
FAPEX: Fractional Amplitude-Phase Expressor for Robust Cross-Subject Seizure Prediction

Ruizhe Zheng^{1§}

Research Institute of Intelligent Complex Systems, Fudan University
rzzheng23@m.fudan.edu.cn

Lingyan Mao^{2§}

Department of Neurology, Zhongshan Hospital, Fudan University
lingyanmao@fudan.edu.cn

Tian Luo⁴

Children’s Hospital of Fudan University
tianluo@fudan.edu.cn

Yi Wang⁴

Children’s Hospital of Fudan University
yiwang@shmu.edu.cn

Dingding Han^{3*}

School of Information Science and Technology, Fudan University
ddhan@fudan.edu.cn

Jing Ding^{2*}

Zhongshan Hospital, Fudan University
jingding@zs-hospital.sh.cn

Yuguo Yu^{1*}

State Key Laboratory of Brain Function and Disorders and MOE Frontiers Center for Brain Science, Research Institute of Intelligent Complex Systems and Institutes of Brain Science, Fudan University, Shanghai Artificial Intelligence Laboratory Shanghai 200232, China
yuyuguo@fudan.edu.cn

Abstract

Precise, generalizable subject-agnostic seizure prediction (SASP) remains a fundamental challenge due to the intrinsic complexity and significant spectral variability of electrophysiological signals across individuals and recording modalities. We propose FAPEX, a novel architecture that introduces a learnable *fractional neural frame operator* (FrNFO) for adaptive time–frequency decomposition. Unlike conventional models that exhibit spectral bias toward low frequencies, our FrNFO employs fractional-order convolutions to capture both high and low-frequency dynamics, achieving approximately 10% improvement in F1-score and sensitivity over state-of-the-art baselines. The FrNFO enables the extraction of *instantaneous phase and amplitude representations* that are particularly informative for preictal biomarker discovery and enhance out-of-distribution generalization. FAPEX further integrates structural state-space modeling and channelwise attention, allowing it to handle heterogeneous electrode montages. Evaluated across 12 benchmarks spanning species (human, rat, dog, macaque) and modalities (Scalp-EEG, SEEG,

*Corresponding authors.

ECoG, LFP), FAPEX consistently outperforms 23 supervised and 10 self-supervised baselines under nested cross-validation, with gains of up to 15% in sensitivity on complex cross-domain scenarios. It further demonstrates superior performance in several external validation cohorts. To our knowledge, these establish FAPEX as the first epilepsy model to show consistent superiority in SASP, offering a promising solution for discovering epileptic biomarker evidence supporting the existence of a distinct and identifiable preictal state for and clinical translation.

1 Introduction

Epilepsy is a common, heterogeneous set of neurological disorders characterized by recurrent, hypersynchronous discharges that disrupt normal cognition and behavior. Affecting over 50 million people worldwide [57], its diagnosis and monitoring rely fundamentally on electrophysiological recordings—whether invasive (e.g., electrocorticography (ECoG), stereo-electroencephalography (SEEG), local field potential (LFP)) or non-invasive (scalp EEG) [38, 17]. Although seizures have long been viewed as abrupt and unpredictable events, a growing body of work demonstrates the existence of a preictal stage marked by subtle neural and behavioral changes, offering an actionable window for intervention.

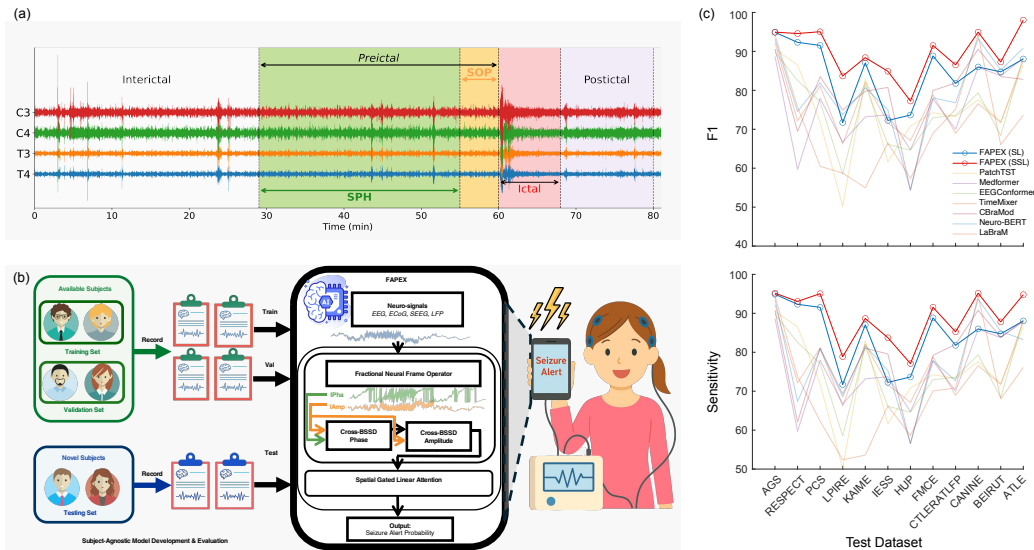


Figure 1: Summary of our work. (a) Definition of the different brain activity stages for the predictive analysis of epileptic seizure (a scalp-EEG record is shown for the purpose of illustration). (b) Overview of the FAPEX development and validation pipeline. (c) Comparative evaluation results demonstrate that FAPEX consistently outperforms state-of-the-art (SOTA) supervised and self-supervised approaches across 12 diverse benchmarks in terms of F1 and sensitivity, demonstrating superior performance and generalization.

Seizure prediction systems seek to detect these preictal alterations and raise alarms sufficiently in advance. As shown in Fig. 1 (a), within the established framework of ictogenesis - which delineates interictal, ictal, and postictal phases - the preictal interval offers a crucial target for clinical interventions ranging from simple alerts aimed at mitigating injury risk to sophisticated closed-loop neuromodulation devices. To formalize practical deployment, seizure prediction systems are typically evaluated with respect to two time parameters: the Seizure Prediction Horizon (SPH), which defines the minimum interval between a raised alarm and seizure onset to allow meaningful intervention, and the Seizure Occurrence Period (SOP), a predefined window during which a seizure is expected following an alarm.

Why subject-agnostic seizure prediction (SASP)? Despite remarkable advances in seizure prediction achieved by pioneering studies [61, 39, 73, 9, 15, 74, 49], the field remains constrained by two fundamental limitations: the reliance on subject-specific modeling paradigms and limited scalability. Subject-specific approaches, while often achieving impressive performance on individual patients, require extensive labeled data collection for each new patient and cannot leverage knowledge across diverse patient populations. This impedes large-scale clinical adoption and negates the potential advantages of aggregating data to identify generalizable seizure biomarkers.

Beyond subject specificity, additional obstacles include narrow EEG modality ranges [69, 39, 72], inconsistent preprocessing pipelines [8, 23], and dependence on rigid electrode configurations. Together, these factors highlight the urgent need for truly subject-agnostic predictive algorithms capable of operating robustly across various patient and recording configurations. Specific challenges include:

(1) Capturing refined high- and low-frequency biomarkers. Clinical evidence [46, 14, 27, 50, 47] shows pathological high-frequency oscillations and low-frequency fluctuations serve as crucial epileptogenesis biomarkers. These subtle, non-stationary features are easily obscured by artifacts. Conventional CNNs [4] and Transformers [36, 53] exhibit spectral biases toward low frequencies, struggling to preserve transient HFO signatures [63, 40].

(2) Modeling phase–amplitude interactions. Epilepsy exhibits abnormal phase-amplitude coupling and (de)synchronization [1, 35, 26, 2, 76, 64] during seizure initiation and propagation across frequency bands. These interactions provide critical clues for distinguishing ictal from interictal states. Current clinical models typically utilize amplitude information in time or frequency domains separately, rarely integrating both. Our architecture captures these fundamental aspects of neural oscillations, leveraging their complementary insights into seizure evolution.

(3) Handling heterogeneous channel layouts.

Seizure onset zones and preictal activity distribution vary significantly across individuals, with predictive features appearing on different electrodes from patient to patient. Implantation strategies, montage configurations, and regional coverage introduce further variability in channel characteristics. Naively pooling signals across channels obscures patient-specific biomarkers and amplifies noise, compromising generalization in subject-agnostic contexts.

Present work. To overcome these challenges, we propose FAPEX, a unified

model for effective generalization across heterogeneous EEG settings, electrode configurations, and clinical subtypes. Our approach integrates three key innovations: (1) *fractional neural frame operator* (FrNFO): a learnable bank of Weyl-Heisenberg filters for adaptive time-frequency decomposition. FrNFO extracts high-fidelity features through fractional-order convolutions with minimal spectral leakage, capturing both high and low-frequency components of epileptic signals. (2) *amplitude-phase cross-encoding* (APCE): A bidirectional state-space architecture processing phase and amplitude representations, learning time-varying relationships to extract seizure evolution patterns. (3) *Spatial correlation aggregation* (SCA): Channel-wise attention mechanisms modeling inter-electrode dependencies to identify predictive spatial patterns. Together, these components enable FAPEX to learn multi-scale representations that capture phase-amplitude coupling while handling non-stationarity and channel heterogeneity. **Together**, these components enable FAPEX to learn rich, multi-scale representations that capture subtle changes in phase-amplitude coupling across frequencies while adaptively handling non-stationarity and channel heterogeneity. As illustrated in Fig. 2, FrNFO serves as the foundational component, addressing low-frequency bias by preserving fragile yet critical high-frequency oscillations while providing fine-grained decomposition of amplitude and phase features for a more comprehensive picture of neural activity in seizure prediction.

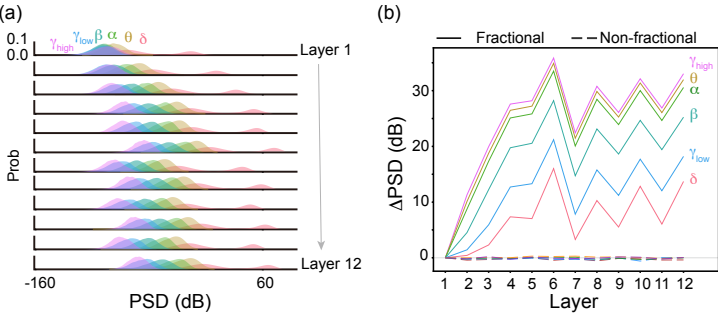


Figure 2: Interpretability of FAPEX. (a) Kernel density estimates of power spectral density (PSD) responses for FrNFO filters across layers and brain frequency subbands. As depth increases, the operator progressively refines its discrimination among subbands, maintaining the natural low-frequency, high-energy and high-frequency, low-energy distribution, with energy gradually stabilizing after intermediate layers. (b) Layer-wise frequency-specific gain relative to the initial layer. Unlike non-fractional operators, FrNFO consistently amplifies both low- and high-frequency components, achieving balanced cross-frequency representations, indicating its ability to capture both fast and slow neural dynamics essential for seizure prediction.

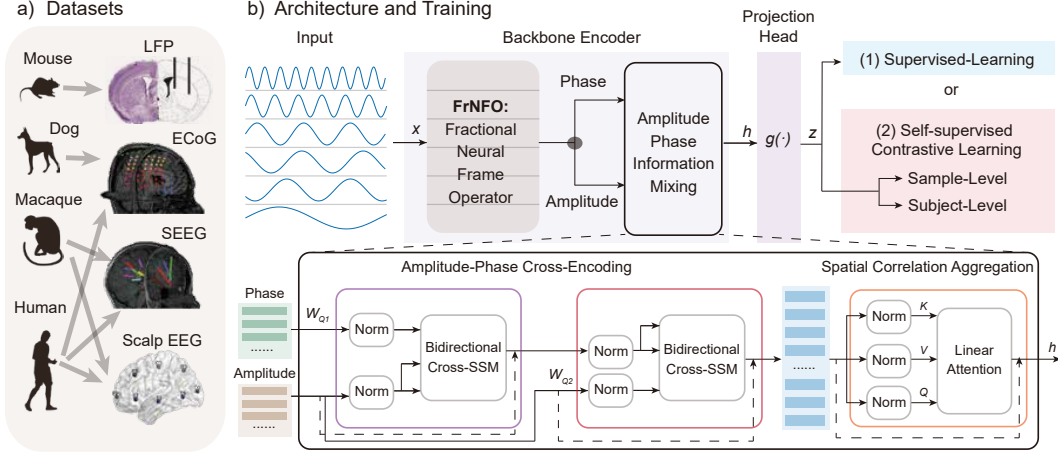


Figure 3: Datasets and network architecture summarization. (a) We used LFP, ECoG, SEEG and Scalp EEG data across species (humans, dogs, rats, and macaques), to validate our model. (b) The network structure and training pipeline of our FAPEX framework. The input signals will be encoded by the backbone encoder that is consisted of our FrNFO, naturally separated into phase and amplitude sections, then go through a Amplitude-Phase information mixing procedure which deals with the two sections interactively using 2 bidirectional Cross-SSM modules, and use a linear attention module for spatial correlation aggregation.

Our main innovations are: **(1)** FAPEX, a subject-agnostic framework integrating our novel Fractional Neural Frame Operator (FrNFO), amplitude-phase cross-encoding, and spatial correlation aggregation to anticipate seizures across diverse modalities. **(2)** FrNFO, a learnable bank of Weyl–Heisenberg filters that performs fractional-order time-frequency decomposition, mitigates low-frequency bias and preserves high-frequency oscillations with provable robustness. **(3)** Extensive validation on 12 benchmarks across species and recording modalities shows FAPEX consistently outperforms 32 baselines, establishing a new standard for seizure prediction and revealing meaningful preictal biomarkers.

2 Method

Problem formulation. Epileptologists classify seizure dynamics into three phases: interictal, preictal, and ictal. **Interictal Phase:** Periods between seizures (typically > 30 minutes) with generally normal brain activity, though occasional interictal epileptiform discharges may occur. **Preictal Phase:** The period preceding a seizure, marked by subtle brain activity changes that may predict an impending seizure. **Ictal Phase:** The seizure event itself, characterized by ictal epileptiform discharges. Formally, a neuroelectrical segment is a set of time series $\{\mathbf{x}^{(i)}\}_{i=1}^C$, where C is the number of channels, and each $\mathbf{x}^{(i)} \in \mathbb{R}^T$ represents a channel with T timestamps. A seizure predictor constructs a function f_{model} that maps $\{\mathbf{x}^{(i)}\}_{i=1}^C$ to a binary label \hat{y}_i , distinguishing interictal from preictal states. The model is trained to align predictions with clinical annotations y_i , enabling prediction in unseen subjects.

Input patchifying. Given a multichannel neural signal segment $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C represents the number of electrode channels and T is the total number of time samples, we aim to establish a preprocessing pipeline that is robust to variations in electrode count and placement. To this end, the continuous data is first partitioned into fixed-duration, non-overlapping patches. Specifically, each channel signal $\mathbf{x}_c \in \mathbb{R}^T$ is segmented using a window of length τ , resulting in $N = \lfloor \frac{T}{\tau} \rfloor$ patches per channel. Each patch is then projected into a common feature space using a channel-shared linear embedding $\mathbf{W} \in \mathbb{R}^{d_{\text{model}} \times \tau}$ and bias vector $\mathbf{b} \in \mathbb{R}^{d_{\text{model}}}$, resulting in the embedded tensor $\tilde{\mathbf{X}} \in \mathbb{R}^{C \times N \times d_{\text{model}}}$ (d_{model} is dimension of the model). This process ensures that the subsequent layers can operate independently of electrode count and spatial arrangement.

2.1 Fractional neural frame operator (FrNFO)

Motivation. Nonstationary signals, such as those encountered in neuroelectrical recordings in epilepsy patients, present significant challenges due to their highly variable time-frequency content

and variability in both amplitude and phase. The *Fractional Fourier Transform (FrFT)* has emerged as a powerful tool for analyzing such signals, providing a flexible, continuous interpolation between the time domain ($\theta = 0$) and the frequency domain ($\theta = \frac{\pi}{2}$) via a fractional order parameter θ . Formally, FrFT generalizes the Fourier transform with a fractional order $\theta \in (0, \pi)$, defined for a signal $f \in L^2(\mathbb{R})$ as:

$$\mathcal{F}_\theta(f)(x) = \frac{1}{\sqrt{|\sin \theta|}} \int_{\mathbb{R}} f(t) \exp[\pi i ((t^2 + x^2) \cot \theta - 2xt \csc \theta)] dt. \quad (1)$$

This transform supports operators like θ -shift, $T_s^\theta f(t) = \exp(2\pi i s(t-s) \cot \theta) f(t-s)$, and θ -modulation, $M_s^\theta f(t) = \exp(\pi i (s^2 \cot \theta + 2st \csc \theta)) f(t)$, enabling θ -fractional convolution [3, 71]:

$$(f \star_\theta g)(x) = \frac{1}{\sqrt{|\sin \theta|}} \int_{\mathbb{R}} f(s)(T_s^\theta g)(x) ds, \quad (2)$$

which offers an alternative to traditional Fourier transform-based convolutions. While the fractional Fourier transform (FrFT) provides a flexible framework for interpolating between time and frequency domains, its practical implementations face two critical challenges that limit their effectiveness for neuroelectrical signals: **(1) Chirp response constraint:** traditional FrFT relies on a fixed chirp function, imposing a globally isotropic structure that poorly adapts to the diverse, localized frequency characteristics of real-world data [22]. This restricts FrFT's expressiveness essential for nuanced phase-amplitude representation. While recent methods have introduced trainable fractional orders [68, 25], they inherit this fundamental limitation, lacking the flexibility to accommodate rapid spectral transitions and localized nonstationarities. **(2) Deformation Sensitivity:** Despite its adaptability in fractional order, FrFT remains sensitive to small deformations, including time shifts, scaling variations, and localized perturbations, which are especially prevalent in neural signals [42, 29]. These limitations underscore the need for more expressive, adaptive frameworks that can capture the intricate amplitude-phase representation.

A neural approach for fine-grained amplitude-phase representation. To overcome these limitations, we propose the *fractional neural frame operator*, which integrates neural implicit representations to learn a parameterization of θ -fractional version of nonstationary Weyl-Heisenberg frame [20, 45, 18], defined as:

$$\Psi_\theta = \left\{ M_{lp_0^{(j)}}^\theta T_{sq_0^{(j)}}^\theta I\Phi_j : s \in \mathbb{R}, l \in \mathbb{Z}, j \in \{1, \dots, N\} \right\}, \quad (3)$$

where p_0^j, q_0^j are positive constants adjusting the scale. It involves θ -modulation $M_{lp_0^{(j)}}^\theta(t) = e^{\pi i ((lp_0^{(j)})^2 \cot \theta + 2lp_0^{(j)} t \csc \theta)}$ and θ -shift $T_{sq_0^{(j)}}^\theta(t) = e^{2\pi i sq_0^{(j)}(t-sq_0^{(j)}) \cot \theta} \Phi_j(t-sq_0^{(j)})$. Ψ_θ presents a redundant set of basis functions that can be used to represent or analyze a signal on the fractional domain. Unlike FrFT, it is equipped with adaptive windows Ψ_j over each scale j to capture a wide range of signal behaviors. Building upon this, we propose fractional neural frame operator.

The core of the FrNFO is an implicit multilayer perceptron (MLP) [34, 62] designed to generate adaptive window function for the frame filters. Given temporal samples N and feature channels d_{model} , the implicit MLP defines the window kernel $\Phi \in \mathbb{C}^{N \times d_{\text{model}}}$ for $j = 1, \dots, N$, $k = 1, \dots, d_{\text{model}}$ as

$$\Phi^{j,k}(t_j) = \left(\sum_{i=1}^M w_{i,k} \exp(-j(b_{i,k} t_j + c_{i,k})) \right) \cdot \left(\sum_{n=0}^K a_{n,k} H_n(t_j) \right), \quad (4)$$

where $w_{i,k}, b_{i,k}, c_{i,k}, a_{n,k}$ are trainable parameters optimized through gradient descent. The basis functions $H_n(t) = (-1)^n e^{t^2} \frac{d^n}{dt^n} e^{-t^2}$ are Hermite polynomials, embedding prior knowledge of localized oscillatory behavior, while the sine activation functions promote smooth and periodic kernel characteristics essential for identifying quasiperiodic activities in brain.

FrNFO further introduces a learnable fractional order $\theta = [\theta_1, \dots, \theta_{d_{\text{model}}}] \in (0, \pi)^{d_{\text{model}}}$, which governs the time-frequency representation for each feature channel independently. Given an input neural embedding $\mathbf{X} \in \mathbb{C}^{N \times d_{\text{model}}}$, employing the fractional convolution theorem [3, 71], the output feature for channel k is defined as:

$$\hat{\mathbf{X}}_{:,k} = \exp(-\pi i \omega^2 \cot \theta_k) \odot \mathcal{F}_{\theta_k}(\mathbf{X}_{:,k}) \odot \mathcal{F}_{\theta_k}(\Psi_{:,k}), \quad k = 1, \dots, d_{\text{model}}, \quad (5)$$

where $\Psi_{:,k}$ is the frame filter kernel equipped with learnable window kernel, \odot denotes the Hadamard product, and ω represents the frequency grid. The phase adjustment factor $\exp(\pi i \omega^2 \cot \theta_k)$ ensures proper alignment and interpretation of fractional frequency components. This adaptive formulation allows FrNFO to dynamically adjust frequency resolution.

FrNFO is a provably robust amplitude representator. As previously formulated, as a neural fractional-order filterbank, FrNFO naturally yields complex-valued signal representation that can be easily formulated into phases and amplitudes across different scales and fractional orders. We further highlight that it also provides a provably robust amplitude representation, which is the main information source in many applications, from the perspective of scattering transform. Refer to further discussion and proof in App. A.

2.2 Amplitude-phase encoding

Amplitude-phase cross encoding (APCE). We introduce APCE to capture heterogeneous, cross-frequency dependencies between amplitude and phase embeddings produced by FrNFO. Inspired by recent advances in selective state space model, proposed first in Mamba [? ? ? ?], we adopt a bidirectional state-space mechanism building on Mamba blocks with cross-attention-like mechanism [59], as shown in Fig. 3. Formally, given amplitude embeddings \mathbf{Amp} and phase embeddings \mathbf{Pha} , we normalize them as: $\mathbf{Amp} = \text{RMSNorm}(\mathbf{Amp})$, $\mathbf{Pha} = \text{RMSNorm}(\mathbf{Pha})$. These normalized embeddings are then processed by the dual cross-Mamba module, which operates in a channel-independent manner to capture amplitude-phase interactions using a bidirectional state-space model (BSSM), comprising two sequential blocks: *phase BSSM* and *amplitude BSSM*. In the phase BSSM block, the normalized phase embeddings $\mathbf{Pha} \in \mathbb{R}^{B \times M \times D}$ are projected into a latent space via two shared linear mappings:

$$\mathbf{X}^P = \mathbf{W}^x \mathbf{Pha}, \quad \mathbf{Z}^P = \mathbf{W}^z \mathbf{Pha}, \quad (6)$$

where $\mathbf{W}^x, \mathbf{W}^z \in \mathbb{R}^{D \times E}$ are learnable projection matrices, and E denotes the number of latent SSM states. The projected embeddings undergo causal and anti-causal convolutions followed by a SiLU activation:

$$\mathbf{X}_o = \text{SiLU}(\text{Conv1d}_o(\mathbf{X}^P)), \quad o \in \{\text{forward}, \text{backward}\}. \quad (7)$$

Using the normalized amplitude embeddings \mathbf{Amp} , we compute state-space parameters:

$$\mathbf{B}_o = \mathbf{W}^B \mathbf{Amp}, \quad \mathbf{C}_o = \mathbf{W}^C \mathbf{Amp}, \quad \mathbf{\Delta}_o = \text{Softplus}(\mathbf{W}^\Delta \mathbf{X}_o + \mathbf{b}^\Delta), \quad (8)$$

where $\mathbf{W}^B, \mathbf{W}^C \in \mathbb{R}^{D \times N}$ are shared across directions, and $\mathbf{W}^\Delta \in \mathbb{R}^{E \times E}$, $\mathbf{b}^\Delta \in \mathbb{R}^E$ are shared scaling parameters. The time-varying transition parameters are then defined as:

$$\bar{\mathbf{A}}_o = \mathbf{\Delta}_o \otimes \mathbf{A}, \quad \bar{\mathbf{B}}_o = \mathbf{\Delta}_o \otimes \mathbf{B}_o, \quad (9)$$

where $\mathbf{A} \in \mathbb{R}^{E \times N}$ is a shared, direction-agnostic transition matrix, and \otimes denotes element-wise multiplication. The output sequence is computed using the SSM kernel:

$$\mathbf{Y}_o = \text{SSM}(\bar{\mathbf{A}}_o, \bar{\mathbf{B}}_o, \mathbf{C}_o)(\mathbf{X}_o). \quad (10)$$

The final phase-to-amplitude representation, which captures phase-informative patterns, is gated as:

$$\mathbf{Y}^P = (\mathbf{Y}_{\text{forward}} + \mathbf{Y}_{\text{backward}}) \odot \text{SiLU}(\mathbf{Z}^P). \quad (11)$$

In the amplitude BSSM, the roles are swapped: the phase-informative \mathbf{Y}^P provides the context, and the amplitude \mathbf{Amp} serves as queries. A residual connection combines the block output \mathbf{Y}^A with the original amplitude embeddings \mathbf{Amp} to produce the final APCE encoding:

$$\tilde{\mathbf{X}} = \mathbf{Y}^A + \mathbf{Amp}, \quad \tilde{\mathbf{X}} \in \mathbb{R}^{B \times M \times D}. \quad (12)$$

Spatial correlation aggregation (SCA). During the preictal interval, epilepsy is marked by dynamic shifts in inter-electrode interdependencies that reflect the spread of pathological activity across brain regions. Accurate seizure forecasting from multichannel recordings therefore hinges on modeling these spatial dependencies. To this end, given neuroelectrical embeddings $\mathbf{X} \in \mathbb{R}^{C \times N \times d}$, SCA models global cross-spatial dependencies of different electrodes while integrating local spatiotemporal

patterns. Formally, linear attention aims to use $\phi(\mathbf{q}_i)\phi(\mathbf{k}_j)^\top$ to approximate softmax attention kernel at linear complexity, where the feature map $\phi(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^d$ is applied row-wise to the query and key matrices. As a result, the c -th row of attention output $\mathbf{a}_t \in \mathbb{R}^d$ can be rewritten as

$$\mathbf{o}_c = \mathbf{a}_c \odot \text{Sigmoid}(\mathbf{g}_c), \quad \mathbf{a}_c = \frac{\sum_{i=1}^C \phi(\mathbf{q}_c)\phi(\mathbf{k}_i)^\top \mathbf{v}_i}{\sum_{j=1}^C \phi(\mathbf{q}_c)\phi(\mathbf{k}_j)^\top} = \frac{\phi(\mathbf{q}_c) \sum_{i=1}^C \phi(\mathbf{k}_i)^\top \mathbf{v}_i}{\phi(\mathbf{q}_c) \sum_{j=1}^C \phi(\mathbf{k}_j)^\top}, \quad (13)$$

where \mathbf{g}_c is the c -th row of $\mathbf{G} := \text{RMSNorm}[\text{DepthwiseConv2d}(\mathbf{X})]$ implemented with a 3×3 depthwise convolutional kernel to aggregate neighborhood spatiotemporal information with RMSNorm to improve stability. The feature map ϕ is made as a one-layer MLP as $\phi_{\text{MLP}}(\mathbf{x}) := \exp(\mathbf{W}_1^\top \mathbf{x})$, where the matrix $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d \times d}$.

3 Experiments

We conducted empirical investigations to address the following **Research Questions**: **RQ1**: How does FAPEX perform in SASP relative to supervised baselines? **RQ2**: Does self-supervised pretraining improve performance of FAPEX in SASP relative to self-supervised baselines? **RQ3**: How well does FAPEX generalize to different cohorts (*e.g.*, species, institution)? **RQ4**: What is the contribution of each design choice within FAPEX?

3.1 Experimental settings

We evaluate FAPEX across diverse settings spanning supervised learning (**RQ1**), self-supervised pretraining-finetuning (**RQ2**), and cross-cohort transfer (**RQ3**). This section outlines the baseline, evaluation protocols, and other basic implementation setups common to all experiments. See details of training protocols in App. G. Full implementation details are provided in App. H.

Datasets. We compile 12 benchmarking datasets spanning four species (human, rat, dog, macaque) and multiple acquisition modalities (Scalp-EEG, ECoG, SEEG, LFP) for evaluation, as summarized in Tab. 1. All recordings are resampled and segmented to standardized lengths, then harmonized to 64 effective channels via channel rejection and duplication, enabling consistent input formatting across all models. See detailed descriptions and preprocessing procedures in App. F. Note that we apply channel alignment during preprocessing to facilitate consistent training across diverse datasets for both our model and a broad range of baselines. In short, FAPEX itself is inherently agnostic to the number and configuration of input channels.

Table 1: Summary of datasets. The datasets span several species (human, rat, dog, macaque) and acquisition modalities (Scalp-EEG, ECoG, SEEG, LFP).

Dataset	Confidentiality	Species	# Subj.	Modality	# Ch.	# Samples	Duration	SOP	SPH	ID/IV	OOD/EV
FMCE	Public	Human	65	ECoG/SEEG ¹	64	32,323	4 s	30 s	1 min	✓	✗
HUP	Public	Human	73	ECoG/SEEG	64	53,323	4 s	30 s	5 min	✓	✗
RESPECT	Public	Human	6	ECoG	64	17,214	4 s	30 s	5 min	✓	✗
BEIRUT	Public	Human	6	Scalp-EEG	64	35,941	4 s	1 min	30 min	✓	✓
CTLE-RATLFP	Public	Rat	7	LFP	64	11,732	2 s	30 s	5 min	✓	✗
LPIRE	Public	Rat	15	LFP	64	159,715	2 s	30 s	5 min	✓	✓
CANINE	Public	Dog	6	ECoG	64	382,278	4 s	5 min	4 hr	✓	✓
ATLE	Private	Human	5	Scalp-EEG	64	11,536	4 s	5 min	30 min	✓	✗
AGS	Private	Human	5	Scalp-EEG	64	32,323	4 s	5 min	30 min	✓	✓
IESS	Private	Human	17	Scalp-EEG	64	48,986	4 s	5 min	30 min	✓	✓
KAIME	Private	Macaque	3	Scalp-EEG & SEEG ²	64	36,092	4 s	5 min	30 min	✓	✓
PCS	Private	Human	5	Scalp-EEG	64	29,679	4 s	5 min	30 min	✓	✗
TUEG	Public	Human	14,987	Scalp-EEG	64	1,030,090	32 s	Used for Pretraining Only			
CCEP	Public	Human	74	ECoG	64	52,337	32 s	Used for Pretraining Only			
PPE	Public	Human	30	Scalp-CEEP	64	13,434	32 s	Used for Pretraining Only			

¹ For “ECoG/SEEG” datasets each subject has *either* sub-dural ECoG grids/strips *or* SEEG depth electrodes, never both.

² KAIME comprises simultaneous scalp-EEG and SEEG depth recordings from three adult rhesus macaques (*Macaca mulatta*).

Baselines. We compare our method against the following baselines, including 22 supervised baselines for **RQ1** and 10 self-supervised ones for **RQ2**. The supervised baselines include (1) *Convolutional models* (5 baselines): ModernTCN [10], MRConv [7], MultiresNet [43], Omni-Scale [48], and SPaRCNet [21]; (2) *Token mixers* (6 baselines): EEGConformer [44], iTransformer [32], Nonformer [31], PatchTST [37], Pathformer [6], and SeizureFormer [13]; (3) *Time-frequency mixers* (4 baselines): ATFNet [66], FrETS [67], NFM [24], and TSLANet [12]; (4) *Multiscale token mixers* (7 baselines): AdaWaveNet [?], Medformer [55], MTST [75], Pyraformer [30], SimpleTM [5], TimesNet [60], and TimeMixer [54]. Self-supervised baselines include 6 *non-contrastive* models:

Table 3: Median Performance Across In-House Datasets. Top-1, Top-2, and Top-3 results are highlighted in red, blue, and green, respectively, within both supervised (SL) and self-supervised (SSL) groups. FAPEX demonstrates consistently strong performance, achieving top-1 TO 3 rankings on the majority of datasets and metrics, reflecting its generalization and adaptability. For detailed results and statistical analysis, refer to App. C.

Model	AGS			ATLE			IESS			KAIME			PCS		
	SEN	FI	ROC	SEN	FI	ROC	SEN	FI	ROC	SEN	FI	ROC	SEN	FI	ROC
ModernTCN	87.0	85.0	93.2	91.7	90.2	100.0	73.4	73.4	67.2	83.4	73.2	87.3	85.9	85.4	86.3
MRConv	91.3	90.3	95.2	86.6	96.1	100.0	68.8	68.7	66.9	81.1	68.5	85.0	83.0	84.1	83.7
MultiresNet	90.1	88.8	96.1	85.4	84.3	100.0	72.1	70.4	68.7	80.4	63.7	82.5	69.2	64.4	83.9
Omni-Scale	91.7	90.9	95.2	87.8	98.6	99.9	67.9	68.7	67.2	81.0	68.8	83.0	80.0	79.6	80.9
SPaRCNet	89.1	87.5	93.4	84.0	81.7	99.8	60.7	64.9	61.4	82.0	77.1	86.5	85.5	84.4	91.0
EEGConformer	89.8	88.5	94.4	88.5	91.2	100.0	66.1	67.9	67.0	81.4	73.4	87.1	77.1	78.8	84.3
EEGMamba	93.8	93.5	96.8	88.2	85.0	100.0	69.6	70.0	68.8	80.4	69.4	83.4	70.8	73.3	85.6
iTransformer	89.5	87.8	95.3	54.9	2.9	99.8	53.4	54.5	66.4	81.3	63.4	87.1	74.3	73.0	83.4
Nonformer	93.2	92.7	96.7	84.7	97.5	99.8	69.7	74.7	68.9	79.5	90.3	81.6	68.8	63.7	84.1
PatchTST	90.5	89.3	95.5	86.6	93.0	100.0	61.6	63.8	67.5	83.0	73.8	88.5	71.9	71.2	73.8
Pathformer	92.5	91.8	96.7	88.7	95.3	100.0	71.3	72.1	68.6	80.6	67.4	85.3	78.7	80.9	83.7
SeizureFormer	92.1	91.3	95.4	86.3	97.9	99.9	69.7	69.9	66.7	77.3	53.4	85.9	58.6	62.4	59.7
ATFNet	85.2	84.1	90.8	83.1	97.0	99.8	59.7	56.2	68.2	65.0	45.7	71.9	74.7	73.5	84.6
FreTS	88.7	87.0	93.0	70.2	70.8	77.8	42.7	32.8	67.0	54.3	56.4	73.8	70.5	72.7	77.8
NFM	88.7	87.0	93.0	71.3	71.8	81.7	52.4	56.2	61.2	76.8	64.4	80.4	73.6	76.9	83.0
TSLANet	94.4	94.2	97.3	91.4	91.6	100.0	73.8	72.9	66.2	82.2	88.4	82.4	84.6	84.0	84.9
AdaWaveNet	89.0	87.6	95.1	82.8	96.6	99.8	70.0	70.7	66.3	70.5	77.7	84.8	72.4	74.2	81.3
MedFormer	88.7	88.0	96.1	88.2	98.9	99.9	73.7	73.1	66.9	73.2	45.2	72.3	77.9	77.1	96.3
MTST	91.6	90.7	98.0	84.1	97.2	99.8	60.3	56.4	70.1	59.0	60.2	74.2	72.8	70.3	74.1
Pyraformer	92.0	91.3	96.5	85.1	97.7	99.8	71.4	70.2	66.9	83.2	56.6	86.4	76.8	79.9	82.3
SimpleTM	85.1	83.5	88.2	90.8	90.4	99.9	66.7	68.8	64.4	80.0	81.1	84.6	76.0	75.0	84.5
TimesNet	89.7	88.3	94.3	82.1	96.3	99.8	59.9	63.6	66.1	80.0	81.1	84.6	77.7	81.0	84.6
TimeMixer	92.3	91.6	96.6	87.9	95.3	100.0	71.7	71.0	68.6	82.1	90.0	85.2	81.1	83.5	85.4
FAPEX-Small (SL)	94.1	93.7	98.4	87.2	98.4	99.9	70.8	70.4	71.7	86.9	92.1	89.3	81.0	81.2	94.1
FAPEX-Base (SL)	94.9	94.6	99.5	88.0	98.8	99.9	72.3	72.4	71.4	87.0	95.6	90.1	91.5	91.5	96.3
Brant	93.2	92.7	96.6	87.9	83.0	99.9	68.0	67.7	69.5	74.5	74.8	74.4	83.1	82.3	95.8
CBraMod	90.8	90.6	98.7	87.9	82.8	99.9	79.6	80.7	76.2	81.0	79.8	83.7	81.1	83.5	85.4
EEGPT	93.4	92.9	98.5	88.2	83.2	100.0	74.2	73.9	71.4	78.4	77.3	78.6	85.5	84.4	91.0
Neuro-BERT	93.7	93.2	96.8	83.3	90.8	100.0	75.3	75.0	71.5	81.7	80.7	83.6	80.8	81.9	96.8
VQ_MTM	87.9	85.5	94.6	81.7	79.5	99.9	72.8	72.9	69.8	62.8	64.7	78.2	81.0	81.2	94.1
COMETS	93.9	93.4	98.2	87.7	83.2	99.8	67.6	68.2	79.3	80.6	80.1	84.0	80.8	81.9	96.8
MF-CLR	91.2	90.0	97.0	84.4	82.8	100.0	79.7	80.8	75.6	80.9	79.8	86.3	79.2	77.4	97.2
TS2Vec	94.7	94.5	97.5	62.7	76.1	73.6	72.4	73.6	72.9	76.1	76.2	76.7	69.0	65.9	96.4
FAPEX-Small (SSL)	94.1	93.7	98.4	94.0	92.8	100.0	81.5	83.7	83.4	87.4	87.1	89.3	91.0	91.0	96.7
FAPEX-Base (SSL)	95.2	94.9	99.7	94.8	98.0	100.0	83.7	84.9	85.9	88.7	88.4	91.4	95.0	95.0	97.5

of FAPEX-Base over Neuro-BERT and CBraMod in median F1. FAPEX-Base consistently achieves positive $\Delta\%$ in F1 across diverse cases. It excels in the SOT setup, with $\Delta\%$ often exceeding 30%, highlighting its strong generalization without target supervision relative to other models. In more informative setups like CDAC and MME, where SOTA models improve with target data, FAPEX-Base still outperforms or matches them in most cases, despite the narrowing gap and occasional dataset-specific underperformance. This resilience underscores its robust architecture and clinical potential in label-scarce settings and adaptivity to different finetuning techniques. See App. C for full results.

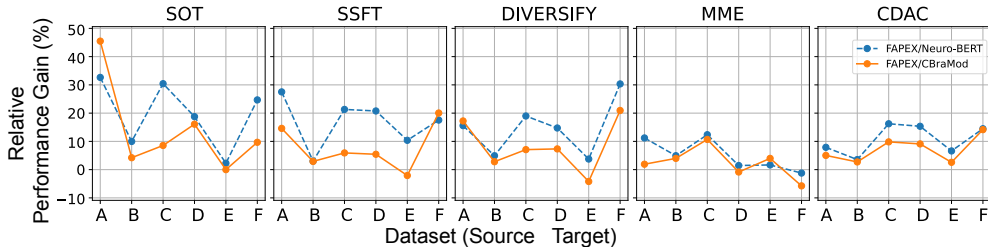


Figure 4: Relative improvement in F1-score medians ($\Delta\%$) of FAPEX-Base over Neuro-BERT and CBraMod across five distinct transfer learning setups for six source-target dataset pairs. FAPEX-Base demonstrates consistent performance gains for most cases, under both weak (SOT) and stronger supervision regimes (CDAC). A: KAIME \rightarrow AGS, B: AGS \rightarrow BEIRUT, C: IESS \rightarrow BEIRUT, D: LPIRE \rightarrow AGS, E: LPIRE \rightarrow IESS, F: LPIRE \rightarrow KAIME). FAPEX-Base consistently achieves superior performance.

Ablation study and further analysis (RQ4). To evaluate the contributions of each component within FAPEX, we conduct comprehensive ablation experiments. These studies isolate the effects of core modules—FrNFO, APCE, and SCA—on seizure prediction performance, providing insights into their individual and collective impacts (see App. D). We further explored the representational

characteristics and interpretability of FAPEX (see App. B). These analyses offer deeper insights into the model’s decision-making processes and its alignment with known neural patterns.

4 Conclusion

We present FAPEX, a compact yet powerful neural architecture that integrates fractional frame theory directly into its core operators. Unlike the trend toward ever-larger models, FAPEX strategically leverages fractional neural frame operators to jointly encode amplitude and phase, achieving provable robustness against deformation and superior preservation of high-frequency biomarkers essential for precise seizure prediction. Extensive evaluations across fully supervised, self-supervised, and multi-cohort, multi-species out-of-domain settings consistently demonstrate that FAPEX surpasses specialized baselines and even large foundation models under comparable data regimes. These results establish FAPEX as a significant step forward in AI for healthcare with strong potential for improving clinical epilepsy management. Future work will aim to expand clinical datasets through collaboration with medical centers, incorporate complementary neuroimaging modalities, and explore deployment on wearable devices and closed-loop neurostimulation systems. Additionally, further theoretical analysis of phase–amplitude disentanglement and interpretability will be prioritized to enhance clinical trust and impact.

5 Acknowledgments

This work was supported by the Science and Technology Innovation 2030 - Brain Science and Brain-Inspired Intelligence Project (Grant No. 2021ZD0201301), the National Key Research and Development Program of China under Contract (Grant No. 2024YFA1610900), the National Natural Science Foundation of China (Grant Nos. 9257020, U20A20221, 12147101), and the Shanghai Municipal Science and Technology Committee of Shanghai Outstanding Academic Leaders Plan (Grant No. 21XD1400400). We thank the Shanghai Institute for Mathematics and Interdisciplinary Sciences (SIMIS)

for financial support (Grant No. SIMIS-ID-2025-NC). The computations were performed on the CFFF platform of Fudan University.

Reference

Bibliography

- [1] M. Amiri, B. Frauscher, and J. Gotman. Phase-amplitude coupling features for seizure detection. *Clinical Neurophysiology*, 127(9):e173, 2016.
- [2] M. Amiri, B. Hatem, and J. Gotman. Phase-amplitude coupling localizes pathologic brain with aid of behavioral staging in sleep. *Journal of Neuroscience*, 40(12):2400–2412, 2020.
- [3] M. H. A. Biswas, P. Massopust, and R. Ramakrishnan. Deep convolutional neural networks and data approximation using the fractional fourier transform. *arXiv preprint arXiv:2408.06757*, 2024. URL <https://arxiv.org/abs/2408.06757>.
- [4] J. O. Caro, Y. Ju, R. Pyle, S. Dey, W. Brendel, F. Anselmi, and A. Patel. Local convolutions cause an implicit bias towards high frequency adversarial examples. *arXiv preprint arXiv:2006.11440*, 2020.
- [5] H. Chen, V. Luong, L. Mukherjee, and V. Singh. SimpleTM: A simple baseline for multivariate time series forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=oANkBaVci5>.
- [6] P. Chen, Y. Zhang, Y. Cheng, Y. Shu, Y. Wang, Q. Wen, B. Yang, and C. Guo. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. In *ICLR*, 2024.
- [7] H. J. Cunningham, G. Giannone, M. Zhang, and M. P. Deisenroth. Reparameterized multi-resolution convolutions for long sequence modelling. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

- [8] A. De, A. Konar, A. Samanta, S. Biswas, and P. Basak. Seizure prediction using low frequency eeg waves from wag/rij rats. In *2017 2nd International Conference for Convergence in Technology (I2CT)*, pages 244–249, 2017. doi: 10.1109/I2CT.2017.8226129.
- [9] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes. Geometric deep learning for subject independent epileptic seizure prediction using scalp eeg signals. *IEEE Journal of Biomedical and Health Informatics*, 26(2):527–538, 2021.
- [10] L. donghao and wang xue. ModernTCN: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*, 2024.
- [11] J. Duan, W. Zheng, Y. Du, W. Wu, H. Jiang, and H. Qi. MF-CLR: Multi-frequency contrastive learning representation for time series. In *Forty-first International Conference on Machine Learning*, 2024. URL <https://openreview.net/forum?id=ec07W0I1MD>.
- [12] E. Eldele, M. Ragab, Z. Chen, M. Wu, and X. Li. Tslanet: Rethinking transformers for time series representation learning. In *International Conference on Machine Learning*, 2024.
- [13] T. Feng, J. Ni, E. Gleichgerricht, and W. Jin. Seizureformer: A transformer model for ica-based seizure risk forecasting, 2025. URL <https://arxiv.org/abs/2504.16098>.
- [14] T. Ferrari-Marinho, P. Perucca, M. Amiri, F. Dubeau, J. Gotman, and L. O. Caboclo. High-frequency oscillations in the scalp eeg of intensive care unit patients with altered level of consciousness. *Journal of Clinical Neurophysiology*, 37(3):246–252, 2020.
- [15] Y. Gao, A. Liu, L. Wang, R. Qian, and X. Chen. A self-interpretable deep learning model for seizure prediction using a multi-scale prototypical part network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1847–1856, 2023.
- [16] H. GUI, X. Li, and X. Chen. Vector quantization pretraining for EEG time series with random projection and phase alignment. In *Forty-first International Conference on Machine Learning*, 2024. URL <https://openreview.net/forum?id=7uWLVFvpis>.
- [17] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993. doi: 10.1103/RevModPhys.65.413.
- [18] D. Jain, M. Kumari, and S. Kumari. A note on fractional gabor systems in l_2 (). *Journal of Function Spaces*, 2024(1):7170692, 2024.
- [19] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu. Large brain model for learning generic representations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024. URL <https://arxiv.org/abs/2405.18765>.
- [20] D. Jindal and L. K. Vashisht. Nonstationary frames of translates and frames from the weyl–heisenberg group and the extended affine group. *Journal of Physics A: Mathematical and Theoretical*, 56(34):345204, 2023.
- [21] J. Jing, W. Ge, S. Hong, M. B. Fernandes, Z. Lin, C. Yang, S. An, A. F. Struck, A. Herlopian, I. Karakis, et al. Development of expert-level classification of seizures and rhythmic and periodic patterns during eeg interpretation. *Neurology*, 100(17):e1750–e1762, 2023.
- [22] M. Kepesi and L. Weruaga. Adaptive chirp-based time–frequency analysis of speech signals. *Speech communication*, 48(5):474–492, 2006.
- [23] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener. Focal onset seizure prediction using convolutional networks. *IEEE Transactions on Biomedical Engineering*, 65(9):2109–2118, 2017.
- [24] M. Kim, Y. Hioka, and M. Witbrock. Neural fourier modelling: A highly compact approach to time-series analysis. *arXiv preprint arXiv:2410.04703*, 2024.
- [25] E. Koç, T. Alikışıfoğlu, A. C. Aras, and A. Koç. Trainable fractional fourier transform. *IEEE Signal Processing Letters*, 31:751–755, 2024.

- [26] K. Lehnertz, S. Bialonski, M.-T. Horstmann, D. Krug, A. Rothkegel, M. Staniek, and T. Wagner. Synchronization phenomena in human epileptic brain. *Journal of Physics: Conference Series*, 186:012013, 2009.
- [27] F. Li, Y. Liang, L. Zhang, C. Yi, Y. Liao, Y. Jiang, Y. Si, Y. Zhang, D. Yao, L. Yu, et al. Transition of brain networks from an interictal to a preictal state preceding a seizure revealed by scalp eeg network analysis. *Cognitive neurodynamics*, 13:175–181, 2019.
- [28] J. Li, G. Li, Y. Shi, and Y. Yu. Cross-domain adaptive clustering for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2505–2514, 2021.
- [29] L. Liu, J. Wu, D. Li, L. Senhadji, and H. Shu. Fractional wavelet scattering network and applications. *IEEE Transactions on Biomedical Engineering*, 66(2):553–563, 2019. doi: 10.1109/TBME.2018.2850356.
- [30] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. *International conference on learning representations*, 2021.
- [31] Y. Liu, H. Wu, J. Wang, and M. Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems*, 35:9881–9893, 2022.
- [32] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long. itransformer: Inverted transformers are effective for time series forecasting. In *ICLR*, 2024.
- [33] W. Lu, J. Wang, X. Sun, Y. Chen, and X. Xie. Out-of-distribution representation learning for time series classification. *arXiv preprint arXiv:2209.07027*, 2022.
- [34] A. Molaei, A. Aminimehr, A. Tavakoli, A. Kazerouni, B. Azad, R. Azad, and D. Merhof. Implicit neural representation in medical imaging: A comparative survey. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2381–2391, 2023.
- [35] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz. Seizure prediction: The long and winding road. *Brain*, 130(2):314–333, 2007.
- [36] T. Nguyen, T. Nguyen, and R. Baraniuk. Mitigating over-smoothing in transformers via regularized nonlocal functionals. *Advances in Neural Information Processing Systems*, 36: 80233–80256, 2023.
- [37] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *ICLR*, 2023.
- [38] E. Niedermeyer and F. H. L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 4 edition, 1998.
- [39] E. D. Pontes, M. Pinto, F. Lopes, and C. Teixeira. Concept-drifts adaptation for machine learning eeg epilepsy seizure prediction. *Scientific Reports*, 14(1):8204, 2024.
- [40] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of neural networks. In *International conference on machine learning*, pages 5301–5310. PMLR, 2019.
- [41] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8050–8058, 2019.
- [42] J. Shi, Y. Zhao, W. Xiang, V. Monga, X. Liu, and R. Tao. Deep scattering network with fractional wavelet transform. *IEEE Transactions on Signal Processing*, 69:4564–4579, 2021. doi: 10.1109/TSP.2021.3098936.
- [43] J. Shi, K. A. Wang, and E. B. Fox. Sequence modeling with multiresolution convolutional memory. In *International Conference on Machine Learning*, 2023.

- [44] Y. Song, Q. Zheng, B. Liu, and X. Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023.
- [45] M. Speckbacher and P. Balazs. The continuous nonstationary gabor transform on lca groups with applications to representations of the affine weyl-heisenberg group. *arXiv preprint arXiv:1407.6830*, 2014.
- [46] W. Stacey, M. Le Van Quyen, F. Mormann, and A. Schulze-Bonhage. What is the present-day eeg evidence for a preictal state? *Epilepsy research*, 97(3):243–251, 2011.
- [47] C. Stamoulis, L. J. Gruber, D. L. Schomer, and B. S. Chang. High-frequency neuronal network modulations encoded in scalp eeg precede the onset of focal seizures. *Epilepsy & Behavior*, 23(4):471–480, 2012.
- [48] W. Tang, G. Long, L. Liu, T. Zhou, M. Blumenstein, and J. Jiang. Omni-scale cnns: a simple and effective kernel size configuration for time series classification, 2022. URL <https://arxiv.org/abs/2002.10061>.
- [49] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, and O. Kavehei. A generalised seizure prediction with convolutional neural networks for intracranial and scalp electroencephalogram data analysis. *CoRR*, abs/1707.01976, 2017. URL <http://arxiv.org/abs/1707.01976>.
- [50] S. M. Usman, S. Khalid, R. Akhtar, Z. Bortolotto, Z. Bashir, and H. Qiu. Using scalp eeg and intracranial eeg signals for predicting epileptic seizures: Review of available methodologies. *Seizure*, 71:258–269, 2019.
- [51] G. Wang, W. Liu, Y. He, C. Xu, L. Ma, and H. Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [52] J. Wang, S. Zhao, Z. Luo, Y. Zhou, H. Jiang, S. Li, T. Li, and G. Pan. CBramod: A criss-cross brain foundation model for EEG decoding. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=NPNUHgHF2w>.
- [53] P. Wang, W. Zheng, T. Chen, and Z. Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. *arXiv preprint arXiv:2203.05962*, 2022.
- [54] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *ICLR*, 2024.
- [55] Y. Wang, N. Huang, T. Li, Y. Yan, and X. Zhang. Medformer: A multi-granularity patching transformer for medical time-series classification. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=jfkid2HwNr>.
- [56] Y. Wang, Y. Liu, J. Li, and J. Zhou. Comets: Contrastive pre-training of multivariate time series. *arXiv preprint arXiv:2401.12345*, 2024.
- [57] World Health Organization. Epilepsy: Key facts. <https://www.who.int/news-room/fact-sheets/detail/epilepsy>, 2023. Accessed: 2025-04-18.
- [58] D. Wu, S. Li, J. Yang, and M. Sawan. Neuro-bert: Rethinking masked autoencoding for self-supervised neurological pretraining. *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [59] D. Wu, Y. Wang, X. Wu, and T. Qu. Cross-attention inspired selective state space models for target sound extraction, 2024. URL <https://arxiv.org/abs/2409.04803>.
- [60] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *ICLR*, 2023.
- [61] J. Xiang, Y. Li, X. Wu, Y. Dong, X. Wen, and Y. Niu. Synchronization-based graph spatio-temporal attention network for seizure prediction. *Scientific Reports*, 15(1):4080, 2025.

- [62] D. Xu, P. Wang, Y. Jiang, Z. Fan, and Z. Wang. Signal processing for implicit neural representations. *Advances in Neural Information Processing Systems*, 35:13404–13418, 2022.
- [63] Z.-Q. J. Xu, Y. Zhang, and T. Luo. Overview frequency principle/spectral bias in deep learning. *Communications on Applied Mathematics and Computation*, pages 1–38, 2024.
- [64] T. Yanagisawa, R. Fukuma, S. Yamamoto, and H. Kishima. Detection of epileptic seizures using phase-amplitude coupling in intracranial electroencephalography. *Scientific Reports*, 11: 12345, 2021.
- [65] C. Yang, M. Westover, and J. Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36, 2024.
- [66] H. Ye, J. Chen, S. Gong, F. Jiang, T. Zhang, J. Chen, and X. Gao. Atfnet: Adaptive time-frequency ensembled network for long-term time series forecasting. *arXiv preprint arXiv:2404.05192*, 2024.
- [67] K. Yi, Q. Zhang, W. Fan, S. Wang, P. Wang, H. He, N. An, D. Lian, L. Cao, and Z. Niu. Frequency-domain mlps are more effective learners in time series forecasting. In *NeurIPS*, 2023.
- [68] H. Yu, J. Huang, L. Li, M. Zhou, and F. Zhao. Deep fractional fourier transform. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36. Curran Associates, Inc., 2023.
- [69] Z. Yuan, D. Zhang, J. Chen, G. Gu, and Y. Yang. Brant-2: Foundation model for brain signals. *CoRR*, 2024.
- [70] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu. Ts2vec: Towards universal representation of time series. In *AAAI*, 2022.
- [71] A. I. Zayed. A convolution and product theorem for the fractional fourier transform. *IEEE Signal processing letters*, 5(4):101–103, 1998.
- [72] D. Zhang, Z. Yuan, Y. Yang, J. Chen, J. Wang, and Y. Li. Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36:26304–26321, 2023.
- [73] J. Zhang, S. Zheng, W. Chen, G. Du, Q. Fu, and H. Jiang. A scheme combining feature fusion and hybrid deep learning models for epileptic seizure detection and prediction. *Scientific Reports*, 14(1):16916, 2024.
- [74] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo. Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network. *IEEE journal of biomedical and health informatics*, 24(2):465–474, 2019.
- [75] Y. Zhang, L. Ma, S. Pal, Y. Zhang, and M. Coates. Multi-resolution time-series transformer for long-term forecasting. In *International Conference on Artificial Intelligence and Statistics*, pages 4222–4230. PMLR, 2024.
- [76] X. Zhao, Y. Zheng, J. Zhang, X. Li, and Y. Wang. Altered phase-amplitude coupling in epileptic brain networks: A stereo-eeg study. *Brain Communications*, 5(2):fcad089, 2023.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: We made a clear presentation of our main innovation and contributions.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: we discussed it in the supplementary materials.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Yes, we presented it in the methods section and supplementary materials.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We revealed them in the supplementaries.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Although in-house datasets are only available upon request so far, we present extensive experiments on public datasets. Moreover, the code of our work will be released upon publication.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We report these details in the supplementaries.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: See supplementaries for performance estimation and statistical significance test. We report median along with interquartile range. Nonparametric variance analysis is conducted using Friedman test with Iman-Davenport correction, while multiple comparison is conducted using Bayesian Wilcoxon signed-rank test.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We provide detailed information in the appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We are informed of code of ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We provide discussions on limitations, societal impacts, among others.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [\[Yes\]](#)

Justification: We discussed it in the supplementaries.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [\[Yes\]](#)

Justification: We properly cited models and datasets of others with download links provided.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset’s creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes] .

Justification: We provide them in the appendix.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: [TODO]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [Yes]

Justification: We collected data from patients and experimental animals with written consent and approval by human subjects and ethics committees.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA] .

Justification: LLM was used for editing and formatting.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.