp 27003–15
[40] Fanuel M, Alaíz C M, Fernández Á and Suykens J A K 2018 Magnetic eigenmaps for the visualization of directed networks *Appl. Comput. Harmon. Anal.* **44** 189–99
[41] Guibas J, Mardani M, Li Z, Tao A, Anandkumar A and Catanzaro B 2021 Efficient token mixing for transformers via adaptive Fourier neural operators *Int. Conf. on Learning Representations*
[42] Liu Li, Wu J, Li D, Senhadji L and Shu H 2019 Fractional wavelet scattering network and applications *IEEE Trans. Biomed. Eng.* **66** 553–63
[43] Chen W, Ye J, Zhao C and Huang Y 2025 MFFCNN: multi-scale fractional fourier transform convolutional neural network for multivariate time series forecasting *J. Supercomput.* **81** 416
[44] Yan F-J, Gao W-B and Li B-Z 2020 Windowed fractional Fourier transform on graphs: Fractional translation operator and Hausdorff-Young inequality *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC)* pp 255–9
[45] Yan and Li 2023 Spectral graph fractional fourier transform for directed graphs and its application *Signal Process.* **210** 109099
[46] Gao J and Ribeiro B 2022 On the equivalence between temporal and static equivariant graph representations *Proc. 39th Int. Conf. on Machine Learning (ICML)* pp 7052–76

[47] Kotoge R, Chen Z, Kimura T, Matsubara Y, Yanagisawa T, Kishima H and Sakurai Y 2025 EvoBrain: dynamic multi-channel EEG graph modeling for time-evolving brain networks *Advances in Neural Information Processing Systems (NeurIPS 2025)* vol 38 pp 1–15
[48] Tang S, Dunnmon J A, Qu L, Saab K, Lee-Messer C and Rubin D 2022 Spatiotemporal modeling of multivariate signals with graph neural networks and structured state space models (arXiv:2211.11176)
[49] Chen G, Zhao Y and Mouthaan K 2025 Joint spatial-frequency scattering network for unsupervised sar image change detection *IEEE Geosci. Remote Sens. Lett.* **22** 1–5
[50] Mao L, Zheng G, Cai Y, Luo W, Zhang Q, Peng W, Ding J and Wang X 2022 Frontotemporal phase lag index correlates with seizure severity in patients with temporal lobe epilepsy *Frontiers Neurol.* **13** 855842
[51] Bowyer S M 2016 Coherence a measure of the brain networks: past and present *Neuropsychiatric Electrophysiol.* **2** 1
[52] Adamovich T, Zakharov I, Tabueva A and Malykh S 2022 The thresholding problem and variability in the EEG graph network parameters *Sci. Rep.* **12** 18659
[53] Shoeibi A, Ghassemi N, Khodatars M, Moridian P, Alizadehsani R, Zare A, Khosravi A, Subasi A, Rajendra Acharya U and Gorriz J M 2022 Automatic diagnosis of epileptic seizures in EEG signals using deep learning algorithms: a review *Comput. Biol. Med.* **144** 105352
[54] Sharma M, Patel V, Tiwari J and Acharya U R 2022 Fractional fourier transform aided computerized framework for alcoholism identification in EEG *Int. Conf. on Computational Intelligence in Data Science (ICCIDS)* (Springer) pp 119–33
[55] Ozaktas H M, Arikan O, Alper Kutay M and Bozdagt G 1996 Digital computation of the fractional fourier transform *IEEE Trans. Signal Process.* **44** 2141–50
[56] Vinck M, Oostenveld R, van Wingerden M, Battaglia F and Pennartz C M A 2011 An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias *Neuroimage* **55** 1548–65
[57] Granger C W J 1969 Investigating causal relations by econometric models and cross-spectral methods *Econometrica: J. Econometric Soc.* **37** 424–38
[58] Dhamala M, Rangarajan G and Ding M 2008 Estimating Granger causality from fourier and wavelet transforms of time series data *Phys. Rev. Lett.* **100** 018701
[59] Li A *et al* 2021 Neural fragility as an EEG marker of the seizure onset zone *Nat. Neurosci.* **24** 1465–74
[60] Kini L G *et al* 2019 Virtual resection predicts surgical outcome for drug-resistant epilepsy *Brain* **142** 3892–905
[61] Stevenson N J, Tapani K, Lauronen L and Vanhatalo S 2019 A dataset of neonatal EEG recordings with seizure annotations *Sci. Data* **6** 190039
[62] Zheng R, Mao L, Han D, Luo T, Wang Y, Ding J and Yu Y 2025 FAPEX: fractional amplitude-phase expressor for robust cross-subject seizure prediction *Advances in Neural Information Processing Systems*
[63] Yan and Li 2022 Multi-dimensional graph fractional Fourier transform and its application to recommendation systems *Digit. Signal Process.* **129** 103683
[64] Zhang H, Li C, Wei B, Han J and Li G 2025 Graph fractional Hilbert transform: theory and application (arXiv:2509.16910)
[65] Wei D and Yun S 2024 Hermitian random walk graph Fourier transform for directed graphs and its applications *Digit. Signal Process.* **155** 104751
[66] Chung F 2005 Laplacians and the Cheeger inequality for directed graphs *Ann. Comb.* **9** 1–19
[67] Birman M S and Solomyak M Z 1967 Operator integration, Lipschitz functions and operator ideals *Uspekhi Mat. Nauk* **22** 27–66
[68] Bhatia R 1997 *Matrix Analysis (Graduate Texts in Mathematics)* vol 169 (Springer)
[69] Higham N J 2008 *Functions of Matrices: Theory and Computation* (Springer)
[70] Zheng R, Li J, Wang Y, Luo T and Yu Y 2023 Scatterformer: locally-invariant scattering transformer for patient-independent multispectral detection of epileptiform discharges *Proc. AAAI Conf. on Artificial Intelligence* vol 37 pp 148–58
[71] Tang L and Zhao M 2024 Epileptic seizure detection in neonatal EEG using a multi-band graph neural network model *Appl. Sci.* **14** 9712
[72] Hogan R, Mathieson S R, Luca A, Ventura S, Griffin S, Boylan G B and O'Toole J M 2025 Scaling convolutional neural networks achieves expert level seizure detection in neonatal EEG *npj Digit. Med.* **8** 17
[73] Webb L, Kauppila M, Roberts J A, Vanhatalo S and Stevenson N J 2021 Automated detection of artefacts in neonatal EEG with residual neural networks *Comput. Methods Programs Biomed.* **208** 106194
[74] Gomez-Quintana S, O'Shea A, Factor A, Popovici E and Temko A 2022 A method for AI assisted human interpretation of neonatal EEG *Sci. Rep.* **12** 10932
[75] Wang Y, Yuan S, Liu J-X, Hu W, Jia Q and Xu F 2024 Combining EEG features and convolutional autoencoder for neonatal seizure detection *Int. J. Neural Syst.* **34** 2450040
[76] Gramacki A and Gramacki J 2022 A deep learning framework for epileptic seizure detection based on neonatal EEG signals *Sci. Rep.* **12** 13010
[77] O'Leary S *et al* 2024 Integrating data across oscillatory power bands predicts the epileptogenic zone: the frequency range explorer epileptogenic zone (FREEZ) identification algorithm *bioRxiv Preprint* pp 2024–05
[78] Thomas J *et al* 2024 Development of a stereo-EEG based seizure matching system for clinical decision making in epilepsy surgery *J. Neural Eng.* **21** 056025
[79] Bernabei J M *et al* 2022 Normative intracranial EEG maps epileptogenic tissues in focal epilepsy *Brain* **145** 1949–61
[80] Conrad E C *et al* 2024 Interictal intracranial EEG asymmetry lateralizes temporal lobe epilepsy *Brain Commun.* **6** fcae284
[81] Tapani K T, Vanhatalo S and Stevenson N J 2019 Time-varying EEG correlations improve automated neonatal seizure detection *Int. J. Neural Syst.* **29** 1850030
[82] O'Shea A, Lightbody G, Boylan G and Temko A 2020 Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture *Neural Netw.* **123** 12–25
[83] Diykh M, Miften F S, Abdulla S, Deo R C, Siuly S, Green J H and Oudahb A Y 2022 Texture analysis based graph approach for automatic detection of neonatal seizure from multi-channel EEG signals *Measurement* **190** 110731
[84] Zhou W, Zheng W, Feng Y and Li X 2024 Lma-eegnet: a lightweight multi-attention network for neonatal seizure detection using EEG signals *Electronics* **13** 2354
[85] Nelson M, Rajendran S, Khalaf O I and Hamam H 2024 Deep-learning-based intelligent neonatal seizure identification using spatial and spectral gnn optimized with the aquila algorithm *AIMS Math.* **9** 19645–69
[86] Yu Y *et al* 2021 Interneuron origin and molecular diversity in the human fetal brain *Nat. Neurosci.* **24** 1745–56
[87] Scheid B H *et al* 2022 Intracranial electroencephalographic biomarker predicts effective responsive neurostimulation for epilepsy prior to treatment *Epilepsia* **63** 652–62

[88] Bernabei J M 2021 Quantitative methods for guiding Epilepsy surgery from intracranial EEG *PhD Thesis* (University of Pennsylvania)
[89] Tyner K 2024 Multimodal Investigations for the Identification of Surgically Relevant Brain Areas *Ph.D. Dissertation* (University of Nebraska Medical Center),
[90] Scheid B H *et al* 2021 Synchronizability predicts effective responsive neurostimulation for epilepsy prior to treatment *medRxiv Preprint* pp 2021–02
[91] Sheybani L, Vivekananda U, Sinha N, Conrad E, Litt B, Burgess N, Bush D and Walker M C 2025 Dynamic interplay between wake slow waves and epileptiform discharges in the epileptogenic zone *Neurology* **105** e214001
[92] Shen Y, Shen Z, Huang Y, Wu Z, Ma Y, Hu F and Shu K 2025 Multimodal quantitative analysis guides precise preoperative localization of epilepsy *J. Neurol.* **272** 1–15
[93] Millán A P, van Straaten E C W, Stam C J, Nissen I A, Idema S, Van Mieghem P and Hillebrand A 2024 Individualized epidemic spreading models predict epilepsy surgery outcomes: a pseudo-prospective study *Netw. Neurosci.* **8** 437–65
[94] Janca R, Jahodova A, Hlinka J, Jezdik P, Svobodova L, Kudr M, Kalina A, Marusic P, Krsek P and Jiruska P 2021 Ictal gamma-band interactions localize ictogenic nodes of the epileptic network in focal cortical dysplasia *Clin. Neurophysiol.* **132** 1927–36
[95] Bernabei J M, Li A, Revell A Y, Smith R J, Gunnarsdottir K M, Ong I Z, Davis K A, Sinha N, Sarma S and Litt B 2023 iEEG Epilepsy Dataset *OpenNeuro* (https://doi.org/10.18112/openneuro.ds004100.v1.1.3)