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

Anonymous Author(s)

Affiliation

Address

email

1 A Mathematical details

2 In this section, we discuss the mathematical details of our approach and present a proof of the
3 robustness of our amplitude representation towards noise deformations using a generalized scattering
4 operator theory.

5 A.1 Fractional Fourier transform and convolution

6 We discuss fractional convolution operator in connection with the Fractional Fourier Transform
7 (FrFT). The FrFT is a special case of the more general Special Affine Fourier Transform (SAFT),
8 where the parameter matrix is given by $\Lambda_\theta = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \end{bmatrix}$. This corresponds to $a = \cos \theta$,
9 $b = \sin \theta$, $c = -\sin \theta$, $d = \cos \theta$, and $p = q = 0$.

$$\mathcal{F}_A[f](x) := \frac{1}{\sqrt{|b|}} \int_{\mathbb{R}} f(t) e^{\frac{\pi i}{b}(at^2 + 2pt - 2xt + dx^2 + 2(bq - dp)x)} dt, \quad x \in \mathbb{R}, \quad (1)$$

10 where A stands for the set of all real parameters $\{a, b, c, d, p, q\}$ with $b \neq 0$ and $ad - bc = 1$. Taking
11 $A = \{\cos \theta, \sin \theta, -\sin \theta, \cos \theta, 0, 0\}$. This naturally derives FrFT:

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

12 In the following, we present the key definitions and theorems specialized for the FrFT.

13 **Definition A.1** (Chirp Modulation for FrFT). *The chirp modulation functions modulate the input*
14 *signal $f(t)$ with a quadratic phase factor:*

$$1. \text{ Forward modulation: } \mathcal{C}_\theta f(t) = \exp\left(j \frac{\cot \theta}{2} t^2\right) f(t)$$

$$16 \quad 2. \text{ Conjugate modulation: } \check{\mathcal{C}}_\theta f(t) = \exp\left(-j \frac{\cot \theta}{2} t^2\right) f(t)$$

17 where $\cot \theta = \cos \theta / \sin \theta$, and $\sin \theta \neq 0$ is assumed to avoid singularities.

18 **Definition A.2** (Fractional Convolution). *Let $f, g \in L^1(\mathbb{R})$. Then, the fractional convolution of f*
19 *and g is defined by*

$$(f \star_\theta g)(x) := \frac{1}{|\sin \theta|^{1/2}} \int_{\mathbb{R}^n} f(s) T_s^\theta g(x) ds \quad (3)$$

20 Furthermore, one has that

$$\mathcal{C}_\theta (f \star_\theta g) = \frac{1}{|\sin \theta|^{1/2}} (\mathcal{C}_\theta f \star \mathcal{C}_\theta g) \quad (4)$$

21 where

$$f \star g(x) := \int_{\mathbb{R}} f(x-t)g(t)dt$$

22 is the usual convolution of two functions.

23 **Definition A.3** (Fractional Convolution). *Let $f, g \in L^1(\mathbb{R})$. Then, the fractional convolution of f*
24 *and g is defined by*

$$(f \star_\theta g)(x) := \frac{1}{|\sin \theta|^{1/2}} \int_{\mathbb{R}^n} f(s) T_s^\theta g(x) ds \quad (5)$$

25 Furthermore, one has that

$$\mathcal{C}_\theta (f \star_\theta g) = \frac{1}{|\sin \theta|^{1/2}} (\mathcal{C}_\theta f \star \mathcal{C}_\theta g) \quad (6)$$

26 where

$$f \star g(x) := \int_{\mathbb{R}} f(x-t)g(t)dt$$

27 is the usual convolution of two functions.

28 **Theorem A.1** (Fractional Convolution Theorem). *For the Fractional convolution defined in A.3*

$$\mathcal{F}_\theta[f *_\theta g](\omega) = \exp(-j\pi\omega^2 \cot \theta) \widehat{f}_\theta(\omega) \widehat{g}_\theta(\omega) \quad (7)$$

29 where \mathcal{F}_θ denotes the FrFT with angle θ , $\widehat{f}_\theta(\omega) = \mathcal{F}_{\text{FrFT}, \theta}[f](\omega)$ is the FrFT of $f(t)$, and $e^{-j\frac{\cot \theta}{2}\omega^2}$
30 is the phase factor.

31 **Theorem A.2** (FrFT Product Theorem). *For a product signal $h(t) = e^{j\frac{\cot \theta}{2}t^2} f(t)g(t)$:*

$$\mathcal{F}_\theta[h](\omega) = (\widehat{f} *_{-\theta} \widehat{g})(\omega)$$

32 where $\widehat{f} = \mathcal{F}_{\text{FrFT}, \theta}[f]$, $\widehat{g} = \mathcal{F}_{\text{FrFT}, \theta}[g]$, $*_{-\theta}$ is the fractional convolution with angle $-\theta$.

33 A.2 Fractional frame and scattering transform

34 Following notations in [6], it is proved in [6] that

Theorem A.3. *A semi-discrete system $\{T_s^\theta \check{g}_\lambda : s \in \mathbb{R}^n, \lambda \in \Lambda\}$ is a semi-discrete frame for $L^2(\mathbb{R}^n)$ with bounds C_1, C_2 if and only if*

$$\frac{C_1}{|\sin \theta|^n} \leq \sum_{\lambda \in \Lambda} |\mathcal{F}_\theta(g_\lambda)(\omega)|^2 \leq \frac{C_2}{|\sin \theta|^n}, \quad \text{for a.e. } \omega \in \mathbb{R}^n.$$

Definition A.4. *Let $\{i\}_{k \in \mathbb{N}}$ be a sequence of countable index sets. For $k \in \mathbb{N}$, let $\Psi_k^\theta = \{T_s^\theta \check{g}_i : s \in \mathbb{R}^n, i \in i_k\}$ be a collection of semi-discrete frames for $L^2(\mathbb{R}^n)$. Further, assume that $\mathcal{M}_k^\theta : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ and $P_k^\theta : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ be Lipschitz continuous operators such that*

$$\mathcal{M}_k^\theta f = P_k^\theta f = 0, \quad \text{for } f = 0, k \in \mathbb{N}.$$

Then, the sequence of triples

$$\Omega^\theta := \{(\Psi_k^\theta, \mathcal{M}_k^\theta, P_k^\theta)\}$$

35 is called a θ -module sequence.

36 **Definition A.5.** *For $s > 0$, we define the θ -dilation operator connected with the FrFT as*

$$D_s^\theta := \mathcal{C}_{-\theta} D_s \mathcal{C}_\theta.$$

37 A sequence of real numbers $\{s_k : s_k > 1\}$ called polling factors. For simplicity in notation, we write
38 D_k^θ instead of $D_{s_k}^\theta$.

Definition A.6. *Let $\Omega^\theta = \{(\Psi_k^\theta, \mathcal{M}_k^\theta, P_k^\theta)\}$ be a θ -module sequence. Assume that $\{g_i : i \in i_k\}$ be the atoms of the semi-discrete frame Ψ_k^θ and $s_k > 1$ be the pooling factors associated with the k^{th} network layer. Then, we define the operator U_k^θ associated to the k^{th} layer of the network as*

$$\begin{aligned} U_k^\theta : i \times L^2(\mathbb{R}^n) &\rightarrow L^2(\mathbb{R}^n) \\ U_k^\theta(i, f) &:= U^\theta[i] f := D_k^\theta P_k^\theta \mathcal{M}_k^\theta(f \star_\theta g_i) \end{aligned}$$

39 In order to see that the operator U_k^θ is well-defined, we consider

$$\|U^\theta[i] f\|^2 \leq \|P_k^\theta \mathcal{M}_k^\theta(f \star_\theta g_i)\|^2 \leq R_k^2 L_k^2 \|f \star_\theta g_i\|^2 \leq B_k R_k^2 L_k^2 \|f\|^2$$

40 where B_k is an upper frame bound for the semi-discrete frame $\{T_s^\theta g_i : s \in \mathbb{R}^n, i \in i_k\}$ and L_k, R_k are
41 the Lipschitz constants for \mathcal{M}_k^θ and P_k^θ , respectively.

42 For $1 \leq k < \infty$, define the set

$$\Lambda^k := \Lambda_1 \times \cdots \times \Lambda_k.$$

43 An ordered tuple $q = (\lambda_1, \dots, \lambda_k) \in \Lambda^k$ is called a path.

44 Let $\mathcal{Q} := \bigcup_{k=1}^{\infty} \Lambda^k$. We define the operator $U^\theta : \mathcal{Q} \times L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ by

$$U^\theta[q]f := U^\theta[(\lambda_1, \dots, \lambda_i)]f := U_k^\theta[\lambda_i] \cdots U_1^\theta[\lambda_1]f,$$

45 where $q = (\lambda_1, \dots, \lambda_i) \in \Lambda^k$ and $f \in L^2(\mathbb{R}^n)$.

46 Let $\Omega^\theta = \{(\Psi_k^\theta, \mathcal{M}_k^\theta, \mathcal{P}_k^\theta)\}$ be a θ -module sequence. Let B_k^θ be an upper frame bound for Ψ_k^θ and
 47 L_k^θ, R_k^θ be the Lipschitz constants for \mathcal{M}_k^θ and \mathcal{P}_k^θ , respectively. Then, the condition

$$\max \left\{ B_k^\theta, B_k^\theta (L_k^\theta)^2 (R_k^\theta)^2 \right\} \leq 1, \text{ for all } k \in \mathbb{N}$$

48 is called the admissibility condition and the corresponding module sequence Ω^θ is called an admissible
 49 θ -module sequence. It can be directly drawn that the FrFNO proposed in our work is a special case
 50 of the operator.

51 A fundamental theorem for fractional frames, established in [6], guarantees the energy preservation
 52 property of fractional operators. This property ensures robust retention of mid-to-high frequency
 53 components, mitigating spectral bias in neural networks and enhancing performance in tasks sensitive
 54 to high-frequency patterns.

55 **Theorem A.4.** Let $\Omega^\theta = \{(\Psi_k^\theta, \mathcal{M}_k^\theta, \mathcal{P}_k^\theta)\}_{k \in \mathbb{N}}$ be an admissible θ -module sequence with pooling
 56 factors $s_k \geq 1$. Assume that the operators \mathcal{M}_k^θ and \mathcal{P}_k^θ commute with the time-translation operator
 57 T_t^θ . Suppose the output-generating atoms satisfy

$$|\omega \mathcal{F}_\theta(\varphi_k)(\omega)| \leq K_1 \quad \text{for almost every } \omega \in \mathbb{R}^n, \text{ for all } k \in \mathbb{N},$$

58 for some constant $K_1 > 0$. Then, there exists a constant $K > 0$ such that

$$\begin{aligned} & \left\| e^{-\pi i k t^2 \cot \theta} \Phi_{\Omega^\theta}^{k, \theta}(T_t^\theta f) - \Phi_{\Omega^\theta}^{k, \theta}(f) \right\|^2 \\ & \leq \pi^2 K^2 \|f\|^2 \left[\frac{t^4}{(s_1 \cdots s_k)^4} \cot^2 \theta + t^4 \left(\sum_{j=1}^k \frac{1}{s_j^2} \right)^2 \cot^2 \theta + \frac{4t^2}{(s_1 \cdots s_k)^2} \csc^2 \theta \right. \\ & \quad \left. + 4 \frac{|t|^3}{(s_1 \cdots s_k)^3} |\cot \theta \cdot \csc \theta| + 4|t|^3 \left(\sum_{j=1}^k \frac{1}{s_j^2} \right) |\cot \theta \cdot \csc \theta| \right], \end{aligned} \quad (8)$$

59 for all $f \in L^2(\mathbb{R}^n)$ and $t \in \mathbb{R}$.

60 In contrast, Biswas et al. [6] do not address a fractional version of this theorem, as established in
 61 Wiatowski and Bölcskei [57]. The fractional formulation, tied to the energy conservation properties
 62 of the classical scattering transform, preserves high-frequency components, mitigating the spectral
 63 bias of neural networks and enhancing robustness to noise [57]. Therefore, we prove the following
 64 theorem for the fractional case:

Theorem A.5. $\Omega^\theta = \{(\Psi_k^\theta, \mathcal{M}_k^\theta, \mathcal{P}_k^\theta)\}$ be an admissible θ -module. The corresponding feature
 extractor $\Phi_{\Omega^\theta} : L^2(\mathbb{R}^d) \rightarrow (L^2(\mathbb{R}^d))^{\mathcal{Q}}$ is Lipschitz-continuous with Lipschitz constant $L_\Omega = 1$,
 i.e.,

$$\|\Phi_{\Omega^\theta}(f) - \Phi_{\Omega^\theta}(h)\| \leq \|f - h\|_2, \quad \forall f, h \in L^2(\mathbb{R}^d).$$

65 *Proof.* Let $f_q := U^\theta[q]f$ and $h_q := U^\theta[q]h$ for $f, h \in L^2(\mathbb{R}^d)$. The key idea of the proof now is
 66 similarly to the proof of Wiatowski and Bölcskei [57, Proposition 4]. Consider

$$\|\Phi_{\Omega^\theta}^\theta(f) - \Phi_{\Omega^\theta}^\theta(h)\|^2 = \sum_{k=0}^{\infty} \|\Phi_{\Omega^\theta}^{k,\theta}(f) - \Phi_{\Omega^\theta}^{k,\theta}(h)\|^2 \quad (9)$$

$$= \sum_{q \in \Lambda^k} \|(U^\theta[q]f) \star_\theta \varphi_k - (U^\theta[q]h) \star_\theta \varphi_k\|^2 \quad (10)$$

$$= \sum_{q \in \Lambda^k} \|f_q \star_\theta \varphi_k - h_q \star_\theta \varphi_k\|^2 \quad (11)$$

$$= \sum_{q \in \Lambda^k} \int_{\mathbb{R}^n} |\mathcal{F}_\theta(f_q \star_\theta \varphi_k)(\omega) - \mathcal{F}_\theta(h_q \star_\theta \varphi_k)(\omega)|^2 d\omega \quad (12)$$

$$= \sum_{q \in \Lambda^k} \int_{\mathbb{R}^n} |\mathcal{F}_\theta(f_q)(\omega) \mathcal{F}_\theta(\varphi_k)(\omega) - \mathcal{F}_\theta(h_q)(\omega) \mathcal{F}_\theta(\varphi_k)(\omega)|^2 d\omega. \quad (13)$$

67 From Biswas et al. [6, Theorem 3.16], we get

$$\|\Phi_{\Omega^\theta}^\theta(f) - \Phi_{\Omega^\theta}^\theta(h)\|^2 \leq \sum_{q \in \Lambda^k} \int_{\mathbb{R}^n} |[\mathcal{F}_\theta(f_q)(\omega) - \mathcal{F}_\theta(h_q)(\omega)]|^2 |\mathcal{F}_\theta(\varphi_k)(\omega)|^2 d\omega.$$

68 As $\varphi_k \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, appealing to the Riemann-Lebesgue lemma for the FrFT, we get
 69 $\mathcal{F}_\theta(\varphi_k) \in C_0(\mathbb{R}^n)$. Hence, there exists a $L_G > 0$ such that

$$|\mathcal{F}_\theta(\varphi_k)(\omega)| \leq L_G, \forall \omega \in \mathbb{R}^n.$$

70 Then,

$$\begin{aligned} & \sum_{q \in \Lambda^k} \int_{\mathbb{R}^n} |[\mathcal{F}_\theta(f_q)(\omega) - \mathcal{F}_\theta(h_q)(\omega)]|^2 |\mathcal{F}_\theta(\varphi_k)(\omega)|^2 d\omega \\ & \leq L_G^2 \sum_{q \in \Lambda^k} \int_{\mathbb{R}^n} |[\mathcal{F}_\theta(f_q)(\omega) - \mathcal{F}_\theta(h_q)(\omega)]|^2 d\omega \\ & = L_G^2 \sum_{q \in \Lambda^k} [\|f_q\|^2 + \|h_q\|^2] \end{aligned}$$

71 Finally, proceeding as in the classical case Wiatowski and Bölcskei [57, Appendix I], we arrive at

$$\sum_{q \in \Lambda^k} \|f_q - h_q\|^2 \leq \sum_{q \in \Lambda^{k-1}} \|f_q - h_q\|^2. \quad (14)$$

72 By repeated use of the inequality (14), we obtain

$$\sum_{q \in \Lambda^k} \|f_q - h_q\|^2 \leq \|f - h\|^2.$$

73 from which the results follows. \square

74 Thus, we rigorously demonstrate that the fractional amplitude representation, particularly the FrNFO
 75 in our case, is resilient to noise-induced deformations and preserves the energy conservation property.

76 B Representational and interpretability analysis

77 B.1 Response analysis of fractional neural frame operator kernel

78 In this appendix, we systematically explore how the frequency-domain behavior of our fractional
 79 neural frame operator kernels changes across network depth, focusing on standard EEG frequency
 80 bands. To support neuroscientific analysis, we examine the following subbands: δ (0.25–4 Hz), θ
 81 (4–8 Hz), α (8–13 Hz), β (13–30 Hz), γ_{low} (30–80Hz), γ_{high} (> 80 Hz). We first calculate the power
 82 spectral density (PSD) for each kernel using the multitaper method, combining PSD estimates across
 83 all channels within each frequency band and layer of the pretrained FAPEX-Base model. We then
 84 perform two complementary analyses to characterize the spectral properties:

- 85 1. **Empirical PSD distribution analysis.** We visualize layer-wise histograms and kernel-
86 density estimates of PSD in each band to reveal how spectral emphasis shifts with depth.
- 87 2. **Power-law tail fitting analysis.** For the upper tail of each band’s PSD distribution, we fit a
88 continuous power-law model via maximum-likelihood estimation [10], extracting the scaling
89 exponent α and lower cutoff x_{\min} . This quantifies the heavy-tail behavior characteristic of
90 high-frequency preservation.

91 **Experimental setup.** To reveal how fractional neural operator kernels emphasize or attenuate different
92 frequency bands as network depth increases, we visualize the layer-wise empirical distributions of
93 band-specific PSD magnitudes using combined histogram and kernel density estimates (KDE).
94 Let $\mathcal{D}_{\ell,b} = \{P_k(f) : k \in \text{layer } \ell, f \in \text{band } b\}$ denote the set of PSD values (in dB) collected
95 over all kernels k in layer ℓ and frequencies f in band b . For each layer ℓ we plot, in a single
96 subplot, the normalized histogram of $\mathcal{D}_{\ell,b}$ (density axis) over a fixed binning scheme, together with a
97 Gaussian-kernel density estimate

$$\hat{p}_{\ell,b}(x) = \frac{1}{|\mathcal{D}_{\ell,b}|h} \sum_{y \in \mathcal{D}_{\ell,b}} K\left(\frac{x-y}{h}\right),$$

98 To quantitatively characterize the heavy-tailed behavior observed in the power spectral density (PSD)
99 magnitude distributions of fractional neural frame operator kernels, we fit each layer–band distribution
100 to a continuous power-law model. For each network layer ℓ and each EEG frequency band b , we
101 collect the set of PSD samples

$$\mathcal{D}_{\ell,b} = \{P_{k,c}(f) : k \in \text{kerels}, c \in \text{channels}, f \in \text{band } b\},$$

102 where $P_{k,c}(f)$ is the multitaper spectral estimate in dB, aggregated over all convolutional kernels k
103 and channels c in layer ℓ and band b .

104 To fit a power-law to the right tail of this distribution, we adopt the following model:

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}, \quad x \geq x_{\min}, \quad \alpha > 1,$$

105 where x_{\min} is the lower threshold above which the power-law behavior is assumed to hold, and α is
106 the scaling exponent. These parameters are estimated via maximum likelihood as:

$$\hat{\alpha} = 1 + \left(n^{-1} \sum_{x_i \geq x_{\min}} \ln \frac{x_i}{x_{\min}} \right)^{-1},$$

107 where n is the number of samples above x_{\min} . The optimal x_{\min} is chosen to minimize the Kol-
108 mogorov–Smirnov (KS) distance between the empirical complementary cumulative distribution
109 function (CCDF) and the fitted power-law model,

$$D = \max_{x \geq x_{\min}} \left| S_{\ell,b}(x) - \left(\frac{x}{x_{\min}} \right)^{1-\hat{\alpha}} \right|,$$

110 where $S_{\ell,b}(x)$ is the empirical CCDF, defined as

$$S_{\ell,b}(x) = \frac{1}{N} |\{x_i \geq x\}|, \quad x_i \in \mathcal{D}_{\ell,b}.$$

111 To assess the validity of the power-law fit, we compute the log-likelihood ratio R between the
112 power-law model and alternative heavy-tailed distributions (e.g., exponential, log-normal), and derive
113 a p -value using Vuong’s test 1. A significantly positive R with $p < 0.05$ indicates that the power
114 law is statistically preferred over the competing distribution for the observed data. The final fitted
115 exponents $\hat{\alpha}$ and thresholds \hat{x}_{\min} are annotated directly on log-log plots of the empirical CCDFs,
116 where we plot the empirical CCDF $\Pr(X > x) = \frac{1}{N} |\{x_i > x\}|$, and the fitted power-law tail line
117 $\left(\frac{x}{x_{\min}} \right)^{1-\hat{\alpha}}$. This approach yields a compact, quantitative descriptor of the tail behavior of PSD
118 response distributions for each layer and frequency band, revealing how high-frequency gain and
119 extreme-value behavior evolve across network depth.

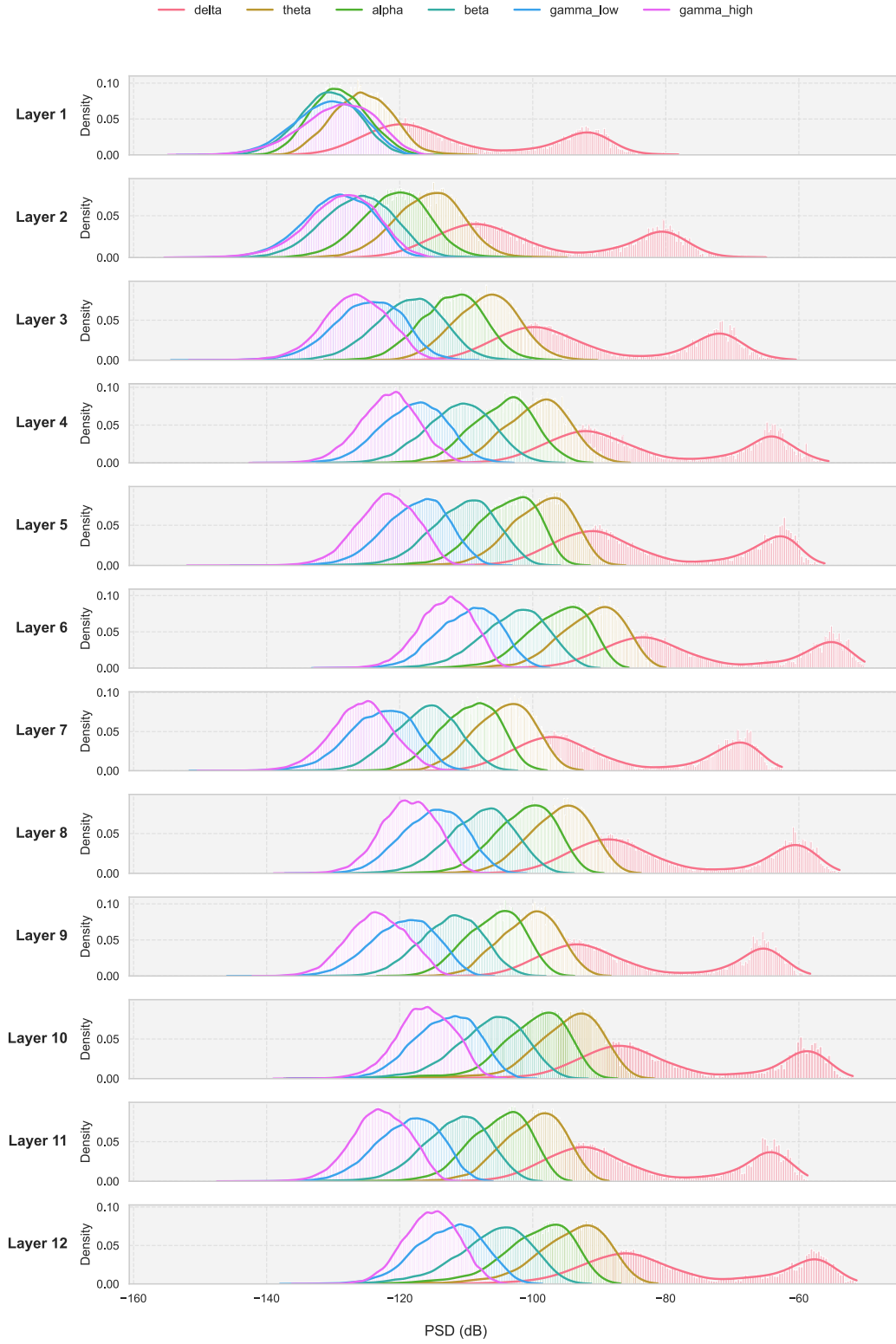


Figure 1: Layer-wise PSD distributions across frequency bands. Progressive separation of power spectral density (PSD) distributions with increasing model depth, reflecting enhanced frequency-specific feature extraction and stabilization deeper within the neural network.

Results and discussion. The PSD distributions and their corresponding power-law analyses are visualized in Fig. 1 and Fig. 2, respectively, revealing how the model adapts its spectral responses to enhance seizure prediction. In Fig. 1, the PSD distributions for distinct frequency bands become increasingly separable with greater layer depth, as evidenced by reduced overlap and sharper, more distinct peaks. This trend suggests that deeper layers effectively isolate frequency-specific neural features, organizing signals into well-defined spectral components. From shallow to intermediate layers, the distributions shift toward higher power, with both low-frequency bands (e.g., δ and θ) and high-frequency bands (e.g., γ) showing amplified responses. This indicates that early layers enhance both slow, rhythmic preictal patterns and transient, high-amplitude ictal oscillations. Beyond a critical depth (e.g., after Layer 6), the distributions stabilize, converging to a steady-state response that maintains a balanced representation of frequency components without further amplification.

The power-law analysis of the CCDFs in Fig. 2 quantitatively validates the trends observed in the PSD distributions in Fig. 1. The power-law exponents (α) exhibit a clear pattern of layer-wise alteration, revealing how the model adapts its spectral responses to enhance seizure prediction.

- **Shallow layers (high α):** In shallow layers, large α values (indicating steeper CCDF tails) reflect a selective response that prioritizes moderate-gain components. This aligns with the amplification of low-frequency preictal slow waves, which exhibit broad spatial synchrony and serve as early seizure indicators. The steeper tails suggest suppression of extreme high-amplitude events, focusing the model on stable, rhythmic patterns.
- **Intermediate layers (decreasing α):** As depth increases, α decreases significantly, particularly in high-frequency bands. This shift to heavier-tailed distributions enables the model to capture rare, high-gain events, such as transient, burst-like oscillations critical for ictal detection. The reduction in α indicates a relaxation of spectral bias, allowing the model to represent both slow and fast oscillatory dynamics effectively.
- **Deep layers (stabilized α):** Beyond a critical depth (e.g., after Layer 6), α values stabilize, mirroring the convergence observed in PSD distributions. This steady-state spectral response balances low- and high-frequency components without further amplification, ensuring robust detection of both sustained preictal dynamics and rapid ictal discharges.

This layer-wise adaptation addresses a critical limitation of conventional neural architectures, which typically exhibit a spectral bias toward low-frequency components, hindering their ability to detect subtle, high-frequency seizure precursors. By dynamically adjusting its spectral response, the model achieves enhanced sensitivity to both preictal and ictal neural dynamics across a broad frequency spectrum, significantly improving the accuracy of seizure prediction.

B.2 Brain region analysis

Experimental Setup. We conducted experiments to assess the importance of different anatomical regions in seizure forecasting. Specifically, we maintain the training stage unchanged, but during testing, we mask channels from specific regions. The trained Ours-Large model is employed for this experiment. We perform this analysis on the **Macaque** dataset, as it minimizes interference from other factors such as data quality. The masked region that induces the greatest performance drop is defined as the most critical. Brain regions include: *Cortical* - Frontal Lobe (FL), Occipital Lobe (OL), Parietal Lobe (PL), Posterior Temporal Lobe (PTL), Temporal Lobe (TL); *Subcortical* - Caudate Nucleus (CN), Central Lateral Nucleus (CLN), Central Medial Nucleus (CMN), Dentate Gyrus (DG), Hippocampus (HC), Parahippocampal Gyrus (PHG), Nucleus Accumbens (NA).

Results and discussion. The results of the brain region importance analysis are summarized in Table ??, with the baseline (*All Regions*) achieving a median balanced accuracy (BA) of 82.23%, sensitivity (SEN) of 87.04%, F1 score of 95.62%, and ROC AUC of 90.08%. Masking specific brain regions during testing reveals their relative contributions to seizure forecasting, as illustrated by the relative performance decreases (RPD) in Figure 3. Among cortical regions, the Posterior Temporal Lobe (PTL) is the most critical, with a 56.0% RPD in F1 score, followed by the Frontal Lobe (FL) with a 39.5% drop in BA. In contrast, the Parietal Lobe (PL) shows the smallest impact, with a 15.7% RPD in PRC. For subcortical regions, the Central Medial Nucleus (CMN) exhibits the most significant decline, with a 59.7% RPD in F1 score, while the Dentate Gyrus (DG) follows with a 50.4% drop in F1. The Parahippocampal Gyrus (PHG) has the least impact, with a 28.0% RPD in BA. These findings suggest that the PTL and CMN are pivotal for seizure prediction, likely due to their

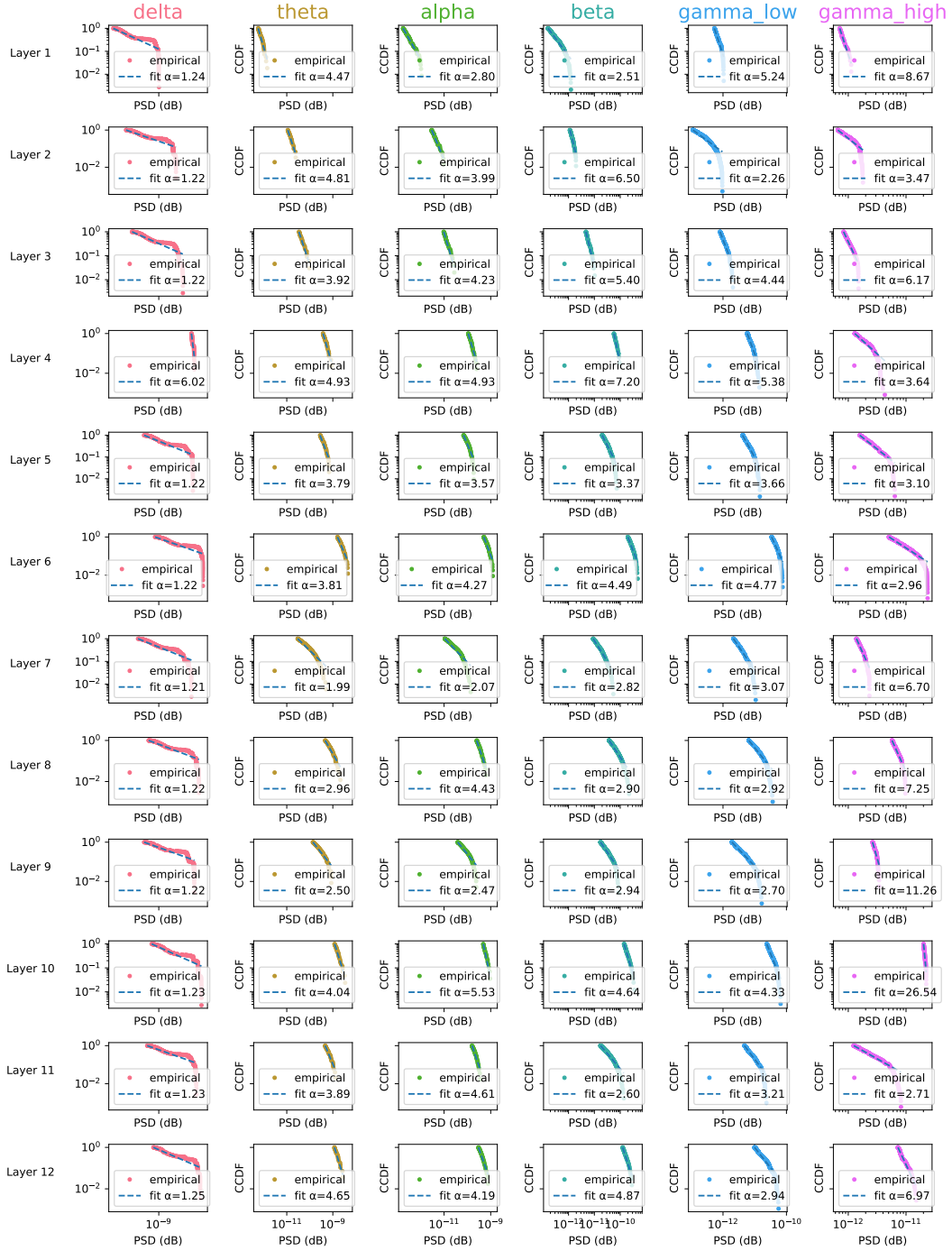


Figure 2: Layer-wise Power-Law Behavior of PSD CCDFs. Complementary cumulative distribution functions (CCDFs) of power spectral densities (PSDs) follow distinct power-law tails across network depth: shallow layers exhibit steep exponents (high α), reflecting suppression of extreme amplitudes and emphasis on rhythmic preictal activity; intermediate layers show reduced α , indicating heavier tails that capture rare, high-gain ictal bursts; beyond a critical depth, α values converge, demonstrating a stabilized spectral response that balances sensitivity to both low- and high-frequency neural dynamics.

roles in temporal pattern processing and thalamocortical synchronization, respectively, while the PL and PHG play secondary roles in the early ictal dynamics captured by the model. Our brain region importance analysis on the Macaque dataset underscores the pivotal roles of the Posterior Temporal Lobe (PTL) and Central Medial Nucleus (CMN) in seizure forecasting, as evidenced by substantial performance drops when these regions are masked. Figure 3). Masking the PTL led to a 56.0% drop in F1 score, reflecting its critical involvement in temporal lobe epilepsy (TLE), a common epilepsy type in macaques and humans. The PTL, encompassing posterior parts of the temporal cortex, is integral to processing temporal patterns in EEG signals, and its dysfunction is linked to seizure initiation, as seen in macaque models with kainic acid-induced hippocampal seizures [8]. Similarly, the CMN, a thalamic nucleus, exhibited a 59.7% F1 score drop, underscoring its role in thalamocortical synchronization. Computational models in mice highlight the thalamus’s modulation of cortical excitability during seizures, supporting the CMN’s significance in forecasting [41]. The Dentate Gyrus (DG), with a 50.4% F1 drop, also plays a notable role, consistent with hippocampal sclerosis observed in TLE, where selective cell loss in the DG disrupts seizure suppression mechanisms [35].

In contrast, the Parietal Lobe (PL) and Parahippocampal Gyrus (PHG) showed the least impact, with performance drops of 15.7% in PRC for PL and 28.0% in BA for PHG. The PL’s primary functions in sensory integration and spatial processing are less directly tied to early ictal dynamics, explaining its minimal contribution to seizure forecasting. The PHG, involved in memory consolidation, likely plays a secondary role in seizure propagation rather than onset, aligning with its modest impact [35]. These findings suggest that our EEG-based model prioritizes regions central to epileptogenic networks, such as the PTL and CMN, over those with peripheral roles.

However, the reliance on EEG channel masking may oversimplify the complex network dynamics of epileptogenesis, which involve connectivity changes, neuroinflammation, and synaptic reorganization [35]. The model’s focus on PTL and CMN may also reflect biases in EEG signal sensitivity to cortical and thalamic activity, potentially underestimating subcortical contributions from regions like the PHG. Future work could integrate multimodal data, such as fMRI or inflammatory biomarkers, to capture broader epileptogenic processes and enhance predictive accuracy across diverse seizure types. Additionally, exploring dynamic masking strategies or graph-based network models could better elucidate inter-regional interactions driving seizure onset.

C Full results

C.1 Full results of supervised learning experiments

Table 1: Comparison with Supervised Baselines on ATLE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (↑)	SEN (↑)	F1 (↑)	ROC (↑)	PRC (↑)	Brier (↓)
Convolutional Mixers	ModernTCN [13]	93.11 (77.83, 94.96)	91.72 (82.63, 96.51)	90.25 (89.85, 92.30)	99.98 (92.80, 100.00)	99.99 (79.26, 100.00)	7.64 (5.19, 13.76)
	MRConv [11]	86.75 (51.77, 87.94)	86.64 (84.44, 91.06)	96.10 (86.73, 100.00)	99.97 (92.16, 100.00)	99.99 (77.44, 100.00)	10.65 (9.69, 19.44)
	MultiresNet [42]	84.34 (53.60, 87.01)	85.41 (71.75, 89.63)	84.28 (74.24, 89.09)	99.98 (92.28, 100.00)	99.99 (77.44, 100.00)	10.69 (8.15, 21.63)
	Omni-Scale [46]	87.91 (55.50, 97.52)	87.81 (85.22, 98.03)	98.59 (96.93, 99.77)	99.87 (92.65, 99.98)	99.92 (79.67, 99.99)	9.24 (2.06, 24.04)
	SPaRCnet [24]	74.19 (51.48, 93.87)	84.01 (68.67, 96.06)	81.73 (67.56, 96.00)	99.84 (92.94, 100.00)	99.91 (81.29, 100.00)	19.17 (4.25, 23.89)
Token Mixers	EEGConformer [43]	89.37 (55.74, 90.46)	88.48 (80.94, 92.86)	91.20 (88.06, 98.88)	99.99 (92.82, 100.00)	99.99 (78.81, 100.00)	8.46 (7.16, 19.16)
	EEGMamba [19]	87.38 (54.50, 98.04)	88.17 (84.37, 98.69)	84.96 (82.87, 98.68)	99.97 (92.41, 100.00)	99.98 (80.04, 100.00)	9.41 (1.86, 25.85)
	iTransformer [32]	50.82 (50.00, 52.81)	54.88 (35.88, 70.23)	2.85 (0.00, 8.54)	99.76 (90.46, 99.95)	99.88 (74.42, 99.98)	19.62 (19.02, 26.07)
	Nonformer [29]	82.54 (51.95, 94.07)	84.74 (81.76, 95.85)	97.50 (93.72, 99.54)	99.85 (92.41, 99.97)	99.90 (78.84, 99.99)	10.05 (3.35, 25.08)
	PatchTST [38]	87.80 (53.38, 88.47)	86.60 (75.86, 91.62)	92.96 (79.64, 100.00)	99.97 (92.44, 100.00)	99.98 (77.92, 100.00)	10.04 (7.94, 20.89)
	Pathformer [9]	90.57 (62.39, 91.61)	88.68 (85.74, 94.19)	95.26 (90.26, 98.77)	99.97 (92.64, 100.00)	99.99 (78.06, 100.00)	8.91 (6.24, 16.44)
	Seizureformer [1]	85.35 (52.73, 95.81)	86.35 (83.23, 96.71)	97.92 (95.02, 99.62)	99.86 (92.54, 99.98)	99.91 (79.19, 99.99)	9.67 (2.78, 24.66)
Time-Frequency Mixers	ATFNet [65]	77.44 (50.62, 93.94)	83.14 (79.56, 95.09)	96.97 (91.87, 99.41)	99.83 (92.20, 99.97)	99.89 (78.38, 99.98)	10.58 (3.94, 25.42)
	PreTS [66]	61.80 (50.40, 74.56)	70.19 (50.30, 88.32)	70.84 (48.15, 83.14)	77.82 (72.76, 81.63)	72.65 (50.69, 89.76)	22.14 (20.41, 28.13)
	NFM [25]	63.75 (51.66, 76.24)	71.31 (54.13, 88.44)	71.83 (53.25, 83.72)	81.73 (72.41, 82.37)	74.62 (56.22, 91.44)	20.79 (19.45, 28.20)
	TSLaNet [16]	92.85 (72.07, 93.92)	91.42 (82.74, 95.79)	91.58 (88.92, 93.92)	99.98 (92.69, 100.00)	99.99 (78.55, 100.00)	7.79 (6.81, 17.94)
Multiscale Token Mixers	AdaWaveNet [68]	81.03 (51.92, 92.79)	82.76 (79.19, 94.34)	96.58 (91.38, 99.31)	99.84 (92.34, 99.97)	99.89 (78.61, 99.98)	10.29 (3.67, 25.22)
	Medformer [54]	88.88 (56.22, 97.96)	88.23 (85.73, 98.54)	98.87 (97.38, 99.88)	99.88 (92.73, 99.98)	99.93 (80.01, 99.99)	8.97 (1.73, 23.76)
	MTS [73]	78.96 (51.79, 94.71)	84.06 (80.81, 95.41)	97.23 (92.84, 99.46)	99.84 (92.31, 99.97)	99.90 (78.67, 99.98)	10.23 (3.58, 25.15)
	Pyraformer [27]	83.74 (52.14, 94.49)	85.14 (82.03, 96.11)	97.66 (94.25, 99.58)	99.85 (92.46, 99.97)	99.90 (78.97, 99.99)	9.89 (3.17, 24.94)
	SimpleTW [3]	91.95 (56.38, 98.52)	90.79 (83.68, 98.64)	90.42 (82.29, 98.64)	99.89 (93.22, 99.99)	99.94 (81.75, 100.00)	6.30 (1.61, 23.69)
	TimesNet [61]	80.23 (51.86, 92.36)	82.14 (78.73, 93.97)	96.34 (90.95, 99.23)	99.83 (92.29, 99.97)	99.89 (78.52, 99.98)	10.41 (3.79, 25.31)
	TimeMixer [51]	88.70 (53.37, 90.10)	87.89 (84.63, 92.17)	95.33 (88.85, 99.38)	99.97 (92.45, 100.00)	99.98 (77.44, 100.00)	8.46 (7.31, 19.68)
	This Work	87.38 [†] (54.50, 98.04)	88.17 [†] (84.37, 98.69)	84.96 [†] (82.87, 98.68)	99.97 [†] (92.41, 100.00)	99.98 [†] (80.04, 100.00)	9.41 [†] (1.86, 25.85)
	PAPEX-BASE	91.95[†] (56.38, 98.52)	90.79[†] (83.68, 98.64)	90.42[†] (82.29, 98.64)	99.89 (93.22, 99.99)	99.94 (81.75, 100.00)	6.30[†] (1.61, 23.69)

Table 2: Comparison with Supervised Baselines on AGS Dataset for Seizure Prediction (Blank Template). Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	80.88 (67.52, 81.85)	86.96 (84.12, 94.10)	84.97 (84.11, 94.25)	93.22 (88.95, 93.87)	68.59 (62.18, 83.15)	14.22 (6.42, 20.78)
	MRConv [12]	78.87 (77.60, 82.99)	91.29 (85.48, 95.29)	90.31 (85.36, 95.25)	95.25 (89.76, 95.72)	68.22 (67.60, 89.31)	12.35 (5.95, 13.78)
	MultiresNet [42]	80.47 (73.96, 82.40)	90.06 (88.75, 96.76)	88.83 (88.62, 96.80)	96.09 (92.46, 97.39)	79.20 (73.57, 89.99)	10.69 (6.12, 14.84)
	Omni-Scale [45]	79.11 (78.84, 83.32)	91.73 (85.50, 95.22)	90.88 (85.86, 95.19)	95.24 (90.09, 95.87)	68.51 (67.79, 89.71)	11.63 (5.88, 13.56)
	SPaRCNet [23]	78.10 (72.33, 82.31)	89.06 (83.34, 94.53)	87.48 (81.89, 94.43)	93.39 (88.12, 95.01)	66.30 (62.65, 83.39)	15.41 (6.11, 15.85)
Token Mixers	EEGConformer [44]	78.90 (74.47, 79.28)	89.80 (83.94, 95.04)	88.52 (84.53, 94.95)	94.41 (88.03, 96.42)	68.65 (66.46, 85.64)	13.95 (6.51, 14.25)
	EEGMamba [71]	80.40 (80.33, 85.14)	93.83 (87.62, 95.79)	93.47 (87.69, 95.83)	96.83 (90.84, 97.46)	74.89 (72.65, 91.96)	8.06 (5.94, 12.05)
	iTransformer [31]	78.27 (72.47, 82.99)	89.49 (85.42, 95.25)	87.85 (85.70, 95.20)	95.30 (89.21, 95.31)	67.21 (66.89, 88.23)	13.76 (5.88, 15.29)
	Nonformer [30]	80.35 (79.75, 82.94)	93.21 (85.70, 95.71)	92.70 (85.77, 95.74)	96.68 (91.05, 97.29)	74.83 (69.86, 91.71)	9.31 (6.12, 13.28)
	PatchTST [39]	78.44 (75.23, 80.19)	90.53 (85.69, 95.08)	89.27 (85.93, 95.06)	95.51 (89.30, 96.44)	68.53 (67.21, 88.73)	13.72 (6.34, 13.78)
	Pathformer [9]	80.92 (80.17, 81.99)	92.51 (85.58, 95.64)	91.84 (84.18, 95.67)	96.70 (91.12, 97.06)	71.54 (69.63, 91.54)	10.20 (5.71, 14.09)
	Seizureformer [2]	79.91 (79.82, 83.84)	92.10 (85.49, 95.17)	91.34 (85.00, 95.16)	95.39 (90.33, 96.08)	68.32 (68.22, 90.21)	10.90 (5.75, 13.88)
Time-Frequency Mixers	ATFNet [65]	78.03 (67.77, 80.62)	85.15 (83.05, 92.90)	84.11 (81.14, 93.01)	90.78 (88.15, 95.11)	69.92 (62.43, 75.54)	14.51 (6.95, 20.90)
	FreTS [67]	77.59 (71.50, 78.74)	88.71 (83.47, 93.74)	87.02 (82.73, 93.76)	92.97 (88.05, 96.42)	67.79 (62.23, 82.14)	15.66 (6.61, 16.03)
	NFM [26]	77.59 (71.50, 78.74)	88.71 (83.47, 93.74)	87.02 (82.73, 93.76)	92.97 (88.05, 96.42)	67.79 (62.23, 82.14)	15.66 (6.61, 16.03)
	TSLANet [15]	80.50 (80.01, 86.78)	94.41 (87.82, 96.16)	94.15 (87.87, 96.17)	97.29 (90.76, 97.51)	76.38 (72.72, 92.99)	7.03 (5.89, 11.88)
Multiscale Token Mixers	AdaWaveNet [68]	79.60 (71.20, 83.55)	89.01 (87.66, 96.24)	87.63 (87.17, 96.27)	95.13 (91.09, 96.13)	73.87 (66.77, 87.36)	12.58 (5.73, 16.80)
	Medformer [54]	80.75 (70.62, 83.16)	88.72 (87.62, 96.49)	87.99 (86.81, 96.50)	96.09 (92.25, 96.94)	79.28 (72.36, 89.21)	10.95 (5.71, 17.02)
	MSTS [73]	78.15 (56.63, 90.98)	91.61 (83.61, 96.55)	90.67 (78.08, 96.42)	96.05 (97.41, 99.20)	93.87 (91.76, 97.71)	9.26 (5.05, 16.62)
	Pyraformer [28]	79.91 (79.65, 81.35)	92.03 (85.79, 95.64)	91.29 (85.65, 95.67)	96.45 (89.78, 96.85)	70.98 (68.23, 90.88)	10.97 (5.88, 13.60)
	SimpleTM [7]	78.60 (65.10, 79.99)	85.09 (82.40, 91.76)	83.54 (81.38, 91.90)	88.23 (87.92, 94.80)	69.63 (61.84, 70.19)	14.97 (7.79, 22.33)
	TimesNet [62]	78.24 (73.97, 78.38)	89.71 (83.68, 94.77)	88.35 (82.61, 94.67)	94.30 (88.22, 96.49)	68.74 (64.15, 85.11)	14.18 (6.49, 15.22)
	TimeMixer [51]	80.69 (78.96, 81.50)	92.29 (85.89, 95.59)	91.62 (86.15, 95.63)	96.59 (91.38, 96.96)	71.76 (69.12, 91.22)	10.27 (5.80, 13.24)
This Work	FAPEX-Small	85.33 (62.34, 92.68)	94.08 (85.77, 97.16)	93.72 (82.04, 97.08)	98.39 (98.15, 99.59)	94.69 (94.47, 98.67)	7.64 (3.90, 13.90)
	FAPEX-Base	87.29 (68.05, 93.72)	94.89 (87.92, 97.56)	94.63 (85.51, 97.49)	99.46 (98.63, 99.82)	97.72 (96.03, 99.36)	5.49 (3.09, 11.83)

Table 3: Comparison with Supervised Baselines on BEIRUT Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	70.52 (69.59, 89.86)	83.37 (66.13, 92.02)	83.11 (68.55, 90.86)	85.03 (79.86, 96.30)	93.62 (60.40, 98.21)	15.26 (8.41, 22.10)
	MRConv [12]	69.61 (61.29, 73.60)	78.64 (75.40, 90.60)	78.21 (73.77, 88.53)	83.49 (67.98, 92.54)	93.17 (53.74, 95.78)	17.37 (16.20, 18.99)
	MultiresNet [42]	64.12 (61.14, 74.11)	73.26 (60.63, 80.69)	72.78 (63.68, 79.34)	74.69 (65.92, 83.11)	86.93 (45.98, 92.50)	19.10 (14.49, 25.28)
	Omni -Scale [45]	69.08 (59.44, 72.43)	72.59 (53.00, 79.61)	71.48 (56.03, 78.27)	83.10 (64.30, 92.10)	94.06 (52.59, 95.81)	19.24 (17.67, 26.47)
	SPaRCnet [23]	71.96 (57.02, 75.09)	71.14 (59.35, 89.44)	71.64 (62.14, 89.19)	79.10 (60.39, 89.07)	92.71 (47.92, 94.81)	19.10 (14.47, 27.04)
Token Mixers	EEGConformer [44]	66.67 (57.97, 69.46)	68.37 (43.65, 78.17)	66.52 (46.64, 75.93)	82.47 (67.58, 87.99)	93.64 (52.74, 94.64)	23.12 (18.54, 30.27)
	EEGMamba [2]	67.39 (58.45, 70.67)	70.00 (48.14, 79.06)	68.48 (51.21, 77.36)	82.58 (65.75, 87.07)	93.54 (53.21, 94.70)	22.05 (18.25, 28.42)
	iTransformer [31]	68.75 (58.93, 71.59)	70.62 (45.51, 79.39)	69.25 (51.44, 78.01)	82.78 (68.07, 87.71)	93.72 (52.73, 94.32)	20.46 (18.03, 28.65)
	Nonformer [30]	63.26 (58.74, 69.91)	68.58 (56.53, 78.28)	63.41 (61.32, 77.21)	75.22 (66.22, 82.44)	87.43 (45.60, 92.37)	20.39 (16.34, 25.68)
	PatchTST [39]	71.70 (57.61, 76.23)	71.86 (62.11, 89.78)	72.50 (64.41, 89.70)	79.28 (60.54, 89.28)	92.83 (48.70, 94.92)	18.85 (14.61, 26.67)
	Pathformer [9]	63.28 (54.97, 68.17)	67.61 (33.93, 73.93)	64.11 (43.18, 69.74)	81.67 (60.45, 85.25)	92.69 (51.69, 94.31)	24.20 (23.32, 31.83)
	Seizureformer [17]	61.57 (52.50, 64.79)	66.99 (52.54, 76.02)	60.18 (57.42, 76.39)	79.81 (64.27, 89.68)	92.53 (51.12, 93.16)	20.28 (15.61, 44.14)
Time-Frequency Mixers	ATFNet [65]	65.46 (60.42, 75.42)	76.01 (40.95, 77.92)	74.60 (38.88, 77.33)	79.43 (64.08, 90.85)	92.75 (51.58, 93.42)	19.36 (16.03, 42.24)
	FreTS [67]	59.59 (52.26, 67.66)	64.66 (32.18, 72.12)	58.65 (40.23, 68.25)	81.85 (59.43, 85.82)	93.20 (51.46, 94.04)	26.34 (23.24, 35.24)
	NFM [26]	65.86 (60.68, 76.31)	77.30 (42.33, 79.04)	75.69 (41.14, 79.24)	79.67 (63.41, 89.61)	92.39 (51.43, 93.03)	19.08 (15.23, 41.10)
	TSLANet [15]	70.52 (69.59, 89.86)	83.37 (66.13, 92.02)	83.11 (68.55, 90.86)	85.03 (79.86, 96.30)	93.62 (60.40, 98.21)	15.26 (8.41, 22.10)
Multiscale Token Mixers	AdaWaveNet [68]	66.20 (54.90, 69.10)	68.04 (36.51, 77.97)	66.13 (46.40, 75.60)	82.66 (60.80, 86.53)	93.45 (52.94, 94.23)	23.36 (19.41, 32.98)
	Medformer [54]	70.20 (68.97, 86.51)	83.75 (62.72, 91.62)	83.09 (65.30, 90.48)	84.44 (80.15, 96.16)	93.51 (59.21, 98.18)	15.63 (8.56, 23.58)
	MSTS [73]	64.77 (61.91, 78.08)	80.34 (52.20, 88.19)	78.42 (53.83, 87.36)	84.07 (79.63, 95.25)	93.74 (61.40, 97.70)	19.17 (12.94, 30.58)
	Pyraformer [28]	67.33 (65.36, 87.54)	82.82 (60.07, 91.40)	81.71 (62.63, 89.66)	85.41 (79.73, 95.36)	93.91 (59.33, 97.72)	17.10 (8.64, 25.13)
	SimpleTM [7]	69.23 (67.70, 87.91)	82.59 (60.61, 91.51)	81.96 (63.12, 90.22)	83.44 (80.92, 95.75)	93.56 (59.61, 97.95)	15.98 (8.69, 24.60)
	TimesNet [62]	71.23 (55.73, 73.25)	70.33 (54.00, 87.99)	70.59 (57.15, 87.95)	78.43 (59.75, 88.08)	92.38 (47.23, 94.23)	19.88 (14.61, 29.30)
	TimeMixer [51]	72.34 (57.42, 75.68)	71.76 (61.39, 89.80)	72.28 (63.81, 89.64)	79.02 (60.68, 89.28)	92.68 (48.41, 94.92)	18.94 (14.36, 26.72)
This Work	FAPEX -Small	72.38 (71.32, 89.34)	83.93 (67.81, 91.65)	83.76 (70.14, 90.92)	85.21 (80.42, 96.37)	93.90 (60.97, 98.27)	14.40 (7.76, 21.21)
	FAPEX -Base	72.45 (70.77, 90.68)	84.74 (66.33, 92.28)	84.34 (68.74, 91.29)	85.83 (83.89, 96.80)	93.96 (66.49, 98.49)	14.78 (7.23, 21.34)

Table 4: Comparison with Supervised Baselines on CANINE Dataset for seizure prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (↑)	SEN (↑)	F1 (↑)	ROC (↑)	PRC (↑)	Brier (↓)
Convolutional Mixers	ModernTCN [34]	61.49 (57.32, 63.03)	84.81 (69.22, 86.64)	84.05 (71.58, 85.03)	73.36 (68.01, 73.96)	40.52 (27.06, 42.38)	19.91 (17.91, 20.98)
	MRConv [12]	60.38 (57.11, 63.52)	83.67 (72.40, 85.59)	83.82 (73.80, 84.18)	72.57 (67.89, 73.69)	38.55 (27.11, 41.58)	20.13 (17.85, 20.55)
	MultiresNet [42]	60.48 (58.35, 64.67)	64.80 (61.08, 84.06)	71.98 (61.74, 82.92)	70.82 (66.87, 74.26)	38.76 (23.24, 42.15)	25.29 (17.01, 27.59)
	Omni-Scale [45]	62.90 (58.10, 65.75)	65.31 (61.85, 79.05)	71.92 (61.90, 80.92)	73.20 (67.84, 74.46)	40.99 (25.49, 42.79)	26.14 (16.91, 26.88)
	SPaRCnet [23]	62.37 (56.28, 65.18)	85.92 (71.06, 87.36)	84.70 (73.07, 85.54)	74.00 (68.08, 74.98)	42.00 (28.78, 42.23)	19.60 (17.85, 20.46)
Token Mixers	EEGConformer [44]	50.38 (50.12, 68.03)	79.35 (78.41, 86.91)	81.42 (69.51, 85.49)	52.90 (50.10, 78.59)	36.59 (20.01, 38.61)	19.70 (17.02, 24.44)
	EEGMamba [71]	50.83 (50.34, 69.77)	79.56 (78.40, 83.10)	81.15 (69.41, 85.79)	53.20 (50.74, 78.88)	37.12 (20.37, 39.47)	19.76 (17.26, 23.76)
	iTransformer [31]	62.48 (52.60, 64.77)	64.22 (60.43, 75.79)	70.44 (60.91, 78.37)	71.24 (66.92, 73.32)	35.33 (23.20, 40.39)	26.69 (19.47, 27.07)
	Nonformer [30]	59.23 (52.74, 64.02)	60.69 (56.33, 69.93)	64.89 (58.55, 73.98)	70.29 (67.38, 72.67)	35.25 (22.32, 39.20)	26.66 (21.53, 28.08)
	PatchTST [39]	60.04 (56.95, 66.56)	77.61 (65.59, 83.98)	80.57 (65.42, 83.77)	72.85 (67.50, 74.07)	40.64 (26.15, 41.93)	21.62 (17.38, 24.03)
	Pathformer [9]	59.17 (55.07, 64.21)	65.94 (61.72, 85.67)	72.26 (61.69, 83.39)	71.83 (67.88, 72.85)	38.30 (24.21, 39.51)	24.36 (18.16, 26.49)
	TimeMachine [2]	56.02 (52.20, 62.16)	56.31 (51.14, 65.76)	62.12 (59.08, 64.21)	70.07 (66.61, 72.71)	36.66 (22.17, 39.66)	26.28 (25.50, 29.28)
Seizureformer [17]	51.11 (50.16, 68.41)	78.82 (78.33, 80.82)	81.77 (69.41, 82.96)	53.52 (50.47, 78.55)	36.89 (20.16, 38.90)	19.73 (17.16, 23.74)	
Time-Frequency Mixers	ATFNet [65]	63.25 (53.82, 64.90)	62.84 (60.16, 77.85)	70.90 (60.81, 80.45)	71.30 (67.95, 73.97)	37.06 (25.82, 39.91)	26.61 (19.93, 27.08)
	FreTS [67]	58.72 (52.90, 62.50)	46.75 (41.35, 61.02)	54.32 (48.75, 61.89)	69.54 (65.98, 75.43)	37.10 (20.96, 42.60)	28.58 (25.68, 34.00)
	TSL [26]	59.09 (55.62, 65.31)	71.48 (66.89, 83.87)	76.88 (66.60, 83.79)	71.96 (67.82, 73.21)	37.85 (26.14, 40.30)	22.10 (17.31, 24.23)
	NFMNet [15]	61.33 (60.48, 62.90)	85.73 (75.32, 87.20)	84.40 (75.27, 85.44)	72.96 (68.12, 73.99)	38.08 (27.24, 41.71)	19.67 (17.30, 20.14) [†]
Multiscale Token Mixers	AdaWaveNet [68]	50.25 (49.62, 66.04)	79.82 (71.21, 89.63)	81.30 (67.94, 86.65)	52.90 (49.49, 78.31)	35.23 (19.84, 37.82)	19.81 (17.24, 25.37)
	Medformer [54]	60.87 (53.43, 64.36)	85.92 (69.79, 87.73) [‡]	84.54 (72.14, 85.69) [‡]	74.14 (68.26, 74.82) [‡]	41.90 (28.27, 42.81) [‡]	20.09 (18.09, 20.81)
	MSTS [73]	63.08 (61.39, 65.60)	67.95 (62.23, 84.17)	74.54 (63.79, 83.84)	72.54 (67.54, 74.19)	39.49 (27.01, 40.58)	24.12 (17.35, 25.82)
	Pyraformer [28]	62.17 (58.11, 64.63)	80.81 (66.55, 84.99)	82.28 (66.64, 84.14)	72.70 (67.67, 74.76)	39.83 (26.85, 42.86)	20.83 (17.29, 23.46)
	SimpleTRF [7]	61.35 (59.68, 62.62)	82.75 (68.30, 86.14)	83.35 (70.42, 84.01)	72.52 (67.86, 74.22)	39.18 (27.08, 41.53)	20.26 (17.69, 21.29)
	TimesNet [62]	62.13 (57.13, 64.46)	67.35 (61.74, 84.66)	74.12 (62.53, 83.53)	72.29 (67.70, 73.17)	39.51 (25.51, 41.35)	25.51 (17.66, 26.76)
	TimeMixer [51]	63.10 (56.50, 65.75) [†]	76.56 (62.49, 84.88)	79.97 (64.25, 83.93)	72.97 (68.14, 73.72)	36.66 (25.99, 41.31)	21.92 (17.60, 23.58)
	FAPEX-Small	61.93 (54.61, 65.05)	85.02 (78.97, 87.90) [†]	84.73 (73.02, 85.74) [†]	74.15 (68.22, 75.24) [†]	42.30 (28.97, 42.56) [†]	19.69 (18.72, 20.42) [†]
	FAPEX-Base	62.16 (54.84, 66.17)	86.00 (69.76, 88.50) [†]	84.71 (72.11, 86.07) [†]	74.53 (68.89, 75.92) [†]	42.99 (29.45, 45.07) [†]	19.51 (18.50, 20.58) [†]

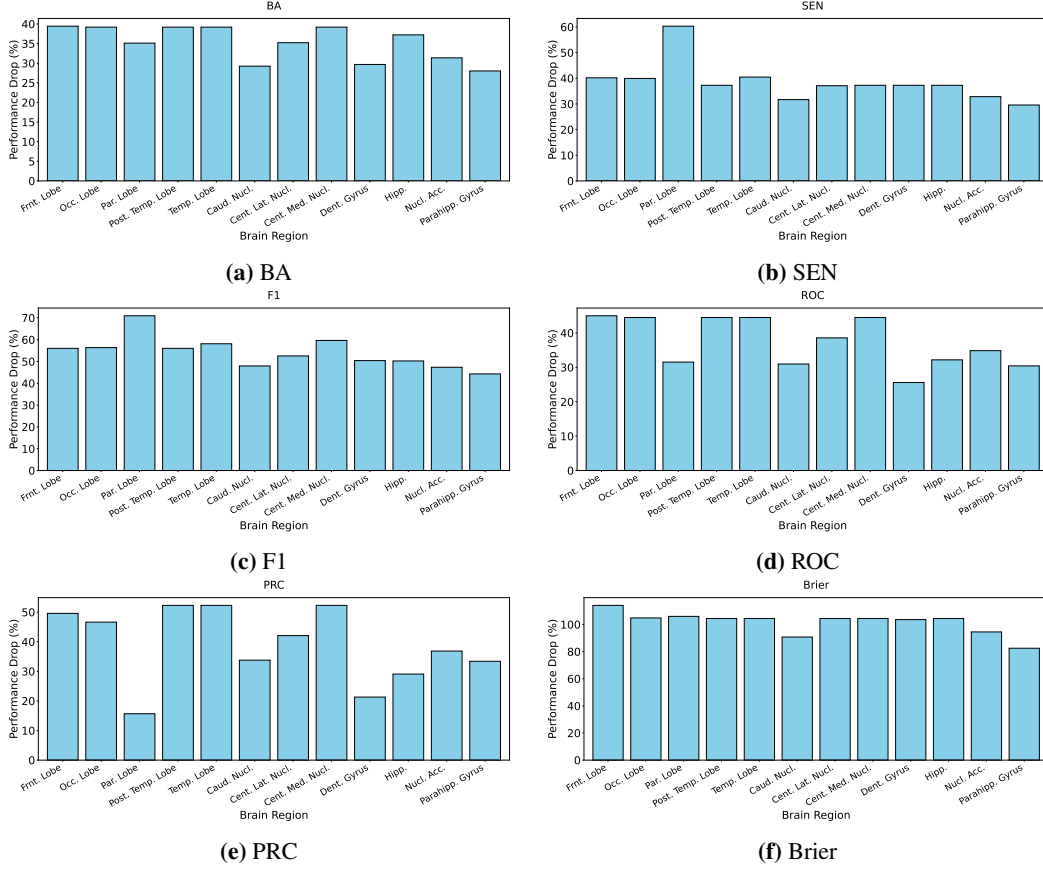


Figure 3: Performance Drop Ratios Across Brain Regions on the Macaque Dataset. Each subfigure shows the relative performance drop ratio (%) for a specific metric when masking individual brain regions compared to the baseline (no masking). The metrics are: (a) Balanced Accuracy (BA), (b) Sensitivity (SEN), (c) F1 Score (F1), (d) ROC AUC (ROC), (e) PRC AUC (PRC), and (f) Brier Score (Brier). Higher drop ratios indicate greater importance of the masked region for model performance. Brain region abbreviations are: Frnt. Lobe (Frontal Lobe), Occ. Lobe (Occipital Lobe), Par. Lobe (Parietal Lobe), Post. Temp. Lobe (Posterior Temporal Lobe), Temp. Lobe (Temporal Lobe), Caud. Nucl. (Caudate Nucleus), Cent. Lat. Nucl. (Central Lateral Nucleus), Cent. Med. Nucl. (Central Medial Nucleus), Dent. Gyrus (Dentate Gyrus), Hipp. (Hippocampus), Nucl. Acc. (Nucleus Accumbens), Parahipp. Gyrus (Parahippocampal Gyrus).

C.2 Full results of self-supervised learning experiments

C.3 Statistical analysis

We conduct Friedman test for all models and datasets, in most cases it result in significant differences ($p < 0.001$). Bayesian comparison further validates the consistent superiority of our approach over baselines. The detailed statistical analysis results are available upon request and will be attached in the final version of the work upon publication.

D Full ablations

We present the complete results of the ablation studies on alternative design variants for FAPEX. This study systematically investigates the impact of our design choices on model performance.

Experimental setup. ensuring a thorough understanding of their roles in performance, including (1) FAPEX-w/o-Fr-1: Removes the FrNFO module completely; (2) FAPEX-w/o-Fr-2: Replaces fractional convolution with usual convolution; (3) FAPEX-w/o-Phase: Removes the phase inputs fed to Cross-BSSM with the original intermediate representation; (4) FAPEX-w/o-SCA-2: removes the

Table 5: Comparison with Supervised Baselines on CTLE-RATLFP Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	76.57 (71.77, 81.38)	69.24 (63.99, 74.74)	74.05 (69.52, 78.42)	89.54 (78.47, 90.79)	98.01 (95.26, 98.29)	21.09 (17.64, 26.08)
	MRCov [12]	65.93 (61.02, 69.51)	73.02 (67.98, 79.75)	76.38 (72.69, 82.03)	73.03 (67.19, 80.40)	93.34 (92.39, 95.93)	20.67 (16.19, 22.86)
	MultiresConv [42]	74.13 (69.27, 78.15)	63.03 (57.38, 65.84)	68.75 (63.39, 71.15)	87.08 (79.11, 89.80)	97.29 (95.55, 98.16)	26.05 (23.57, 29.86)
	Omni-Scale [45]	65.56 (60.95, 69.27)	75.84 (70.87, 81.74)	78.34 (74.64, 83.36)	72.51 (67.82, 80.96)	93.30 (92.60, 96.09)	19.70 (15.95, 22.11)
	SPARCNNet [23]	62.00 (58.51, 65.17)	72.18 (66.61, 77.59)	75.48 (71.44, 79.90)	67.55 (64.06, 74.65)	92.15 (91.41, 94.09)	21.66 (18.13, 24.09)
Token Mixers	EEGConformer [44]	65.16 (60.91, 70.18)	73.55 (68.95, 80.46)	76.75 (73.23, 82.47)	72.33 (67.51, 80.79)	93.22 (92.52, 96.05)	20.34 (15.93, 22.73)
	EEGMamba [71]	76.02 (68.28, 79.39)	62.05 (58.42, 73.18)	67.98 (64.69, 77.26)	89.50 (77.25, 90.04)	97.86 (95.00, 98.09)	25.75 (17.69, 29.53)
	iTransformer [31]	64.78 (62.79, 68.84)	56.18 (46.20, 71.06)	62.59 (52.42, 75.49)	75.63 (73.78, 78.13)	94.83 (93.94, 95.29)	26.45 (19.00, 29.75)
	Nonformer [30]	72.35 (68.97, 79.61)	72.57 (63.92, 78.43)	76.27 (69.55, 81.35)	80.88 (75.50, 87.96)	95.94 (94.62, 97.65)	19.32 (14.41, 22.26)
	PatchTST [39]	64.79 (60.80, 69.53)	73.27 (68.38, 79.68)	76.55 (72.95, 81.92)	72.02 (67.01, 80.40)	93.31 (92.29, 95.93)	20.73 (16.22, 22.80)
	Pathformer [9]	75.68 (69.99, 83.89)	80.42 (69.94, 83.83)	81.35 (74.40, 85.74)	82.85 (78.98, 91.51)	96.37 (95.46, 98.45)	16.67 (11.56, 20.42)
	TimeMachine [2]	69.66 (66.72, 76.94)	70.79 (61.17, 76.49)	74.51 (67.21, 79.66)	79.92 (72.08, 84.21)	95.74 (93.69, 96.90)	21.06 (16.05, 23.49)
	Seizureformer [17]	71.54 (65.24, 73.60)	56.57 (48.70, 60.33)	62.95 (55.16, 66.29)	84.45 (75.17, 87.86)	96.65 (94.75, 97.64)	29.62 (26.66, 35.73)
Time-Frequency Mixers	ATFNet [65]	67.64 (63.35, 70.92)	54.11 (47.30, 58.57)	60.70 (53.81, 64.54)	78.18 (71.72, 89.36)	95.21 (93.80, 97.79)	30.88 (28.37, 36.56)
	FreTS [67]	59.56 (55.60, 62.49)	49.23 (40.26, 60.84)	56.45 (46.50, 66.88)	68.45 (59.12, 73.81)	92.50 (89.65, 94.29)	28.10 (23.72, 33.37)
	NFM [26]	62.81 (58.02, 67.72)	46.50 (34.33, 54.23)	52.49 (37.69, 60.96)	73.96 (65.27, 82.99)	93.95 (91.60, 96.57)	36.52 (31.72, 44.97)
	TSALNet [15]	77.06 (69.80, 79.79)	65.33 (62.84, 75.27)	70.69 (68.62, 78.91)	88.94 (76.86, 91.20)	97.70 (94.68, 98.40)	24.93 (16.66, 27.20)
Multiscale Token Mixers	AdaWaveNet [68]	70.91 (67.42, 72.59)	55.16 (50.37, 62.65)	61.43 (56.66, 68.42)	83.56 (80.13, 90.32)	96.77 (95.88, 98.25)	30.93 (25.84, 34.99)
	Medformer [54]	76.70 (69.99, 79.22)	70.00 (65.12, 79.16)	74.35 (70.56, 81.95)	86.72 (78.25, 89.49)	97.25 (95.44, 97.96)	22.26 (14.54, 24.44)
	MSTS [73]	63.00 (60.69, 66.94)	70.75 (67.71, 76.44)	74.67 (72.22, 79.29)	69.54 (66.09, 76.53)	92.84 (92.09, 94.73)	21.19 (17.43, 23.70)
	Pyraformer [28]	72.85 (68.72, 78.29)	60.91 (58.43, 72.21)	66.93 (64.68, 76.46)	86.46 (78.09, 89.20)	97.23 (95.24, 97.91)	25.47 (18.44, 29.94)
	SimpleTM [7]	64.18 (60.56, 67.12)	72.55 (70.09, 79.91)	76.02 (74.07, 81.99)	70.91 (66.71, 79.04)	93.05 (92.18, 95.52)	20.84 (16.78, 23.06)
	TimesNet [62]	66.16 (63.78, 76.63)	63.43 (54.43, 67.10)	69.03 (61.20, 72.21)	74.25 (70.52, 85.10)	94.11 (93.12, 96.99)	23.49 (18.43, 25.64)
	TimeMixer [51]	75.31 (72.27, 79.00)	68.97 (60.05, 72.78)	73.81 (65.98, 76.95)	87.92 (81.79, 88.97)	97.74 (96.26, 98.00)	21.33 (18.78, 27.95)
This Work	FAPEX- <i>Small</i>	80.22 (74.46, 82.48)	76.86 (67.33, 80.69)	80.24 (72.38, 83.38)	89.87 (83.11, 91.44)	98.05 (96.50, 98.33)	16.61 (13.69, 23.61)
	FAPEX- <i>Base</i>	81.78 (73.36, 83.37)	81.78 (73.75, 83.58)	83.17 (77.37, 85.41)	91.16 (80.62, 92.75)	98.21 (95.88, 98.63)	14.43 (12.73, 21.81)

Table 6: Comparison with Supervised Baselines on FMCE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	72.27 (69.83, 77.15)	79.56 (76.23, 82.41)	89.29 (83.84, 93.22)	88.65 (83.81, 91.37)	71.99 (58.01, 86.88)	16.85 (14.41, 20.79)
	MRCov [12]	72.33 (69.08, 78.72)	78.83 (74.86, 81.25)	89.87 (83.91, 92.28)	88.21 (83.88, 90.53)	72.50 (55.20, 85.07)	17.90 (15.25, 19.61)
	MultiresNet [42]	72.71 (69.99, 79.92)	75.76 (72.83, 82.90)	82.53 (74.48, 88.81)	83.33 (79.87, 89.35)	66.87 (55.27, 81.04)	19.11 (13.77, 21.75)
	Omni-Scale [45]	74.65 (71.50, 82.34)	79.09 (74.36, 83.88)	84.14 (81.64, 87.34)	83.68 (79.86, 90.92)	68.57 (57.81, 82.83)	16.59 (12.73, 19.96)
	SPARCNNet [23]	56.89 (53.53, 63.04)	60.36 (54.37, 65.11)	67.74 (58.69, 85.24)	65.89 (57.43, 71.23)	51.02 (27.06, 60.39)	25.79 (23.24, 30.51)
Token Mixers	EEGConformer [44]	71.74 (67.67, 78.08)	72.96 (70.07, 79.88)	85.03 (73.81, 87.90)	81.38 (76.65, 89.35)	63.36 (49.89, 82.27)	19.38 (14.93, 21.79)
	EEGMamba [71]	61.04 (55.32, 65.05)	68.89 (59.03, 78.92)	95.45 (91.17, 96.95)	85.73 (81.15, 87.21)	65.08 (47.90, 82.26)	29.38 (23.29, 31.94)
	iTransformer [31]	65.89 (61.69, 69.49)	68.39 (63.22, 74.95)	80.04 (67.25, 87.59)	78.31 (69.47, 82.90)	54.65 (40.79, 75.80)	23.91 (21.07, 26.14)
	Nonformer [30]	72.50 (68.49, 75.62)	74.93 (69.62, 78.37)	82.61 (72.46, 87.59)	86.53 (77.01, 88.11)	64.06 (43.68, 81.80)	24.70 (20.55, 30.86)
	PatchTST [39]	65.91 (62.59, 72.16)	74.24 (69.30, 79.06)	91.83 (85.78, 95.97)	86.61 (83.78, 90.36)	69.76 (52.26, 84.33)	22.73 (20.46, 25.27)
	Pathformer [9]	72.05 (67.09, 77.68)	77.91 (73.11, 81.52)	91.17 (85.57, 93.17)	88.86 (84.46, 89.70)	72.02 (54.96, 86.00)	19.41 (15.43, 22.09)
	TimeMachine [2]	67.29 (64.55, 71.98)	70.07 (66.50, 78.70)	79.86 (71.72, 87.98)	78.38 (72.05, 84.84)	55.86 (39.58, 78.23)	22.19 (18.45, 25.88)
	Seizureformer [17]	67.33 (66.34, 71.86)	73.59 (68.83, 77.94)	78.78 (75.60, 87.93)	79.31 (73.91, 84.71)	58.55 (49.07, 77.78)	21.67 (16.91, 23.27)
Time-Frequency Mixers	ATFNet [65]	72.10 (66.09, 74.63)	73.37 (68.40, 74.11)	81.34 (71.13, 84.02)	82.73 (74.72, 88.00)	61.51 (39.62, 75.76)	25.80 (24.42, 31.51)
	FreTS [67]	57.22 (52.49, 60.49)	62.02 (57.93, 68.30)	68.26 (60.33, 77.89)	61.39 (50.23, 66.52)	39.04 (30.86, 57.42)	28.76 (26.65, 32.41)
	NFM [26]	64.97 (61.86, 69.35)	74.44 (66.98, 79.32)	91.84 (84.12, 94.23)	86.42 (82.13, 88.78)	68.57 (52.15, 83.11)	23.85 (19.92, 25.90)
	TSALNet [15]	65.91 (62.59, 72.16)	74.24 (69.30, 79.06)	91.83 (85.78, 95.97)	86.61 (83.78, 90.36)	69.76 (52.26, 84.33)	22.73 (20.46, 25.27)
Multiscale Token Mixers	AdaWaveNet [68]	75.06 (69.71, 78.69)	76.56 (70.80, 80.58)	82.60 (74.86, 88.15)	87.32 (76.08, 90.08)	64.98 (45.45, 83.15)	23.16 (18.68, 29.51)
	Medformer [54]	76.23 (69.77, 79.40)	77.76 (71.66, 79.71)	83.63 (76.04, 88.51)	88.59 (76.39, 90.37)	65.77 (48.08, 84.04)	21.73 (19.37, 28.72)
	MSTS [73]	68.87 (64.44, 73.13)	75.52 (72.19, 78.89)	89.28 (84.33, 96.35)	87.73 (83.77, 89.35)	71.24 (53.52, 85.86)	19.53 (17.58, 22.92)
	Pyraformer [28]	55.50 (51.84, 62.11)	67.01 (55.96, 78.37)	96.31 (91.60, 98.46)	79.78 (75.64, 87.19)	61.03 (45.38, 77.49)	31.26 (25.83, 39.43)
	SimpleTM [7]	70.93 (68.76, 77.09)	74.36 (71.07, 79.03)	80.52 (69.66, 85.95)	80.57 (78.42, 86.37)	63.30 (53.00, 78.43)	20.70 (16.54, 23.08)
	TimesNet [62]	62.92 (58.51, 67.94)	66.98 (55.91, 69.78)	74.61 (57.99, 79.37)	73.36 (63.59, 77.39)	53.71 (30.91, 66.65)	23.68 (21.72, 28.95)
	TimeMixer [51]	75.15 (71.76, 82.78)	78.78 (73.58, 83.96)	82.93 (77.01, 87.63)	84.47 (79.75, 90.31)	69.31 (55.93, 83.27)	17.25 (12.30, 19.80)
This Work	FAPEX- <i>Small</i>	88.02 (79.42, 93.69)	87.42 (75.43, 93.08)	90.56 (64.97, 96.77)	97.04 (91.11, 98.67)	95.58 (88.69, 98.42)	8.61 (5.72, 15.88)
	FAPEX- <i>Base</i>	89.35 (81.49, 95.14)	88.83 (82.07, 94.00)	90.71 (69.69, 97.57)	97.19 (92.79, 99.22)	95.84 (90.47, 99.58)	8.21 (4.99, 12.96)

(5) FAPEX-w/o-SCA: removes the SCA module; (6) FAPEX-w/o-Hermite: removes the Hermite window in the implicit MLP for generation of fractional convolutional kernels.

Results and discussion. The experimental results, presented in Table 25, reveal nuanced trade-offs among the various ablation configurations. Notably, our Official Design consistently outperforms the alternatives, strongly supporting our insights into electrode embedding. These findings validate the synergistic contributions of the fractional neural frame, phase-aware processing, and state-space duality modules in enhancing the model’s ability to extract informative representations. The performance gains are driven by the innovative integration of fractional convolution, phase-amplitude interaction and spatial correlation aggregation modules, which collectively enable robust and generalizable learning. These experiments collectively validate the essential contributions of each architectural component, underscoring the design rationale of FAPEX and its robust performance across diverse scenarios.

Table 7: Comparison with Supervised Baselines on HUP Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]11	65.45 (61.37, 68.07)	71.33 (67.40, 76.16)	70.30 (65.24, 73.03)	67.35 (66.11, 72.36)	58.19 (52.52, 73.35)	21.59 (20.12, 23.78)
	MRConv [12]	64.06 (61.42, 66.54)	69.06 (65.87, 72.34)	67.80 (65.07, 71.23)	66.08 (64.45, 70.80)	57.33 (51.23, 70.52)	22.51 (20.80, 24.31)
	MultiresConv [42]	61.40 (58.07, 62.91)	65.35 (64.27, 66.77)	63.82 (61.26, 65.43)	65.45 (62.89, 69.99)	54.24 (50.03, 62.50)	23.22 (22.43, 24.62)
	Omni-Scale [45]	65.13 (60.67, 66.25)	70.66 (66.59, 73.48)	68.72 (64.45, 70.79)	67.12 (65.10, 71.48)	57.36 (51.73, 72.25)	21.75 (20.28, 24.05)
	SPaRCNet [23]	58.68 (55.59, 61.51)	61.63 (59.14, 63.77)	60.17 (58.98, 64.06)	62.80 (61.30, 68.29)	56.11 (47.76, 60.71)	24.58 (24.33, 26.37)
Token Mixers	EEGConformer [44]	60.98 (57.53, 62.01)	64.61 (61.54, 66.83)	63.34 (62.04, 64.66)	64.56 (59.63, 66.59)	55.82 (41.38, 64.10)	23.20 (23.01, 23.79)
	EEGMamba [71]	59.67 (56.41, 61.69)	63.19 (61.05, 64.20)	61.34 (59.76, 64.15)	62.76 (61.39, 68.17)	56.54 (47.50, 61.52)	24.49 (24.04, 25.14)
	iTransformer [31]	55.91 (52.46, 57.65)	58.27 (53.87, 61.17)	57.67 (54.30, 59.36)	57.38 (56.47, 60.63)	46.84 (39.36, 52.76)	24.65 (24.14, 25.65)
	Nonformer [30]	59.08 (55.82, 61.33)	62.11 (59.25, 65.73)	61.92 (59.67, 63.50)	62.36 (59.42, 65.68)	52.70 (40.28, 62.76)	23.81 (23.54, 23.90)
	PatchTST [39]	66.17 (61.04, 67.35)	70.83 (67.91, 73.43)	69.26 (65.20, 72.24)	67.50 (66.32, 71.90)	57.66 (52.56, 73.09)	21.79 (20.34, 23.89)
	Pathformer [9]	62.35 (60.25, 64.50)	65.73 (62.08, 66.88)	65.20 (61.27, 66.89)	66.30 (63.47, 66.86)	58.12 (50.98, 64.75)	23.85 (23.14, 24.25)
	TimeMachine [2]	61.43 (58.70, 63.11)	65.90 (63.52, 67.53)	63.97 (62.08, 65.96)	68.02 (64.82, 69.68)	58.26 (55.46, 59.58)	23.58 (22.88, 25.36)
	Seizureformer [17]	58.32 (53.69, 59.17)	59.68 (56.56, 61.79)	59.36 (56.14, 61.18)	61.01 (58.32, 62.53)	53.18 (40.53, 58.93)	24.10 (23.43, 25.00)
Time-Frequency Mixers	ATFNet [65]	58.54 (55.92, 60.09)	64.48 (63.60, 65.05)	60.36 (58.14, 63.27)	63.57 (59.53, 65.04)	54.85 (47.34, 62.94)	24.15 (22.77, 25.55)
	FreTS [67]	50.45 (50.06, 53.03)	62.62 (61.02, 64.09)	50.09 (46.81, 58.81)	57.20 (56.18, 60.61)	46.41 (45.05, 49.85)	24.44 (24.32, 24.63)
	NFM [26]	61.17 (59.37, 63.13)	62.70 (60.15, 66.17)	63.39 (61.31, 64.97)	64.43 (63.58, 66.13)	57.22 (48.68, 64.10)	24.48 (23.54, 25.40)
	TSANet [15]	63.02 (59.38, 65.08)	67.86 (66.08, 69.78)	66.47 (63.07, 68.07)	67.54 (64.16, 70.56)	56.59 (51.66, 68.90)	22.54 (21.04, 24.20)
Multiscale Token Mixers	AdaWaveNet [68]	61.28 (55.27, 62.98)	64.00 (63.48, 64.63)	63.22 (59.51, 64.30)	64.38 (58.58, 67.02)	59.01 (45.89, 62.79)	23.63 (22.80, 25.09)
	Medformer [54]	62.28 (56.48, 63.68)	64.75 (64.08, 65.74)	64.34 (60.85, 64.97)	65.00 (58.67, 67.33)	59.30 (47.25, 63.93)	23.63 (22.35, 24.88)
	MTST [73]	61.06 (60.23, 63.16)	65.88 (64.08, 67.45)	64.43 (63.75, 67.04)	66.21 (63.03, 70.20)	55.68 (53.43, 66.87)	23.60 (21.47, 23.52)
	Pyraformer [28]	55.96 (54.43, 56.66)	60.49 (58.87, 61.84)	58.20 (56.79, 61.50)	58.91 (55.30, 60.46)	49.75 (41.52, 50.42)	27.03 (26.47, 27.89)
	SimpleTM [7]	62.07 (60.61, 65.06)	67.78 (64.46, 68.64)	65.24 (64.35, 67.86)	68.52 (66.69, 70.98)	60.44 (59.11, 65.46)	23.09 (22.07, 24.17)
	TimesNet [62]	60.27 (56.05, 60.53)	65.90 (64.68, 66.17)	61.68 (60.03, 64.00)	65.37 (63.66, 66.60)	56.23 (53.59, 62.69)	24.02 (22.74, 26.06)
	[51]	63.15 (61.72, 65.34)	67.12 (65.19, 69.06)	66.15 (63.96, 70.44)	67.34 (64.50, 68.70)	59.55 (55.26, 68.47)	22.66 (21.63, 23.57)
This Work	FAPEX-Small	69.72 (66.25, 73.56)	72.58 (68.95, 75.75)	71.99 (68.30, 75.28)	78.33 (72.79, 79.92)	71.91 (67.77, 75.75)	19.12 (17.57, 20.97)
	FAPEX-Base	69.05 (67.48, 74.31)	73.65 (68.73, 76.88)	72.55 (68.52, 76.08)	79.30 (72.54, 81.49)	73.08 (69.76, 76.91)	18.80 (16.97, 21.23)

Table 8: Comparison with Supervised Baselines on IEES Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	62.37 (60.07, 63.47)	73.39 (66.55, 81.35)	73.37 (71.78, 81.49)	67.24 (66.55, 71.13)	76.30 (17.70, 85.23)	25.39 (22.54, 26.83)
	MRConv [12]	61.08 (58.28, 62.68)	68.76 (50.89, 71.17)	68.65 (54.29, 74.53)	66.87 (66.03, 68.51)	67.38 (21.30, 85.23)	31.43 (29.38, 40.53)
	MultiresNet [42]	60.98 (56.51, 63.66)	72.07 (52.22, 74.24)	70.37 (61.98, 78.80)	68.70 (67.00, 71.73)	73.63 (18.40, 85.18)	27.63 (24.55, 34.87)
	Omni-Scale [45]	61.21 (59.00, 61.94)	67.88 (49.00, 71.36)	68.70 (50.81, 74.94)	67.19 (65.35, 69.69)	71.84 (21.68, 85.24)	31.00 (28.48, 42.04)
	SPaRCNet [23]	58.14 (57.07, 58.23)	60.70 (51.48, 70.88)	64.94 (56.25, 70.99)	61.42 (59.91, 62.87)	60.87 (18.16, 80.95)	36.35 (27.77, 41.91)
Token Mixers	EEGConformer [44]	60.32 (56.93, 62.20)	66.13 (45.07, 69.94)	67.88 (50.92, 72.00)	67.01 (65.39, 71.53)	69.93 (19.28, 83.85)	32.03 (30.14, 44.62)
	EEGMamba [71]	61.14 (57.72, 63.56)	69.60 (45.48, 72.33)	69.98 (52.04, 75.14)	68.80 (66.06, 72.30)	76.23 (21.57, 84.71)	30.79 (28.70, 44.64)
	iTransformer [31]	56.66 (54.96, 60.46)	53.37 (43.00, 62.11)	54.55 (40.33, 64.61)	66.40 (64.98, 69.75)	71.49 (18.73, 83.82)	40.26 (34.52, 48.41)
	Nonformer [30]	62.99 (62.09, 63.72)	69.69 (65.42, 78.15)	74.66 (66.47, 78.05)	68.92 (65.42, 70.85)	68.76 (22.91, 87.02)	29.77 (23.30, 33.78)
	PatchTST [39]	61.30 (58.90, 62.68)	61.63 (46.22, 64.87)	63.78 (44.06, 67.27)	67.49 (66.60, 69.13)	70.10 (22.37, 85.11)	34.28 (32.53, 45.22)
	Pathformer [9]	61.32 (59.36, 63.89)	71.32 (52.76, 74.85)	72.14 (56.61, 77.38)	68.63 (65.47, 70.37)	71.30 (21.24, 86.97)	29.59 (26.58, 39.24)
	TimeMachine [2]	61.27 (59.51, 63.69)	69.90 (52.66, 73.45)	70.08 (55.00, 76.71)	67.38 (65.58, 70.15)	71.82 (23.15, 85.78)	29.66 (27.33, 39.81)
	Seizureformer [17]	60.58 (58.68, 62.13)	69.66 (51.98, 74.76)	69.94 (56.78, 78.37)	66.72 (65.38, 68.60)	72.13 (20.70, 86.30)	29.55 (26.25, 40.01)
Time-Frequency Mixers	ATFNet [65]	55.33 (51.91, 58.48)	59.67 (39.46, 73.16)	56.24 (40.06, 65.89)	68.19 (66.66, 73.15)	75.19 (17.56, 83.89)	35.22 (32.09, 46.34)
	FreTS [67]	51.06 (50.41, 54.66)	42.66 (10.71, 52.18)	32.77 (4.79, 51.17)	66.95 (61.98, 70.33)	66.57 (20.41, 81.34)	50.20 (43.31, 73.35)
	NFM [26]	57.57 (55.53, 60.10)	52.40 (49.31, 59.62)	56.25 (51.02, 66.22)	61.19 (56.70, 63.80)	57.26 (18.36, 81.23)	41.07 (37.27, 42.93)
	TSANet [15]	60.62 (59.39, 62.78)	73.77 (59.57, 80.29)	72.93 (68.57, 79.70)	66.25 (65.70, 71.73)	74.79 (21.26, 87.06)	29.12 (23.47, 34.28)
Multiscale Token Mixers	AdaWaveNet [68]	59.74 (58.67, 62.79)	70.05 (51.13, 72.24)	70.74 (52.22, 76.48)	66.27 (65.26, 70.22)	71.54 (22.74, 84.63)	30.56 (28.04, 40.58)
	Medformer [54]	60.71 (58.37, 64.65)	73.71 (58.10, 76.47)	73.07 (66.85, 79.80)	66.90 (65.38, 69.23)	71.64 (20.78, 86.50)	28.87 (25.01, 35.73)
	MTST [73]	62.92 (59.81, 64.20)	60.27 (45.89, 77.39)	56.37 (54.19, 76.46)	70.09 (67.67, 81.74)	81.84 (22.97, 87.20)	26.84 (22.70, 32.36)
	Pyraformer [28]	60.00 (57.16, 63.08)	71.38 (50.62, 74.64)	70.21 (60.69, 79.33)	66.88 (66.11, 73.83)	72.44 (16.59, 84.32)	26.93 (24.62, 35.17)
	SimpleTM [7]	60.91 (59.17, 64.51)	66.73 (58.25, 73.16)	68.76 (62.80, 73.44)	64.43 (63.02, 68.76)	62.43 (21.70, 82.90)	33.84 (26.28, 37.64)
	TimesNet [62]	60.50 (56.91, 61.26)	59.93 (49.74, 64.73)	63.65 (49.91, 66.67)	66.09 (65.26, 68.54)	67.79 (19.49, 84.80)	35.59 (33.52, 43.53)
	TimeMixer [51]	61.30 (59.07, 64.65)	71.65 (58.57, 78.39)	71.02 (67.77, 81.17)	68.64 (66.71, 72.84)	77.58 (20.25, 85.12)	25.54 (24.14, 30.42)
This Work	FAPEX-Small	65.19 (62.27, 69.83)	70.83 (57.26, 77.92)	70.40 (64.56, 77.86)	71.65 (68.09, 83.63)	82.91 (26.06, 88.90)	21.21 (19.95, 26.19)
	FAPEX-Base	65.67 (63.15, 70.48)	72.26 (59.73, 78.10)	72.41 (66.28, 78.13)	71.44 (68.72, 82.50)	82.96 (23.26, 88.36)	20.43 (19.83, 25.31)

E Baselines

- Pathformer [9] divides the input series into multi-resolution patches and applies dual attention—inter-patch to capture global context and intra-patch to capture fine-grained details. An adaptive pathway controller dynamically routes information through the most relevant scales, allowing the model to flexibly emphasize long- vs. short-term patterns. This design yields state-of-the-art forecasting on both univariate and multivariate benchmarks without blowing up compute. The code is available at <https://github.com/decisionintelligence/pathformer>.
- ModernTCN [34] is a purely convolutional backbone that expands the effective receptive field via stacked dilated convolutions and interleaved cross-variable mixing layers. By carefully scaling kernel sizes and dilation rates, it achieves linear complexity while capturing both local signal details and global trends, making it a versatile drop-in for forecasting, classification, anomaly detection, and more. The code is available at <https://github.com/luodhhh/ModernTCN>.

Table 9: Comparison with Supervised Baselines on KAIME Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	72.81 (66.12, 82.95)	83.40 (67.79, 86.27)	73.18 (54.08, 92.27)	87.32 (73.73, 90.47)	93.09 (79.92, 93.39)	17.45 (12.54, 25.00)
	MRCov [12]	76.28 (65.49, 83.74)	81.08 (67.40, 85.28)	68.51 (57.72, 85.72)	84.98 (72.02, 90.87)	90.32 (76.09, 92.13)	15.52 (12.52, 21.93)
	MultiresNet [42]	74.24 (63.93, 81.48)	80.36 (66.06, 83.15)	63.70 (58.62, 77.93)	82.55 (70.08, 89.86)	90.08 (74.38, 91.04)	16.44 (13.60, 22.11)
	Omni-Scale [45]	76.04 (65.84, 82.12)	80.99 (67.06, 84.36)	68.85 (64.42, 81.93)	83.02 (71.09, 89.99)	89.91 (73.89, 91.05)	16.24 (12.96, 22.05)
	SPaRCNet [23]	75.94 (64.80, 84.60)	81.96 (65.55, 86.23)	77.12 (71.98, 89.03)	86.50 (71.78, 90.61)	92.34 (78.02, 92.58)	15.09 (11.04, 21.91)
Token Mixers	EEGConformer [44]	67.60 (62.86, 78.19)	81.43 (65.08, 81.92)	73.40 (45.31, 91.28)	87.08 (69.32, 89.00)	89.92 (75.79, 92.24)	22.25 (16.56, 34.25)
	EEGMamba [71]	72.69 (60.58, 80.88)	80.39 (62.52, 82.49)	69.42 (53.26, 88.06)	83.41 (76.02, 89.02)	89.93 (81.44, 92.15)	16.86 (13.10, 24.76)
	iTransformer [31]	68.99 (64.76, 78.42)	81.26 (66.52, 81.83)	63.41 (44.28, 77.67)	87.07 (72.22, 90.36)	93.02 (78.84, 93.55)	18.97 (14.23, 25.43)
	Nonformer [30]	74.47 (62.27, 82.14)	79.50 (62.11, 83.46)	90.30 (76.82, 94.97)	81.61 (74.19, 88.02)	89.99 (79.76, 90.69)	16.82 (12.78, 26.14)
	PatchTST [39]	71.26 (64.09, 83.65)	83.00 (66.59, 85.68)	73.82 (59.75, 85.43)	88.46 (70.34, 93.03)	92.72 (77.91, 94.53)	20.06 (12.36, 32.35)
	Pathformer [9]	73.67 (60.90, 82.50)	80.64 (63.61, 84.07)	67.40 (49.54, 85.17)	85.27 (71.15, 89.66)	91.81 (77.01, 92.38)	16.96 (12.41, 23.56)
	TimeMachine [2]	68.11 (60.57, 75.91)	77.57 (63.38, 79.60)	54.26 (42.40, 64.24)	83.23 (69.82, 86.25)	89.01 (75.97, 91.41)	19.29 (15.86, 27.51)
	TimeMixer [51]	65.88 (60.76, 75.41)	77.27 (63.15, 80.39)	53.38 (32.33, 78.77)	85.90 (70.39, 89.54)	89.63 (77.35, 92.71)	25.02 (20.49, 34.51)
	ATFNet [65]	60.03 (57.13, 61.59)	64.97 (59.18, 72.91)	45.74 (24.41, 77.11)	71.94 (64.07, 74.20)	66.62 (65.84, 82.83)	22.71 (22.12, 26.89)
Time-Frequency Mixers	FreTS [66]	57.84 (53.40, 59.60)	54.27 (48.08, 57.41)	56.36 (11.19, 94.40)	73.78 (69.17, 76.99)	74.78 (72.29, 81.29)	25.44 (22.85, 28.88)
	NFM [25]	74.69 (63.23, 75.84)	76.83 (63.64, 79.60)	64.41 (58.40, 69.09)	80.37 (68.99, 82.38)	83.52 (71.09, 88.59)	17.83 (17.72, 23.11)
	TSLANet [16]	72.75 (63.71, 81.55)	82.25 (64.61, 84.77)	88.40 (61.01, 94.44)	82.37 (68.78, 88.35)	89.44 (75.47, 90.56)	18.21 (14.40, 24.62)
Multiscale Token Mixers	AdaWaveNet [68]	65.24 (60.38, 70.36)	70.49 (60.85, 75.05)	77.73 (50.48, 94.04)	84.85 (67.78, 92.33)	91.24 (76.00, 94.06)	28.04 (24.34, 36.61)
	Medformer [54]	60.51 (56.21, 71.19)	73.17 (61.85, 74.59)	45.23 (20.10, 67.42)	72.30 (70.06, 80.38)	80.48 (73.57, 85.60)	20.71 (20.50, 27.17)
	MTST [73]	64.72 (57.52, 65.89)	59.03 (57.97, 62.86)	60.18 (24.76, 81.65)	74.25 (68.90, 78.33)	74.71 (72.49, 83.05)	25.34 (22.85, 26.18)
	Pyrformer [28]	72.52 (63.58, 81.21)	83.16 (66.63, 83.57)	56.59 (52.01, 75.34)	86.43 (72.04, 90.19)	93.07 (78.93, 93.39)	17.37 (12.88, 24.57)
	SimpleTM [7]	76.04 (66.91, 82.45)	79.95 (67.65, 83.97)	81.06 (66.65, 86.66)	84.62 (74.11, 88.23)	87.84 (78.94, 91.88)	16.34 (13.27, 21.00)
	TimesNet [62]	76.04 (66.91, 82.45)	79.95 (67.65, 83.97)	81.06 (66.65, 86.66)	84.62 (74.11, 88.23)	87.84 (78.94, 91.88)	16.34 (13.27, 21.00)
	TimeMixer [51]	75.96 (65.48, 83.39)	82.13 (65.09, 85.29)	89.96 (70.56, 94.35)	85.22 (76.24, 90.35)	91.99 (82.46, 93.46)	15.58 (11.68, 22.85)
	FAPEX-Small	81.46 (72.20, 86.63)	86.91 (72.25, 88.73)	92.12 (74.82, 97.21)	89.32 (79.85, 92.66)	94.64 (85.44, 94.78)	12.29 (10.19, 20.03)
This Work	FAPEX-Base	82.23 (72.24, 86.84)	87.04 (72.43, 88.54)	95.62 (80.63, 98.05)	90.08 (81.15, 93.87)	95.23 (86.49, 95.70)	12.22 (10.31, 19.54)

Table 10: Comparison with Supervised Baselines on LPIRE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Convolutional Mixers	ModernTCN [34]	72.66 (63.96, 75.56)	68.70 (60.86, 71.46)	72.62 (71.14, 77.28)	80.35 (69.80, 83.52)	97.33 (97.23, 97.80)	19.53 (18.83, 23.12)
	MRCov [12]	63.04 (61.86, 70.56)	60.50 (56.86, 68.88)	68.42 (66.92, 77.42)	68.70 (66.58, 78.12)	97.03 (93.25, 97.65)	24.46 (20.04, 25.63)
	MultiresNet [42]	71.96 (63.21, 74.61)	67.08 (56.40, 69.81)	71.68 (67.46, 74.80)	80.12 (69.47, 83.14)	97.49 (97.17, 97.69)	20.34 (18.66, 25.44)
	Omni-Scale [45]	68.90 (60.92, 69.20)	51.92 (47.91, 67.81)	65.07 (59.82, 73.95)	75.70 (66.66, 78.24)	96.89 (95.61, 97.62)	27.58 (20.69, 29.83)
	SPaRCNet [23]	50.94 (49.75, 51.13)	43.58 (41.51, 48.23)	48.17 (47.19, 50.19)	50.84 (50.13, 51.24)	86.65 (79.50, 90.00)	39.78 (34.58, 43.91)
Token Mixers	EEGConformer [44]	62.38 (61.63, 68.99)	58.48 (49.66, 61.34)	65.82 (58.20, 70.16)	70.19 (65.49, 75.91)	97.22 (93.00, 97.78)	26.38 (22.56, 27.73)
	EEGMamba [71]	72.09 (62.62, 74.41)	63.69 (54.82, 68.73)	70.65 (66.13, 72.45)	80.69 (69.11, 83.12)	97.48 (97.13, 97.79)	21.44 (19.25, 25.99)
	iTransformer [31]	49.51 (48.74, 50.79)	40.90 (38.08, 46.91)	45.98 (41.54, 48.65)	49.44 (48.82, 51.44)	86.13 (78.60, 90.25)	50.42 (35.80, 47.57)
	Nonformer [30]	66.76 (61.55, 71.31)	64.25 (59.97, 72.92)	74.40 (71.33, 77.78)	76.14 (66.71, 79.10)	96.89 (95.74, 97.77)	22.55 (19.01, 24.46)
	PatchTST [39]	52.17 (50.70, 52.67)	50.21 (39.76, 53.01)	50.39 (49.33, 57.59)	52.58 (50.89, 52.92)	88.04 (79.61, 90.27)	36.57 (33.01, 39.95)
	Pathformer [9]	72.57 (63.62, 75.27)	68.74 (60.43, 71.39)	73.01 (70.80, 77.08)	80.06 (69.70, 83.07)	97.48 (97.18, 97.75)	19.90 (18.38, 23.53)
	TimeMachine [2]	62.96 (60.58, 73.93)	75.98 (70.23, 78.98)	80.99 (75.30, 83.93)	68.94 (65.73, 82.20)	97.09 (93.07, 97.94)	19.41 (17.89, 21.83)
	Seizformer [51]	62.78 (62.16, 70.03)	65.63 (61.50, 71.77)	73.23 (70.37, 78.86)	68.15 (66.27, 79.02)	97.00 (93.20, 97.65)	22.80 (18.95, 24.15)
	ATFNet [65]	49.81 (49.56, 50.80)	40.04 (37.96, 47.25)	44.92 (43.03, 47.73)	50.04 (49.12, 51.40)	86.75 (79.14, 89.94)	40.98 (34.02, 46.75)
Time-Frequency Mixers	FreTS [67]	49.83 (48.69, 50.40)	34.83 (32.66, 45.31)	38.61 (36.75, 44.18)	49.38 (48.34, 50.00)	86.29 (77.91, 89.72)	43.91 (34.78, 50.10)
	NFM [26]	49.12 (48.90, 49.71)	37.30 (34.97, 47.57)	42.95 (38.75, 48.04)	48.83 (48.08, 50.20)	86.25 (78.50, 89.55)	39.07 (33.66, 48.66)
	TSLANet [15]	72.20 (63.70, 75.30)	68.63 (60.17, 70.23)	72.17 (70.59, 76.43)	79.92 (69.25, 83.32)	97.38 (97.16, 97.71)	19.95 (18.98, 23.68)
Multiscale Token Mixers	AdaWaveNet [68]	69.45 (61.10, 70.37)	54.09 (51.82, 66.24)	66.25 (64.05, 73.80)	76.18 (66.70, 80.59)	96.88 (95.73, 97.84)	26.14 (20.74, 28.46)
	Medformer [54]	72.45 (63.67, 74.98)	66.71 (58.60, 70.38)	72.35 (69.31, 74.66)	80.60 (69.81, 83.07)	97.43 (97.20, 97.77)	20.51 (18.69, 24.44)
	MTST [73]	50.42 (49.34, 51.10)	45.26 (42.14, 46.45)	49.60 (44.30, 50.62)	50.39 (49.31, 51.02)	86.32 (77.70, 90.03)	37.93 (35.71, 42.17)
	Pyrformer [28]	66.50 (61.89, 70.66)	60.25 (56.79, 71.17)	72.12 (67.83, 76.49)	76.19 (67.72, 78.56)	96.98 (95.74, 97.74)	23.40 (19.94, 25.14)
	SimpleTM [7]	51.13 (49.75, 52.24)	47.49 (44.15, 49.07)	51.97 (47.20, 52.46)	51.06 (50.38, 52.91)	86.68 (79.63, 90.05)	40.81 (34.81, 42.29)
	TimesNet [62]	51.21 (49.87, 51.89)	47.12 (41.78, 50.08)	49.38 (47.82, 54.40)	51.54 (50.14, 52.33)	86.83 (78.28, 90.22)	38.44 (36.65, 41.71)
	TimeMixer [51]	72.40 (63.22, 74.82)	66.19 (57.41, 70.26)	72.02 (68.32, 74.18)	80.57 (69.36, 83.10)	97.51 (97.16, 97.76)	20.64 (18.65, 24.71)
	FAPEX-Small	73.42 (64.69, 75.69)	69.33 (63.05, 71.36)	73.38 (72.80, 76.49)	81.22 (70.32, 83.56)	97.36 (97.24, 97.85)	19.27 (18.38, 22.43)
This Work	FAPEX-Base	73.42 (64.72, 75.67)	71.72 (66.57, 72.73)	76.13 (73.73, 78.84)	80.97 (70.29, 83.55)	97.41 (97.28, 97.84)	18.43 (18.10, 21.40)

- Medformer [54] is a specialized multi-granularity patching transformer designed for the classification of medical time series (MedTS). It integrates three innovative mechanisms to harness the distinctive attributes of MedTS data. These mechanisms include cross-channel patching to exploit inter-channel relationships, multi-granularity embedding to capture features across various scales, and a two-stage multi-granularity self-attention process (both intra- and inter-granularity) to learn features and correlations within and between different granularities. This architecture enables the model to generalize effectively in scenarios where patient independence is a critical factor. The code is available at this repository: <https://github.com/DL4mHealth/Medformer>.
- TimeMixer [51] is a fully-MLP model that first down-samples the series into multiple scales, then uses a Past-Decomposable-Mixing (PDM) block to fuse seasonal and trend components across scales, and finally a Future-Multipredictor-Mixing (FMM) block to ensemble per-scale predictors. This two-stage mixing captures both micro and macro temporal patterns with minimal overhead. The code is available at <https://github.com/DL4mHealth/TimeMixer>.

Table 11: Comparison with Supervised Baselines on PCS Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments with different random seeds.

Architecture	Model	Metrics					
		BA (↑)	SEN (↑)	F1 (↑)	ROC (↑)	PRC (↑)	Brier (↓)
Convolutional Mixers	T-WaveNet [34]	55.39 (50.64, 58.68)	70.27 (69.13, 89.66)	61.47 (56.36, 88.62)	71.41 (63.38, 89.40)	76.16 (68.18, 86.61)	24.00 (19.50, 36.90)
	ModernTCN [34]	80.00 (75.38, 86.09)	85.91 (84.06, 87.82)	85.41 (84.27, 87.42)	86.34 (84.74, 87.91)	94.26 (73.00, 95.39)	11.00 (10.56, 11.53)
	MRCov [12]	75.07 (72.76, 80.36)	83.03 (82.27, 86.00)	84.12 (80.84, 86.31)	83.74 (83.27, 86.13)	89.64 (77.79, 93.00)	13.22 (11.07, 15.23)
	MultiresNet [42]	64.23 (62.48, 71.26)	69.21 (53.38, 72.26)	64.41 (56.89, 72.31)	83.88 (74.01, 91.24)	89.27 (40.97, 94.73)	23.09 (19.49, 28.46)
	Omni-Scale [45]	74.57 (64.71, 77.75)	80.03 (77.54, 81.18)	79.57 (76.73, 83.16)	80.86 (77.69, 82.10)	81.92 (74.33, 93.48)	16.63 (13.26, 19.38)
	SPaRCnet [23]	74.78 (59.73, 78.58)	85.49 (73.56, 91.52)	84.36 (71.95, 87.47)	90.96 (69.81, 98.29)	94.17 (66.18, 99.07)	16.94 (13.04, 22.00)
Token Mixers	EEGConformer [44]	71.02 (66.89, 73.73)	77.10 (70.27, 83.04)	78.81 (65.30, 84.63)	84.26 (78.78, 86.80)	90.65 (50.62, 93.65)	18.55 (14.32, 20.41)
	EEGMamba [71]	72.71 (67.09, 73.95)	70.85 (67.53, 80.24)	73.31 (67.45, 79.99)	85.59 (76.90, 86.38)	89.90 (46.10, 93.11)	18.62 (14.30, 22.38)
	iTransformer [31]	65.24 (64.89, 71.89)	74.28 (69.56, 83.31)	73.01 (58.14, 85.42)	83.38 (75.44, 87.52)	89.87 (44.47, 93.25)	18.23 (17.35, 25.33)
	Nonformer [30]	60.52 (56.96, 66.35)	68.84 (62.14, 78.25)	63.67 (56.43, 80.65)	84.12 (74.27, 89.02)	83.86 (37.61, 94.70)	25.83 (20.77, 36.74)
	PatchTST [39]	72.71 (59.29, 78.54)	71.87 (71.22, 89.97)	71.23 (68.13, 86.73)	73.75 (71.32, 88.81)	93.86 (67.72, 97.91)	21.74 (13.76, 26.73)
	Pathformer [9]	73.93 (67.89, 74.93)	78.66 (71.92, 82.19)	80.93 (70.79, 83.82)	83.70 (76.48, 86.76)	88.69 (50.74, 93.33)	17.18 (12.34, 19.19)
	TimeMachine [2]	64.57 (63.42, 71.70)	69.39 (56.45, 74.06)	65.31 (55.44, 76.24)	82.52 (75.23, 89.20)	89.13 (39.63, 93.98)	21.20 (18.22, 24.75)
	xlSTM-Mixer [51]	59.46 (53.04, 61.99)	58.59 (47.79, 72.23)	62.37 (35.06, 67.00)	59.69 (42.84, 72.40)	71.01 (65.37, 76.62)	24.91 (22.74, 44.53)
Time-Frequency Mixers	ATFNet [65]	65.90 (65.17, 71.99)	74.71 (69.78, 81.90)	73.55 (58.17, 84.53)	84.64 (81.73, 87.19)	91.94 (52.41, 92.76)	18.34 (15.87, 24.26)
	FreTS [67]	68.21 (63.70, 70.25)	70.52 (52.48, 76.16)	72.65 (58.82, 75.97)	77.85 (71.75, 83.06)	52.03 (39.54, 86.09)	19.35 (17.61, 31.86)
	NFM [26]	70.10 (66.79, 73.67)	73.59 (69.60, 78.82)	76.86 (64.05, 81.86)	82.97 (73.72, 85.97)	86.97 (48.04, 91.19)	19.10 (14.89, 21.50)
	TSLANet [15]	76.34 (75.17, 84.32)	84.63 (82.46, 84.96)	83.98 (82.70, 84.86)	84.93 (82.77, 85.31)	91.94 (73.86, 93.41)	12.46 (12.14, 13.08)
Multiscale Token Mixers	AdaWaveNet [68]	66.42 (65.36, 72.04)	72.36 (67.02, 84.45)	74.21 (58.74, 84.84)	81.34 (74.70, 86.40)	87.39 (41.50, 91.47)	19.54 (17.16, 23.75)
	Medformer [54]	78.22 (60.54, 79.85)	77.94 (68.56, 85.52)	77.09 (67.49, 84.39)	96.30 (56.89, 98.29)	97.63 (67.03, 99.07)	16.14 (13.01, 36.89)
	MTST [73]	65.82 (61.16, 67.22)	72.76 (67.27, 85.30)	70.33 (61.99, 86.17)	74.14 (67.69, 85.64)	77.73 (72.36, 90.00)	21.83 (18.30, 24.79)
	Pyraformer [28]	74.60 (68.75, 76.02)	76.77 (71.65, 83.72)	79.88 (70.06, 83.91)	82.31 (77.02, 88.47)	88.85 (52.10, 92.98)	17.05 (11.64, 19.38)
	SimpleTM [7]	68.22 (65.51, 72.87)	75.99 (69.87, 81.59)	74.99 (60.39, 84.35)	84.46 (75.84, 86.97)	88.94 (49.80, 93.71)	18.38 (17.03, 22.34)
	TimesNet [62]	77.01 (69.19, 79.59)	77.71 (71.40, 85.24)	81.02 (77.49, 85.02)	84.56 (74.40, 90.20)	87.42 (53.85, 92.40)	15.64 (12.44, 20.59)
	TimeMixer [51]	78.97 (69.61, 80.96)	81.09 (73.36, 84.84)	83.51 (75.09, 87.29)	85.36 (80.27, 88.07)	91.08 (55.35, 93.28)	14.48 (12.88, 19.07)
This Work	FAPEX-Small	81.16 (64.29, 87.70)	80.97 (70.78, 87.30)	81.20 (70.25, 87.01)	94.09 (65.49, 97.94)	97.10 (72.54, 98.99)	13.05 (10.00, 27.00)
	FAPEX-Base	90.74 (80.87, 96.05)	91.53 (80.54, 95.85)	91.53 (81.09, 95.85)	96.32 (88.07, 98.09)	98.17 (91.80, 98.97)	7.45 (3.87, 13.60)

Table 12: Comparison with Supervised Baselines on RESP Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Architecture	Model	Metrics					
		BA (↑)	SEN (↑)	F1 (↑)	ROC (↑)	PRC (↑)	Brier (↓)
Convolutional Mixers	T-Wavelet [20]	54.81 (49.95, 73.26)	66.67 (49.80, 77.96)	77.01 (66.10, 78.65)	56.72 (52.10, 81.04)	32.42 (0.39, 76.69)	23.80 (19.37, 36.11)
	ModernTCN [34]	68.51 (58.63, 76.08)	70.05 (43.89, 88.80)	74.95 (51.67, 89.37)	80.01 (54.92, 90.41)	49.01 (10.51, 76.13)	24.03 (13.53, 39.63)
	MRCov [12]	66.35 (54.16, 68.61)	71.13 (46.12, 89.97)	72.70 (41.17, 90.03)	73.79 (53.14, 88.87)	49.08 (15.63, 68.05)	25.13 (14.28, 40.84)
	MultiresConv [42]	58.18 (50.39, 74.13)	62.37 (52.24, 77.62)	75.22 (68.27, 78.98)	61.73 (50.02, 83.01)	33.43 (0.45, 81.27)	24.84 (21.55, 29.67)
	Omni-Scale [45]	67.39 (56.74, 74.50)	69.18 (44.82, 88.73)	73.26 (44.00, 89.17)	75.67 (56.76, 90.31)	51.55 (6.54, 70.44)	25.28 (13.12, 41.65)
	SPaRCnet [23]	60.90 (51.94, 73.18)	73.65 (51.77, 82.30)	81.29 (67.90, 82.68)	62.98 (50.30, 82.92)	35.17 (1.02, 83.53)	23.13 (14.86, 25.32)
Token Mixers	EEGConformer [44]	70.02 (53.42, 75.04)	82.54 (45.42, 92.79)	84.59 (32.35, 91.85)	86.46 (74.87, 93.02)	70.36 (11.41, 82.64)	14.07 (10.12, 45.47)
	EEGMamba [71]	65.12 (53.48, 69.10)	79.58 (57.74, 88.10)	80.02 (72.30, 91.12)	73.04 (64.97, 81.72)	46.94 (0.83, 79.35)	21.97 (16.63, 24.27)
	iTransformer [31]	74.99 (53.70, 78.41)	80.48 (45.83, 94.71)	84.66 (35.50, 93.91)	87.87 (75.33, 96.03)	78.00 (11.76, 85.35)	14.56 (8.77, 43.51)
	Nonformer [30]	66.55 (54.97, 80.10)	74.09 (45.26, 82.36)	78.69 (37.49, 87.30)	81.57 (61.95, 90.74)	71.40 (5.14, 79.37)	19.34 (12.87, 45.00)
	PatchTST [39]	67.66 (50.30, 72.73)	86.51 (45.30, 89.90)	86.94 (28.58, 91.60)	82.15 (71.04, 92.13)	66.63 (7.57, 79.56)	13.25 (10.96, 46.65)
	Pathformer [9]	74.21 (61.17, 79.28)	76.90 (46.27, 89.68)	79.86 (60.41, 90.23)	82.73 (61.72, 90.98)	51.69 (9.56, 79.61)	16.86 (12.30, 38.92)
	TimeMachine [2]	64.91 (50.95, 68.50)	83.61 (44.95, 91.76)	86.00 (28.88, 90.86)	79.92 (68.34, 90.35)	65.31 (3.00, 76.66)	14.84 (11.06, 44.53)
	Seizureformer [17]	68.10 (52.80, 72.20)	83.07 (64.61, 94.94)	85.02 (64.79, 94.20)	74.28 (71.99, 85.64)	72.18 (0.76, 75.62)	17.16 (11.37, 31.63)
Time-Frequency Mixers	ATFNet [65]	75.98 (63.90, 84.48)	78.41 (50.35, 91.21)	82.93 (62.68, 91.35)	85.74 (71.42, 91.90)	53.78 (14.34, 85.53)	17.46 (11.20, 37.21)
	FreTS [67]	62.78 (51.74, 76.82)	63.96 (45.99, 79.23)	72.52 (32.37, 85.63)	78.36 (64.23, 86.84)	66.35 (3.59, 71.66)	22.89 (13.83, 44.25)
	NFM [26]	68.31 (60.51, 75.90)	75.04 (44.96, 89.13)	77.05 (54.41, 89.57)	80.81 (62.15, 92.09)	56.21 (9.37, 68.04)	20.28 (11.97, 43.48)
	TSLANet [15]	62.96 (58.99, 72.24)	74.33 (46.82, 83.18)	78.66 (48.56, 85.07)	76.16 (62.89, 87.34)	58.10 (1.58, 79.37)	19.17 (13.02, 34.61)
Multiscales Token Mixers	AdaWaveNet [68]	56.12 (50.00, 65.45)	78.23 (64.80, 81.60)	84.13 (51.70, 87.77)	66.63 (59.18, 75.08)	35.89 (0.82, 69.38)	20.91 (16.22, 24.73)
	Medformer [54]	60.36 (52.51, 69.21)	59.70 (43.95, 83.66)	68.02 (40.33, 83.79)	70.50 (52.32, 75.53)	38.92 (1.00, 82.32)	29.26 (15.35, 35.72)
	MTST [73]	76.29 (68.32, 79.73)	79.43 (69.76, 91.79)	82.89 (76.15, 93.34)	82.25 (74.01, 90.02)	70.01 (0.91, 74.77)	16.29 (10.12, 24.77)
	Pyraformer [28]	62.78 (51.74, 76.82)	63.96 (45.99, 79.23)	72.52 (32.37, 85.63)	78.36 (64.23, 86.84)	66.35 (3.59, 71.66)	22.89 (13.83, 44.25)
	SimpleTM [7]	61.28 (53.24, 69.36)	74.52 (48.65, 84.93)	77.71 (60.93, 86.08)	69.98 (60.49, 76.33)	38.72 (0.79, 84.22)	24.58 (15.02, 30.51)
	TimesNet [62]	62.60 (55.46, 68.19)	64.20 (37.11, 93.06)	63.41 (45.97, 91.86)	76.86 (72.09, 88.43)	68.38 (0.68, 74.80)	33.61 (13.69, 58.91)
	TimeMixer [51]	69.91 (60.04, 76.75)	72.14 (45.81, 89.66)	76.49 (60.78, 90.05)	81.91 (61.90, 91.43)	56.21 (10.05, 79.51)	21.98 (12.04, 39.00)
This Work	FAPEX-Small	77.32 (51.49, 84.85)	86.21 (45.72, 90.36)	87.11 (36.68, 93.93)	89.31 (77.26, 95.05)	78.46 (9.60, 90.48)	11.82 (7.96, 40.29)
	FAPEX-Base	82.03 (76.86, 90.08)	92.32 (80.21, 96.18)	92.30 (81.52, 97.14)	91.60 (82.17, 97.20)	80.18 (67.23, 95.76)	7.22 (6.58, 16.41)

Table 13: Comparison with Self-Supervised Baselines on ATLE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (↑)	SEN (↑)	F1 (↑)	ROC (↑)	PRC (↑)	Brier (↓)
Non-Contrastive	Brant [70]	Baseline	77.44 (50.62, 93.94)	87.94 (72.64, 95.79)	83.00 (72.10, 95.72)	99.92 (92.70, 99.99)	99.96 (78.96, 100.00)	15.98 (4.03, 23.65)
	CBraMod [50]	Baseline	82.73 (51.88, 96.35)	87.89 (78.87, 97.28)	82.81 (78.82, 97.26)	99.86 (92.98, 100.00)	99.93 (80.50, 100.00)	11.98 (2.81, 23.08)
	EEGT [49]	Baseline	85.54 [†] (51.24, 97.46)	88.23 [†] (82.35, 98.07)	83.16 (82.32, 98.06)	99.97 (92.99, 100.00)	99.98 (81.45, 100.00)	10.70 (2.11, 24.57)
	LaBraM [22]	Baseline	62.67 (56.86, 77.25)	76.11 (67.69, 79.64)	73.64 (63.79, 79.21)	99.96 (91.78, 100.00)	99.98 (74.71, 100.00)	12.72 (11.67, 24.07)
	Neuro-BERT [59]	Baseline	84.34 (53.60, 87.01)	85.41 (71.75, 89.63)	84.28 [†] (74.24, 89.09)	99.98[†] (92.28, 100.00)	99.99[†] (77.44, 100.00)	10.69 (8.15, 21.63)
	VQ_MTM [18]	Baseline	79.33 (53.55, 96.11)	81.71 (69.92, 97.39)	79.47 (71.24, 97.37)	99.89 (92.81, 100.00)	99.94 (79.99, 100.00)	14.83 (2.73, 24.02)
Contrastive	BIOT [63]	Baseline	78.00 (52.23, 95.87)	87.57 (73.29, 97.24)	83.18 (72.81, 97.21)	99.94 (92.12, 100.00)	99.97 (79.01, 100.00)	15.49 (2.90, 23.99)
	COMETS [52]	Baseline	84.03 (52.32, 96.56)	87.70 (80.63, 97.64)	83.16 (80.65, 97.63)	99.84 (92.88, 99.99)	99.92 (80.19, 99.99)	11.71 (2.60, 24.71)
	MF-CLR [14]	Baseline	80.20 (51.67, 85.59)	84.41 (76.11, 87.74)	82.84 (78.52, 85.38)	99.98 [†] (92.26, 100.00)	99.99 [†] (77.07, 100.00)	10.53 [†] (9.91, 20.52)
	TS2Vec [69]	Baseline	74.19 (51.48, 93.87)	84.01 (68.67, 96.06)	81.73 (67.56, 96.00)	99.84 (92.94, 100.00)	99.91 (81.29, 100.00)	19.17 (4.25, 23.89)
This Work	FAPEX-Small	Proposed	94.85 (88.60, 95.50)	94.05 (87.26, 96.76)	92.77 (90.48, 93.23)	99.97 (92.65, 100.00)	99.98 (78.66, 100.00)	6.15 (4.72, 16.38)
	FAPEX-Base	Proposed	95.33 (68.23, 99.15)	94.80 (85.22, 99.25)	98.03 (95.74, 99.45)	99.97 (92.65, 99.99)	99.98 (79.34, 100.00)	3.77 (1.11, 17.45)

Table 14: Comparison with Self-Supervised Baselines on AGS Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	82.74 (80.20, 82.85)	93.21 (85.92, 95.90)	92.69 (85.15, 95.89)	96.59 (89.70, 96.85)	71.08 (70.07, 91.57)	9.33 (5.72, 13.92)
	CBraMod [50]	Baseline	84.59 (79.25, 86.01)	90.76 (90.58, 98.23)	90.58 (89.53, 98.23)	98.68 (94.19, 98.76)	83.79 (82.91, 96.64)	8.89 (4.22, 12.01)
	EEGPT [49]	Baseline	83.14 (60.73, 92.68)	93.39 (85.16, 97.16)	92.89 (80.98, 97.08)	98.46 (98.37, 99.56)	95.06 (94.88, 97.22)	13.14 (7.82, 14.57)
	LaBraM [22]	Baseline	81.69 (80.81, 86.43)	94.49 (87.23, 96.70)	94.19 (87.53, 96.69)	97.64 (93.02, 97.64)	80.39 (72.97, 93.89)	7.16 (5.54, 11.65)
	Neuro-BERT [59]	Baseline	84.10 (80.89, 84.33)	93.66 (86.52, 95.82)	93.25 (86.87, 95.84)	96.81 (90.63, 97.09)	71.46 (71.19, 91.85)	8.54 (5.08, 12.26)
Contrastive	VQ_MTM [18]	Baseline	81.54 (68.19, 82.03)	87.90 (84.34, 95.44)	85.53 (85.21, 95.48)	94.57 (89.82, 95.13)	69.60 (64.52, 85.78)	13.60 (6.31, 18.78)
	BIOT [63]	Baseline	82.39 (58.54, 92.68)	93.11 (84.33, 97.16)	92.55 (79.47, 97.08)	98.17 (97.81, 99.49)	94.15 (93.10, 98.38)	8.38 (4.16, 15.80)
	COMETS [52]	Baseline	83.97 (82.65, 86.14)	93.87 (92.17, 98.47)	93.41 (91.73, 98.38)	98.16 (96.16, 99.14)	90.36 (87.81, 95.26)	8.79 (3.49, 9.34)
	MF-CLR [14]	Baseline	83.90 (77.64, 86.80)	91.15 (89.31, 97.80)	90.04 (89.45, 97.80)	97.03 (93.56, 98.10)	80.74 (80.17, 92.79)	9.40 (4.08, 13.10)
	TS2Vec [69]	Baseline	81.20 (80.51, 86.98)	94.73 (88.62, 96.72)	94.46 (88.51, 96.71)	97.55 (93.16, 97.72)	80.40 (74.06, 93.69)	7.07 (5.41, 11.65)
This Work	FAPEX-Small	Proposed	85.33 (62.34, 92.68)	94.08 (85.77, 97.16)	93.72 (82.04, 97.08)	98.39 (98.15, 99.59)	94.69 (94.47, 98.67)	7.64 (3.90, 13.90)
	FAPEX-Base	Proposed	87.96 (69.55, 95.44)	95.17 (88.49, 98.28)	94.94 (86.35, 98.25)	99.67 (99.45, 99.98)	98.92 (97.75, 99.93)	5.32 (2.08, 12.28)

Table 15: Comparison with Self-Supervised Baselines on BEIRUT Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	71.88 (56.52, 75.03)	71.01 (58.63, 88.66)	71.66 (61.46, 88.49)	78.94 (59.57, 87.14)	92.58 (48.04, 93.76)	19.03 (14.75, 27.74)
	CBraMod [50]	Baseline	71.14 (69.98, 89.98)	83.86 (64.97, 92.14)	83.54 (67.49, 91.01)	85.51 (80.72, 96.09)	93.80 (61.50, 98.17)	14.98 (7.79, 22.36)
	EEGPT [49]	Baseline	86.93 (45.58, 92.50)	63.88 (60.07, 74.51)	73.26 (59.74, 79.87)	71.90 (63.43, 78.36)	74.69 (66.00, 83.11)	19.20 (14.59, 25.32)
	LaBraM [22]	Baseline	66.76 (57.94, 69.55)	68.02 (42.04, 78.15)	66.02 (46.88, 75.89)	83.04 (68.27, 90.20)	94.31 (52.67, 94.89)	23.36 (18.63, 30.53)
	Neuro-BERT [60]	Baseline	73.68 (71.33, 91.96)	85.36 (70.13, 92.14)	85.19 (72.28, 91.21)	86.94 (83.23, 97.67)	94.02 (66.33, 98.90)	14.17 (6.42, 19.02)
Contrastive	VQ_MTM [18]	Baseline	73.13 (59.79, 78.88)	74.47 (62.99, 90.24)	74.95 (71.51, 89.81)	82.71 (65.27, 92.69)	93.93 (52.04, 96.66)	17.19 (12.84, 23.65)
	BIOT [63]	Baseline	73.88 [†] (61.06, 82.36)	76.90 (74.36, 89.87)	78.58 (77.44, 89.65)	83.14 (67.32, 92.74)	94.42 [†] (52.99, 96.78)	15.93 (12.72, 19.72)
	COMETS [55]	Baseline	73.85 [†] (71.13, 91.95)	86.16 [†] (78.68, 92.43)	86.00 [†] (78.76, 91.67)	87.35 [†] (81.02, 97.55)	94.28 (63.82, 98.83)	13.65 [†] (6.14, 17.27)
	MF-CLR [14]	Baseline	69.35 (60.70, 72.31)	78.21 (73.35, 91.10)	76.54 (73.68, 88.30)	83.80 (68.25, 92.01)	92.51 (53.79, 95.96)	17.37 (16.87, 19.35)
	TS2Vec [69]	Baseline	57.15 (50.63, 61.93)	66.99 (53.44, 77.88)	58.72 (56.75, 77.09)	74.27 (51.05, 82.79)	86.98 (44.96, 92.44)	22.00 (17.38, 25.38)
This Work	FAPEX-Small	Proposed	75.17[†] (73.90, 92.67)	87.49[†] (74.90, 93.39)	86.68[†] (76.42, 93.39)	90.02[†] (83.84, 98.19)	94.72[†] (69.09, 99.14)	12.60[†] (5.59, 16.95)
	FAPEX-Base	Proposed	78.47[†] (73.60, 94.50)	87.79[†] (74.17, 94.32)	87.28[†] (76.03, 94.37)	89.92[†] (89.34, 98.64)	94.60[†] (76.20, 99.36)	12.29[†] (4.47, 15.78)

Table 16: Comparison with Self-Supervised Baselines on CANINE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	56.36 (53.21, 70.45)	85.84[†] (66.79, 89.69)	84.91[†] (66.90, 88.04)	67.90 (54.54, 83.86)	23.38 (20.36, 71.19)	20.07 (16.67, 21.76)
	CBraMod [50]	Baseline	55.18 (51.63, 58.09)	80.51[†] (70.09, 87.70)	82.24[†] (70.61, 83.85)	67.01 (54.47, 76.60)	22.48 (19.67, 62.24)	19.67 [†] (17.01, 21.26)
	EEGPT [49]	Baseline	54.55 (51.45, 68.01)	78.87 (70.85, 86.85)	81.44 [†] (70.78, 86.22)	66.72 (53.30, 75.62)	21.46 (19.54, 52.44)	19.60 (17.02, 22.28)
	LaBraM [22]	Baseline	60.42 (51.46, 63.85)	73.92 (69.39, 85.17)	72.42 (69.96, 83.18)	76.20 (55.23, 77.79)	38.46 (22.82, 45.28)	20.43 (17.38, 23.47)
	Neuro-BERT [59]	Baseline	61.27 [†] (51.18, 62.01)	74.54 (69.60, 84.95)	73.53 (70.14, 83.31)	76.35 [†] (55.37, 77.70)	38.17 (22.99, 47.75)	20.34 (17.21, 23.47)
Contrastive	VQ_MTM [18]	Baseline	54.40 (53.12, 67.12)	72.47 (69.57, 87.27)	76.63 (69.65, 86.13)	66.50 (56.29, 76.01)	21.98 (19.02, 52.33)	20.71 (17.06, 23.60)
	BIOT [63]	Baseline	60.08 (51.54, 63.31)	71.97 (69.31, 85.29)	72.67 (69.89, 83.19)	75.99 (54.99, 77.58)	38.08 (22.89, 40.85)	20.54 (17.56, 23.23)
	COMETS [52]	Baseline	53.29 (52.73, 68.52)	72.41 (69.09, 86.94)	76.27 (69.33, 86.04)	65.74 (56.28, 76.17)	21.88 (18.33, 52.61)	20.99 (17.17, 23.64)
	MF-CLR [14]	Baseline	61.82[†] (51.16, 64.33)	75.52 (68.33, 84.85)	72.49 (69.23, 83.22)	75.96 (55.08, 77.25)	38.50 [†] (22.65, 39.49)	20.79 (16.75, 23.66)
	TS2Vec [69]	Baseline	52.63 (51.83, 69.29)	72.37 (68.36, 84.26)	77.51 (68.81, 84.90)	65.37 (54.88, 76.02)	22.05 (18.42, 52.93)	21.01 (17.57, 23.53)
This Work	FAPEX-Small	Proposed	60.59 (51.14, 70.25)	78.55 (74.40, 85.07)	78.88 (74.51, 83.20)	78.00[†] (55.85, 86.24)	38.89[†] (23.28, 55.71)	17.37[†] (15.48, 23.35)
	FAPEX-Base	Proposed	66.16[†] (51.19, 70.87)	78.88 [†] (75.05, 83.57)	79.17 (75.54, 83.07)	79.42[†] (56.36, 86.23)	42.98[†] (23.39, 55.54)	16.19[†] (15.43, 23.55)

Table 17: Comparison with Self-Supervised Baselines on CTLE-RATLFP Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	74.36 (69.13, 79.47)	72.07 (64.42, 79.84)	76.29 (69.95, 82.55)	83.60 (75.84, 88.25)	96.54 (94.91, 97.74)	19.42 (14.02, 21.63)
	CBraMod [50]	Baseline	73.89 (68.63, 78.92)	82.03 [†] (72.32, 84.59)	81.91 [†] (76.11, 86.26)	82.82 (77.47, 88.83)	96.30 (94.91, 97.92)	17.00 [†] (11.81, 20.46)
	EEGPT [49]	Baseline	70.51 (66.35, 75.28)	58.68 (52.30, 65.35)	64.60 (59.14, 70.76)	86.23 (74.00, 88.84)	97.01 (94.47, 97.83)	27.59 (23.42, 33.91)
	LaBraM [22]	Baseline	70.14 (66.16, 80.76)	70.81 (64.84, 73.29)	75.17 (70.31, 77.18)	78.71 (74.07, 88.37)	95.37 (93.81, 97.88)	19.61 (15.90, 23.45)
	Neuro-BERT [59]	Baseline	76.46 (70.94, 81.37)	72.65 (69.33, 81.86)	76.79 (74.01, 84.14)	85.47 (77.20, 89.63)	96.93 (94.94, 98.05)	20.86 (14.44, 23.09)
Contrastive	VQ_MTM [18]	Baseline	63.77 (59.39, 68.03)	72.62 (70.84, 78.46)	75.81 (74.48, 80.97)	69.71 (65.36, 78.51)	92.71 (91.85, 95.31)	20.66 (16.98, 23.23)
	BIOT [63]	Baseline	74.60 (66.15, 78.16)	61.64 (56.67, 70.54)	67.63 (62.71, 75.13)	85.29 (73.39, 90.84)	96.93 (94.15, 98.31)	28.16 (19.86, 29.38)
	COMETS [52]	Baseline	75.22 (69.35, 82.40)	74.23 (67.33, 80.04)	77.43 (72.37, 82.73)	83.71 (78.17, 90.34)	96.51 (95.45, 98.24)	18.10 (13.31, 21.01)
	MF-CLR [14]	Baseline	77.38 [†] (71.24, 79.64)	66.62 (61.96, 76.60)	71.75 (67.82, 79.97)	88.55 [†] (78.92, 89.31)	97.72 [†] (95.09, 97.94)	23.78 (16.41, 25.98)
	TS2Vec [69]	Baseline	69.56 (65.36, 77.05)	71.98 (63.64, 76.74)	75.30 (69.30, 80.03)	77.79 (70.76, 84.74)	95.07 (93.33, 97.00)	21.51 (16.26, 22.29)
This Work	FAPEX-Small	Proposed	80.45[†] (71.89, 83.48)	84.83[†] (81.50, 88.81)	86.24[†] (83.50, 89.59)	90.43[†] (81.33, 91.40)	98.07[†] (96.13, 98.34)	12.82[†] (11.62, 18.19)
	FAPEX-Base	Proposed	82.00[†] (75.80, 85.40)	85.24[†] (79.05, 89.68)	86.53[†] (81.67, 89.99)	90.60[†] (82.96, 93.64)	98.26[†] (96.49, 98.89)	13.26[†] (10.22, 18.83)

Table 18: Comparison with Self-Supervised Baselines on FMCE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	68.19 (66.02, 74.26)	75.21 (72.00, 78.61)	74.70 (71.89, 78.91)	86.84 (83.29, 89.68)	71.47 [†] (53.25, 85.05)	19.08 [†] (17.71, 21.39)
	CBraMod [50]	Baseline	76.85 (71.62, 82.11)	79.22 (75.16, 83.10)	79.94 (75.11, 83.31)	88.81 [†] (79.31, 91.42)	67.28 (51.47, 85.58)	20.64 (16.88, 25.94)
	EEGPT [49]	Baseline	63.30 (59.98, 68.27)	68.81 (59.51, 73.29)	68.94 (62.98, 74.01)	74.70 (67.15, 78.16)	58.40 (31.36, 68.01)	22.74 (19.45, 28.98)
	LaBraM [22]	Baseline	65.16 (61.12, 68.82)	70.06 (64.78, 75.24)	68.94 (64.35, 76.71)	77.65 (67.37, 82.07)	59.72 (34.19, 71.01)	23.05 (18.67, 27.38)
	Neuro-BERT [59]	Baseline	76.04 (70.28, 81.49)	77.87 (72.36, 82.65)	78.06 (72.75, 82.82)	87.73 (77.17, 90.78)	65.55 (48.83, 84.09)	21.60 (17.26, 28.03)
Contrastive	VQ_MTM [18]	Baseline	72.84 (69.97, 79.79)	73.71 (68.47, 80.44)	74.32 (70.86, 80.46)	82.31 (77.26, 89.25)	64.63 (54.23, 81.79)	19.89 (14.10, 22.23)
	BIOT [63]	Baseline	70.61 (68.88, 76.96)	74.07 (71.67, 79.89)	75.09 (71.67, 80.01)	81.31 (78.31, 88.05)	65.25 (52.29, 79.61)	20.85 (14.59, 21.86)
	COMETS [52]	Baseline	75.00 (68.19, 78.77)	76.13 (70.76, 79.85)	76.33 (71.25, 80.10)	87.14 (76.08, 89.32)	65.46 (44.40, 81.95)	23.29 (19.22, 29.85)
	MF-CLR [14]	Baseline	77.45 [†] (72.20, 82.70)	79.45 [†] (74.06, 83.68)	80.06 [†] (74.50, 83.90)	88.53 (79.16, 92.19)	66.85 (52.42, 85.90)	20.06 (16.53, 25.92)
	TS2Vec [69]	Baseline	68.87 (64.44, 73.13)	75.52 (72.19, 78.89)	74.55 (72.16, 78.33)	87.73 (83.77, 89.35)	71.24 (53.52, 85.86)	19.53 (17.58, 22.92)
This Work	FAPEX-Small	Proposed	89.96 [†] (81.10, 95.38)	89.52 [†] (83.57, 93.04)	89.56 [†] (83.54, 93.54)	97.52 [†] (93.49, 99.41)	96.67 [†] (91.48, 99.67)	7.27 [†] (5.12, 11.61)
	FAPEX-Full	Proposed	91.51 [†] (84.47, 97.10)	91.85 [†] (85.40, 96.48)	91.52 [†] (85.12, 96.68)	98.05 [†] (94.35, 99.51)	97.15 [†] (91.90, 99.75)	6.43 [†] (2.95, 10.80)

Table 19: Comparison with Self-Supervised Baselines on HUP Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	60.42 (57.34, 61.83)	64.04 (61.12, 66.56)	63.16 (61.40, 64.36)	63.88 (59.31, 66.39)	54.90 (40.71, 64.10)	23.41 (23.22, 23.67)
	CBraMod [50]	Baseline	61.93 (61.22, 64.10)	64.40 (63.59, 66.11)	64.51 (62.21, 66.66)	66.18 (63.88, 67.32)	58.92 (53.19, 66.45)	23.76 (21.97, 24.58)
	EEGPT [49]	Baseline	56.09 (55.51, 57.01)	61.26 (59.78, 62.27)	58.81 (57.55, 62.31)	58.78 (57.52, 60.30)	48.66 (41.03, 51.55)	27.02 (26.16, 27.51)
	LaBraM [22]	Baseline	55.25 (52.99, 55.46)	58.65 (58.14, 59.50)	57.39 (56.39, 59.41)	57.50 (55.10, 58.03)	45.32 (39.63, 47.86)	27.24 (27.03, 28.07)
	Neuro-BERT [59]	Baseline	66.36 [‡] (53.17, 73.65)	67.58 [‡] (45.46, 87.01)	72.36 [‡] (34.75, 87.92)	71.97 [‡] (54.23, 87.11)	49.44 (2.63, 65.84)	27.10 (13.05, 44.34)
Contrastive	VQ_MTM [18]	Baseline	61.90 (59.25, 63.46)	62.08 (59.52, 65.43)	63.52 (60.12, 65.00)	64.94 (63.84, 66.21)	57.92 (47.66, 63.46)	24.47 (23.09, 25.93)
	BIOT [63]	Baseline	61.41 (60.58, 61.94)	64.90 (61.04, 66.04)	63.04 (61.56, 65.12)	65.08 (63.52, 65.94)	57.50 (50.22, 65.61)	23.68 (22.58, 24.59)
	COMETS [52]	Baseline	62.31 (61.66, 64.00)	65.24 (62.76, 67.53)	64.09 (62.83, 68.55)	66.42 (64.72, 68.49)	59.57 [‡] (55.58, 68.15)	23.48 (21.74, 23.95)
	MF-CLR [14]	Baseline	63.15 (61.72, 65.34)	67.12 (65.19, 69.06)	66.15 (63.96, 70.44)	67.34 (64.50, 68.70)	59.55 (55.26, 68.47)	22.66 [‡] (21.63, 23.57)
	TS2Vec [69]	Baseline	51.26 (49.78, 58.72)	58.94 (51.14, 61.79)	51.64 (48.67, 58.60)	56.70 (47.71, 62.71)	44.35 (37.74, 56.47)	26.83 (25.02, 27.68)
This Work	FAPEX-Small	Proposed	71.73[‡] (67.31, 74.56)	71.97[‡] (69.28, 76.60)	72.09[‡] (69.00, 76.29)	79.39[‡] (74.58, 81.61)	73.05[‡] (70.42, 77.69)	18.61[‡] (17.12, 20.29)
	FAPEX-Base	Proposed	70.87[‡] (68.07, 74.69)	74.31[‡] (70.36, 77.95)	73.46[‡] (69.69, 76.73)	80.69[‡] (76.34, 82.96)	74.37[‡] (70.50, 78.77)	17.94[‡] (16.75, 19.55)

Table 20: Comparison with Self-Supervised Baselines on IESS Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	62.29 (59.16, 65.57)	68.01 (61.83, 76.27)	67.68 (60.60, 78.29)	69.47 (60.25, 74.75)	83.39 (66.72, 88.42)	23.27 (21.42, 26.13)
	CBraMod [50]	Baseline	62.92 (61.46, 77.10)	79.56 (64.40, 86.65)	80.74 (68.18, 87.34)	76.21 (66.96, 85.30)	83.00 (35.36, 94.81)	20.56 (14.58, 23.50)
	EEGPT [49]	Baseline	63.55 (61.60, 68.35)	74.20 (64.33, 79.43)	73.87 (63.27, 79.30)	71.38 (67.66, 78.36)	86.42 (68.69, 90.49)	21.34 (19.52, 25.55)
	Neuro-BERT [59]	Baseline	64.32 (61.93, 68.69)	75.32 (65.25, 79.65)	74.97 (65.07, 79.45)	71.54 (67.79, 78.68)	86.24 (68.16, 90.16)	20.40 (19.54, 23.88)
	LaBraM [22]	Baseline	63.12 (58.85, 64.50)	66.19 (56.55, 72.10)	66.45 (59.06, 76.30)	69.77 (59.82, 74.25)	82.36 (66.36, 85.77)	24.78 (22.13, 27.75)
Contrastive	BIOT [63]	Baseline	55.00 (53.48, 61.05)	70.83 (61.42, 77.09)	69.16 (60.39, 76.27)	66.05 (63.75, 69.24)	81.76 (48.79, 86.15)	29.10 (26.42, 33.97)
	COMETS [52]	Baseline	64.79 (49.51, 74.04)	67.57 (52.30, 82.58)	68.20 (52.48, 81.85)	79.33 (58.79, 83.31)	76.49 (24.70, 93.75)	23.22 (18.49, 32.48)
	MF-CLR [14]	Baseline	60.24 (56.32, 73.08)	79.69 (60.77, 85.24)	80.76 (62.54, 85.92)	75.62 (62.78, 80.32)	81.05 (31.31, 94.44)	19.03 (15.60, 26.44)
	TS2Vec [69]	Baseline	61.75 (57.91, 71.92)	72.36 (65.33, 77.20)	73.57 (65.13, 81.82)	72.89 (65.58, 82.67)	77.99 (63.65, 91.29)	21.13 (17.01, 22.30)
This Work	FAPEX-Small	Proposed	69.55 (65.43, 78.97)	81.52 (73.04, 87.36)	83.65 (73.92, 88.52)	83.38 (72.23, 89.89)	90.81 (42.86, 95.26)	15.69 (13.98, 23.32)
	FAPEX-Base	Proposed	72.35 (65.70, 81.96)	83.74 (73.13, 87.58)	84.89 (74.37, 88.55)	85.94 (75.12, 90.84)	90.04 (47.96, 95.10)	17.14 (13.26, 20.43)

Table 21: Comparison with Self-Supervised Baselines on KAIME Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments. .

Pretraining Type	Model	Setup	Metrics					
			BAC (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	68.78 (64.50, 77.51)	74.55 (65.93, 79.72)	74.80 (66.06, 79.47)	74.41 (70.81, 85.49)	86.26 (74.06, 87.84)	19.48 (15.80, 22.93)
	CBraMod [50]	Baseline	70.88 (65.87, 83.76)	81.03 (68.25, 86.81)	79.76 (67.96, 86.41)	83.70 (73.10, 90.39)	92.27 (81.36, 93.45)	16.88 (11.61, 21.10)
	EEGPT [49]	Baseline	68.64 (64.63, 78.37)	78.43 (65.87, 80.19)	77.34 (64.43, 79.77)	78.57 (76.71, 89.83)	87.33 (82.14, 91.78)	18.34 (14.21, 23.93)
	LaBraM [22]	Baseline	63.45 (54.20, 64.10)	53.58 (53.31, 63.81)	54.98 (51.61, 61.07)	73.76 (58.78, 79.58)	78.62 (63.72, 84.56)	24.42 (21.67, 27.14)
	Neuro-BERT [59]	Baseline	72.52 (66.49, 84.22)	81.67 [‡] (68.63, 87.04)	80.69 [‡] (68.39, 86.71)	83.63 (73.15, 90.37)	92.21 (81.40, 93.43)	16.54 [‡] (11.53, 20.99)
Contrastive	VQ_MTM [18]	Baseline	68.78 (60.84, 73.07)	62.78 (58.97, 70.46)	64.72 (59.45, 70.43)	78.20 (65.51, 83.37)	85.87 (68.95, 89.24)	21.78 (18.14, 24.83)
	BIOT [63]	Baseline	67.27 (59.40, 71.17)	61.50 (57.86, 71.21)	63.50 (58.37, 70.79)	75.52 (62.86, 80.95)	80.90 (67.24, 85.52)	22.59 (20.13, 25.31)
	COMETS [52]	Baseline	73.85 [‡] (63.74, 82.64)	80.61 (64.66, 83.37)	80.13 (60.98, 83.47)	84.05 (71.55, 89.44)	91.01 (77.77, 92.46)	16.57 (13.38, 23.92)
	MF-CLR [14]	Baseline	72.31 (63.53, 82.42)	80.93 (64.85, 83.39)	79.84 (63.94, 83.30)	86.34 [‡] (73.38, 90.63)	92.68 [‡] (80.76, 93.68)	18.23 (12.53, 24.99)
	TS2Vec [69]	Baseline	68.35 (63.42, 79.82)	76.13 (66.01, 80.52)	76.17 (66.03, 80.58)	76.68 (75.02, 87.84)	87.63 (77.12, 90.53)	19.62 (14.21, 22.23)
This Work	FAPEX-Small	Proposed	82.34[‡] (72.42, 86.53)	87.37[‡] (72.70, 88.46)	87.14[‡] (71.57, 88.24)	89.35[‡] (79.50, 92.66)	94.35[‡] (85.04, 94.98)	12.02[‡] (10.37, 19.49)
	FAPEX-Base	Proposed	83.56[‡] (72.99, 88.23)	88.68[‡] (73.40, 89.99)	88.40[‡] (72.34, 89.82)	91.39[‡] (80.51, 95.04)	95.92[‡] (86.05, 96.43)	11.02[‡] (8.55, 18.82)

Table 22: Comparison with Self-Supervised Baselines on LPIRE Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	68.15 (61.12, 71.12)	56.83 (52.02, 68.02)	68.60 (64.73, 74.66)	75.76 (66.93, 80.42)	96.89 (95.58, 97.82)	25.17 (20.28, 25.94)
	CBraMod [50]	Baseline	62.60 (61.24, 66.58)	69.30 [‡] (65.92, 79.82)	74.94 [‡] (74.16, 83.83)	68.41 (65.77, 80.24)	97.00 (93.06, 97.73)	21.96 (16.01, 22.48)
	EEGPT [49]	Baseline	72.56 (63.06, 74.62)	65.10 (56.21, 69.59)	71.49 (67.32, 73.25)	80.45 (69.52, 82.93)	97.47 [‡] (97.17, 97.74)	21.25 (18.82, 25.18)
	LaBraM [22]	Baseline	51.54 (50.31, 52.04)	52.40 (50.94, 57.15)	58.73 (56.27, 62.01)	50.92 (49.66, 52.57)	86.04 (79.48, 89.84)	35.60 (32.29, 37.08)
	Neuro-BERT [59]	Baseline	73.49[‡] (63.96, 75.60)	68.57 (63.86, 71.76)	73.75 (73.39, 75.99)	81.45[‡] (69.42, 83.45)	97.31 (97.17, 97.89)	19.44 [‡] (18.19, 22.50)
Contrastive	VQ_MTM [18]	Baseline	59.70 (59.18, 63.09)	50.70 (47.53, 62.23)	55.54 (53.84, 72.77)	63.66 (62.52, 66.89)	95.95 (92.52, 96.62)	34.25 (32.49, 38.44)
	BIOT [63]	Baseline	73.10 [‡] (63.74, 75.61)	68.94 (62.51, 71.22)	73.09 (72.44, 76.58)	80.55 [‡] (69.53, 83.46)	97.37 (97.18, 97.79)	19.78 (18.44, 22.78)
	COMETS [52]	Baseline	61.81 (61.17, 70.85)	53.37 (48.23, 59.44)	61.56 (57.17, 68.58)	68.59 (66.09, 79.26)	97.01 (93.19, 97.64)	28.02 (23.07, 29.11)
	MF-CLR [14]	Baseline	50.85 (49.86, 51.84)	47.59 (43.52, 49.34)	51.26 (49.14, 53.00)	50.57 (50.34, 52.08)	86.75 (79.26, 90.01)	39.95 (35.83, 41.88)
	TS2Vec [69]	Baseline	51.58 (49.92, 53.34)	59.67 (53.52, 66.16)	64.12 (59.88, 69.69)	51.78 (49.96, 53.52)	87.16 (79.55, 89.86)	33.03 (27.48, 35.10)
This Work	FAPEX-Small	Proposed	73.68[‡] (65.33, 76.20)	72.02[‡] (66.79, 73.26)	75.80[‡] (74.15, 79.16)	81.18[‡] (71.21, 83.67)	97.51[‡] (97.39, 97.89)	18.38[‡] (18.01, 21.26)
	FAPEX-Base	Proposed	68.02 (60.08, 71.11)	78.86[‡] (74.11, 86.28)	83.67[‡] (77.63, 89.50)	75.22 (68.27, 82.44)	97.61[‡] (93.26, 97.99)	18.17[‡] (15.62, 20.87)

Table 23: Comparison with Self-Supervised Baselines on PCS Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (↑)	SEN (↑)	F1 (↑)	ROC (↑)	PRC (↑)	Brier (↓)
Non-Contrastive	Brant [70]	Baseline	78.88 (61.27, 81.43)	83.09 (75.01, 90.05)	82.27 (72.34, 86.74)	95.80 (61.88, 98.29)	97.64 (71.77, 99.07)	15.28 (12.82, 25.43)
	CBraMod [50]	Baseline	78.97 (69.61, 80.96)	81.09 (73.36, 84.84)	83.51 (75.09, 87.29)	85.36 (80.27, 88.07)	91.08 (55.35, 93.28)	14.48 (12.88, 19.07)
	EEGPT [49]	Baseline	74.78 (59.73, 78.58)	85.49 (73.56, 91.52)	84.36 (71.95, 87.47)	90.96 (69.81, 98.29)	94.17 (66.18, 99.07)	16.94 (13.04, 22.00)
	LaBraM [22]	Baseline	58.20 (52.21, 78.06)	62.27 (55.86, 85.14)	60.49 (49.64, 83.94)	96.64 (45.26, 98.29)	98.10 (62.74, 99.07)	35.84 (13.44, 38.87)
	Neuro-BERT [59]	Baseline	70.66 (55.88, 78.34)	80.82 (69.10, 88.26)	81.90 (66.48, 85.95)	96.79 [‡] (50.56, 98.29)	97.90 (64.27, 98.88)	22.48 (13.34, 24.85)
Contrastive	VQ_MTM [18]	Baseline	81.16 (64.29, 87.70)	80.97 (70.78, 87.30)	81.20 (70.25, 87.01)	94.09 (65.49, 97.94)	97.10 (72.54, 98.99)	13.05 (10.00, 27.00)
	BIOT [63]	Baseline	81.46 [‡] (65.78, 89.45)	87.26 [‡] (69.74, 88.90)	86.50 [‡] (70.32, 88.83)	95.40 (68.56, 98.44)	97.45 (76.10, 99.31)	9.95 [‡] (8.31, 25.78)
	COMETS [52]	Baseline	70.66 (55.88, 78.34)	80.82 (69.10, 88.26)	81.90 (66.48, 85.95)	96.79 [‡] (50.56, 98.29)	97.90 (64.27, 98.88)	22.48 (13.34, 24.85)
	MF-CLR [14]	Baseline	71.40 (57.20, 78.40)	79.18 (69.85, 89.98)	77.39 (66.02, 86.65)	97.16 (54.45, 98.29)	98.34 (66.19, 99.10)	18.24 (13.25, 30.71)
	TS2Vec [69]	Baseline	69.05 (54.23, 78.20)	69.03 (62.65, 85.23)	65.93 (61.73, 84.05)	96.38 (45.76, 98.47)	97.88 (61.56, 98.99)	21.85 (12.65, 36.81)
This Work	FAPEX-Small	Proposed	89.12 [‡] (79.06, 95.67)	91.04 [‡] (79.35, 95.51)	90.97 [‡] (79.90, 95.51)	96.74 (85.34, 98.29)	98.41 [‡] (88.44, 98.96)	7.44 [‡] (3.98, 14.69)
	FAPEX-Base	Proposed	94.61 [‡] (86.39, 96.54)	95.05 [‡] (90.30, 96.39)	95.05 [‡] (90.91, 96.41)	97.47 [‡] (94.57, 98.25)	98.60 [‡] (97.05, 99.24)	4.72 [‡] (3.66, 8.89)

Table 24: Comparison with Self-Supervised Baselines on RESPECT Dataset for Seizure Prediction. Metrics are reported as median and interquartile range (IQR) in percentage across experiments.

Pretraining Type	Model	Setup	Metrics					
			BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
Non-Contrastive	Brant [70]	Baseline	59.46 (58.15, 62.81)	70.57 (54.95, 73.08)	70.36 (49.46, 80.94)	76.01 (68.76, 77.83)	43.47 (0.58, 74.31)	22.01 (18.98, 33.10)
	CBraMod [50]	Baseline	61.26 (50.00, 65.79)	62.06 (45.72, 75.74)	69.45 (34.09, 82.18)	67.60 (48.56, 73.89)	48.77 (1.16, 61.82)	29.87 (17.40, 41.07)
	EEGPT [49]	Baseline	53.37 (50.00, 61.86)	76.83 (64.80, 80.33)	82.23 (50.96, 87.60)	60.37 (55.81, 72.67)	34.50 (0.72, 69.85)	22.46 (16.21, 24.95)
	LaBraM [22]	Baseline	60.02 (53.55, 67.18)	73.76 (54.91, 84.54)	75.76 (66.96, 86.05)	66.00 (64.42, 80.82)	51.37 (0.57, 74.97)	20.63 (18.42, 28.24)
	Neuro-BERT [59]	Baseline	60.30 (51.72, 68.44)	67.35 (43.23, 90.69)	74.57 (56.07, 90.03)	82.41 (62.84, 87.97)	58.31 (2.79, 93.02)	25.53 (8.66, 41.18)
	VQ_MTM [18]	Baseline	56.33 (50.19, 64.73)	66.81 (33.77, 89.77)	67.32 (40.14, 89.07)	81.73 (57.69, 86.87)	56.17 (1.31, 90.02)	28.37 (9.83, 49.93)
Contrastive	BIOT [63]	Baseline	66.69 [†] (61.97, 74.55)	77.64 (47.27, 87.30)	80.17 (51.61, 88.12)	84.07 [†] (64.43, 89.62)	61.64 (4.98, 80.42)	16.09 [†] (12.51, 31.80)
	COMETS [52]	Baseline	66.28 (63.84, 67.64)	76.30 (73.44, 78.42)	81.52 (73.71, 85.95)	75.63 (71.53, 81.12)	48.32 (0.69, 71.74)	17.11 (15.75, 19.83)
	MF-CLR [14]	Baseline	63.51 (61.07, 66.63)	76.29 (74.38, 80.85)	82.17 (74.02, 86.89)	76.79 (70.47, 80.51)	45.52 (0.60, 75.39)	17.30 (15.91, 18.41)
	TS2Vec [69]	Baseline	58.10 (51.26, 67.83)	79.59 [†] (45.27, 90.90)	82.64 [†] (28.82, 90.15)	78.48 (65.74, 89.17)	64.97 [†] (4.10, 79.43)	16.46 (13.38, 46.14)
This Work	FAPEX-Small	Proposed	79.29 [†] (61.69, 82.83)	78.96 [†] (45.39, 93.53)	82.35 [†] (54.73, 93.12)	89.81 [†] (71.48, 93.26)	77.99 [†] (69.82, 86.49)	14.30 [†] (10.10, 48.58)
	FAPEX-Base	Proposed	84.23[†] (76.25, 90.39)	93.06[†] (88.64, 96.64)	94.59[†] (88.92, 97.14)	95.17[†] (82.23, 98.22)	79.03[†] (36.09, 96.32)	7.76[†] (6.23, 16.18)

Table 25: Ablation study. Performance metrics for seven alternative design variants evaluated on 6 benchmark datasets.

Dataset	Case	BA (\uparrow)	SEN (\uparrow)	F1 (\uparrow)	ROC (\uparrow)	PRC (\uparrow)	Brier (\downarrow)
FMCE ^b	Base	89.35 (81.49, 95.14)	88.83 (82.07, 94.00)	90.71 (69.69, 97.57)	97.19 (92.79, 99.22)	95.84 (90.47, 99.58)	8.21 (4.99, 12.96)
	Ab1. 1	74.34 (67.05, 80.68)	70.56 (59.86, 79.54)	71.78 (62.00, 80.43)	89.47 (77.01, 95.28)	86.52 (63.96, 96.94)	17.65 (12.47, 22.23)
	Ab1. 2	80.45 (70.10, 84.26)	76.36 (66.51, 80.63)	77.12 (66.72, 81.96)	91.96 (79.74, 97.37)	88.62 (80.82, 96.25)	14.47 (11.74, 21.10)
	Ab1. 3	81.56 (70.93, 87.50)	75.18 (69.79, 83.83)	76.20 (69.71, 85.16)	93.81 (82.53, 95.99)	89.48 (80.06, 97.05)	14.11 (10.36, 18.45)
	Ab1. 4	84.47 (73.28, 89.64)	82.63 (73.00, 86.03)	83.23 (72.86, 87.53)	93.88 (83.88, 97.12)	90.73 (81.85, 98.23)	10.96 (9.86, 18.40)
	Ab1. 5	85.50 (76.16, 89.79)	80.88 (73.11, 89.93)	81.77 (74.19, 90.52)	95.42 (85.53, 97.45)	92.49 (82.49, 98.59)	10.56 (8.35, 17.22)
	Ab1. 6	85.72 (78.03, 91.33)	82.40 (76.31, 90.52)	83.18 (76.50, 91.14)	96.10 (87.56, 97.68)	92.68 (82.74, 98.43)	9.81 (7.33, 15.49)
HUP ^a	Base	69.05 (67.48, 74.31)	73.65 (68.73, 76.88)	72.55 (68.52, 76.08)	79.30 (72.54, 81.49)	73.08 (69.76, 76.91)	18.80 (16.97, 21.23)
	Ab1. 1	65.68 (62.08, 68.71)	68.16 (57.01, 73.01)	67.06 (56.20, 70.93)	73.50 (70.81, 76.91)	68.57 (64.48, 73.30)	21.55 (20.86, 25.52)
	Ab1. 2	66.82 (64.09, 69.88)	70.63 (65.85, 73.04)	68.96 (65.79, 71.60)	77.88 (73.41, 81.20)	71.82 (67.04, 76.73)	20.83 (18.93, 22.29)
	Ab1. 3	67.42 (64.34, 71.34)	69.17 (63.95, 74.39)	68.05 (64.27, 73.83)	75.87 (71.84, 78.98)	69.59 (66.93, 74.15)	20.04 (18.05, 22.74)
	Ab1. 4	67.61 (63.06, 71.36)	65.66 (60.67, 72.01)	65.72 (60.74, 72.07)	75.49 (72.57, 78.42)	69.80 (65.39, 74.75)	20.86 (19.51, 23.84)
	Ab1. 5	67.64 (64.53, 72.48)	71.00 (65.68, 75.91)	69.88 (65.81, 74.21)	77.63 (74.93, 79.47)	73.00 (68.57, 75.66)	19.77 (18.21, 22.09)
	Ab1. 6	69.16 (63.50, 71.46)	67.36 (60.68, 74.04)	67.42 (60.98, 73.41)	76.27 (74.32, 79.22)	72.65 (66.27, 75.39)	20.52 (18.78, 23.00)
KAIME ^b	Base	82.23 (72.24, 86.84)	87.04 (72.43, 88.54)	86.84 (70.78, 88.36)	90.08 (81.15, 93.87)	95.23 (86.49, 95.70)	12.22 (10.31, 19.54)
	Ab1. 1	61.59 (59.46, 64.73)	67.01 (62.79, 72.44)	66.49 (59.35, 70.42)	72.75 (69.83, 73.23)	70.80 (68.86, 82.85)	21.59 (21.20, 24.57)
	Ab1. 2	73.67 (60.90, 82.50)	80.64 (63.61, 84.07)	80.03 (63.05, 83.75)	85.27 (71.15, 89.66)	91.81 (77.01, 92.38)	16.96 (12.41, 23.56)
	Ab1. 3	75.96 (65.48, 83.39)	82.13 (65.09, 85.29)	81.86 (61.79, 85.15)	85.22 (76.24, 90.35)	91.99 (82.46, 93.46)	15.58 (11.68, 22.85)
	Ab1. 4	67.59 (57.41, 72.39)	68.27 (57.45, 72.79)	69.08 (52.34, 73.67)	74.44 (62.94, 81.10)	82.82 (69.85, 88.06)	20.34 (18.41, 27.98)
	Ab1. 5	68.83 (61.76, 76.09)	71.56 (62.73, 75.72)	72.32 (63.00, 75.37)	76.85 (73.53, 85.29)	86.34 (78.41, 88.45)	20.11 (16.50, 24.47)
	Ab1. 6	71.81 (66.44, 81.48)	80.22 (68.32, 83.66)	79.39 (67.21, 83.43)	81.61 (77.86, 89.50)	90.50 (79.44, 91.70)	17.09 (13.39, 21.55)
RESPECT ^a	Base	84.12 (54.27, 86.63)	85.14 (46.20, 92.10)	87.82 (31.08, 92.61)	91.83 (76.63, 95.07)	73.34 (10.56, 88.87)	11.56 (7.45, 41.60)
	Ab1. 1	78.60 (69.39, 88.46)	89.10 (79.50, 91.56)	90.76 (80.71, 92.40)	91.42 (82.14, 96.67)	80.88 (66.34, 95.09)	12.01 (7.48, 16.37)
	Ab1. 2	78.72 (76.19, 88.01)	82.53 (77.96, 90.48)	90.04 (76.92, 90.37)	93.50 (82.06, 96.30)	80.13 (74.16, 94.55)	13.01 (8.24, 16.84)
	Ab1. 3	78.78 (76.53, 89.76)	82.14 (78.23, 89.64)	88.91 (77.27, 91.04)	93.02 (82.07, 96.62)	80.10 (76.30, 95.04)	13.37 (8.36, 16.64)
	Ab1. 4	79.10 (76.58, 89.27)	82.93 (78.10, 90.43)	90.28 (77.33, 90.41)	93.57 (81.99, 96.23)	80.05 (78.24, 94.41)	12.81 (7.99, 16.55)
	Ab1. 5	77.95 (76.10, 88.83)	78.94 (77.27, 90.18)	87.88 (76.77, 90.13)	93.79 (82.19, 96.17)	80.25 (76.97, 94.28)	15.07 (8.19, 16.65)
	Ab1. 6	80.19 (76.48, 89.37)	84.88 (78.09, 90.84)	90.73 (78.20, 91.67)	91.22 (82.10, 96.47)	80.42 (70.04, 94.65)	12.89 (7.80, 16.74)
IESS ^b	Base	65.67 (63.15, 70.48)	72.26 (59.73, 78.10)	72.41 (66.28, 78.13)	71.44 (68.72, 82.50)	82.96 (23.26, 88.36)	20.43 (19.83, 25.31)
	Ab1. 1	64.00 (60.70, 67.55)	64.67 (49.05, 76.75)	62.75 (58.48, 76.79)	71.21 (69.05, 81.34)	81.91 (23.90, 87.06)	22.74 (21.80, 31.16)
	Ab1. 2	63.56 (60.51, 68.15)	64.32 (51.33, 77.15)	62.80 (59.39, 77.14)	70.23 (68.29, 82.75)	81.64 (19.53, 88.01)	23.28 (21.97, 29.03)
	Ab1. 3	64.03 (60.97, 68.17)	65.42 (51.17, 77.20)	63.77 (59.84, 76.99)	70.18 (69.01, 82.82)	82.45 (21.24, 88.23)	22.70 (21.30, 29.07)
	Ab1. 4	64.22 (62.30, 68.38)	65.12 (50.68, 77.51)	64.59 (60.08, 77.30)	71.22 (68.40, 78.03)	82.00 (24.08, 84.58)	21.94 (21.35, 30.86)
	Ab1. 5	64.46 (61.33, 69.06)	67.88 (52.01, 78.36)	65.67 (60.84, 78.03)	71.69 (68.92, 84.18)	81.65 (25.71, 89.06)	22.79 (20.78, 29.61)
	Ab1. 6	63.94 (62.51, 69.45)	67.26 (51.26, 77.07)	66.15 (60.45, 77.33)	70.75 (68.86, 81.39)	81.80 (21.65, 87.24)	21.71 (20.81, 29.28)
CANINE ^a	Base	62.16 (54.84, 66.17)	86.00 (69.76, 88.50)	84.71 (72.11, 86.07)	74.53 (68.89, 75.92)	42.99 (29.45, 45.07)	19.51 (18.59, 20.58)
	Ab1. 1	54.07 (52.93, 67.22)	70.24 (68.34, 86.64)	76.18 (68.91, 85.67)	65.47 (56.23, 75.98)	23.52 (17.94, 52.65)	21.03 (17.26, 23.76)
	Ab1. 2	52.63 (51.83, 69.29)	72.37 (68.36, 84.26)	77.51 (68.81, 84.90)	65.37 (54.88, 76.02)	22.05 (18.42, 52.93)	21.01 (17.57, 23.53)
	Ab1. 3	52.63 (51.83, 69.29)	72.37 (68.36, 84.26)	77.51 (68.81, 84.90)	65.37 (54.88, 76.02)	22.05 (18.42, 52.93)	21.01 (17.57, 23.53)
	Ab1. 4	53.36 (52.06, 67.40)	70.25 (69.08, 86.05)	75.84 (69.77, 85.58)	64.70 (55.90, 74.93)	23.13 (17.60, 53.10)	20.82 (18.15, 23.90)
	Ab1. 5	60.42 (51.46, 63.85)	73.92 (69.39, 85.17)	72.42 (69.96, 83.18)	76.20 (55.23, 77.79)	38.46 (22.82, 45.28)	20.43 (17.38, 23.47)
	Ab1. 6	59.39 (51.24, 62.63)	75.43 (68.37, 85.42)	71.91 (69.18, 83.07)	75.64 (54.92, 76.93)	37.02 (22.51, 38.24)	20.73 (17.67, 23.56)

^aPublic dataset ^bIn-house dataset

5. TSLANet [15] replaces heavy Transformer layers with a lightweight Adaptive Spectral Block—using learnable Fourier filters to extract noise-robust frequency features—and an Interactive Convolution Block for local context. A self-supervised pretraining objective aligns representations across scales, enabling superior generalization on forecasting, classification, and anomaly detection. The code is available at this repository: <https://github.com/emadeldeen24/TSLANet>.
6. EEGMamba [71] is a universal EEG backbone combining a Spatio-Temporal-Adaptive convolutional frontend, bidirectional “Mamba” state-space blocks to capture sequential dynamics, and a task-aware Mixture-of-Experts head. This lets one model handle seizure detection, emotion recognition, sleep staging, and more—across variable channel counts and sequence lengths. We replicate the model based on the descriptions and configurations in [71].
7. Nonformer (Non-stationary Transformers) [30] augments any Transformer with two plug-in modules—Series Stationarization to normalize non-stationary inputs and De-stationary Attention to learn re-introducing trends—so that the backbone can focus on residual patterns. This simple add-on reduces forecasting error by up to 30% without retraining the entire model. The code is available at https://github.com/thuml/Nonstationary_Transformers.
8. TimeMachine [2] is a state-space model for long-term forecasting, built on the PatchTST framework and leveraging the Mamba module. It effectively captures long-term dependencies in multivariate series while maintaining linear scalability and a small memory footprint. The model produces salient contextual cues at multiple scales and uses an integrated quadruple-Mamba architecture to handle both channel-mixing and channel-independent situations. The code is available at <https://github.com/Atik-Ahamed/TimeMachine>.
9. Pyraformer [28] introduces pyramidal attention by constructing a multi-resolution temporal graph: nodes represent time tokens at different scales, with inter-scale edges capturing long-range links. A locality-aware routing mechanism ensures each time step only attends to a sparse subset, giving subquadratic complexity and strong performance on very long sequences. The code is available at this repository: <https://github.com/ant-research/Pyraformer>.
10. MultiresConv [42] introduces a Multiresolution Convolutional Memory layer that mimics wavelet-style analysis via shared, dilated convolutions at logarithmically spaced scales. By reusing convolutional weights across scales and blending their outputs, it captures both local details and long-range dependencies in $O(N \log N)$ time and memory. The code is available at this repository: <https://github.com/thjashin/multires-conv>.
11. AdaWaveNet [68] integrates a lifting-scheme-based adaptive wavelet decomposition into a deep MLP framework, enabling multi-scale analysis of non-stationary time series. Its learnable wavelet filters and residual mixing layers yield state-of-the-art accuracy on forecasting, imputation, and super-resolution tasks with low parameter count and fast inference. The code is available at this repository: <https://github.com/comp-well-org/AdaWaveNet>.
12. SeizureFormer [17] is a Transformer-based model for long-term seizure risk forecasting using clinically derived IEA and long-episode biomarkers. It employs CNN patch embeddings, multi-head self-attention, and squeeze-and-excitation blocks to model both short-term dynamics and long-term seizure cycles, achieving ROC AUC ≈ 0.79 and PR AUC ≈ 0.76 across 1–14 day horizons. We replicate the model based on the descriptions and configurations in [17].
13. Omni-Scale [45] introduces an OS-block that uses a set of prime-numbered kernels to cover receptive fields across all relevant scales, allowing simple 1D-CNNs to adaptively capture both local details and global context with minimal hyperparameter tuning, and achieving SOTA on univariate and multivariate classification benchmarks. Code is available at this repository: <https://github.com/Wensi-Tang/OS-CNN>.
14. MRConv [12] reparameterizes global convolutional kernels as learnable sums of low-rank, dilated sub-kernels trained in parallel with batch-norm-based decay, capturing long-range dependencies in $O(N)$ time and memory. It matches or outperforms SSMs and linear-time transformers on LRA, sequential CIFAR, and Speech Commands, and improves ImageNet accuracy when replacing standard convolutions. We replicate the model based on the descriptions and configurations in [12].

- 315 15. EEGConformer [44] is a compact convolutional Transformer for EEG decoding that fuses
316 temporal and spatial convolutional embeddings with multi-head self-attention, enabling
317 unified local-to-global feature learning and class activation mapping for tasks like motor
318 imagery and cognitive attention. The code is available at this repository: <https://github.com/eehysong/EEG-Conformer>.
319
- 320 16. PatchTST [39] segments each channel of a multivariate series into fixed-length patches
321 (“words”) and applies channel-independent self-attention, preserving local semantics while
322 reducing quadratic cost. This patching and shared-weight design yields SOTA long-term
323 forecasting accuracy with linear complexity. The code is available at this repository: <https://github.com/yuqinie98/PatchTST>.
324
- 325 17. TimesNet [62] transforms 1D series into 2D tensors based on learned periods and uses
326 an Inception-style TimesBlock to capture intraperiod and interperiod variations, achieving
327 SOTA across forecasting, imputation, classification, and anomaly-detection tasks through
328 efficient 2D variation modeling. The code is available at this repository: <https://github.com/thuml/TimesNet>.
329
- 330 18. **SPaRCNet** [23] is a CNN-based framework for expert-level detection of seizures and
331 rhythmic/periodic EEG patterns (IIIC) in ICU recordings, matching electrophysiologists in
332 performance and providing Class II evidence for automated EEG interpretation. The code is
333 available at this repository: <https://github.com/bdsp-core/IIIC-SPaRCNet>.
- 334 19. TF-C [72] is a contrastive model designed to address the challenges of pre-training on
335 time series data on the time-frequency domain. While domain adaptation methods can
336 mitigate these shifts, most require direct examples from the target domain, making them
337 less suitable for pre-training scenarios. To overcome this, TF-C introduces the concept of
338 time-frequency consistency, positing that time-based and frequency-based representations
339 of the same example should be embedded close together in the time-frequency space. The
340 model employs a decomposable pre-training approach, where the self-supervised signal is
341 derived from the distance between time and frequency components, each individually trained
342 through contrastive estimation. Code is available at this repository <https://github.com/mims-harvard/TFC-pretraining>.
343
- 344 20. ATFNet [65] combines a time-domain module and a frequency-domain module to jointly
345 capture local (short-term) and global (long-term) dependencies in multivariate series. It
346 uses Dominant Harmonic Series Energy Weighting to dynamically balance these modules
347 based on input periodicity, an Extended DFT for frequency alignment, and Complex-
348 valued Spectrum Attention to model inter-frequency relationships, achieving state-of-the-
349 art long-horizon forecasts. Code is available at this repository: <https://github.com/YHYHYHYHYHY/ATFNet>.
350
- 351 21. **iTransformer** [31] inverts the standard Transformer by treating each variate (feature channel)
352 across time as a token, so self-attention captures inter-channel correlations and the feed-
353 forward network operates on each channel independently. This simple inversion uses native
354 Transformer components without modification, scales to long lookbacks, and yields more
355 meaningful attention patterns. Code is available at this repository: <https://github.com/thuml/iTransformer>.
356
- 357 22. **NFM** [26] performs time series analysis entirely in the Fourier domain. It introduces
358 Learnable Frequency Tokens for adaptive frequency interpolation/extrapolation and Implicit
359 Neural Fourier Filters for continuous global convolution—delivering a highly compact (<
360 40K parameters) yet expressive model for forecasting, anomaly detection, and classification.
361 Code is available at this repository: <https://github.com/minkiml/NFM>.
- 362 23. **FreTS** [67] moves MLPs into the frequency domain for forecasting. Its two-stage pipeline
363 converts time-domain signals via DFT and then applies simple MLPs to learn real and
364 imaginary components of frequency representations, capturing global dependencies with
365 minimal architectural complexity. Code is available at this repository: <https://github.com/aikunyi/FreTS>.
366
- 367 24. **BIOT** [63] tokenizes each biosignal channel into fixed-length segments, rearranges them
368 into a unified “sentence,” and uses channel and relative-position embeddings to preserve
369 spatio-temporal structure. This enables pre-training and fine-tuning across heterogeneous

EEG, ECG, and activity datasets with mismatched channels, variable lengths, and missing values. Code is available at this repository: <https://github.com/ycq091044/BIOT>.

25. **COMET** [52] is a hierarchical contrastive framework for medical time series that applies contrastive losses at four levels—observation, sample, trial, and patient. This multi-level strategy learns robust representations that transfer effectively in low-label settings for tasks like ECG myocardial infarction detection and EEG-based Alzheimer’s or Parkinson’s classification. Code is available at this repository: <https://github.com/DL4mHealth/COMET>.
26. **MF-CLR** [14] introduces Multi-Frequency Contrastive Learning Representation, contrasting subseries at adjacent sampling rates in a hierarchical manner to enforce cross-frequency consistency. It yields versatile embeddings for forecasting, classification, anomaly detection, and imputation across diverse scales. Code is available at this repository: <https://github.com/duanjufang/MF-CLR>.
27. **TS2Vec** [69] learns universal representations by hierarchically contrasting both instance-level and temporal views over augmented context windows. It produces a timestamp-level embedding that can be pooled to represent arbitrary sub-sequences, enabling strong performance in classification, forecasting, and anomaly detection. Code is available at this repository: <https://github.com/zhihanyue/ts2vec>.
28. **Neuro-BERT** [59] is a BERT-style masked autoencoding framework for neurological signals (e.g., EEG, EMG) in the Fourier domain. It masks portions of the input, uses a Fourier Inversion Prediction task to reconstruct the masked segments, and learns frequency-phase structures without needing contrastive augmentations—boosting downstream tasks like sleep staging and gesture recognition. Code is available at this repository: <https://github.com/Westlake-AI/OpenBioSeq>.
29. **CBraMod** [50] is an EEG foundation model that uses a criss-cross transformer backbone to separately model spatial and temporal dependencies via parallel attention pathways, coupled with an asymmetric conditional positional encoding scheme for format-agnostic adaptation. Pre-trained on a large corpus of EEG data through patch-based masked reconstruction, it achieves state-of-the-art performance across up to 10 downstream BCI tasks, demonstrating strong generalizability. Code is available at this repository: <https://github.com/wjq-learning/CBraMod>.
30. **EEGPT** [49] is a 10-million-parameter pretrained Transformer designed for universal EEG feature extraction. It treats each electrode as a token in an autoregressive pre-training setup, uses mask-based dual self-supervision with spatio-temporal representation alignment, and employs a multi-task transfer learning paradigm with a shared electrode graph network, achieving SOTA across 12 benchmarks. Code is available at this repository: <https://github.com/BINE022/EEGPT>.
31. **LaBraM** [22] is a unified EEG foundation model that segments signals into channel-wise patches and uses vector-quantized neural spectrum prediction to learn a compact neural tokenizer. A Transformer is then pre-trained to reconstruct masked neural codes, leveraging 2,500 hours of diverse EEG data to excel on abnormal detection, event classification, emotion recognition, and gait prediction. Code is available at this repository: <https://github.com/935963004/LaBraM>.
32. **VQMTM** [18] is a BERT-style self-supervised model for EEG time series that combines a random-projection quantization module (to derive well-defined semantic units under the Johnson–Lindenstrauss lemma) with a Fourier-guided phase-alignment module. It generates robust tokens for masked modeling of corrupted, periodic signals, achieving consistent gains on seizure detection and classification with low complexity. Code is available at this repository: https://github.com/HaokunGUI/VQ_MTM.
33. **Brant** [70] is a foundation model for intracranial neural signals, pre-trained on a large corpus (1.01 TB) of SEEG data with over 500 M parameters. Its Transformer architecture jointly captures long-term temporal dependencies, spatial correlations, and time-frequency representations, delivering SOTA performance on downstream tasks such as signal forecasting, frequency-phase prediction, imputation, and seizure detection. Code is available at this repository: <https://zju-brainnet.github.io/Brant.github.io/>.

- 425 34. **MTST** [73] is a Multi-resolution Time-Series Transformer for Long-term Forecasting. It
 426 employs a multi-branch architecture where each branch tokenizes the input into patches
 427 of a specific size (e.g., short vs. long windows) and applies self-attention with relative
 428 positional encoding to capture both high-frequency local patterns and low-frequency global
 429 trends. The branch outputs are then fused to produce accurate long-horizon forecasts
 430 while keeping computational complexity subquadratic. Code is available at this repository:
 431 <https://github.com/Yitiann/MTST>.
- 432 35. **SimpleTM** [7] is a lightweight baseline specialized for multivariate time-series forecasting.
 433 It tokenizes each series using classical signal-processing ideas (e.g. wavelet-inspired win-
 434 dowing), then applies a single self-attention layer augmented with basic geometric-algebra
 435 operations (beyond standard dot-product) to capture cross-time and cross-dimension depen-
 436 dencies. Despite its simplicity, SimpleTM matches or outperforms heavier Transformer
 437 models on several benchmarks while keeping computation and memory costs minimal. Code
 438 is available at this repository: <https://github.com/vsingh-group/SimpleTM>.

439 F Datasets and data processing

440 F.1 Publicly available datasets for evaluation

441 **FMCE.** The Fragility Multi-Center iEEG (FMCE) dataset, publicly available on the OpenNeuro
 442 website, includes intracranial EEG recordings [? ?]. Our study organizes iEEG and EEG data from
 443 five centers, with a total of 100 subjects. Due to data sharing restrictions, only data from four centers
 444 are included. The dataset comprises both ECoG and SEEG acquisitions, where each run represents a
 445 distinct snapshot of EEG data from a subject’s session. Each subject has a raw EDF file. In seizure
 446 sessions, each run corresponds to an EEG snapshot around a different seizure event. The primary
 447 objective of this dataset is to investigate neural fragility and identify biomarkers for seizure onset.

448 The dataset is available at <https://openneuro.org/datasets/ds003876> and
 449 <https://openneuro.org/datasets/ds003029>.

450 **HUP.** The HUP dataset [?] is derived from the Hospital of the University of Pennsylvania,
 451 encompassing intracranial EEG signal data from 55 patients, comprising 55 subjects, all of whom
 452 were diagnosed with drug-resistant epilepsy and had undergone surgical interventions. The patients
 453 were treated with stereotactically placed depth electrodes (SEEG) or subdural grid, strip, and depth
 454 electrodes (ECoG), with the number of electrodes implanted varying according to each patient’s
 455 specific condition. The dataset provides comprehensive information, including the patients’ handed-
 456 ness, the type of treatment administered, the surgical target point, the lesion status on MRI, and the
 457 surgical outcomes in terms of seizure recurrence. For the 36 patients who underwent SEEG electrode
 458 implantation, the dataset includes 10 minutes of interictal data as well as data from seizure periods
 459 of varying durations. Additionally, the dataset contains electrophysiologic data for both interictal
 460 and ictal periods, electrode localizations in ICBM152 MNI space, clinically-determined seizure
 461 onset channels, and channels that overlap with the resection/ablation zone, which was meticulously
 462 determined by segmenting the resection cavity.

463 The dataset is available at <https://openneuro.org/datasets/ds004100/versions/1.1.1>.

464 **RESP.** The RESP dataset [47] is an intracranial EEG dataset collected at the University Medical
 465 Center Utrecht, Netherlands. It includes data from 12 patients: 6 with intraoperative electrocor-
 466 ticography (acute ECoG) recordings and 6 with long-term recordings (3 with ECoG and 3 with
 467 stereo-encephalography, SEEG). For seizure prediction, we use only the 6 patients with long-term
 468 recordings. The data is organized according to the Brain Imaging Data Structure (BIDS) specification,
 469 a community-driven standard for structuring neurophysiology data and associated metadata. Each
 470 patient has a dedicated folder (e.g., sub-RESP0280) containing their intracranial EEG recordings and
 471 metadata necessary for interpreting the raw data and event timing. The BIDS structure is implemented
 472 differently for intraoperative and long-term recordings. Intraoperative ECoG recordings, conducted
 473 during surgery, are categorized into three main scenarios, logically grouped into BIDS sessions.

474 The dataset is available at <https://openneuro.org/datasets/ds003848/versions/1.0.3>.

475 **BEIRUT.** This dataset includes the EEG of 6 epileptic patients recorded at the Epilepsy monitoring
476 unit of the American university of Beirut Medical Center between January 2014 and July 2015.
477 The data represents measurements from 21 scalp electrodes, following the 10-20 electrode system,
478 sampled at 500 Hz . All channels have been bandpass filtered between 1/1.6 Hz and 70Hz while
479 filtering out the 50Hz (electrical utility frequency). Some channels have been omitted from specific
480 recordings due to artifact constraints.

481 The dataset is available at <https://data.mendeley.com/datasets/5pc2j46cbc/1>.

482 **CTLE-RATLFP.** This dataset [64] is derived from a longitudinal intracranial local field potential
483 (LFP) recording study conducted on six pilocarpine-induced Sprague-Dawley rat models of chronic
484 temporal lobe epilepsy (cTLE). Recordings were obtained using a custom-designed multi-site acquisition
485 system at 1 kHz sampling rate across depth electrodes implanted bilaterally in hippocampal and
486 limbic regions, including dentate gyrus (DG), CA1 subfield of the hippocampus (CA1), CA3 subfield
487 of the hippocampus (CA3) and anterior nucleus of the thalamus (ANT). All animals underwent
488 chronic monitoring over a period of 1–4 months, with behavioral and electrographic seizure activity
489 continuously recorded. For each animal, two distinct stages—early and late—were identified based
490 on seizure frequency and temporal distance from initial status epilepticus. In total, 12–40 generalized
491 seizures (Racine scale 5) were annotated per subject using a combination of expert visual inspection
492 and quantitative time-frequency markers. Each annotated seizure includes precise onset timing,
493 seizure onset zone (SOZ), seizure onset pattern (SOP), and inter-channel latency, enabling both
494 within-subject and cross-subject analyses. All seizures were accompanied by time-aligned video and
495 underwent artifact inspection.

496 The dataset is publicly available at <https://dandiarchive.org/dandiset/001212?search=seizure&pos=2>.

497 **LPIRE.** The Lithium-Pilocarpine Induced Rat Epilepsy (LPIRE) dataset [37] is derived from a
498 longitudinal study on temporal lobe epilepsy (TLE) in rats, focusing on multi-site local field potential
499 (LFP) recordings obtained from 15 pilocarpine-induced rats exhibiting chronic TLE. The LFP
500 data was recorded using a high-density multi-electrode array at a 1 kHz sampling rate, capturing
501 electrophysiological activity across 12 sites in the Papez circuit. The rats were monitored over a
502 period of 3–6 months, with both spontaneous seizures and interictal activity continuously recorded
503 during this period. Each rat underwent a chronic phase of monitoring post-status epilepticus, with
504 seizure activity recorded and annotated by expert analysis. Each seizure event is accompanied by
505 precise timing data and is further annotated with detailed characteristics such as the seizure onset
506 pattern (SOP), frequency, and related behavioral observations, all of which were cross-verified using
507 time-aligned video and automated detection methods.

508 The dataset is publicly available at <https://dandiarchive.org/dandiset/001044?search=epilepsy&pos=4>.

509 **CANINE.** This dataset is designed to advance seizure forecasting using intracranial EEG (iEEG)
510 recordings from both human and canine subjects with naturally occurring epilepsy. It includes data
511 from five dogs and two humans, with recordings segmented into 10-minute clips labeled as preictal
512 (pre-seizure) or interictal (non-seizure). Each clip is stored in MATLAB format, containing raw
513 iEEG signals sampled at 400 Hz (decimated to 200 Hz in some analyses) across 16 channels for
514 canine subjects and variable channels for humans. The dataset, supported by the National Institutes
515 of Health and NeuroVista, provides metadata on subject-specific channel configurations and seizure
516 events, enabling machine learning research for seizure prediction algorithms.

517 The dataset is publicly available at <https://www.kaggle.com/competitions/seizure-prediction>.

518 **F.2 In-house dataset for evaluation**

519 **KAIME.** The Kainic Acid-Induced Macaque Epilepsy (KAIME) dataset is an in-house dataset
520 comprising continuous SEEG and Scalp-EEG recordings from three macaques (*Rhesus macaques*)
521 with pharmacologically induced temporal lobe epilepsy. This dataset is based on the well-established
522 Kainic Acid (KA) model of epilepsy, which is widely used to study TLE in non-human primates. The
523 KA model involves the administration of kainic acid, an excitatory amino acid, through unilateral
524 hippocampal injections, leading to the induction of seizures, including status epilepticus (SE) and
525 subsequent spontaneous recurrent seizures. These seizures closely mimic the development and
526 progression of TLE in humans, making the KAIME dataset a valuable resource for studying seizure

mechanisms and exploring potential treatments. The injections typically induced SE, a state of continuous or rapidly recurrent seizures. Following SE, the animals experienced SRS, which are spontaneous seizures occurring after a period of latency. This model allows for the investigation of both acute seizure dynamics and the longer-term consequences of chronic seizure activity.

We collected continuous overnight joint EEG and SEEG recordings from three macaques, annotating EEG events and physiological states in collaboration with clinical electroencephalographers. For each raw session, the data was downsampled to 512 Hz to ensure consistency across sessions. Corrupted channels were manually identified and excluded. The sessions were subsequently epoched based on the annotations provided by the clinical electroencephalographers, with interictal and preictal epochs selected for further analysis. These epochs were segmented into 4-second samples, each containing 2048 time points, resulting in a total of 36,092 4-second samples at 512 Hz. A bandpass filter was applied in the range of 0.5 Hz to 250 Hz, and a notch filter was utilized to eliminate frequency harmonics at 50 Hz and 60 Hz. Additionally, each channel was normalized to ensure standardization, and a notch filter was applied at 60 Hz, 120 Hz, 180 Hz, and 240 Hz to remove line noise.

PCS. The Pediatric Clonic Seizure (PCS) dataset is an in-house collection of scalp-EEG recordings from five pediatric patients undergoing routine video-EEG monitoring for epilepsy evaluation. The dataset captures clonic seizure episodes, characterized by rhythmic muscle contractions, and includes both ictal and interictal periods. Each patient underwent continuous monitoring in a clinical setting, with data recorded at 1 kHz and subsequently downsampled to 512 Hz for consistency. Expert neurologists annotated seizure onset, offset, and duration, with each event further characterized by seizure type, clinical manifestations, and corresponding video evidence. To ensure signal quality, data was preprocessed using bandpass filtering (0.5–250 Hz) and notch filtering to eliminate powerline noise. Each session was segmented into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

IESS. The Infantile Epileptic Spasm Syndrome (IESS) dataset is an in-house collection from 17 pediatric patients diagnosed with IESS, a severe form of childhood epilepsy characterized by hypsarrhythmia and frequent spasms. Recordings were obtained using high-density scalp-EEG systems at 1 kHz, capturing the complex, chaotic cortical discharges typical of this syndrome. Data preprocessing included bandpass filtering (0.5–250 Hz), notch filtering to eliminate powerline interference, and artifact rejection to ensure high-quality signal retention. Each session was segmented into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

ATLE. The ATLE dataset is an in-house collection of scalp-EEG recordings from five adult patients with chronic temporal lobe epilepsy (TLE). This dataset focuses on seizure events recorded during routine outpatient monitoring, capturing both focal and secondary generalized seizures. The data was collected at a sampling rate of 1 kHz and subsequently downsampled to 512 Hz to ensure consistency across sessions. Each seizure event was manually annotated for onset, offset, and duration, with additional clinical metadata such as seizure frequency and patient history included. Data preprocessing involved bandpass filtering (0.5–250 Hz) to capture the full range of cortical oscillations relevant to TLE, notch filtering to eliminate powerline interference, and channel normalization to reduce inter-session variability. Each session was segmented into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

AGS.

AGS. The AGS dataset is an in-house collection of scalp-EEG recordings from five adult patients diagnosed with generalized epilepsy. This dataset captures seizure events recorded during routine outpatient monitoring, primarily focusing on primary generalized seizures, including tonic-clonic and absence seizures. The data was acquired at a sampling rate of 1 kHz and downsampled to 512 Hz for consistency with other datasets. Each seizure event was manually annotated by clinical experts for onset, offset, and duration, with additional metadata including seizure type, frequency, and relevant patient medical history. Preprocessing steps included bandpass filtering (0.5–250 Hz) to isolate cortical oscillations pertinent to generalized epilepsy, notch filtering to remove powerline interference, and channel normalization to minimize variability across recording sessions. Each session was segmented into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

F.3 Publicly available datasets for pretraining

In this section, we provide a comprehensive overview of the publicly available datasets used for pretraining, which form the foundation for evaluating FAPEX’s self-supervised learning capabilities. To ensure a fair comparison with existing pretrained models for brain signal analysis and generalized time series representation, these datasets are selected to cover a wide range of temporal dynamics and physiological signals. This approach mitigates potential biases arising from variations in dataset configurations and ensures that performance gains are attributable to the underlying model design rather than data-specific optimizations.

TUEG. The Temple University Hospital EEG Corpus (TUEG) is one of the largest publicly available clinical EEG databases, comprising over 29,000 sessions from more than 16,000 subjects collected between 2002 and 2020. It includes scalp EEG recordings in EDF format, paired with de-identified clinician reports in text format, detailing patient history, medications, and clinical impressions. The dataset is organized by patient and session, with metadata including age, gender, and session details, facilitating research into epilepsy, seizure detection, and other neurological conditions. The TUH-EEG includes subsets like the TUH Abnormal EEG Corpus (TUAB) and TUH Seizure Corpus (TUSZ), with annotations for pathological events such as seizures and epileptiform discharges. It supports machine learning applications for automated EEG analysis and is accessible via the Neural Engineering Data Consortium, promoting advancements in neuroscience and clinical diagnostics. 14,987 subjects are used in this work.

The dataset is publicly available at <https://www.nedcdata.org>.

INTERICTAL iEEG-HFO. This dataset originates from an intracranial EEG (iEEG) study investigating high-frequency oscillations (HFOs) during slow-wave sleep (SWS) in epilepsy patients. It includes 20 subjects, with recordings structured in BIDS format and annotated for HFO events. Each subject folder contains raw iEEG data, anatomical labels, and metadata detailing resection status. The dataset was converted into BIDS format using MNE-BIDS, with event annotations stored in .tsv files for easy integration with mne-python. Channels included in the resected region and those excluded from analysis are documented in the clinical Excel file under the sourcedata/ directory. The dataset enables research into epilepsy biomarkers and surgical outcome predictions, supporting broader neuroscience investigations. It is openly accessible via platforms such as Brainlife.io, NEMAR, and CBRAIN for advanced neuroimaging analysis.

The dataset is publicly available at <https://openneuro.org/datasets/ds003498/versions/1.1.1>.

iEEG NEURAL SPATIAL VOLATILITY. This dataset is part of a collection of intracranial electroencephalography (iEEG) recordings designed to investigate neural spatial volatility. The dataset includes recordings from 166 participants across two sessions, with data acquired using standardized protocols for high-resolution neural activity monitoring. The dataset is structured following the BIDS format, ensuring compatibility with various neuroimaging analysis tools. Available modalities include iEEG, and the dataset is hosted on OpenNeuro. Researchers can access metadata, participant information, and raw signal recordings for further analysis. Ethical approvals for data collection were obtained from Wayne State University and Tohoku University. The dataset is licensed under CC0, allowing unrestricted use for secondary analysis and research purposes. The dataset is publicly available at <https://openneuro.org/datasets/ds006107/versions/1.0.0>.

iEEG-CHILDRENGAMEPLAY. This dataset is derived from an intracranial EEG (iEEG) study conducted on ten human epilepsy patients performing a visuospatial working memory task. The subjects, undergoing intracranial monitoring for seizure localization, participated in 60 trials (five sessions) of the Memory Matrix game on the Lumosity platform. Recordings were collected during interictal iEEG sessions and structured into individual subject folders containing raw iEEG data in .edf format. Each file includes continuous iEEG signals recorded during gameplay, with detailed annotations specifying stimulus onset timing and finger-tapping responses (successful and failed attempts). Baseline periods were also established, consisting of non-overlapping 2,000-ms time windows of spontaneous, resting, eye-open wakefulness preceding game sessions. The dataset enables analyses of neural activity during cognitive tasks in epilepsy patients and provides structured metadata for research applications. The dataset is publicly available at <https://openneuro.org/datasets/ds004770/versions/1.0.0>.

632 **OPEN iEEG DATASET.** This dataset comprises de-identified interictal intracranial EEG (iEEG)
633 recordings collected from 185 subjects during sleep at University of California Los Angeles Mattel
634 Children’s Hospital and Children’s Hospital of Michigan, Detroit. Each subject folder contains raw
635 iEEG data, including anatomical labels and resection status. The dataset also includes processed
636 derivatives for high-frequency oscillation (HFO) detection and classification, utilizing RMS and
637 MNI detectors. Metadata summarizing subject outcomes and background information is available
638 in the ‘participants.tsv’ file. The dataset supports epilepsy research by providing structured iEEG
639 recordings for biomarker analysis and surgical outcome prediction. The dataset is publicly available
640 at (<https://openneuro.org/datasets/ds005398/versions/1.0.1>).

641 **KAIME.** The Kainic Acid-Induced Macaque Epilepsy (KAIME) dataset is an in-house dataset
642 comprising continuous SEEG and Scalp-EEG recordings from three macaques (*Rhesus macaques*)
643 with pharmacologically induced temporal lobe epilepsy. This dataset is based on the well-established
644 Kainic Acid (KA) model of epilepsy, which is widely used to study TLE in non-human primates. The
645 KA model involves the administration of kainic acid, an excitatory amino acid, through unilateral
646 hippocampal injections, leading to the induction of seizures, including status epilepticus (SE) and
647 subsequent spontaneous recurrent seizures. These seizures closely mimic the development and
648 progression of TLE in humans, making the KAIME dataset a valuable resource for studying seizure
649 mechanisms and exploring potential treatments. The injections typically induced SE, a state of
650 continuous or rapidly recurrent seizures. Following SE, the animals experienced SRS, which are
651 spontaneous seizures occurring after a period of latency. This model allows for the investigation of
652 both acute seizure dynamics and the longer-term consequences of chronic seizure activity.

653 We collected continuous overnight joint EEG and SEEG recordings from three macaques, annotating
654 EEG events and physiological states in collaboration with clinical electroencephalographers. For each
655 raw session, the data was downsampled to 512 Hz to ensure consistency across sessions. Corrupted
656 channels were manually identified and excluded. The sessions were subsequently epoched based on
657 the annotations provided by the clinical electroencephalographers, with interictal and preictal epochs
658 selected for further analysis. These epochs were segmented into 4-second samples, each containing
659 2048 time points, resulting in a total of 36,092 4-second samples at 512 Hz. A bandpass filter was
660 applied in the range of 0.5 Hz to 250 Hz, and a notch filter was utilized to eliminate frequency
661 harmonics at 50 Hz and 60 Hz. Additionally, each channel was normalized to ensure standardization,
662 and a notch filter was applied at 60 Hz, 120 Hz, 180 Hz, and 240 Hz to remove line noise.

663 **PCS.** The Pediatric Clonic Seizure (PCS) dataset is an in-house collection of scalp-EEG recordings
664 from five pediatric patients undergoing routine video-EEG monitoring for epilepsy evaluation. The
665 dataset captures clonic seizure episodes, characterized by rhythmic muscle contractions, and includes
666 both ictal and interictal periods. Each patient underwent continuous monitoring in a clinical setting,
667 with data recorded at 1 kHz and subsequently downsampled to 512 Hz for consistency. Expert
668 neurologists annotated seizure onset, offset, and duration, with each event further characterized by
669 seizure type, clinical manifestations, and corresponding video evidence. To ensure signal quality, data
670 was preprocessed using bandpass filtering (0.5–250 Hz) and notch filtering to eliminate powerline
671 noise. Each session was segmented into 4-second epochs, with precise seizure onset and offset
672 annotations provided by expert clinicians.

673 **IESS.** The Infantile Epileptic Spasm Syndrome (IESS) dataset is an in-house collection from
674 17 pediatric patients diagnosed with IESS, a severe form of childhood epilepsy characterized by
675 hypsarrhythmia and frequent spasms. Recordings were obtained using high-density scalp-EEG
676 systems at 1 kHz, capturing the complex, chaotic cortical discharges typical of this syndrome.
677 Data preprocessing included bandpass filtering (0.5–250 Hz), notch filtering to eliminate powerline
678 interference, and artifact rejection to ensure high-quality signal retention. Each session was segmented
679 into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

680 **ATLE.** The ATLE dataset is an in-house collection of scalp-EEG recordings from five adult patients
681 with chronic temporal lobe epilepsy (TLE). This dataset focuses on seizure events recorded during
682 routine outpatient monitoring, capturing both focal and secondary generalized seizures. The data was
683 collected at a sampling rate of 1 kHz and subsequently downsampled to 512 Hz to ensure consistency
684 across sessions. Each seizure event was manually annotated for onset, offset, and duration, with
685 additional clinical metadata such as seizure frequency and patient history included. Data preprocessing

involved bandpass filtering (0.5–250 Hz) to capture the full range of cortical oscillations relevant to TLE, notch filtering to eliminate powerline interference, and channel normalization to reduce inter-session variability. Each session was segmented into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

AGS. The AGS dataset is an in-house collection of scalp-EEG recordings from five adult patients diagnosed with generalized epilepsy. This dataset captures seizure events recorded during routine outpatient monitoring, primarily focusing on primary generalized seizures, including tonic-clonic and absence seizures. The data was acquired at a sampling rate of 1 kHz and downsampled to 512 Hz for consistency with other datasets. Each seizure event was manually annotated by clinical experts for onset, offset, and duration, with additional metadata including seizure type, frequency, and relevant patient medical history. Preprocessing steps included bandpass filtering (0.5–250 Hz) to isolate cortical oscillations pertinent to generalized epilepsy, notch filtering to remove powerline interference, and channel normalization to minimize variability across recording sessions. Each session was segmented into 4-second epochs, with precise seizure onset and offset annotations provided by expert clinicians.

F.4 Data preprocessing

Neuroelectrical signals are inherently susceptible to various noise sources and physiological artifacts that can obscure underlying neural dynamics, particularly within the high-frequency spectrum. Given our focus on extracting and utilizing diverse frequency information from EEG—especially high-frequency components vital for understanding transient neural dynamics in epilepsy—it is imperative to conduct meticulous preprocessing to preserve signal fidelity.

Data acquisition and filtering. Across public and in-house datasets, sessions of records were acquired with a sampling rate ranging from 500 to 2000 Hz, contingent upon the specific experimental setup and hardware capabilities, sometimes even different individuals. Given the susceptibility of EEG signals to various artifacts—particularly in the high-frequency bands critical for our analyses—it is imperative to implement a rigorous preprocessing pipeline to ensure data integrity. For consistency, we resampled all data to 512 Hz before preprocessing on the MNE v1.9.0 platform.

Power-line interference mitigation. Power-line interference, presenting as sinusoidal noise at 50 or 60 Hz and its harmonics, is a common issue in EEG recordings, particularly in unshielded settings. To mitigate this interference, we employed the MNE v1.9.0 platform by applying a notch filter targeting the power-line frequency (50/60 Hz) and its harmonics. Specifically, we used the `mne.filter.notch_filter` function, which allows precise attenuation of specified frequencies. The filter was designed with a bandwidth of 1 Hz around the target frequencies (e.g., 50 Hz, 100 Hz) to effectively suppress the interference while preserving the integrity of the EEG signal.

Band-pass filtering for signal isolation. Following power-line interference removal, a band-pass filter was applied with cut-off frequencies at 0.1 Hz and 200 Hz. This filtering range was selected to retain the physiologically relevant components of the EEG signal, encompassing δ (0.5–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (13–30 Hz), and low γ (30–80 Hz) rhythms, as well as high γ /high-frequency oscillations (HFOs) above 80 Hz, which are critical for understanding transient neural dynamics in epilepsy. The lower cut-off at 0.1 Hz was chosen to minimize the attenuation of slow cortical potentials, which are essential for capturing the full spectrum of neural activity. The upper cut-off at 200 Hz was selected to preserve high-frequency components while eliminating further high-frequency noise and artifacts.

It is important to note that the application of high-pass filters with cut-off frequencies at or below 0.1 Hz can introduce low-frequency artifacts, such as baseline drifts. To address this, careful consideration was given to the filter design to minimize such artifacts. Additionally, an artifact subspace reconstruction (ASR) method was employed with a burst criterion of 10 standard deviations, effectively reconstructing the signal by removing transient high-variance artifacts while preserving the underlying neural dynamics.

To further address physiological artifacts, we applied independent component analysis (ICA) using the Picard algorithm, which offers faster convergence compared to traditional methods like Infomax. This approach allowed us to decompose the EEG signals into independent components, facilitating

the identification of artifacts. Components associated with eye movements, muscle bursts, and heartbeat were identified via temporal, spectral, and spatial inspection. We employed both manual identification by clinical electroencephalographers and ICLabel, an automatic classification algorithm within EEGLAB v2025.0.0 to ensure a consistent and objective preprocessing process.

Rereferencing and final quality ccontrol. Re-referencing was performed based on dataset characteristics to enhance the signal-to-noise ratio (SNR) for spatially localized high-frequency patterns. Common average referencing (CAR) was adopted as the default approach, except for datasets that were pre-referenced during acquisition. In these cases, the original referencing scheme was preserved to maintain consistency with the data’s inherent structure. To ensure data quality, we conducted comprehensive visual inspections and applied statistical quality control metrics to validate the retention of high-frequency content without introducing spurious amplification or phase distortion. This step is essential for preserving the spectral integrity of the signals, allowing for accurate downstream feature extraction and physiological interpretation.

Overall, our preprocessing pipeline prioritizes the preservation of physiologically meaningful features across different EEG frequency subbands while robustly eliminating confounding artifacts, providing a reliable foundation for subsequent analyses.

F.5 Data augmentation

To enhance model robustness and generalization in both supervised and self-supervised training approaches, we implement a suite of data augmentation techniques following [56, 55]. During each training iteration’s forward pass, a single augmentation method is randomly selected from the available options with equal probability. The augmentation techniques include temporal flipping, temporal masking, frequency masking, channel masking, jittering, and dropout, with the potential for future expansion. Below, we provide a detailed delineation of each method.

1. **Temporal Flipping:** This technique involves reversing the EEG data along the time axis. The likelihood of applying this augmentation is governed by a probability parameter, *prob*, set to 0.5 by default.
2. **Temporal Masking:** In this method, random time points are masked across all EEG channels. The fraction of time points masked is determined by a ratio parameter, with a default value of 0.1.
3. **Frequency Masking:** This method transforms EEG data into the frequency domain, randomly masks selected frequency bands, and then reverts the data to the time domain. The proportion of frequency bands masked is controlled by the ratio parameter, defaulting to 0.1.
4. **Channel Masking:** This method randomly masks entire channels across all time points. The percentage of channels masked is set by the ratio parameter, with a default of 0.1.
5. **Jittering:** Random noise, uniformly distributed between 0 and 1, is added to the raw EEG data. The noise amplitude is modulated by a scale parameter, which defaults to 0.1.
6. **Dropout:** Analogous to dropout in neural networks, this method randomly sets a fraction of EEG data values to zero. The proportion of values dropped is controlled by the ratio parameter, with a default value of 0.1.

G Training recipe

In this section, we describe the self-supervised pretraining framework for the **FAPEX** model, which leverages *sample-level contrasting* and *subject-level contrasting* to learn robust feature representations from unlabeled EEG data. These representations enhance performance on downstream tasks.

For each EEG input sample \mathbf{x}_i , where i denotes the sample index, we apply two augmentation techniques, a and b , to generate augmented views \mathbf{x}_i^a and \mathbf{x}_i^b . These views are processed by a backbone encoder $f(\cdot)$ to produce representations $\mathbf{h}_i^a = f(\mathbf{x}_i^a)$ and $\mathbf{h}_i^b = f(\mathbf{x}_i^b)$. A projection head $g(\cdot)$ further transforms these into denser representations, $\mathbf{z}_i^a = g(\mathbf{h}_i^a)$ and $\mathbf{z}_i^b = g(\mathbf{h}_i^b)$, which are used for contrastive learning. After pretraining, the projection head is discarded, and only the encoder $f(\cdot)$ is retained for downstream tasks.

787 **Sample-Level Contrastive Loss** Sample-level contrastive learning encourages the model to learn
788 generalizable EEG patterns by treating augmented views of the same sample as positive pairs and
789 views from other samples as negative pairs. This approach minimizes reliance on labeled data while
790 improving downstream task performance.

791 The sample-level InfoNCE contrastive loss for a batch \mathcal{B} is defined as:

$$\mathcal{L}_{\text{Sam}} = \mathbb{E}_{\mathbf{x}_i} \left[-\log \frac{\exp(\text{sim}(\mathbf{z}_i^a, \mathbf{z}_i^b)/\tau)}{\sum_{j \in \mathcal{B}} \exp(\text{sim}(\mathbf{z}_i^a, \mathbf{z}_j^b)/\tau)} \right], \quad (15)$$

792 where $\text{sim}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$ denotes cosine similarity, and $\tau > 0$ is a temperature parameter control-
793 ling the scale of similarities. This loss ensures that the model learns discriminative representations,
794 preserving similarities within a sample’s augmented views while distinguishing between different
795 samples.

796 **Subject-Level Contrastive Loss** Subject-level contrastive learning extends sample-level contrasting
797 by incorporating the stable medical state of subjects. Samples from the same subject are treated as
798 positive pairs, while those from different subjects are negative pairs. This approach reduces subject-
799 specific noise and enhances generalizable feature learning. The subject-level InfoNCE contrastive
800 loss is defined as:

$$\mathcal{L}_{\text{Sub}} = \mathbb{E}_{\mathbf{x}_i} \left[\mathbb{E}_{\mathbf{x}_k | s_k = s_i} \left[-\log \frac{\exp(\text{sim}(\mathbf{z}_i^a, \mathbf{z}_k^b)/\tau)}{\sum_{j \in \mathcal{B}} \exp(\text{sim}(\mathbf{z}_i^a, \mathbf{z}_j^b)/\tau)} \right] \right], \quad (16)$$

801 where \mathbf{x}_k represents samples from the same subject as \mathbf{x}_i (i.e., subject IDs satisfy $s_k = s_i$). This loss
802 promotes consistency in representations across samples from the same subject while minimizing the
803 influence of unrelated subject-specific features, improving generalization across subjects.

804 H Further implementation details and experimental settings

805 In this section, we delineate the comprehensive framework employed in our study, encompassing the
806 computational environment, model initialization protocols, and the rigorous hyperparameter tuning,
807 model selection and evaluation methodologies adopted to ensure the robustness and reproducibility
808 of our findings. By providing detailed insights into our implementation strategies, we aim to facilitate
809 the replication of our experiments and foster transparency in our research process.

810 H.1 Basic environment

811 All experiments are conducted utilizing 4 Nvidia A100 40GB GPUs, leveraging the computational
812 capabilities of PyTorch 2.4 [40] and CUDA 12.0. Upon publication, we will release the source
813 code to the public to promote transparency and facilitate reproducibility. In our experimental
814 framework, we employ the official implementations of all baseline models, where available, to ensure
815 methodological consistency and enhance the validity of comparative analyses. This strategy mitigates
816 potential discrepancies that may arise from custom implementations, reinforcing the integrity of our
817 experimental results.

818 H.2 Model configurations

819 We employ two variants of the FAPEX architecture: FAPEX-Small and FAPEX-Base. FAPEX-Small
820 consists of 4 layers with a hidden dimension of 128, while FAPEX-Base comprises 6 layers with a
821 hidden dimension of 256. Each layer in both variants includes two consecutive Frequency-aware
822 Neural Fourier Operators (FrNFOs). For supervised baselines, we set the hidden dimension to 256
823 across all models to ensure consistency. Self-supervised baselines adopt their respective recommended
824 configurations, as detailed in their original publications.

825 H.3 Subject agnosticity

826 In the patient-agnostic setting, samples from the same patient may appear in both the training and
827 testing sets, which is unrealistic for clinical applications. This overlap can lead to inflated performance

estimates, as the model might memorize patient-specific patterns instead of learning generalized features. In contrast, the patient-independent setting—central to our approach—ensures that no data from the same patient is used in both training and testing phases. This partitioning mirrors real-world scenarios, where models are expected to predict outcomes for new patients whose data was not used during training.

Challenges in model selection & evaluation Model selection and evaluation are critical stages in developing predictive models, particularly in complex domains such as healthcare. A common approach in the literature, as reviewed in [58], involves partitioning the dataset into a separate "holdout" or "test set" that contains unseen samples excluded from the model training process. This strategy typically uses a random selection method, such as a triplefold "training-validation-test" split of the original dataset, as adopted in works like [54, 53]. While this method provides a straightforward means of evaluating model performance, it is not without limitations. In particular, it fails to address potential validation bias, which can result in overly optimistic estimates of model performance. Cross-validation and resampling techniques, such as bootstrapping, offer more robust alternatives by providing more reliable estimates of out-of-sample performance—*i.e.*, a model's ability to generalize to new, unseen data. However, despite their advantages, these methods come with their own set of strengths and weaknesses. The choice of cross-validation strategy must be carefully tailored to the task at hand, as different variations of cross-validation are suited to different types of data and research objectives.

On the other hand, the challenges related to the data itself are equally significant. High-quality medical data, especially in domains like epilepsy research, is notoriously scarce and expensive to obtain. Data collection and annotation in medical research require substantial financial investment and time. In addition, many individual datasets suffer from limited sample sizes, which can significantly reduce the statistical power of the analysis. These issues are compounded by class imbalance, where certain categories, such as healthy individuals, outnumber patients, further exacerbating the risk of overfitting during model training. Overfitting occurs when the model becomes too attuned to the idiosyncrasies of the training data, resulting in poor generalization to new data and suboptimal performance in real-world applications. Addressing these challenges is crucial for developing robust and reliable predictive models.

Additionally, medical datasets often exhibit considerable heterogeneity due to multi-subject and multi-center data collection. Datasets gathered from different subjects and institutions may display substantial differences in distribution, arising from variations in data collection protocols, equipment, and patient demographics. This heterogeneity poses a significant obstacle in applying models trained on one dataset to others, as models trained on one distribution may fail to generalize across different institutional datasets. This challenge is particularly evident in diseases like epilepsy, where understanding the variability in disease manifestations is essential for accurate diagnosis and effective treatment. Consequently, ensuring that models can account for and adapt to this heterogeneity is a fundamental concern for researchers working with clinical data.

Subject-agnostic nested cross-validation (SANCV). In light of the aforementioned challenges, we propose *Patient-independent nested cross-validation* (SANCV) as a robust framework for model selection and evaluation. SANCV is employed throughout this study to ensure rigorous model assessment while maintaining patient independence at all stages. The complete procedure, including hyperparameter tuning, is detailed in Algorithm 1.

Unlike conventional data partitioning techniques, which divide a specific dataset into separate training, validation, and testing subsets, SANCV introduces an additional evaluation layer to mitigate data leakage and overfitting. This is particularly crucial in clinical settings, where patient-specific characteristics could unduly influence model performance. By enforcing strict patient independence—ensuring that no data from the same patient appear in both the training and testing stages—SANCV provides a more reliable estimate of a model's ability to generalize to unseen patients. Furthermore, as a variant of nested cross-validation, SANCV avoids biased metric estimation, which can otherwise arise from improper validation procedures. Compared to standard K -fold cross-validation, the nested design eliminates validation bias, thereby preventing overly optimistic estimates of model performance and generalizability.

In contrast to standard nested cross-validation, SANCV streamlines the evaluation process by subdividing the development set into a larger training subset and a smaller validation subset within each

outer fold. This design simplifies the procedure while preserving patient independence, thereby improving the robustness and practical applicability of the model in real-world clinical workflows. Each outer training fold consists of patient data strictly separate from the corresponding test fold, ensuring that hyperparameter tuning occurs exclusively within the inner cross-validation process.

Beyond its structural advantages, SANCV accommodates flexible hyperparameter optimization strategies, which are particularly valuable given the limited number of available subjects in many clinical studies. Traditional performance estimation techniques often rely on large external datasets, which may not always be accessible. By ensuring patient independence in both training and testing, and by incorporating a test set in each outer fold consisting solely of unseen patient data, SANCV produces realistic and unbiased performance estimates. This framework mitigates overfitting and provides more reliable evaluation metrics, even in the absence of external validation datasets.

Algorithm 1 Subject-Agnostic Nested Cross-Validation (SANCV)

Require: K_{outer} : number of outer folds

Require: $D = \{(X_i, y_i, s_i)\}_{i=1}^N$: dataset (features X_i , labels y_i , patient IDs s_i)

Require: P_{grid} : hyperparameter grid

Require: M : learning algorithm

Ensure: Outer test errors $\{E_{\text{test}}^i\}_{i=1}^{K_{\text{outer}}}$ and selected configs $\{p_{\text{opt}}^i\}_{i=1}^{K_{\text{outer}}}$

```

1: Output: Estimate of model performance  $\bar{E}_{\text{test}} = \frac{1}{K_{\text{outer}}} \sum_i E_{\text{test}}^i$ 
2: for  $i \leftarrow 1$  to  $K_{\text{outer}}$  do
3:   Split  $D$  into  $D_{\text{dev}}^{(i)}$  (development) and  $D_{\text{test}}^{(i)}$ , ensuring  $\{s_i\}$  disjoint
4:   Initialize best validation error  $E_{\text{val}}^* \leftarrow +\infty$ 
5:   for all  $p$  in  $P_{\text{grid}}$  do
6:     Split  $D_{\text{dev}}^{(i)}$  into  $D_{\text{train}}^{(i)}$  and  $D_{\text{val}}^{(i)}$ , ensuring patient-ID disjointness
7:     Train  $M$  on  $D_{\text{train}}^{(i)}$  with hyperparameters  $p$ , applying early stopping on  $D_{\text{val}}^{(i)}$ 
8:     Compute validation error

$$E_{\text{val}}^{(i,p)} = \mathcal{L}(M_p(D_{\text{train}}^{(i)}), D_{\text{val}}^{(i)})$$

9:     if  $E_{\text{val}}^{(i,p)} < E_{\text{val}}^*$  then
10:        $E_{\text{val}}^* \leftarrow E_{\text{val}}^{(i,p)}$ 
11:        $p_{\text{opt}}^{(i)} \leftarrow p$ 
12:       Save model checkpoint  $\hat{M}^{(i)} \leftarrow M_p$ 
13:     end if
14:   end for
15:   Evaluate  $\hat{M}^{(i)}$  on  $D_{\text{test}}^{(i)}$  to obtain

```

$$E_{\text{test}}^{(i)} = \mathcal{L}(\hat{M}^{(i)}, D_{\text{test}}^{(i)})$$

16: **end for**

17: **return** $\{E_{\text{test}}^i\}_{i=1}^{K_{\text{outer}}}$, $\{p_{\text{opt}}^i\}_{i=1}^{K_{\text{outer}}}$, and \bar{E}_{test}

894 H.4 Hyperparameter tuning strategy

895 Our hyperparameter tuning process is designed to be systematic and comprehensive while reducing
896 the risk of overfitting. We prioritize the hyperparameters and configurations recommended in each
897 model’s respective source code and/or paper. This ensures that our tuning process builds upon the
898 established knowledge and best practices for each model, providing a solid foundation for further
899 optimization. Based on the recommended hyperparameter settings, we conduct a grid search to
900 identify the optimal values for critical hyperparameters. The specific hyperparameters and their
901 search ranges are listed as follows:

- 902 1. **Learning rate:** We search for the optimal learning rate from the set $\{1 \times 10^{-2}, 5 \times 10^{-3}, 1 \times$
903 $10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}\}$. The learning rate is a critical hyperparameter that
904 controls the step size during optimization, and finding the right value is essential for
905 achieving stable and efficient training.

- 906 2. **Weight decay rate:** The optimal weight decay rate is identified from the set $\{1 \times 10^{-3}, 5 \times$
907 $10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}, 1 \times 10^{-5}\}$. Weight decay helps prevent overfitting by adding a
908 penalty term to the loss function, and tuning this hyperparameter is crucial for models prone
909 to overfitting.
- 910 3. **Dropout rate:** We search for the optimal dropout rate from the set $\{5 \times 10^{-2}, 1 \times 10^{-1}, 2 \times$
911 $10^{-1}\}$. Dropout is a regularization technique that randomly deactivates a fraction of neurons
912 during training, and the dropout rate significantly affects the model’s ability to generalize.
- 913 4. **Batch size:** We search for optimal batch size from the set $\{16, 24, 32, 48\}$.

914 We cease hyperparameter tuning when there is a consistent performance drop along one dimension of
915 the hyperparameter set on the validation set in each run of patient-independent nested cross-validation.
916 This approach allows us to limit the search to a subset of the full grid, reducing computational costs
917 while still identifying effective hyperparameter configurations.

918 H.5 Optimization strategy

919 **Optimizer.** We employ the AdamW optimizer for all models with $\beta_1 = 0.9$, $\beta_2 = 0.999$ [33].
920 AdamW is a variant of the Adam optimizer that decouples weight decay from gradient-based
921 parameter updates, providing more effective regularization and improved generalization. The choice
922 of AdamW is motivated by its ability to achieve stable and efficient training, particularly for large-
923 scale deep learning models.

924 **Early stopping.** To avoid overfitting and ensure computational efficiency, we adopt early stopping
925 throughout all experiments. Early stopping monitors the performance of the model on a validation
926 set and halts the training process if the performance begins to degrade, indicating that the model is
927 starting to overfit. Specifically, we set the total number of epochs to 50 and the patience parameter
928 to 5, meaning that training will stop if there is no improvement in validation performance over the
929 course of 5 consecutive epochs. This approach helps balance model training time with performance,
930 ensuring the model generalizes well without unnecessary overfitting.

931 **Exponential weight averaging.** In all experiments, we employ an Exponential Moving Average
932 (EMA) with a decay rate of 0.995. This technique is utilized to enhance the model’s robustness
933 against perturbations, thereby reducing the calibration error for deep learning models [36], and is
934 beneficial for inter-subject generalization in representational learning [21].

935 H.6 Metrics

936 In this section, we introduce the details of the metrics used in the paper.

- 937 1. **Balanced Accuracy** is a performance metric that considers the accuracy of each class in
938 imbalanced datasets, defined as the arithmetic mean of the true positive rate (TPR) and true
939 negative rate (TNR), thereby balancing the demand for higher recall and less false alarms in
940 seizure prediction.
- 941 2. **Sensitivity** measures the proportion of actual positive cases that are correctly identified by
942 the model. In seizure prediction, it quantifies the model’s ability to detect preictal segments.
943 Higher sensitivity indicates fewer missed seizure events, which is critical for reliable early
944 warning systems.
- 945 3. **Weighted F1** is the weighted mean of class-specific F1 scores, with each F1 score multiplied
946 by its class support (number of true instances) and then normalized by the total sample size.
947 It balances precision and recall across imbalanced classes.
- 948 4. **AUROC** is the area under the receiver operating characteristic curve, summarizing the
949 trade-off between TPR and false positive rate (FPR) across all classification thresholds.
- 950 5. **AUPRC** is the area under the precision-recall curve, capturing the trade-off between preci-
951 sion and recall—especially informative when positive events (preictal segments) are rare.
- 952 6. **Stratified Brier Score** is the average of squared differences between predicted probabil-
953 ities and true binary outcomes, computed within each class and then aggregated by class
954 prevalence. This metric penalizes both miscalibration and inaccuracy, while preserving the
955 influence of each class.

H.7 Performance estimation and statistical analysis

In clinical decision-making models, performance metrics often exhibit substantial variability due to the inherent heterogeneity of medical data. This variability stems from diverse patient populations, differing data collection methods, and the complex, non-linear nature of clinical datasets. Consequently, traditional summary statistics, such as the mean and standard deviation, may inadequately capture the central tendency and dispersion of these metrics, particularly in the presence of outliers or skewed distributions. To address this, we advocate for the use of the median and interquartile range (IQR) as primary descriptors of model performance. The median provides a robust measure of central tendency, less sensitive to extreme values common in clinical data. Similarly, the IQR offers a reliable measure of variability, encapsulating the spread of the middle 50% of the data and providing insights into the consistency of model performance across diverse patient subsets.

Given the non-normality, heteroscedasticity, and interdependence observed in clinical datasets and experimental results, nonparametric hypothesis tests are well-suited for assessing statistical significance. These challenges are compounded in our experiments by data heterogeneity within and across cohorts, as well as the limited number of performance metric results due to the computationally intensive training of our proposed model and multiple baselines.

Following established practices [48, 5, 4], we employ the Friedman test with the Iman-Davenport correction to assess the performance differences across multiple baseline models. For *post-hoc* pairwise comparisons, we utilize the Bayesian Wilcoxon signed-rank test [5], which accommodates non-normal distributions and accounts for data interdependence. This Bayesian approach yields a posterior distribution of the performance differences between models, enabling probabilistic conclusions regarding model superiority or equivalence.

I Ethics statement, social impacts, limitations and future work

I.1 Ethics statement

The development and evaluation of our seizure prediction model, FAPEX, utilize Scalp-EEG, ECoG and SEEG datasets, comprising both publicly available and proprietary collections from human and macaque subjects. All relevant ethical guidelines have been followed, and any necessary IRB and/or ethics committee approvals have been obtained. All human data were obtained from de-identified public repositories or proprietary datasets collected under rigorous ethical oversight, with approval from relevant institutional review boards (IRBs). Informed consent was secured from human participants or their legal guardians, and the appropriate institutional forms have been archived, ensuring adherence to ethical standards, including the Declaration of Helsinki.

Any patient/participant/sample identifiers included were not known to anyone (e.g., hospital staff, patients or participants themselves) outside the research group so cannot be used to identify individuals. For macaque data, experiments complied with the National Institutes of Health (NIH) guidelines and were approved by institutional animal care and use committees (IACUCs), emphasizing animal welfare and minimizing distress.

The FAPEX model aims to advance seizure prediction to improve epilepsy management and patient quality of life. Nevertheless, we recognize the ethical challenges of deploying predictive models in clinical contexts, particularly the risk of over-reliance on automated predictions, which could influence clinical decision-making. To address this, we underscore that FAPEX is designed as a decision-support tool, complementing rather than supplanting medical expertise. Data privacy remains a critical priority. All datasets were anonymized prior to analysis, and no personally identifiable information was accessed during model development. Robust safeguards have been implemented to ensure responsible data handling and prevent misuse of sensitive neurophysiological data. Furthermore, we advocate for equitable access to advanced medical technologies and are committed to disseminating our findings to support global epilepsy research, particularly in underserved regions with limited access to neurological care.

I.2 Limitations, societal impacts, and future directions

Despite the robust performance of FAPEX in seizure prediction across diverse EEG, ECoG, and SEEG datasets with varying electrode configurations, several limitations merit attention. First, the

1007 model’s generalizability may be constrained by inconsistencies in data acquisition protocols, such as
1008 differences in sampling rates and preprocessing methods across public and proprietary datasets. To
1009 mitigate this, we applied minimal preprocessing to all datasets, reducing dataset-specific artifacts
1010 and ensuring consistency. Second, physiological differences related to seizure subtypes and species
1011 may not fully translate to human epilepsy, potentially limiting applicability to certain patient groups
1012 due to the scarcity of well-annotated, seizure-specific data. Third, while our frequency band analysis
1013 provides valuable insights, it does not yet account for cross-frequency coupling or dynamic inter-
1014 band interactions, which could further elucidate seizure mechanisms. Finally, the computational
1015 demands of FAPEX, driven by its fractional neural framework and phase-aware processing, may hinder
1016 deployment on resource-constrained devices, such as wearable EEG systems.

1017 The societal implications of FAPEX are profound. Accurate seizure prediction could revolutionize
1018 epilepsy management by enabling timely interventions, enhancing patient autonomy, and improving
1019 quality of life. By providing early warnings, FAPEX empowers individuals with epilepsy to make
1020 informed decisions, such as avoiding high-risk activities during predicted pre-ictal periods. However,
1021 risks such as false positives, which may cause unnecessary stress or medication adjustments, and
1022 false negatives, which could undermine trust, must be addressed. We advocate for rigorous clinical
1023 validation and integration with clinician oversight to mitigate these concerns. While the model’s
1024 development using diverse datasets promotes inclusivity, ensuring equitable access to such technology
1025 in low-resource settings remains a critical challenge.

1026 Future research will prioritize several objectives. First, we will enhance model generalizability
1027 by incorporating additional datasets, including those from pediatric and drug-resistant epilepsy
1028 populations, to capture a wider range of seizure phenotypes. Second, we will explore cross-frequency
1029 coupling and multimodal data integration, such as combining EEG with functional MRI or behavioral
1030 data, to improve predictive accuracy. Third, we aim to optimize FAPEX for low-power hardware,
1031 enabling real-time, wearable applications to broaden accessibility. Finally, longitudinal clinical studies
1032 will validate the model’s real-world performance, facilitating regulatory approval and widespread
1033 adoption. These efforts seek to bridge the gap between research and clinical practice, ultimately
1034 alleviating the global burden of epilepsy.

1035 **Reference**

1036 **Bibliography**

- 1037 [1] M. A. Ahamed and Q. Cheng. Timemachine: A time series is worth 4 mambas for long-term
1038 forecasting. *arXiv preprint arXiv:2403.09898*, 2024.
- 1039 [2] M. A. Ahamed and Q. Cheng. Timemachine: A time series is worth 4 mambas for long-term
1040 forecasting. *arXiv preprint arXiv:2403.09898*, 2024.
- 1041 [3] N. N. Author. Suppressed for anonymity, 2021.
- 1042 [4] A. Benavoli, G. Corani, and F. Mangili. Should we really use post-hoc tests based on mean-
1043 ranks? *The Journal of Machine Learning Research*, 17(1):152–161, 2016.
- 1044 [5] A. Benavoli, G. Corani, J. Demšar, and M. Zaffalon. Time for a change: a tutorial for comparing
1045 multiple classifiers through bayesian analysis. *Journal of Machine Learning Research*, 18(77):
1046 1–36, 2017.
- 1047 [6] M. H. A. Biswas, P. Massopust, and R. Ramakrishnan. Deep convolutional neural networks and
1048 data approximation using the fractional fourier transform. *arXiv preprint arXiv:2408.06757*,
1049 2024. URL <https://arxiv.org/abs/2408.06757>.
- 1050 [7] H. Chen, V. Luong, L. Mukherjee, and V. Singh. Simpletm: A simple baseline for multivariate
1051 time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2025.
1052 URL <https://openreview.net/forum?id=oANkBaVci5>.
- 1053 [8] N. Chen, C. Liu, N. Yan, W. Hu, J.-G. Zhang, Y. Ge, and F.-G. Meng. A cynomolgus monkey
1054 model of temporal lobe epilepsy. *PLoS ONE*, 8(8):e72336, 2013. doi: 10.1371/journal.pone.
1055 0072336.
- 1056 [9] P. Chen, Y. Zhang, Y. Cheng, Y. Shu, Y. Wang, Q. Wen, B. Yang, and C. Guo. Pathformer:
1057 Multi-scale transformers with adaptive pathways for time series forecasting. In *International
1058 Conference on Learning Representations (ICLR)*, 2024. URL [https://openreview.net/
1059 forum?id=1Jk0CMP2aW](https://openreview.net/forum?id=1Jk0CMP2aW).
- 1060 [10] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data.
1061 *SIAM Review*, 51(4):661–703, Nov. 2009. ISSN 1095-7200. doi: 10.1137/070710111. URL
1062 <http://dx.doi.org/10.1137/070710111>.
- 1063 [11] H. J. Cunningham, G. Giannone, M. Zhang, and M. P. Deisenroth. Reparameterized multi-
1064 resolution convolutions for long sequence modelling. In *The Thirty-eighth Annual Conference
1065 on Neural Information Processing Systems*, 2024.
- 1066 [12] H. J. Cunningham, G. Giannone, M. Zhang, and M. P. Deisenroth. Reparameterized multi-
1067 resolution convolutions for long sequence modelling. In *NeurIPS*, 2024. URL [https://arxiv.
1068 org/abs/2408.09453](https://arxiv.org/abs/2408.09453).
- 1069 [13] L. donghao and wang xue. ModernTCN: A modern pure convolution structure for general time
1070 series analysis. In *The Twelfth International Conference on Learning Representations*, 2024.
- 1071 [14] J. Duan, W. Zheng, Y. Du, W. Wu, H. Jiang, and H. Qi. Mf-clr: Multi-frequency contrastive
1072 learning representation for time series. In *International Conference on Machine Learning
1073 (Poster)*, 2024. URL <https://openreview.net/forum?id=ec07W0I1MD>.
- 1074 [15] E. Eldele, M. Ragab, Z. Chen, M. Wu, and X. Li. Tslanet: Rethinking transformers for time
1075 series representation learning. In *International Conference on Machine Learning (ICML)*, 2024.
1076 URL <https://arxiv.org/abs/2404.08472>.
- 1077 [16] E. Eldele, M. Ragab, Z. Chen, M. Wu, and X. Li. Tslanet: Rethinking transformers for time
1078 series representation learning. In *International Conference on Machine Learning*, 2024.

- [17] T. Feng, J. Ni, E. Gleichgerrcht, and W. Jin. Seizureformer: A transformer model for ica-based seizure risk forecasting. *arXiv preprint arXiv:2504.16098*, 2025. URL <https://arxiv.org/abs/2504.16098>.
- [18] H. Gui, X. Li, and X. Chen. Vector quantization pretraining for eeg time series with random projection and phase alignment. In *International Conference on Machine Learning (ICML)*, 2024. URL <https://openreview.net/forum?id=7uwLvFvpis>.
- [19] Y. Gui, M. Chen, Y. Su, G. Luo, and Y. Yang. Eegmamba: Bidirectional state space model with mixture of experts for eeg multi-task classification, 2024. URL <https://arxiv.org/abs/2407.20254>.
- [20] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and F. Petitjean. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020.
- [21] P. Izmailov, D. Podoprikin, T. Garipov, D. Vetrov, and A. G. Wilson. Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407*, 2018.
- [22] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In *International Conference on Learning Representations (ICLR)*, 2024. URL <https://openreview.net/forum?id=QzTpTRVtrP>.
- [23] J. Jing and et al. Development of expert-level classification of seizures and rhythmic and periodic patterns during eeg interpretation. *Neurology*, 100(17):e1750–e1762, 2023. doi: 10.1212/WNL.0000000000207127.
- [24] 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.
- [25] M. Kim, Y. Hioka, and M. Witbrock. Neural fourier modelling: A highly compact approach to time-series analysis. *arXiv preprint arXiv:2410.04703*, 2024.
- [26] M. Kim, Y. Hioka, and M. Witbrock. Neural fourier modelling: A highly compact approach to time-series analysis. *arXiv preprint arXiv:2410.04703*, 2024. URL <https://arxiv.org/abs/2410.04703>.
- [27] 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.
- [28] 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. In *International Conference on Learning Representations (ICLR)*, 2022. URL <https://openreview.net/forum?id=Pyraformer>.
- [29] 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.
- [30] Y. Liu, H. Wu, J. Wang, and M. Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. URL <https://arxiv.org/abs/2205.14415>.
- [31] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023. URL <https://arxiv.org/abs/2310.06625>.
- [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] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=Bkg6RiCqY7>.

- [34] D. Luo and X. Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *International Conference on Learning Representations (ICLR) Spotlight*, 2024. URL <https://openreview.net/forum?id=vpJMJerXHU>.
- [35] W. Löscher. Animal models of acquired epilepsy: insights into mechanisms of human epileptogenesis. *Neuropathology and Applied Neurobiology*, 44(1):7–24, 2018. doi: 10.1111/nan.12467.
- [36] D. Morales-Brotons, T. Vogels, and H. Hendrikx. Exponential moving average of weights in deep learning: Dynamics and benefits, 2024. URL <https://arxiv.org/abs/2411.18704>.
- [37] H. Ni, Y. Yang, F. Zhang, Y. Sun, Y. Zheng, J. Zhu, and K. Xu. Dataset of long-term multi-site lfp activity with spontaneous chronic seizures in temporal lobe epilepsy rats. *Scientific Data*, 12(1):709, 2025.
- [38] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *ICLR*, 2023.
- [39] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *ICLR*, 2023. URL <https://arxiv.org/abs/2211.14730>.
- [40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, pages 8024–8035, 2019.
- [41] M. L. Saggio, J. Schmalz, and V. K. Jirsa. An exploratory computational analysis in mice brain networks of widespread epileptic seizure onset locations along with potential therapeutic targets. *Frontiers in Neurology*, 14:1098315, 2023. doi: 10.3389/fneur.2023.1098315.
- [42] J. Shi, K. A. Wang, and E. B. Fox. Sequence modeling with multiresolution convolutional memory. In *International Conference on Machine Learning*, 2023.
- [43] 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.
- [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. doi: 10.1109/TNSRE.2022.3230250.
- [45] 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. *arXiv preprint arXiv:2002.10061*, 2020. URL <https://arxiv.org/abs/2002.10061>.
- [46] 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>.
- [47] van Blooij D., D. M., Z. W, H. D., L. F., and Z. M. "dataset clinical epilepsy ieeg to bids - respect_{longterm}_{ieeg}", 2021.
- [48] J. van Doorn, A. Ly, M. Marsman, and E.-J. Wagenmakers. Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and spearman’s ρ . *Journal of Applied Statistics*, 47(16): 2984–3006, 2020.
- [49] 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 *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/4540d267eeec4e5dbd9dae9448f0b739-Abstract-Conference.html.
- [50] 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>.

- [51] 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 *International Conference on Learning Representations (ICLR)*, 2024. URL <https://openreview.net/forum?id=7oLshfEIC2>.
- [52] Y. Wang, Y. Han, H. Wang, and X. Zhang. Contrast everything: A hierarchical contrastive framework for medical time-series. In *NeurIPS 2023*, 2023. URL <https://arxiv.org/abs/2310.14017>.
- [53] Y. Wang, Y. Han, H. Wang, and X. Zhang. Contrast everything: A hierarchical contrastive framework for medical time-series. *Advances in Neural Information Processing Systems*, 36, 2024.
- [54] Y. Wang, N. Huang, T. Li, Y. Yan, and X. Zhang. Medformer: A multi-granularity patching transformer for medical time-series classification. *Advances in Neural Information Processing Systems*, 37:36314–36341, 2024.
- [55] Y. Wang, Y. Liu, J. Li, and J. Zhou. Comets: Contrastive pre-training of multivariate time series. *arXiv preprint arXiv:2401.12345*, 2024.
- [56] Y. Wang, N. Huang, N. Mammone, M. Cecchi, and X. Zhang. Lead: Large foundation model for eeg-based alzheimer’s disease detection. *arXiv preprint arXiv:2502.01678*, 2025.
- [57] T. Wiatowski and H. Bölcskei. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory*, 64(3):1845–1866, 2018. doi: 10.1109/TIT.2017.2776228.
- [58] D. Wilimitis and C. G. Walsh. Practical considerations and applied examples of cross-validation for model development and evaluation in health care: Tutorial. *JMIR AI*, 2:e49023, Dec 2023. ISSN 2817-1705. doi: 10.2196/49023. URL <https://ai.jmir.org/2023/1/e49023>.
- [59] D. Wu, S. Li, J. Yang, and M. Sawan. Neuro-bert: Rethinking masked autoencoding for self-supervised neurological pretraining. *arXiv preprint arXiv:2204.12440*, 2022. URL <https://arxiv.org/abs/2204.12440>.
- [60] 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.
- [61] 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.
- [62] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *ICLR*, 2023. URL <https://github.com/thuml/TimesNet>.
- [63] C. Yang, M. B. Westover, and J. Sun. Biot: Cross-data biosignal learning in the wild. *arXiv preprint arXiv:2305.10351*, 2023. URL <https://arxiv.org/abs/2305.10351>.
- [64] Y. Yang, F. Zhang, X. Gao, L. Feng, and K. Xu. Progressive alterations in electrophysiological and epileptic network properties during the development of temporal lobe epilepsy in rats. *Epilepsy & Behavior*, 141:109120, 2023.
- [65] 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. URL <https://arxiv.org/abs/2404.05192>.
- [66] 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.
- [67] K. Yi, Q. Zhang, W. Fan, S. Wang, P. Wang, H. He, D. Lian, N. An, L. Cao, and Z. Niu. Frequency-domain mlps are more effective learners in time series forecasting. *arXiv preprint arXiv:2311.06184*, 2023. URL <https://arxiv.org/abs/2311.06184>.
- [68] H. Yu, P. Guo, and A. Sano. Adawavenet: Adaptive wavelet network for time series analysis. *arXiv preprint arXiv:2405.11124*, 2024. URL <https://arxiv.org/abs/2405.11124>.
- [69] 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.

- [20] D. Zhang, Z. Yuan, Y. Yang, J. Chen, J. Wang, and Y. Li. Brant: Foundation model for intracranial neural signal. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/535915d26859036410b0533804cee788-Paper-Conference.pdf.
- [24] F. N. H. Zhang and Eegmamba: Bidirectional state space models with mixture of experts for universal eeg classification. *arXiv preprint arXiv:2407.20254*, 2024.
- [22] X. Zhang, Z. Zhao, T. Tsiligkaridis, and M. Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. *arXiv preprint arXiv:2206.08496*, 2022. URL <https://arxiv.org/abs/2206.08496>.
- [23] 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.