MEREDITH: LEARNING PHYSIOLOGICAL RHYTHMS AND DELAYED CHANNEL INTERACTIONS FOR MEDICAL TIME SERIES CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate medical time series classification is important for identifying pathological conditions and supporting clinical diagnosis and disease management. Recent deep learning methods have achieved remarkable progress in this task, with their core designs largely centered on temporal representation learning and cross-channel modeling. However, prevailing strategies, such as generic multi-scale modeling for temporal representations and attention mechanisms or graph learning for cross-channel relationships, often lack explicit consideration of the intrinsic characteristics of physiological signals, potentially limiting the identification of diagnostically relevant patterns. To address this limitation, we draw upon physiological characteristics of medical time series to guide temporal representation learning and cross-channel modeling. First, we introduce *Rhythm Mixer*, which incorporates physiological rhythms into representation learning through learnable frequency decomposition and adaptive frequency-band aggregation, enabling medical-specific temporal representations beyond generic multi-scale modeling. Second, motivated by the temporal delays that can arise from the propagation and interaction of physiological activities, we propose the *Delay-Aware Channel Enhancement Network* (DACE Net), which accounts for different temporal delays to enhance each channel representation with delay-aware information from other channels. Building around these two core components, we develop *Meredith*, a physiology-informed framework for medical time series classification. Extensive experiments on five medical time series benchmarks demonstrate the strong and robust performance of *Meredith* against 10 competitive baselines under both Fixed Subject Split and Monte Carlo Cross-Validation protocols.

1 INTRODUCTION

Medical time series, such as electrocardiogram (ECG) and electroencephalogram (EEG) recordings, capture physiological states and dynamics of the human body through sequential measurements (Wang et al., 2025; Jovic et al., 2025; Badr et al., 2024). Different physiological and pathological conditions can manifest as distinct patterns in these signals. For example, atrial fibrillation alters cardiac electrical activity and rhythm reflected in ECG recordings (Joglar et al., 2024), while epilepsy can produce abnormal brain electrical activity captured by EEG (Islam et al., 2023). Accurately recognizing such patterns and classifying medical time series are therefore important for clinical diagnosis, disease monitoring, and treatment decision-making.

Deep learning has achieved substantial progress in medical time series classification by learning discriminative representations from complex physiological signals. Representative methods, such as Medformer (Wang et al., 2024b), MedGNN (Fan et al., 2025), and TeCh (Yu et al., 2026), adopt diverse architectures and modeling strategies, yet commonly focus on temporal representation learning and channel dependency modeling, two fundamental challenges for accurate medical time series classification. For temporal representation learning, a prevalent strategy is to construct multi-scale representations from the same input, such as representations with different temporal granularities (Wang et al., 2024b) or resolutions (Fan et al., 2025; Zhang et al., 2026), typically through different patch lengths or convolutional kernels with varying receptive fields. While effective, these temporal scales are typically manually specified without explicitly considering the intrinsic physio-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

logical characteristics of medical signals, such as the canonical rhythms in EEG (e.g., delta and theta rhythms) and the characteristic waveform components in ECG (e.g., P and T waves). Consequently, such strategies remain broadly applicable to general multivariate time series (Wang et al., 2024a) rather than being specifically tailored to the medical domain. For channel dependency modeling, attention mechanisms (Wang et al., 2024b) and graph learning (Fan et al., 2025) have demonstrated their effectiveness in capturing interactions among physiological channels. However, modeling these interactions directly from observed representations may overlook how they are shaped by the physiological processes that generate and propagate the recorded signals (Kléber & Rudy, 2004; Fu et al., 2024; Muller et al., 2018). Encouragingly, recent research has begun to place greater emphasis on medical-specific modeling. TeCh (Yu et al., 2026), for example, argues that the decentralized nature of Transformer attention is inconsistent with the inherently centralized organization of medical time series and accordingly introduces a centralized MLP-based module to better accommodate this characteristic. Nevertheless, the intrinsic physiological characteristics of medical time series remain insufficiently explored and leveraged in existing model designs. This raises a fundamental question: *how can we design classification models that better leverage these characteristics?*

To answer this question, we revisit medical time series from a physiological perspective and focus on two characteristics that are closely related to temporal representation learning and cross-channel modeling. First, physiological signals exhibit characteristic temporal structures. EEG signals, for instance, contain physiologically meaningful delta, theta, alpha, beta, and gamma rhythms (Zebhauser et al., 2026; Wang & Mengoni, 2022), while ECG signals comprise characteristic waveform components, such as the P wave, QRS complex, and T wave, each exhibiting distinct temporal and spectral characteristics (Anbalagan et al., 2023). Motivated by these structures, we introduce *Rhythm Mixer*, which incorporates physiological priors into temporal modeling through learnable frequency decomposition and adaptive frequency-band information fusion. Specifically, it initializes a Rhythm Bank comprising a set of frequency bands based on the physiological characteristics of different signals, extracts the information within each band accordingly, independently transforms the extracted components into band-specific representations, and subsequently adaptively aggregates these representations. Second, the propagation and interaction of physiological activities can involve temporal delays. For example, cardiac excitation propagates sequentially through the cardiac conduction system and myocardium (Kléber & Rudy, 2004), while neural activities across different brain regions can exhibit temporally lagged interactions (Muller et al., 2018). These physiological dynamics can manifest as delayed relationships among the recorded channels, making their dependencies difficult to fully characterize through conventional channel dependency modeling. Motivated by this characteristic, we introduce the *Delay-Aware Channel Enhancement Network* (DACE Net), which explicitly incorporates temporal delays to enhance each channel representation with delay-aware information from other channels. Specifically, DACE Net adaptively weights different delays up to a predefined maximum delay for each channel pair and aggregates the resulting delay-aware cross-channel information to refine channel representations.

Building upon these two components, we develop *Meredith*, a physiology-informed framework for medical time series classification, where *Rhythm Mixer* captures physiologically meaningful temporal structures and *DACE Net* enhances channel representations with delay-aware information from other channels. The outputs of these two components are then integrated and further modeled by a Transformer encoder (Vaswani et al., 2017) with cross-attention before being passed to the classification head. Extensive experiments on five medical time series benchmarks demonstrate the strong performance of *Meredith* compared with 10 competitive baselines under both Fixed Subject Split (Wang et al., 2024b; Fan et al., 2025) and Monte Carlo Cross-Validation (Huang et al., 2026) protocols. In summary, our main contributions are as follows:

- We introduce *Rhythm Mixer*, which incorporates physiological rhythms into representation learning through learnable frequency decomposition and adaptive frequency-band aggregation, enabling medical-specific temporal representations beyond generic multi-scale modeling.
- We propose *Delay-Aware Channel Enhancement Network* (DACE Net), which accounts for temporal delays to enhance each channel representation with delay-aware information from other channels, thereby facilitating subsequent channel dependency modeling.
- We develop *Meredith*, a unified physiology-informed framework for medical time series classification built around two core components, *Rhythm Mixer* and *DACE Net*. Extensive experiments on five benchmarks under both Fixed Subject Split and Monte Carlo Cross-Validation protocols demonstrate its strong and robust performance against 10 competitive baselines.

2 RELATED WORK

2.1 TEMPORAL REPRESENTATION LEARNING FOR MEDICAL TIME SERIES

Learning informative temporal representations is fundamental to medical time series classification, as physiological signals often contain temporal patterns occurring at different time scales (Ma et al., 2026; Qamar & Abibullaev, 2026). Accordingly, recent studies have explored multi-granularity or multi-resolution temporal modeling. Medformer (Wang et al., 2024b) adopts different patch lengths to construct representations at multiple temporal granularities. Similarly, MedGNN (Fan et al., 2025) and MedMamba (Zhang et al., 2026) employ temporal convolutions with different kernel sizes to capture local patterns at different temporal resolutions. Despite their effectiveness, these methods primarily capture temporal diversity through generic architectural choices such as patch lengths or convolutional receptive fields, without explicitly exploiting the characteristic physiological rhythms inherent in medical signals. In contrast, our Rhythm Mixer explicitly incorporates physiologically meaningful rhythms into temporal representation learning through learnable frequency decomposition and adaptive integration of rhythm-specific information.

2.2 CHANNEL DEPENDENCY MODELING FOR TIME SERIES

Modeling dependencies among channels is essential for multivariate time series analysis, as interactions across channels provide complementary information beyond their individual temporal dynamics. Mainstream approaches to channel dependency modeling can be broadly categorized into Transformer-based and graph-based methods. Transformer-based methods typically capture channel interactions through attention mechanisms. Crossformer (Zhang & Yan, 2023) and CARD (Zhou et al., 2024) model cross-channel interactions at the segment or patch level, while iTransformer (Liu et al., 2024) embeds each channel's observations into a channel token and applies self-attention over these tokens to capture cross-channel correlations. For medical time series, Medformer (Wang et al., 2024b) introduces cross-channel patching to capture cross-channel correlations within multi-granularity temporal representations. Graph-based methods instead represent channels as nodes and model their dependencies through learned graph structures and information propagation. AGCRN (Bai et al., 2020) and MTGNN (Wu et al., 2020) adaptively learn inter-channel graphs to propagate and aggregate information across channels, while FourierGNN (Yi et al., 2023) extends such interactions to a hypervariate graph across channels and timestamps. For medical time series, MedGNN (Fan et al., 2025) learns adaptive channel graphs at multiple temporal resolutions, while MedMamba (Zhang et al., 2026) combines sample-adaptive directed graph learning with channel-wise Mamba modeling. Despite these advances, existing methods rarely account for temporal delays associated with the propagation and interaction of physiological activities, which can lead to delayed relationships across recording channels (Kariman et al., 2026; Chiarion et al., 2023). To address this limitation, our Delay-Aware Channel Enhancement Network accounts for different temporal delays to enhance each channel representation with delay-aware information from other channels, thereby facilitating subsequent channel dependency modeling.

3 PRELIMINARIES

3.1 PROBLEM FORMULATION

The medical time series classification task aims to distinguish different clinical conditions from multichannel physiological signals. Given a multichannel medical time series $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C and T denote the number of channels and the sequence length, respectively, the task is to identify its corresponding clinical condition $y \in \{1, \ldots, N\}$, where N is the number of classes. Specifically, a classification model f_θ maps the input sample $\mathbf{X}$ to classification logits $\hat{\mathbf{y}} = f_\theta(\mathbf{X}) \in \mathbb{R}^N$, where θ denotes the model parameters.

3.2 EVALUATION SETTINGS

We evaluate the model under two subject-independent protocols to provide a comprehensive assessment of its classification performance.

Fixed Subject Split. Following the commonly adopted evaluation protocol (Wang et al., 2024b; Fan et al., 2025; Yu et al., 2026), we use a predefined subject split, where the subject IDs assigned to the training, validation, and test sets are fixed and mutually exclusive. For each dataset, we run the model five times with different random seeds while keeping the subject split unchanged. The mean and standard deviation over the five runs are reported as the final results.

Monte Carlo Cross-Validation. To account for the variability introduced by different subject partitions, we further adopt Monte Carlo cross-validation (Huang et al., 2026). For each random

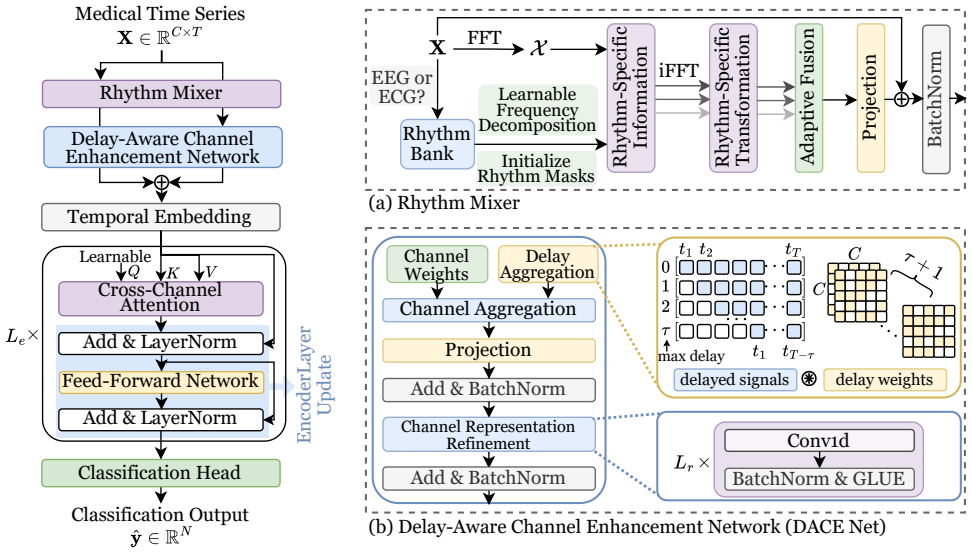


Figure 1: Overall architecture of *Meredith*. The framework comprises two parallel physiology-informed modules, (a) *Rhythm Mixer* for rhythm-aware representation learning and (b) *Delay-Aware Channel Enhancement Network* (DACE Net) for delay-aware channel enhancement, followed by cross-channel encoding and classification. In (a), physiologically meaningful rhythm components are adaptively extracted through learnable frequency decomposition and subsequently fused according to their varying contributions. In (b), delayed signals are first constructed at different temporal offsets, followed by selective delay and channel aggregation and subsequent representation refinement. The resulting features are subsequently integrated and processed by a Transformer encoder before being mapped to the class space by the classification head.

seed, subjects are grouped by their class labels and randomly shuffled within each class. The subjects from each class are then partitioned into the training, validation, and test sets at a ratio of 6:2:2, with the resulting sets containing mutually exclusive subjects. This procedure is repeated using five different random seeds, resulting in five independent subject splits. The model is trained and evaluated separately on each split, and the mean and standard deviation across the five runs are reported as the final results.

4 Method

In this section, we present our proposed method, *Meredith*, for medical time series classification. The central idea is to incorporate physiologically grounded inductive biases into the model design by jointly exploiting characteristic rhythms and temporal delays underlying physiological signals.

4.1 Overview

Meredith is a physiology-informed framework for medical time series classification built around two core components, *Rhythm Mixer* and *Delay-Aware Channel Enhancement Network* (DACE Net). Given an input medical time series $\mathbf{X} \in \mathbb{R}^{C \times T}$, both Rhythm Mixer and DACE Net directly process $\mathbf{X}$, with Rhythm Mixer capturing physiologically meaningful rhythms and DACE Net enhancing channel representations by incorporating delay-aware information from other channels. Their outputs are then integrated and further modeled by a Transformer encoder with learnable-query cross-channel attention, followed by a linear classification head to produce the final classification logits.

4.2 Rhythm Mixer

To incorporate physiological rhythms into temporal representation learning, we introduce *Rhythm Mixer*, which begins with learnable frequency decomposition. Effective frequency decomposition relies on an appropriate *Rhythm Bank*, which provides physiologically informed frequency ranges for initialization across different types of medical signals. For EEG, the Rhythm Bank consists of the canonical delta, theta, alpha, beta, and gamma bands (Zebhauser et al., 2026; Wang & Mengoni, 2022). For ECG, we combine the characteristic spectral ranges of major waveform components, including the P wave, QRS complex, and T wave (Anbalagan et al., 2023), with general frequency partitions to capture broader spectral variations. Accordingly, for each signal type, we define the Rhythm Bank as $\mathcal{R} = \{[r_k^b, r_k^e]\}_{k=1}^K$, where K denotes the number of rhythm bands, and r_k^b and r_k^e

denote the lower and upper frequency bounds of the k-th band, respectively. Based on the predefined Rhythm Bank $\mathcal{R}$, we next perform learnable frequency decomposition to extract rhythm-specific spectral components from the input signal.

Given an input medical time series $\mathbf{X} \in \mathbb{R}^{C\times T}$, we first transform it into the frequency domain using the real-valued fast Fourier transform (rFFT) (Sneddon, 1995; Sorensen et al., 1987), yielding $\mathcal{X} = \mathrm{rFFT}(\mathbf{X}) \in \mathbb{C}^{C\times F}$, where $F = \lfloor T/2 \rfloor + 1$ denotes the number of frequency bins. For the k-th rhythm in $\mathcal{R}$, the corresponding frequency range $[r_k^b, r_k^e]$ is used to initialize its center frequency μ_k and bandwidth σ_k as $\mu_k = \frac{r_k^b + r_k^e}{2}, \sigma_k = \frac{r_k^e - r_k^b}{2}$. Rather than using the predefined frequency ranges as fixed boundaries, μ_k and σ_k are further optimized during training. Based on these learnable parameters, we construct a Rhythm Mask $\mathcal{M} \in \mathbb{R}^{K\times F}$, whose k-th row is given by $\mathcal{M}_k = \exp\left(-\frac{1}{2}\left(\frac{\mathbf{f}-\mu_k}{\sigma_k}\right)^2\right) \in \mathbb{R}^F$, where $\mathbf{f} \in \mathbb{R}^F$ denotes the frequencies corresponding to the F frequency bins. For the k-th rhythm, $\mathcal{M}_k$ extracts the physiological spectral information associated with this rhythm from $\mathcal{X}$, and the resulting spectral representation is subsequently mapped back to the time domain through the inverse real-valued fast Fourier transform (irFFT) (Sneddon, 1995; Sorensen et al., 1987) as follows:

$$\mathcal{X}_k = \mathcal{X} \odot \mathcal{M}_k \in \mathbb{C}^{C\times F}, \quad \mathbf{X}_k = \mathrm{irFFT}(\mathcal{X}_k) \in \mathbb{R}^{C\times T}, \tag{1}$$

where $\odot$ denotes element-wise multiplication. The rhythm-specific components $\{\mathbf{X}_k\}_{k=1}^K$ capture distinct signal characteristics, motivating dedicated representation learning for each rhythm. We therefore employ an independent transformation $\phi_k(\cdot)$ for each rhythm, yielding a set of rhythm-specific representations $\{\mathbf{H}_k\}_{k=1}^K$, with $\mathbf{H}_k = \phi_k(\mathbf{X}_k) \in \mathbb{R}^{C_e\times T}$. Here, C_e denotes the representation dimension, and $\phi_k(\cdot)$ consists of two 1D convolutional layers, each followed by batch normalization (Ioffe & Szegedy, 2015) and GELU activation.

To jointly leverage the discriminative information captured by different rhythms, we integrate the learned rhythm-specific representations while adaptively accounting for their varying contributions. Specifically, each rhythm-specific representation $\mathbf{H}_k$ ($k = 1, \ldots, K$) is first temporally aggregated through global average pooling and then processed by a lightweight gating function to produce a rhythm score $s_k = g\left(\mathrm{AvgPool}(\mathbf{H}_k)\right) \in \mathbb{R}$, where $g(\cdot)$ comprises two 1D convolutional layers and an intermediate GELU activation. The resulting scores $\{s_k\}_{k=1}^K$ are normalized across all rhythms via a softmax function to obtain the fusion weights $\{\alpha_k\}_{k=1}^K$, based on which the rhythm-specific representations are integrated as $\widetilde{\mathbf{H}} = \sum_{k=1}^K \alpha_k \mathbf{H}_k \in \mathbb{R}^{C_e\times T}$. Finally, $\widetilde{\mathbf{H}}$ is processed by a 1D convolution layer $\psi(\cdot)$, followed by a residual connection and batch normalization, yielding the Rhythm Mixer output $\mathbf{X}_{\mathrm{rhy}} = \mathrm{BatchNorm}\left(\psi(\widetilde{\mathbf{H}}) + \mathbf{X}\right) \in \mathbb{R}^{C\times T}$

4.3 Delay-Aware Channel Enhancement Network

The propagation and interaction of physiological activities can involve temporal delays, which may manifest as delayed relationships among the recorded channels (Muller et al., 2018). To account for such delays, we introduce the *Delay-Aware Channel Enhancement Network* (DACE Net), which enhances each channel representation with delay-aware information from other channels.

Given an input medical time series $\mathbf{X} \in \mathbb{R}^{C\times T}$, DACE Net first constructs delayed versions of each channel to expose its information across different temporal delays. Specifically, for the c-th channel signal $\mathbf{x}_c \in \mathbb{R}^T$, we consider temporal delays $d = 0, \ldots, \tau$, where d denotes the delay in time steps and τ is the maximum temporal delay. For each delay d, we apply a shift operator $\mathcal{S}_d(\cdot)$ that shifts the signal by d time steps and pads zeros at the beginning to preserve the temporal length T. This produces $\{\mathcal{S}_d(\mathbf{x}_c)\}_{d=0}^\tau$ for each channel, and stacking the delayed signals from all C channels yields $\mathcal{D} \in \mathbb{R}^{C\times(\tau+1)\times T}$. The delayed signals in $\mathcal{D}$ provide candidate information across different temporal delays, whose contributions to channel enhancement may vary for different channel pairs. We therefore introduce learnable delay weights $\mathbf{W}^{\mathrm{delay}} \in \mathbb{R}^{C\times C\times(\tau+1)}$, where $\mathbf{W}_{ij}^{\mathrm{delay}} \in \mathbb{R}^{\tau+1}$ denotes the delay weights from source channel j to target channel i, normalized by a softmax over the delay dimension. Using these weights, the delay-aware information from source channel j for enhancing target channel i is obtained by adaptively aggregating its delayed signals: $\widetilde{\mathbf{x}}_{ij} = \sum_{d=0}^\tau W_{ijd}^{\mathrm{delay}} \mathcal{S}_d(\mathbf{x}_j)$, where $\widetilde{\mathbf{x}}_{ij} \in \mathbb{R}^T$. Collecting $\widetilde{\mathbf{x}}_{ij}$ over all channel pairs yields the pairwise delay-aware representations $\widetilde{\mathbf{X}} \in \mathbb{R}^{C\times C\times T}$.

At this stage, the delay-aware information for each target channel remains distributed across multiple source-channel representations. To consolidate this information into a unified representation for each target channel, we perform channel aggregation. Accordingly, we introduce learnable channel weights $\mathbf{W}^{\text{channel}} \in \mathbb{R}^{C\times C}$, where the i-th row $\mathbf{W}_i^{\text{channel}} \in \mathbb{R}^C$ contains the softmax-normalized weights assigned to different source channels for target channel i. The enhanced representation of the i-th channel is then obtained as $\mathbf{p}_i = \sum_{j=1}^{C} W_{i,j}^{\text{channel}} \widetilde{\mathbf{x}}_{ij} \in \mathbb{R}^T$. Stacking the representations of all channels yields the delay-aware channel representation $\mathbf{P} \in \mathbb{R}^{C\times T}$. Subsequently, $\mathbf{P}$ is transformed by a 1D convolutional projection $\varphi(\cdot)$ and combined with the input $\mathbf{X}$ through a residual connection, followed by batch normalization, yielding $\widetilde{\mathbf{P}} = \text{BatchNorm}\left(\varphi(\mathbf{P}) + \mathbf{X}\right) \in \mathbb{R}^{C\times T}$.

Although the preceding delay and channel aggregation selectively combines information across different delays and channels, the resulting representation is primarily a weighted aggregation. Further refinement is therefore necessary to reorganize the aggregated information and strengthen its nonlinear expressive capability. To this end, we introduce a set of refinement layers on top of $\widetilde{\mathbf{P}}$. Let L_r denote the number of refinement layers, and let $\rho_\ell(\cdot)$ denote the nonlinear refinement transformation in the ℓ-th layer, implemented as a 1D convolutional layer equipped with batch normalization and GELU activation. Starting from $\widetilde{\mathbf{P}}$, the representation is progressively refined through residual updates. Specifically, we set $\widetilde{\mathbf{P}}^{(0)} = \widetilde{\mathbf{P}}$ and compute $\widetilde{\mathbf{P}}^{(\ell)} = \widetilde{\mathbf{P}}^{(\ell-1)} + \rho_\ell\left(\widetilde{\mathbf{P}}^{(\ell-1)}\right)$ for $\ell = 1, \ldots, L_r$. Finally, the refined representation is combined with the original input $\mathbf{X}$ through a residual connection, followed by batch normalization, yielding the DACE Net output $\mathbf{X}_{\text{delay}} = \text{BatchNorm}\left(\widetilde{\mathbf{P}}^{(L_r)} + \mathbf{X}\right) \in \mathbb{R}^{C\times T}$.

4.4 Transformer Encoder with Learnable-Query Cross-Channel Attention

After rhythm-specific and delay-aware enhancement, the outputs of Rhythm Mixer and DACE Net are integrated through element-wise addition and projected into a D-dimensional latent space, yielding $\mathbf{Z} = \text{Linear}\left(\mathbf{X}_{\text{rhy}} + \mathbf{X}_{\text{delay}}\right) \in \mathbb{R}^{C\times D}$. The resulting representations are then processed by a Transformer encoder (Vaswani et al., 2017) for cross-channel modeling. Inspired by TQNet (Lin et al., 2025), we introduce learnable queries into the attention mechanism while deriving the keys and values from the sample-specific channel representations. Let $\mathbf{Z}^{(\ell-1)} \in \mathbb{R}^{C\times D}$ denote the input to the ℓ-th encoder layer, with $\mathbf{Z}^{(0)} = \mathbf{Z}$. For each layer, we introduce a learnable query matrix $\mathbf{Q}^{(\ell)} \in \mathbb{R}^{C\times D}$ shared across samples, while the keys and values are obtained from $\mathbf{Z}^{(\ell-1)}$. For an encoder with L_e layers, the attention output and layer-wise update are given by

$$\begin{aligned} \widetilde{\mathbf{Z}}^{(\ell)} &= \text{CrossChannelAttn}\left(\mathbf{Q}^{(\ell)}, \mathbf{Z}^{(\ell-1)}, \mathbf{Z}^{(\ell-1)}\right) \in \mathbb{R}^{C\times D}, \\ \mathbf{Z}^{(\ell)} &= \text{EncoderLayerUpdate}\left(\mathbf{Z}^{(\ell-1)}, \widetilde{\mathbf{Z}}^{(\ell)}\right) \in \mathbb{R}^{C\times D}, \qquad \ell = 1, \ldots, L_e. \end{aligned} \tag{2}$$

4.5 Classification Head

The output of the Transformer encoder, $\mathbf{Z}^{(L_e)} \in \mathbb{R}^{C\times D}$, contains the final cross-channel representations learned for classification. We first flatten these channel representations into a single feature vector to integrate information across all channels, and then map it to N-dimensional classification logits through a linear classification head. Specifically, $\widehat{\mathbf{y}} = \text{Linear}\left(\text{Flatten}\left(\mathbf{Z}^{(L_e)}\right)\right) \in \mathbb{R}^N$, where N denotes the number of classes. The predicted class is determined from the resulting logits.

5 Experiments

5.1 Experimental Setup

Baselines. We compare Meredith with 10 representative baselines, including TeCh (Yu et al., 2026), TS-Fingerprint (Li et al., 2026), MedGNN (Fan et al., 2025), Medformer (Wang et al., 2024b), iTransformer (Liu et al., 2024), Nonformer (Liu et al., 2022), PatchTST (Nie et al., 2022), Autoformer (Wu et al., 2021), FEDformer (Zhou et al., 2022), and Informer (Zhou et al., 2021).

Datasets. We evaluate Meredith on five widely used medical time series classification benchmarks, comprising three EEG datasets, APAVA (Escudero et al., 2006), TDBRAIN (van Dijk et al., 2022), and ADFD (Miltiadous et al., 2023), and two ECG datasets, PTB (PhysioBank, 2000) and PTB-XL (Wagner et al., 2020). To further examine its generalizability beyond medical time series, we additionally evaluate the model on two human activity recognition (HAR) datasets, FLAAP (Kumar & Suresh, 2022) and UCI-HAR (Anguita et al., 2013). Detailed descriptions of the baselines, datasets, and implementation details are provided in Appendices A.1, A.2, and A.3, respectively.

Table 1: Classification performance (%, mean±std) on medical time series classification benchmarks under the subject-independent Fixed Subject Split protocol (Wang et al., 2024b). The best and second-best results are highlighted in **red** and *blue*, respectively.

Datasets	Models	Accuracy	Precision	Recall	F1 score	AUROC	AUPRC	Avg
ADFTD (3-Classes)	Autoformer (2021)	45.25±1.48	43.67±1.94	42.96±2.03	42.59±1.85	61.02±1.82	43.10±2.30	46.43±1.90
	FEDformer (2022)	46.30±0.59	46.05±0.76	44.22±1.38	43.91±1.37	62.62±1.75	46.11±1.44	48.20±1.22
	Informer (2021)	48.45±1.96	46.54±1.68	46.06±1.84	45.74±1.38	65.87±1.27	47.60±1.30	50.04±1.57
	iTransformer (2024)	52.60±1.59	46.79±1.27	47.28±1.29	46.79±1.13	67.26±1.16	49.53±1.21	51.71±1.28
	Nonformer (2022)	49.95±1.05	47.71±0.97	47.46±1.50	46.96±1.35	66.23±1.37	47.33±1.78	50.94±1.34
	PatchTST (2022)	44.37±0.95	42.40±1.13	42.06±1.48	41.97±1.37	60.08±1.50	42.49±1.79	45.56±1.37
	Medformer (2024b)	53.27±1.54	51.02±1.57	50.71±1.55	50.65±1.51	70.93±1.19	51.21±1.32	54.63±1.45
	MedGNN (2025)	**56.12**±0.11	**55.07**±0.09	**55.47**±0.34	**55.00**±0.24	**74.68**±0.33	**56.67**±0.38	**58.84**±0.25
	TeCh (2026)	54.54±0.70	53.02±0.87	49.25±1.01	48.84±1.72	68.67±1.05	50.62±1.26	54.16±1.10
	TS-Fingerprint (2026)	53.91±0.54	51.89±0.49	52.10±1.09	51.79±0.74	72.11±0.73	-	-
	Meredith (ours)	*54.88*±0.38	*54.19*±0.42	*54.65*±0.39	*54.05*±0.41	*72.55*±0.45	*55.17*±0.39	*57.58*±0.41
APAVA (2-Classes)	Autoformer (2021)	68.64±1.82	68.48±2.10	68.77±2.27	68.06±1.94	75.94±3.61	74.38±4.05	70.71±2.63
	FEDformer (2022)	74.94±2.15	74.59±1.50	73.56±3.55	73.51±3.39	83.72±1.97	82.94±2.37	77.21±2.49
	Informer (2021)	73.11±4.40	75.17±6.06	69.17±4.56	69.47±5.06	70.46±4.91	70.75±5.27	71.36±5.04
	iTransformer (2024)	74.55±1.66	74.77±2.10	71.76±1.72	72.30±1.79	85.59±1.55	84.39±1.57	77.23±1.73
	Nonformer (2022)	71.89±3.81	71.80±4.58	69.44±3.56	69.74±3.84	70.55±2.96	70.78±4.08	70.70±3.81
	PatchTST (2022)	67.03±1.65	78.76±1.28	59.91±2.02	55.97±3.10	65.65±0.28	67.99±0.76	65.89±1.52
	Medformer (2024b)	78.74±0.64	81.11±0.84	75.40±0.66	76.31±0.71	83.20±0.91	83.66±0.92	79.74±0.78
	MedGNN (2025)	82.60±0.35	*87.70*±0.22	78.93±0.09	80.25±0.16	85.93±0.26	87.35±0.22	83.79±0.22
	TeCh (2026)	*86.86*±1.09	86.85±1.29	*86.10*±1.00	*86.30*±1.06	*94.02*±0.52	*93.79*±0.56	*88.99*±0.92
	TS-Fingerprint (2026)	80.29±0.32	81.62±0.44	77.61±0.49	78.50±0.31	86.49±1.00	-	-
	Meredith (ours)	**91.70**±0.68	**93.12**±0.78	**90.23**±0.41	**91.12**±0.44	**98.02**±0.32	**98.08**±0.40	**95.88**±0.38
TDBrain (2-Classes)	Autoformer (2021)	87.33±3.79	88.06±3.56	87.33±3.79	87.26±3.84	93.81±2.26	93.32±2.42	89.52±3.28
	FEDformer (2022)	78.13±1.98	78.52±1.91	78.13±1.98	78.04±2.01	86.56±1.86	86.48±1.99	80.98±1.96
	Informer (2021)	89.02±2.50	89.43±2.14	89.02±2.50	88.98±2.54	96.64±0.68	96.75±0.63	91.64±1.83
	iTransformer (2024)	74.67±1.06	74.71±1.06	74.67±1.06	74.65±1.06	83.37±1.14	83.73±1.27	77.63±1.11
	Nonformer (2022)	87.88±2.48	88.86±1.84	87.88±2.48	87.78±2.56	97.05±0.68	96.99±0.68	91.07±1.79
	PatchTST (2022)	79.25±3.79	79.60±4.09	79.25±3.79	79.20±3.77	87.95±4.96	86.36±6.67	81.94±4.51
	Medformer (2024b)	89.62±0.81	89.68±0.78	89.62±0.81	89.62±0.81	96.41±0.35	96.51±0.33	91.91±0.65
	MedGNN (2025)	91.04±0.09	91.15±0.12	91.04±0.20	91.04±0.08	96.74±0.04	96.38±0.09	92.90±0.10
	TeCh (2026)	*93.21*±0.61	93.39±0.58	*93.21*±0.61	*93.20*±0.61	*98.68*±0.19	*98.72*±0.17	*95.07*±0.46
	TS-Fingerprint (2026)	-	-	-	-	-	-	-
	Meredith (ours)	**94.98**±1.21	**95.10**±1.18	**94.98**±1.18	**94.97**±1.22	**99.10**±0.55	**99.11**±0.54	**96.37**±0.98
PTB (2-Classes)	Autoformer (2021)	73.35±2.10	72.11±2.89	63.24±3.17	63.69±3.84	78.54±3.48	74.25±3.53	70.86±3.17
	FEDformer (2022)	76.05±2.54	77.58±3.61	66.10±3.55	67.14±4.37	85.93±4.31	82.59±5.42	75.90±3.97
	Informer (2021)	78.69±1.68	82.87±1.02	69.19±2.90	70.84±3.47	92.09±0.53	90.02±0.60	80.62±1.70
	iTransformer (2024)	83.89±0.71	88.25±1.18	76.39±1.01	79.06±1.06	91.18±1.16	90.93±0.98	84.95±1.02
	Nonformer (2022)	78.66±0.49	82.77±0.86	69.12±0.87	70.90±1.00	89.37±2.51	86.67±2.38	79.58±1.35
	PatchTST (2022)	74.74±1.62	76.94±1.51	63.89±2.71	64.36±3.38	88.79±0.91	83.39±0.96	75.35±1.85
	Medformer (2024b)	83.50±2.01	85.19±0.94	77.11±3.39	79.18±3.31	92.81±1.48	90.32±1.54	84.69±2.11
	MedGNN (2025)	84.53±0.28	87.35±0.45	77.90±0.66	80.40±0.62	93.31±0.46	92.88±0.39	86.06±0.48
	TeCh (2026)	*85.96*±2.52	*89.92*±0.74	79.43±4.13	81.97±4.07	*94.57*±0.70	*94.36*±0.66	*87.70*±2.14
	TS-Fingerprint (2026)	85.39±0.63	86.01±0.72	*80.87*±0.91	*82.41*±1.11	92.99±0.98	-	-
	Meredith (ours)	**87.50**±0.43	**90.64**±0.39	**81.76**±0.46	**84.39**±0.68	**96.72**±0.33	**96.40**±0.35	**89.57**±0.44
PTB-XL (5-Classes)	Autoformer (2021)	61.68±2.72	51.60±1.64	49.10±1.52	48.85±2.27	82.04±1.44	51.93±1.71	57.53±1.88
	FEDformer (2022)	57.20±9.47	52.38±6.09	49.04±7.26	47.89±8.44	82.13±4.17	52.31±7.03	56.83±7.08
	Informer (2021)	71.43±0.32	62.64±0.60	59.12±0.47	60.44±0.43	88.65±0.09	64.76±0.17	67.84±0.35
	iTransformer (2024)	69.28±0.22	59.59±0.45	54.62±0.18	56.20±0.19	86.71±0.10	60.27±0.21	64.45±0.23
	Nonformer (2022)	70.56±0.55	61.57±0.66	57.75±0.72	59.10±0.66	88.32±0.36	63.40±0.79	66.78±0.62
	PatchTST (2022)	73.23±0.25	65.70±0.64	60.82±0.76	62.61±0.34	89.74±0.19	*67.32*±0.22	69.90±0.40
	Medformer (2024b)	72.87±0.23	64.14±0.42	60.60±0.46	62.02±0.37	89.66±0.13	66.39±0.22	69.28±0.31
	MedGNN (2025)	*73.87*±0.18	66.26±0.29	61.13±0.23	62.54±0.20	90.21±0.15	67.21±0.23	*70.20*±0.21
	TeCh (2026)	73.53±0.07	65.92±0.52	60.61±0.59	62.44±0.27	90.03±0.12	67.19±0.25	69.95±0.30
	TS-Fingerprint (2026)	**73.98**±0.21	*66.45*±0.38	*62.21*±0.21	*63.51*±0.34	*90.41*±0.11	-	-
	Meredith (ours)	73.24±0.12	**66.48**±0.23	**62.55**±0.17	**64.17**±0.24	**90.71**±0.10	**68.88**±0.19	**71.01**±0.18

5.2 Main Results

Table 1 reports the classification performance under the subject-independent Fixed Subject Split protocol (Wang et al., 2024b; Fan et al., 2025), while the results under the Monte Carlo Cross-Validation protocol (Huang et al., 2026) are presented in Table 7. Under the Fixed Subject Split protocol, Meredith achieves the best average performance on four of the five medical benchmarks and ranks second on ADFTD, demonstrating consistently strong performance across both EEG and ECG datasets. We attribute this performance to the complementary physiological priors introduced by the two core components. Rhythm Mixer explicitly captures diagnostically informative rhythm characteristics, while DACE Net incorporates temporally delayed information across channels, enabling Meredith to jointly exploit frequency-structured temporal patterns and delay-dependent channel information that are characteristic of physiological signals. The consistent advantage across datasets with different physiological signal types and class configurations indicates that explicitly modeling domain-relevant temporal and inter-channel characteristics can offer a stronger inductive bias for medical time series classification than relying solely on generic representation learning.

5.3 Model Analysis

Full Ablation Studies. As shown in Table 2, removing either Rhythm Mixer *(w/o Rhythm Mixer)* or DACE Net *(w/o DACE Net)* consistently degrades the overall classification performance, confirming that both physiological rhythm modeling and delay-aware channel enhancement contribute substantially to Meredith. The degradation caused by removing Rhythm Mixer is particularly pronounced on APAVA, TDBrain, and PTB datasets, indicating that explicitly characterizing physiologically meaningful rhythmic patterns provides important discriminative information for both EEG and ECG signals. Removing DACE Net also leads to clear performance drops across all datasets, demonstrating that temporally delayed information across channels offers complementary cues that cannot be fully recovered by temporal modeling alone.

Table 2: Full ablation results (%) of Meredith on five benchmarks under the Fixed Subject Split protocol. "w/o Both" denotes simultaneously removing Rhythm Mixer and DACE Net, while the other variants remove the corresponding component individually.

Datasets	Model Designs	Accuracy	Precision	Recall	F1 score	AUROC	AUPRC
ADFTD	**original**	**54.88**	**54.19**	**54.65**	**54.05**	**72.55**	**55.17**
	w/o Rhythm Mixer	47.07	46.77	42.62	43.07	63.92	46.03
	w/o DACE Net	45.40	47.44	47.52	45.49	67.05	49.31
	w/o Both	39.39	44.16	33.51	29.43	50.29	33.47
	w/o Learnable Query	47.95	48.04	49.27	47.57	67.71	52.18
APAVA	**original**	**91.70**	**93.12**	**90.23**	**91.12**	**98.02**	**98.08**
	w/o Rhythm Mixer	73.95	76.27	70.06	70.55	78.41	78.79
	w/o DACE Net	87.83	89.96	85.74	86.68	95.58	95.71
	w/o Both	71.38	72.46	67.41	67.58	81.77	80.97
	w/o Learnable Query	90.92	92.82	89.17	90.18	97.52	97.61
TDBrain	**original**	**94.98**	**95.10**	**94.98**	**94.97**	**99.10**	**99.11**
	w/o Rhythm Mixer	81.98	82.22	81.98	81.95	90.83	91.19
	w/o DACE Net	91.42	91.52	91.43	91.56	94.34	94.44
	w/o Both	68.02	70.21	68.02	66.77	80.32	79.54
	w/o Learnable Query	94.67	94.86	94.67	94.66	99.15	99.15
PTB	**original**	**87.50**	**90.64**	**81.76**	**84.39**	**96.72**	**96.40**
	w/o Rhythm Mixer	75.06	78.28	64.03	64.64	90.31	87.80
	w/o DACE Net	78.28	85.23	67.76	69.32	94.59	93.47
	w/o Both	81.62	86.35	73.11	75.53	93.66	92.60
	w/o Learnable Query	82.75	88.86	74.22	76.90	92.80	92.57
PTB-XL	**original**	**73.24**	**66.48**	**62.55**	**64.17**	**90.71**	**68.88**
	w/o Rhythm Mixer	70.74	61.74	57.83	59.42	87.94	62.43
	w/o DACE Net	71.24	64.23	59.70	61.22	88.54	66.29
	w/o Both	45.46	9.09	20.00	12.50	50.19	20.01
	w/o Learnable Query	59.17	41.04	39.88	39.67	76.33	41.09

The effect becomes more evident when Rhythm Mixer and DACE Net are removed simultaneously *(w/o Both)*. In most cases, the resulting variant suffers substantially larger degradation than removing either component individually, with an especially severe collapse on PTB-XL dataset. This suggests that the two modules are not merely redundant sources of improvement: Rhythm Mixer emphasizes informative temporal structures within physiological signals, whereas DACE Net enriches each channel with delay-aware information from other channels, allowing Meredith to exploit complementary physiological characteristics from temporal and channel perspectives. The learnable-query design provides a further refinement to cross-channel modeling. Removing it generally reduces performance, with particularly noticeable drops on ADFTD, PTB, and PTB-XL datasets, although its effect is relatively modest on APAVA and TDBrain datasets. These results indicate that learnable queries act primarily as an auxiliary mechanism for stabilizing and enriching cross-channel representations, while the principal performance gains stem from the physiology-informed Rhythm Mixer and DACE Net.

Effect of Rhythm Mask $\mathcal{M}$ Shape. Table 3 compares two functional forms of the learnable Rhythm Mask, while keeping the center frequency μ_k and bandwidth σ_k identical for each rhythm. Meredith adopts the Gaussian-shaped mask. Although a soft-rectangular form may appear more intuitive because it more closely resembles a conventional band-pass region with relatively uniform emphasis inside the selected frequency range, the Gaussian-shaped mask consistently achieves better performance across all three datasets and evaluation metrics. These results suggest that the shape of the frequency weighting is itself important, rather than merely the location and width of the selected band. Compared with the flatter soft-rectangular profile, the Gaussian form assigns progressively larger weights to frequencies closer to the learned center and smoothly sup-

Table 3: Effect of Rhythm Mask $\mathcal{M}$ shape in Rhythm Mixer on classification performance (%). Gaussian-shaped and soft-rectangular masks use identical learnable center frequencies μ_k and bandwidths σ_k for the k-th rhythm, with only the mask shape varied for comparison.

Datasets	Learnable Rhythm Mask	Accuracy	Precision	Recall	F1 score	AUROC
APAVA	**Gaussian-shape**	**91.70**	**93.12**	**90.23**	**91.12**	**98.02**
	Soft-Rectangular	89.57	91.48	87.69	88.76	97.18
TDBrain	**Gaussian-shape**	**94.98**	**95.10**	**94.98**	**94.97**	**99.10**
	Soft-Rectangular	88.75	89.41	88.75	88.69	95.90
PTB	**Gaussian-shape**	**87.50**	**90.64**	**81.76**	**84.39**	**96.72**
	Soft-Rectangular	83.62	88.53	75.81	78.52	95.53

presses more distant components, thereby providing a graded notion of rhythm relevance instead of treating frequencies within the band more uniformly. Such smooth weighting may better reflect the fact that physiological rhythms are not confined to sharply bounded frequency intervals, while also reducing sensitivity to small shifts around the band boundaries. The superior results therefore indicate that a soft, center-aware spectral emphasis is more suitable for Rhythm Mixer than an approximately rectangular selection.

Necessity of the Transformer Encoder. Although DACE Net incorporates information from other channels through delay-aware aggregation, its primary role is channel enhancement rather than explicit cross-channel modeling. We therefore examine whether the subsequent Transformer encoder remains necessary by removing it while keeping Rhythm Mixer and DACE Net unchanged. As shown in Table 4, this removal consistently degrades performance on both ADFTD and PTB datasets across all metrics. These results indicate that DACE Net and the Transformer encoder serve complementary roles: the former enriches channel representations with delay-aware information, while the latter further captures dependencies among the enhanced channels.

Table 4: Assessing the necessity of the Transformer encoder for classification performance (%).

Datasets	Model Designs	Accuracy	Precision	Recall	F1 score	AUROC
ADFTD	**w Transformer Encoder**	**54.88**	**54.19**	**54.65**	**54.05**	**72.55**
	w/o Transformer Encoder	44.50	43.50	43.43	43.11	61.95
PTB	**w Transformer Encoder**	**87.50**	**90.64**	**81.76**	**84.39**	**96.72**
	w/o Transformer Encoder	78.41	81.00	69.32	71.07	92.71

5.4 Efficiency Analysis

Beyond classification performance, computational efficiency is also important for medical time series models, particularly for practical deployment under constrained computational resources. We therefore evaluate Meredith in terms of parameter count, MACs, training time, and memory usage. The efficiency evaluation is conducted on a single NVIDIA RTX 4090 GPU with 24 GB of memory. To ensure a fair comparison, the batch size is fixed to 32 for all models on the APAVA dataset (16 channels). As shown in Table 5, Meredith achieves the highest Accuracy and AUROC while using the fewest parameters among all compared methods, together with moderate computational and memory costs. In particular, its 556.01K parameters and 28.66M MACs are substantially lower than those of most medical-specific baselines (e.g. Medformer, MedGNN, TeCh), indicating that the performance gains are achieved without relying on excessive model capacity or computational complexity. Additional efficiency evaluations on the ADFTD dataset (19 channels) and the PTB dataset (15 channels) are reported in Tables 9 and 10, respectively, further demonstrating the efficiency of Meredith across different medical time series scenarios.

Table 5: Efficiency comparison between Meredith and baselines on the APAVA dataset under the subject-independent Fixed Subject Split protocol. *Time* denotes the average *training time* required per epoch.

Model	Parameters	MACs	Time	Memory Usage	Accuracy (%)	AUROC (%)
Autoformer	861.95 K	204.64 M	7.6 s	2642 MB	68.64	75.94
FEDformer	1.46 M	204.64 M	43.7 s	2380 MB	74.94	83.72
Informer	1.07 M	208.31 M	6.3 s	2234 MB	73.11	70.46
iTransformer	832.13 K	13.23 M	6.4 s	1888 MB	74.55	85.59
Nonformer	943.23 K	204.87 M	5.3 s	3536 MB	71.89	70.55
PatchTST	928.26 K	407.44 M	5.1 s	3012 MB	67.03	65.65
Medformer	7.24 M	565.68 M	29.5 s	4008 MB	78.74	83.20
MedGNN	11.48 M	720.50 M	14.1 s	3088 MB	82.60	85.93
TeCh	6.0 M	810.39 M	4.8 s	2350 MB	86.86	94.02
Meredith (ours)	556.01 K	28.66 M	6.4 s	2018 MB	91.70	98.02

6 Conclusion and Limitation

Meredith is built on a simple principle: medical time series should not be treated merely as generic multivariate sequences, but as physiological signals with characteristic rhythms and delayed inter-channel responses. Guided by this view, Rhythm Mixer and DACE Net introduce complementary inductive biases into temporal representation learning and channel enhancement, while the subsequent Transformer encoder further models cross-channel dependencies. The consistent improvements across EEG and ECG benchmarks, together with the ablation and efficiency studies, support the value of incorporating physiologically meaningful structure directly into model design rather than relying solely on generic representation learning. Despite these results, several limitations remain. The physiological priors used by Meredith are still simplified: Rhythm Mixer relies on predefined rhythm ranges for initialization, while DACE Net models temporal delays within a finite discrete range, which may not fully capture more complex or time-varying physiological dynamics. In addition, the current design focuses on regularly sampled multichannel signals, and extending these ideas to irregular sampling, missing observations, or richer clinical settings remains an important direction for future work.

AI USE STATEMENT

In this work, generative AI tools were used solely to assist with grammar checking, spell checking, and language refinement. They were not used for substantive research activities, such as formulating research ideas, designing methodologies or experiments, implementing the proposed methods, conducting theoretical analyses, interpreting experimental results, or deriving research findings and conclusions. All AI-assisted revisions were carefully reviewed and approved by the authors, who take full responsibility for the final content of this work.

ETHICS STATEMENT

This work uses only publicly available datasets that have been widely adopted in prior research. We did not collect any new data or conduct any new studies involving human participants. All datasets were used in accordance with their publicly available terms and for research purposes only.

REPRODUCIBILITY STATEMENT

To support reproducibility, we provide comprehensive descriptions of the proposed methods in the main text, with implementation details, experimental settings, and dataset descriptions documented in the appendix. The source code is included in the supplementary materials and will be made publicly available upon publication.

REFERENCES

Thivya Anbalagan, Malaya Kumar Nath, Dhurairajan Vijayalakshmi, and Archana Anbalagan. Analysis of various techniques for ecg signal in healthcare, past, present, and future. *Biomedical Engineering Advances*, 6:100089, 2023.

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, pp. 3–4, 2013.

Yara Badr, Usman Tariq, Fares Al-Shargie, Fabio Babiloni, Fadwa Al Mughairbi, and Hasan Al-Nashash. A review on evaluating mental stress by deep learning using eeg signals. *Neural Computing and Applications*, 36(21):12629–12654, 2024.

Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33:17804–17815, 2020.

Giovanni Chiarion, Laura Sparacino, Yuri Antonacci, Luca Faes, and Luca Mesin. Connectivity analysis in eeg data: a tutorial review of the state of the art and emerging trends. *Bioengineering*, 10(3):372, 2023.

J Escudero, Daniel Abásolo, Roberto Hornero, Pedro Espino, and Miguel López. Analysis of electroencephalograms in alzheimer's disease patients with multiscale entropy. *Physiological measurement*, 27(11):1091, 2006.

Wei Fan, Jingru Fei, Dingyu Guo, Kun Yi, Xiaozhuang Song, Haolong Xiang, Hangting Ye, and Min Li. Towards multi-resolution spatiotemporal graph learning for medical time series classification. In *Proceedings of the ACM on Web Conference 2025*, pp. 5054–5064, 2025.

Zhenyin Fu, Ruiqing Dong, Huanyong Zheng, Zefeng Wang, Boyang Cao, Jinghui Bai, Mingxia Ma, Zhanchun Song, Fuzhi Pan, Ling Xia, et al. Progress of conductivity and conduction velocity measured in human and animal hearts. *Reviews in Cardiovascular Medicine*, 25(10):364, 2024.

Nan Huang, Haishuai Wang, Zihuai He, Marinka Zitnik, and Xiang Zhang. Repurposing foundation model for generalizable medical time series classification. In *International Conference on Learning Representations*, volume 2026, pp. 83809–83831, 2026.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. pmlr, 2015.

Md Rabiul Islam, Xuyang Zhao, Yao Miao, Hidenori Sugano, and Toshihisa Tanaka. Epileptic seizure focus detection from interictal electroencephalogram: a survey. *Cognitive neurodynamics*, 17(1):1–23, 2023.

José A Joglar, Mina K Chung, Anastasia L Armbruster, Emelia J Benjamin, Janice Y Chyou, Edmond M Cronin, Anita Deswal, Lee L Eckhardt, Zachary D Goldberger, Rakesh Gopinathannair, et al. 2023 acc/aha/accp/hrs guideline for the diagnosis and management of atrial fibrillation: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. *Circulation*, 149(1):e1–e156, 2024.

Alan Jovic, Nikolina Frid, Karla Brkic, and Mario Cifrek. Interpretability and accuracy of machine learning algorithms for biomedical time series analysis–a scoping review. *Biomedical Signal Processing and Control*, 110:108153, 2025.

Mohammadreza Kariman, Bettina Pretterklieber, Michael L Pretterklieber, Niels Hammer, Jordi Heijman, Gernot Plank, and Christoph M Augustin. The cardiac conduction system: A narrative review. *Annals of Anatomy-Anatomischer Anzeiger*, pp. 152880, 2026.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

André G Kléber and Yoram Rudy. Basic mechanisms of cardiac impulse propagation and associated arrhythmias. *Physiological reviews*, 84(2):431–488, 2004.

Prabhat Kumar and S Suresh. Flaap: An open human activity recognition (har) dataset for learning and finding the associated activity patterns. *Procedia Computer Science*, 212:64–73, 2022.

Huayu Li, ZhengXiao He, Xiwen Chen, Jingjing Wang, Siyuan Tian, Jinghao Wen, and Ao Li. Learning fingerprints for medical time series with redundancy-constrained information maximization. *arXiv preprint arXiv:2605.00130*, 2026.

Shengsheng Lin, Haojun Chen, Haijie Wu, Chunyun Qiu, and Weiwei Lin. Temporal query network for efficient multivariate time series forecasting. *arXiv preprint arXiv:2505.12917*, 2025.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in neural information processing systems*, 35: 9881–9893, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 11116–11140, 2024.

Jingying Ma, Feng Wu, Qika Lin, Yucheng Xing, Chenyu Liu, Ziyu Jia, and Mengling Feng. Codebrain: Bridging decoupled tokenizer and multi-scale architecture for eeg foundation model. In *International Conference on Learning Representations*, volume 2026, pp. 83526–83574, 2026.

Andreas Miltiadous, Katerina D Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G Tsalikakis, Pantelis Angelidis, Markos G Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, et al. A dataset of scalp eeg recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine eeg. *Data*, 8(6):95, 2023.

Lyle Muller, Frédéric Chavane, John Reynolds, and Terrence J Sejnowski. Cortical travelling waves: mechanisms and computational principles. *Nature reviews neuroscience*, 19(5):255–268, 2018.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

PhysioToolkit PhysioBank. Physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

Wasi Ur Rehman Qamar and Berdakh Abibullaev. Multi-scale eeg feature decoding with swin transformers for subject independent motor imagery bcis. *Scientific Reports*, 16(1):2503, 2026.

Ian Naismith Sneddon. *Fourier transforms*. Courier Corporation, 1995.

H V Sorensen, D Jones, Michael Heideman, and C Burrus. Real-valued fast fourier transform algorithms. *IEEE Transactions on acoustics, speech, and signal processing*, 35(6):849–863, 1987.

Hanneke van Dijk, Guido van Wingen, Damiaan Denys, Sebastian Olbrich, Rosalinde van Ruth, and Martijn Arns. The two decades brainclinics research archive for insights in neurophysiology (tdbrain) database. *Scientific data*, 9(1):333, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, and Tobias Schaeffter. Ptb-xl, a large publicly available electrocardiography dataset. *Scientific data*, 7(1):1–15, 2020.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. In *International conference on learning representations*, volume 2025, pp. 75310–75346, 2025.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 38626–38652, 2024a.

Yihe Wang, Nan Huang, Taida Li, Yujun Yan, and Xiang Zhang. Medformer: A multi-granularity patching transformer for medical time-series classification. *Advances in Neural Information Processing Systems*, 37:36314–36341, 2024b.

Ziwei Wang and Paolo Mengoni. Seizure classification with selected frequency bands and eeg montages: a natural language processing approach. *Brain Informatics*, 9(1):11, 2022.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 753–763, 2020.

Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective. *Advances in neural information processing systems*, 36:69638–69660, 2023.

Guoqi Yu, Juncheng Wang, Chen Yang, Jing Qin, Angelica I Aviles-Rivero, and Shujun Wang. Decentralized attention fails centralized signals: Rethinking transformers for medical time series. *arXiv preprint arXiv:2602.18473*, 2026.

Paul Theo Zebhauser, Henrik Heitmann, Peter Henningsen, and Markus Ploner. Rethinking eeg biomarkers of brain disorders: a transdiagnostic dimensional view. *Translational Psychiatry*, 16 (1):316, 2026.

Da Zhang, Bingyu Li, Zhiyuan Zhao, Hongyuan Zhang, Junyu Gao, and Xuelong Li. Medmamba: Multi-view state space models with adaptive graph learning for medical time series classification. *arXiv preprint arXiv:2605.24961*, 2026.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pp. 27268–27286. PMLR, 2022.

Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, Rong Jin, et al. Card: Channel aligned robust blend transformer for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 11245–11283, 2024.

APPENDIX CONTENTS

A Experimental Details **15**

A.1 Baselines 15

A.2 Datasets 16

A.2.1 EEG datasets 16

A.2.2 ECG datasets 16

A.2.3 Human Activity Recognition (HAR) Datasets 17

A.3 Implementation Details 17

B Additional Results **17**

B.1 Subject-Independent Generalization under Monte Carlo Cross-Validation 17

B.2 Generalization Beyond Medical Time Series 17

B.3 Parameter sensitivity analysis 18

B.3.1 Sensitivity to Sampling Rate 18

B.3.2 Sensitivity to the Number of Channel Refinement Layers L_r 19

B.3.3 Sensitivity to the Maximum Temporal Delay 19

B.3.4 Sensitivity to d_{model} 20

B.3.5 Sensitivity to d_{ff} 20

B.3.6 Sensitivity to the Number of Transformer Encoder Layers L_e 20

B.4 Further efficiency analysis 21

B.5 Visualization Analysis 21

B.5.1 Visualization of Delay-Aware Channel Interactions in DACE Net 21

B.6 Visualization of Learned Channel Interactions in DACE Net 22

A Experimental Details

A.1 Baselines

We compare Meredith with 10 representative baselines, including TeCh (Yu et al., 2026), TS-Fingerprint (Li et al., 2026), MedGNN (Fan et al., 2025), Medformer (Wang et al., 2024b), iTransformer (Liu et al., 2024), Nonformer (Liu et al., 2022), PatchTST (Nie et al., 2022), Autoformer (Wu et al., 2021), FEDformer (Zhou et al., 2022), and Informer (Zhou et al., 2021). A brief description of each baseline is provided below:

TeCh (Yu et al., 2026) employs a centralized core-token aggregation–redistribution mechanism to capture temporal dependencies and channel interactions efficiently in medical time series. The official code is available at `https://github.com/Levi-Ackman/TeCh`. The TeCh results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh paper, respectively.

TS-Fingerprint (Li et al., 2026) learns interpretable and disentangled medical time series representations through a fingerprint-token bottleneck optimized with reconstruction and redundancy-reduction objectives. The TS-Fingerprint results in Table 1 and Table 8 are adopted from Table 1 of the original TS-Fingerprint paper. Since the official implementation is not publicly available, TS-Fingerprint is not evaluated on the TDBrain dataset.

MedGNN (Fan et al., 2025) integrates multi-resolution adaptive graph learning, difference-based temporal attention, and frequency-domain feature extraction to capture multi-scale spatiotemporal dependencies in medical time series. The official code is available at `https://github.com/aikunyi/MedGNN`. The MedGNN results in Table 1 follow those reported in Table 2 of the original MedGNN paper. The results in Table 8 are obtained by running the official code five times and reporting the average results.

Medformer (Wang et al., 2024b) employs cross-channel patching, multi-granularity embeddings, and two-stage multi-granularity attention to capture multi-scale temporal patterns and inter-channel correlations in medical time series. The official code is available at `https://github.com/DL4mHealth/Medformer`. The Medformer results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

iTransformer (Liu et al., 2024) inverts the conventional Transformer architecture by treating individual series as variate tokens, enabling attention-based multivariate correlation modeling and variate-centric representation learning. The official code is available at `https://github.com/thuml/iTransformer`. The iTransformer results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

Nonformer (Liu et al., 2022) (Non-stationary Transformer) combines series stationarization with de-stationary attention to enhance sequence stationarity while preserving intrinsic non-stationary temporal dynamics. The official code is available at `https://github.com/thuml/Nonstationary_Transformers`. The Nonformer results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

PatchTST (Nie et al., 2022) represents time series as subseries-level patches and adopts a channel-independent Transformer architecture to capture local semantics and long-range temporal dependencies efficiently. The official code is available at `https://github.com/yuqinie98/patchtst`. The PatchTST results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

Autoformer (Wu et al., 2021) integrates progressive series decomposition with a series-level Auto-Correlation mechanism to capture long-range temporal dependencies and periodic patterns more effectively. The official code is available at `https://github.com/thuml/Autoformer`. The Autoformer results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

FEDformer (Zhou et al., 2022) integrates seasonal-trend decomposition with frequency-domain attention to capture both global temporal structures and sparse spectral representations under linear complexity. The official code is available at `https://github.com/MAZiqing/`

Table 6: Statistics of the datasets used in our experiments.

Dataset	Subject	Sample	Class	Channel	Timestamps	Modality
ADFTD	88	69,752	3	19	256	EEG
APAVA	23	5,967	2	16	256	EEG
TDBrain	72	6,240	2	33	256	EEG
PTB	198	64,356	2	15	300	ECG
PTB-XL	17,596	191,400	5	12	250	ECG
FLAPP	8	13,123	10	6	100	IMU (Acc+Gyro)
UCI-HAR	30	10,299	6	9	128	IMU (Acc+Gyro)

`FEDformer`. The FEDformer results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

Informer (Zhou et al., 2021) introduces ProbSparse attention, self-attention distilling, and a generative decoder to efficiently model long-range temporal dependencies with sub-quadratic complexity. The official code is available at `https://github.com/zhouhaoyi/Informer2020`. The Informer results in Table 1 and Table 8 follow those reported in Table 2 and Table 3 of the original TeCh (Yu et al., 2026) paper, respectively.

A.2 Datasets

To comprehensively evaluate Meredith, we conduct experiments on five benchmark datasets for medical time series classification. To further assess its generalizability beyond the medical domain, we additionally evaluate the model on two human activity recognition (HAR) datasets, FLAAP (Kumar & Suresh, 2022) and UCI-HAR (Anguita et al., 2013). The statistics of all datasets are summarized in Table 6, with brief descriptions provided below.

A.2.1 EEG datasets

The **ADFTD** (Miltiadous et al., 2023) dataset is a public EEG benchmark comprising three classes: 36 patients with Alzheimer's disease (AD), 23 patients with Frontotemporal Dementia (FD), and 29 healthy controls (HC). EEG signals are recorded from 19 channels at 500 Hz, with each subject contributing a single trial of approximately 12–14 minutes. Following a 0.5–45 Hz bandpass filter, all recordings are downsampled to 256 Hz and segmented into non-overlapping 1-second samples, each containing 256 timestamps, resulting in 69,752 valid samples.

The **APAVA** (Escudero et al., 2006) dataset is a public EEG dataset comprising 23 subjects, including 12 Alzheimer's disease patients and 11 healthy controls. Each subject contributes an average of 30.0 ± 12.5 trials, with each 5-second trial containing 1,280 timestamps across 16 channels. Following standard scaling, each trial is divided into nine 1-second samples with 50% overlap, each containing 256 timestamps, resulting in 5,967 samples in total. Each sample is associated with the corresponding subject ID.

The **TDBRAIN** (van Dijk et al., 2022) dataset is a large permission-accessible EEG dataset containing recordings from 1,274 subjects across 33 channels, with each subject providing one eye-open and one eye-closed trial. Although the complete dataset includes 60 diagnostic labels, we use a subset of 25 Parkinson's disease patients and 25 healthy controls under the eye-closed condition. Each trial is divided into non-overlapping 1-second samples of 256 timestamps, while segments shorter than 1 second are discarded, yielding a total of 6,240 samples.

A.2.2 ECG datasets

The **PTB** (PhysioBank, 2000) dataset is a public ECG dataset comprising recordings from 290 subjects across 15 channels and eight diagnostic labels. In this work, we use a subset of 198 subjects, including myocardial infarction patients and healthy controls. The signals are originally sampled at 1000 Hz and are subsequently downsampled to 250 Hz and standardized. Heartbeat samples are then obtained by detecting R-peaks across channels, removing outliers, aligning each beat with its corresponding R-peak, and applying zero padding to ensure a uniform length. This preprocessing results in 64,356 heartbeat samples.

The **PTB-XL** (Wagner et al., 2020) dataset is a large public ECG dataset containing recordings from 18,869 subjects across 12 channels and five diagnostic categories. To maintain label consistency, we exclude subjects whose diagnoses are inconsistent across trials, leaving 17,596 subjects for evaluation. Each trial is a 10-second recording provided at both 100 Hz and 500 Hz; in this work, we use the 500 Hz version, downsample it to 250 Hz, and apply standard normalization. The recordings are then divided into non-overlapping 1-second samples of 250 timestamps, with segments shorter than 1 second discarded, yielding a total of 191,400 samples.

A.2.3 Human Activity Recognition (HAR) Datasets

FLAAP (Kumar & Suresh, 2022) is a smartphone-based HAR dataset containing accelerometer and gyroscope recordings for activity pattern recognition. **UCI-HAR** (Anguita et al., 2013) consists of accelerometer and gyroscope data collected using waist-mounted smartphones and is widely adopted for evaluating HAR models. Key dataset statistics, including the number of subjects, channels, and timestamps, are summarized in Table 1. Data preprocessing and dataset splitting follow Medformer (Wang et al., 2024b).

A.3 Implementation Details

All experiments are implemented in PyTorch (Paszke et al., 2019) and conducted on a single NVIDIA RTX 4090 GPU with 24 GB of memory. The model is optimized using Adam (Kingma & Ba, 2014) with the cross-entropy loss. The learning rate is selected from $\{1 \times 10^{-4}, 2 \times 10^{-3}, 1 \times 10^{-3}, 3 \times 10^{-4}\}$, and the batch size is chosen from $\{32, 64, 128\}$. Training is performed for either 10 or 50 epochs, with early stopping based on validation performance and the patience selected from $\{3, 5, 10, 50\}$. The hidden dimensions are selected from $d_{\text{model}} \in \{64, 256\}$ and $d_{\text{ff}} \in \{128, 256\}$. For DACE Net, the predefined maximum temporal delay is fixed as $\tau = 8$, while the number of channel refinement layers L_r is selected from $\{3, 4\}$. The number of Transformer encoder layers L_e is chosen from $\{1, 2, 3\}$. For a fair comparison with prior work (Wang et al., 2024b; Fan et al., 2025), we adopt six widely used classification metrics: Accuracy, Precision, Recall, F1 score, Area Under the Receiver Operating Characteristic Curve (AUROC), and Area Under the Precision-Recall Curve (AUPRC).

B Additional Results

B.1 Subject-Independent Generalization under Monte Carlo Cross-Validation

As shown in Table 7, Meredith maintains strong performance across five independently generated subject-level splits, indicating limited sensitivity to variations in subject partitioning. More importantly, the relatively small standard deviations observed across most datasets suggest that its performance remains stable when the identities of the training, validation, and test subjects change. This provides stronger evidence of subject-independent generalization than a single fixed split, as the model is repeatedly evaluated on previously unseen subjects under different partition configurations. The results therefore suggest that Meredith captures physiological patterns that are transferable across subjects, rather than relying heavily on subject-specific characteristics associated with a particular data split.

B.2 Generalization Beyond Medical Time Series

As shown in Table 8, Meredith remains highly competitive on both HAR benchmarks, achieving the best overall performance on UCI-HAR and the second-best average result on FLAAP. This indicates that the effectiveness of Meredith is not restricted to physiological signals, despite its original design motivation being rooted in medical time series. More importantly, the strong transfer to HAR suggests that the structures captured by Meredith are not exclusively medical-specific: characteristic temporal patterns and delayed interactions across sensor channels are also prevalent in human activity signals. Rhythm Mixer can therefore provide a useful inductive bias for separating activity-related temporal patterns, while DACE Net can model asynchronous dependencies among accelerometer and gyroscope channels. The results suggest that these two designs capture general

Table 7: Classification performance (%, mean ± std) on medical time series classification benchmarks using subject-independent Monte Carlo cross-validation (Huang et al., 2026). The best and second-best results are highlighted in **red** and *blue*, respectively.

Datasets	Models	Accuracy	Precision	Recall	F1 score	AUROC	AUPRC	Avg
ADFTD (3-Classes)	Autoformer (2021)	50.09±2.47	**47.21±2.22**	46.55±2.57	**46.48±2.39**	64.12±2.34	46.45±2.84	50.15±2.47
	FEDformer (2022)	48.43±2.57	45.54±2.07	45.79±1.51	44.94±2.13	64.67±1.49	46.14±2.17	49.25±1.99
	Informer (2021)	49.24±1.07	46.97±1.46	46.00±1.45	46.18±1.40	65.12±1.62	46.30±1.69	49.97±1.45
	iTransformer (2024)	47.35±0.67	42.27±0.63	43.23±0.72	42.54±0.60	60.94±0.63	42.25±0.72	46.43±0.66
	Nonformer (2022)	42.78±1.41	40.35±2.45	39.54±1.58	39.51±1.84	58.57±1.00	41.14±1.22	43.65±1.58
	PatchTST (2022)	41.48±0.49	39.37±0.37	39.36±0.36	39.34±0.37	57.41±0.44	38.28±0.33	42.54±0.39
	Medformer (2024b)	41.93±3.06	32.63±1.34	34.37±1.65	29.53±1.38	57.18±1.25	37.75±1.51	38.90±1.70
	MedGNN (2025)	48.70±1.75	45.31±1.77	44.69±1.92	44.71±1.94	64.44±2.07	*48.19±2.43*	49.34±1.98
	TeCh (2026)	*51.12±0.32*	*46.97±0.96*	*46.91±0.94*	*46.19±1.04*	*65.45±0.37*	47.14±0.22	*50.63±0.64*
	Meredith (ours)	**53.20±1.07**	45.76±2.95	**47.61±2.15**	45.46±2.59	**68.88±1.38**	**51.36±1.40**	**52.05±1.92**
APAVA (2-Classes)	Autoformer (2021)	54.88±2.05	65.48±1.21	61.30±1.31	53.54±2.79	63.31±4.34	61.23±3.30	59.96±2.50
	FEDformer (2022)	54.13±1.60	71.84±0.63	62.15±1.30	51.44±2.09	88.66±1.79	86.75±2.00	69.16±1.57
	Informer (2021)	80.12±2.13	81.70±1.13	82.80±1.58	80.04±2.08	95.52±0.49	95.61±0.50	85.97±1.32
	iTransformer (2024)	*95.65±0.94*	*95.00±1.03*	*96.43±0.76*	*95.51±0.96*	**99.99±0.01**	**99.99±0.01**	*97.10±0.62*
	Nonformer (2022)	76.95±4.97	78.46±3.54	79.17±3.85	76.78±4.81	90.73±2.40	90.54±2.59	82.11±3.69
	PatchTST (2022)	86.30±3.63	86.97±2.58	88.59±3.00	86.20±3.57	98.39±1.78	98.17±2.07	90.77±2.77
	Medformer (2024b)	76.68±2.23	81.04±1.15	80.71±1.83	76.66±2.25	96.76±1.39	95.39±2.81	84.54±1.94
	MedGNN (2025)	75.17±8.41	80.56±5.20	79.26±6.77	75.01±8.40	96.21±1.74	96.27±1.66	83.75±5.36
	TeCh (2026)	95.38±0.66	94.74±0.68	96.15±0.56	95.24±0.67	*99.90±0.05*	*99.89±0.05*	96.88±0.45
	Meredith (ours)	**96.12±1.01**	**95.56±1.07**	**96.60±0.84**	**95.98±1.03**	99.72±0.12	99.72±0.13	**97.28±0.70**
TDBrain (2-Classes)	Autoformer (2021)	94.68±1.68	93.86±1.77	94.81±1.96	94.27±1.83	98.80±0.55	98.61±0.54	95.84±1.39
	FEDformer (2022)	89.96±0.67	89.36±0.45	88.74±1.29	88.99±0.86	96.11±0.57	95.77±0.51	91.49±0.73
	Informer (2021)	94.65±1.25	94.17±1.35	94.22±1.41	94.19±1.36	98.21±0.51	98.06±0.63	95.58±1.09
	iTransformer (2024)	83.00±0.69	82.16±0.77	80.17±0.88	80.95±0.82	89.82±0.70	89.18±0.64	84.21±0.75
	Nonformer (2022)	94.52±0.69	93.93±0.79	94.20±0.70	94.06±0.74	98.47±0.34	98.39±0.34	95.60±0.60
	PatchTST (2022)	93.26±0.86	93.76±0.86	91.61±0.86	92.48±0.86	97.91±0.86	97.54±0.86	94.43±0.86
	Medformer (2024b)	95.94±0.33	95.51±0.50	95.70±0.26	95.60±0.35	98.90±0.11	98.84±0.13	96.75±0.28
	MedGNN (2025)	**96.63±0.83**	**96.57±0.77**	*96.10±1.09*	**96.32±0.92**	*99.37±0.23*	*99.33±0.23*	**97.39±0.68**
	TeCh (2026)	96.24±0.56	*95.91±0.73*	95.94±0.59	95.91±0.60	98.86±0.17	98.84±0.19	96.95±0.47
	Meredith (ours)	*96.38±0.92*	95.90±0.75	**96.27±1.33**	*96.07±1.02*	**99.43±0.40**	**99.37±0.41**	*97.24±0.81*
PTB (2-Classes)	Autoformer (2021)	85.85±1.72	68.90±3.85	65.28±2.29	66.18±1.17	72.84±2.54	67.83±2.51	71.15±2.35
	FEDformer (2022)	86.25±1.09	69.58±2.42	70.43±3.94	69.90±3.11	82.57±3.78	75.60±3.45	75.72±2.97
	Informer (2021)	90.33±0.42	79.68±1.61	74.07±1.12	76.37±0.67	87.08±0.97	81.69±0.77	81.54±0.93
	iTransformer (2024)	89.14±0.80	76.04±1.41	**80.53±1.50**	77.89±0.91	*89.41±1.08*	*82.32±1.25*	82.56±1.16
	Nonformer (2022)	90.06±1.05	78.53±2.81	75.87±2.15	76.97±1.88	84.23±2.22	79.74±2.78	80.90±2.15
	PatchTST (2022)	88.71±0.83	75.12±1.89	74.65±2.25	74.75±1.46	87.18±1.49	76.83±1.98	79.54±1.65
	Medformer (2024b)	87.23±0.71	71.52±1.57	70.70±2.40	71.03±1.81	83.57±0.87	75.70±1.07	76.63±1.41
	MedGNN (2025)	89.56±1.08	77.77±3.56	74.14±3.72	75.35±2.42	84.74±2.38	79.26±2.52	80.14±2.61
	TeCh (2026)	*91.11±0.53*	*80.74±2.02*	*79.30±1.74*	**79.84±0.67**	**90.96±0.79**	82.20±1.83	**84.03±1.26**
	Meredith (ours)	**91.40±0.78**	**82.14±2.02**	77.04±2.10	*79.25±2.05*	89.21±1.53	**82.79±2.09**	*83.64±1.76*
PTB-XL (5-Classes)	Autoformer (2021)	58.60±1.93	47.97±1.29	47.97±1.38	47.34±1.53	79.98±0.83	48.56±1.46	55.07±1.40
	FEDformer (2022)	53.44±12.65	51.70±4.26	47.09±7.48	45.46±9.38	81.21±3.82	51.33±6.12	55.04±7.29
	Informer (2021)	69.04±0.35	58.78±0.42	57.33±0.72	57.90±0.58	86.47±0.32	60.93±0.60	65.24±0.50
	iTransformer (2024)	66.04±0.25	55.33±0.24	53.46±0.37	54.22±0.32	84.03±0.14	55.95±0.18	61.51±0.25
	Nonformer (2022)	67.39±0.16	56.30±0.45	54.52±0.22	55.19±0.32	85.18±0.23	57.71±0.53	62.72±0.32
	PatchTST (2022)	70.46±0.25	60.68±0.34	58.50±0.61	59.40±0.47	87.03±0.14	62.16±0.41	66.37±0.37
	Medformer (2024b)	73.03±0.13	64.35±0.17	61.16±0.12	62.47±0.12	89.75±0.08	66.73±0.18	69.58±0.13
	MedGNN (2025)	72.34±0.23	62.64±0.31	59.48±0.34	60.57±0.30	89.00±0.32	65.24±0.49	68.21±0.33
	TeCh (2026)	*73.75±0.20*	*65.68±0.47*	**61.47±0.58**	**62.92±0.38**	*90.32±0.09*	*68.08±0.39*	*70.37±0.35*
	Meredith (ours)	**74.09±0.17**	**66.03±0.28**	*61.18±0.22*	*62.73±0.28*	**90.98±0.13**	**68.37±0.38**	**70.58±0.24**

structural properties of multichannel time series, which explains why Meredith can retain strong performance even when transferred beyond the medical domain.

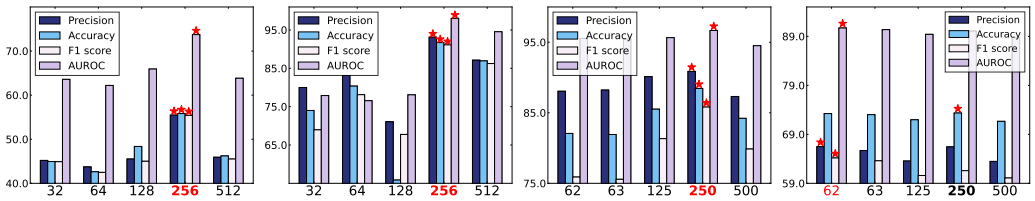
B.3 Parameter sensitivity analysis

B.3.1 Sensitivity to Sampling Rate

Since the sampling rate determines how physical frequencies are mapped into the discrete spectrum, it directly specifies the lower and upper frequency bounds, r_k^b and r_k^e, of each band in the Rhythm Bank $\mathcal{R} = \{[r_k^b, r_k^e]\}_{k=1}^K$. We therefore examine the sensitivity of Meredith to this setting. As shown in Figure 2, Meredith consistently achieves its best performance when the sampling rate used to construct the Rhythm Bank matches the actual sampling rate of the preprocessed signals. This finding provides indirect evidence for the physiological motivation of Rhythm Mixer: the predefined rhythm bands are meaningful only when their frequency boundaries are aligned with the true physi-

Table 8: Classification performance (%, mean ± std) on human activity recognition benchmarks for evaluating generalization beyond medical time series. The best and second-best results are highlighted in **red** and *blue*, respectively.

Datasets	Models	Accuracy	Precision	Recall	F1 score	AUROC	AUPRC	Avg
FLAAP (10-Classes)	Autoformer (2021)	38.93±1.01	38.22±1.31	37.40±1.17	33.51±1.14	74.12±0.35	35.77±0.91	42.99±0.98
	FEDformer (2022)	59.51±9.03	59.84±8.10	58.57±8.97	57.73±9.99	89.75±5.37	60.88±9.63	64.38±8.52
	Informer (2021)	72.87±0.89	73.20±0.97	72.76±0.92	72.59±0.96	95.91±0.24	77.57±1.21	77.48±0.87
	iTransformer (2024)	75.15±0.48	75.09±0.53	75.14±0.47	74.91±0.51	96.64±0.14	80.81±0.60	79.62±0.46
	PatchTST (2022)	56.34±0.31	56.36±0.63	55.29±0.32	55.58±0.45	89.24±0.11	58.92±0.36	61.96±0.36
	Medformer (2024b)	76.44±0.64	76.61±1.13	76.63±1.36	76.25±0.65	95.44±0.26	81.12±1.60	80.42±0.94
	MedGNN (2025)	75.32±0.42	75.68±0.46	75.02±0.60	74.73±0.57	95.24±0.46	79.87±0.64	79.31±0.53
	TeCh (2026)	**80.60**±0.30	**80.29**±0.24	**80.36**±0.32	**80.23**±0.24	**97.67**±0.10	**86.18**±0.31	**84.22**±0.25
	TS-Fingerprint (2026)	77.59±0.83	76.83±0.77	76.76±1.02	77.48±0.89	96.76±0.47	-	-
	Meredith (ours)	*78.32*±0.41	*78.64*±0.48	*78.56*±0.52	*78.47*±0.45	*97.18*±0.32	*83.66*±0.56	*82.47*±0.46
UCI-HAR (6-Classes)	Autoformer (2021)	41.86±2.46	49.62±11.48	44.30±2.55	32.69±2.60	83.72±2.53	58.56±4.67	51.79±4.38
	FEDformer (2022)	76.89±9.59	75.66±9.46	77.56±9.79	75.03±9.77	95.16±4.66	83.28±8.14	80.60±8.57
	Informer (2021)	88.33±1.26	88.28±1.20	88.47±1.20	88.20±1.29	98.36±0.14	94.20±0.33	90.97±0.90
	iTransformer (2024)	92.41±0.63	92.24±0.63	92.33±0.67	92.39±0.64	99.07±0.07	96.01±0.39	94.08±0.51
	PatchTST (2022)	87.67±0.39	88.37±0.43	87.97±0.37	88.02±0.38	98.50±0.09	93.86±0.40	90.73±0.34
	Medformer (2024b)	89.62±0.81	89.70±0.18	89.80±0.14	89.62±0.81	98.11±0.06	94.80±0.72	91.94±0.45
	MedGNN (2025)	90.94±0.75	91.02±0.62	91.03±0.51	90.99±0.53	99.07±0.48	94.87±0.46	92.82±0.56
	TeCh (2026)	*94.15*±0.96	*94.27*±0.96	*94.30*±0.97	*94.26*±0.98	*99.32*±0.05	*96.74*±0.18	*95.51*±0.68
	TS-Fingerprint (2026)	90.87±1.21	90.61±1.01	90.87±1.21	90.92±1.19	98.67±0.79	-	-
	Meredith (ours)	**95.21**±0.78	**95.44**±0.76	**95.17**±0.76	**95.21**±0.77	**99.70**±0.05	**98.80**±0.17	**96.59**±0.55



(a) ADFTD (b) APAVA (c) PTB (d) PTB-XL

Figure 2: Sensitivity to sampling rate on four datasets. The true sampling rate of each dataset is highlighted in red on the horizontal axis, and the best performance for each metric is indicated by a red star ★.

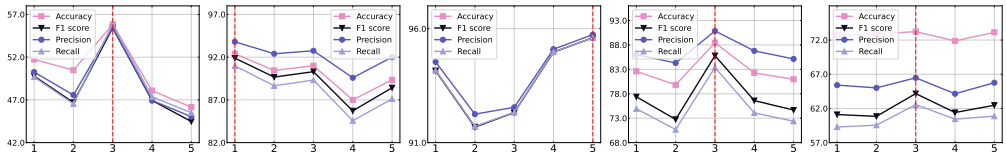
cal frequency scale of the input signals. Once this alignment is disrupted, the encoded rhythm prior becomes less faithful to the underlying physiological rhythms, leading to degraded performance.

B.3.2 Sensitivity to the Number of Channel Refinement Layers L_r

We further investigate the sensitivity of Meredith to the number of channel refinement layers L_r in DACE Net. As shown in Figure 3, performance generally improves when increasing L_r from a shallow setting to a moderate depth, but further stacking does not consistently yield additional gains and can even lead to degradation. This suggests that a sufficient number of refinement stages is necessary to progressively integrate delay-aware information across channels, whereas excessive refinement may introduce redundant interaction or over-smooth channel representations. Moreover, the optimal L_r varies across datasets, indicating that the depth required for effective channel refinement depends on the complexity of cross-channel dependencies inherent in different physiological signals.

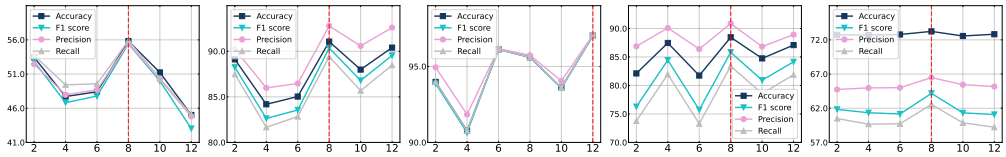
B.3.3 Sensitivity to the Maximum Temporal Delay

We further examine the sensitivity of Meredith to the predefined maximum temporal delay τ in DACE Net. As shown in Figure 4, the performance varies noticeably with τ, and most datasets achieve their best results at a moderate delay range. A small τ restricts DACE Net to short temporal offsets and may fail to capture delayed interactions between physiological channels, whereas an excessively large τ enlarges the candidate delay space and may introduce less relevant or noisy temporal correspondences. The dataset-dependent optimal settings further suggest that effective



(a) ADFTD (b) APAVA (c) TDBrain (d) PTB (e) PTB-XL

Figure 3: Sensitivity to the number of channel refinement layers L_r in DACE Net across five datasets. The red dashed line indicates the setting yielding the best overall performance on each dataset.



(a) ADFTD (b) APAVA (c) TDBrain (d) PTB (e) PTB-XL

Figure 4: Sensitivity to the predefined maximum temporal delay τ in DACE Net across five datasets. The red dashed line indicates the setting yielding the best overall performance on each dataset.

cross-channel interactions occur over different temporal scales across physiological signals, supporting the need to explicitly account for temporal delays rather than assuming strictly synchronous channel relationships.

B.3.4 Sensitivity to d_{model}

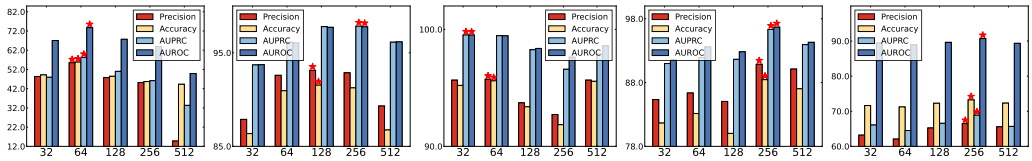
We investigate the sensitivity of Meredith to d_{model}, which denotes the hidden representation dimension of the Transformer encoder. As shown in Figure 5, the optimal setting varies across datasets, but the best performance is generally achieved at a moderate dimensionality rather than at the largest D. A small hidden dimension may limit the capacity of the model to encode discriminative temporal and cross-channel patterns, whereas excessively increasing D introduces additional representational redundancy without consistently improving classification performance. These results suggest that Meredith does not rely on high-dimensional representations for strong performance; instead, an appropriately sized hidden space is sufficient to capture the physiological structures exploited by Rhythm Mixer and DACE Net.

B.3.5 Sensitivity to d_{ff}

The feed-forward dimension d_{ff} controls the capacity of the intermediate transformation within each Transformer encoder layer. As shown in Figure 6, Meredith generally achieves its best performance at a moderate setting of d_{ff}, while either reducing or excessively enlarging this dimension can lead to performance degradation. A smaller d_{ff} may restrict the nonlinear transformation capacity of the encoder, whereas an overly large value introduces additional redundancy without consistently improving discriminative representation learning. The results suggest that Meredith does not rely on an over-parameterized feed-forward network, and that an appropriately sized d_{ff} is sufficient to support effective modeling of the representations produced by Rhythm Mixer and DACE Net.

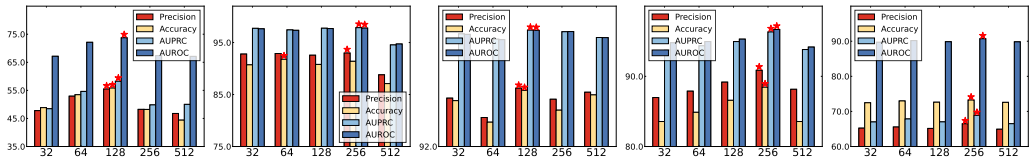
B.3.6 Sensitivity to the Number of Transformer Encoder Layers L_e

The number of Transformer encoder layers L_e determines the depth of high-level representation refinement after the physiologically informed features are extracted. As shown in Figure 7, the optimal depth varies across datasets, while increasing L_e beyond a moderate range does not consistently improve performance and can even cause substantial degradation. This suggests that deeper stacking is not necessarily beneficial once the discriminative temporal and cross-channel structures have already been captured by Rhythm Mixer and DACE Net. A compact Transformer encoder is



(a) ADFTD (b) APAVA (c) TDBrain (d) PTB (e) PTB-XL

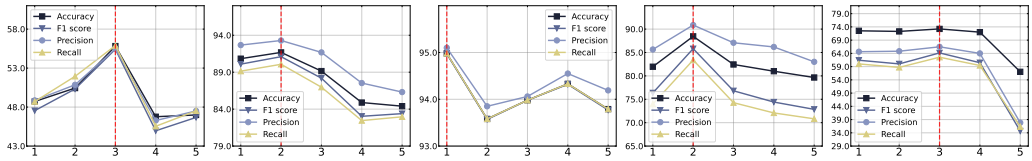
Figure 5: Sensitivity to d_{model} across five datasets. The best performance for each metric is marked with a red star ★.



(a) ADFTD (b) APAVA (c) TDBrain (d) PTB (e) PTB-XL

Figure 6: Sensitivity to d_{ff} across five datasets. The best performance for each metric is marked with a red star ★.

therefore sufficient in most cases, indicating that the effectiveness of Meredith relies more on its physiologically grounded feature modeling than on increasing network depth.



(a) ADFTD (b) APAVA (c) TDBrain (d) PTB (e) PTB-XL

Figure 7: Sensitivity to the number of Transformer encoder layers L_e across five datasets. The red dashed line indicates the setting yielding the best overall performance on each dataset.

B.4 Further efficiency analysis

Tables 9 and 10 report additional efficiency results on ADFTD and PTB datasets, where the batch size is fixed to 128 for all models. Meredith consistently preserves a favorable accuracy-efficiency trade-off across both datasets, further confirming its efficiency across different medical time series scenarios.

B.5 Visualization Analysis

B.5.1 Visualization of Delay-Aware Channel Interactions in DACE Net

Figures 8 and 9 visualize representative slices of the learned delay weights on APAVA and TDBRAIN datasets. The interaction patterns vary noticeably across different delays, showing that the importance assigned to a given channel pair is not fixed but changes with the temporal offset. Moreover, the learned structures differ between the two datasets, suggesting that DACE Net adapts its delay-aware channel modeling to dataset-specific physiological dynamics rather than imposing a uniform dependency pattern. These observations provide qualitative evidence that inter-channel relationships in physiological signals are inherently delay-dependent, supporting the motivation of explicitly modeling temporal offsets in DACE Net.

Table 9: Efficiency comparison between Meredith and baselines on the ADFTD dataset under the subject-independent Fixed Subject Split protocol. *Time* denotes the average *training time* required per epoch.

Model	Parameters	MACs	Time	Memory Usage	Accuracy (%)	AUROC (%)
Autoformer	895.88 K	204.97 M	19.5 s	4762 MB	45.25	61.02
FEDformer	1.49 M	204.97 M	120.0 s	3508 MB	46.30	62.62
Informer	1.1 M	208.63 M	16.1 s	3384 MB	48.45	65.87
iTransformer	835.33 K	15.71 M	11.5 s	2022 MB	52.60	67.26
Nonformer	978.69 K	205.2 M	30.9 s	8820 MB	49.95	66.23
PatchTST	1.03 M	483.91 M	28.9 s	6912 MB	44.37	60.08
Medformer	7.98 M	334.48 M	95.6 s	5604 MB	53.27	70.93
MedGNN	11.5 M	749.01 M	36.2 s	5758 MB	56.12	74.68
TeCh	1.53 M	207.1 M	13.3 s	3042 MB	54.54	68.67
Meredith(ours)	210.97 K	24.4 M	26.3 s	2372 MB	54.88	72.55

Table 10: Efficiency comparison between Meredith and baselines on the PTB dataset under the subject-independent Fixed Subject Split protocol. *Time* denotes the average *training time* required per epoch.

Model	Parameters	MACs	Time	Memory Usage	Accuracy (%)	AUROC (%)
Autoformer	872.83 K	239.69 M	15.5 s	5268 MB	73.35	78.54
FEDformer	1.47 M	239.69 M	106.2 s	3670 MB	76.05	85.93
Informer	1.08 M	243.99 M	14.0 s	3552 MB	78.69	92.09
iTransformer	837.51 K	12.48 M	9.6 s	1990 MB	83.89	91.18
Nonformer	959.5 K	239.95 M	36.4 s	11156 MB	78.66	89.37
PatchTST	939.27 K	441.66 M	23.8 s	6272 MB	74.74	88.79
Medformer	6.0 M	337.19 M	64.2 s	6128 MB	83.50	92.81
MedGNN	11.49 M	808.52 M	33.5 s	6086 MB	84.53	93.31
TeCh	1.56 M	23.44 M	6.5 s	1546 MB	85.96	94.57
Meredith(ours)	992.17 K	47.17 M	24.7 s	2612 MB	87.50	96.72

B.6 Visualization of Learned Channel Interactions in DACE Net

Figure 10 visualizes the learned channel weights on three datasets. The resulting interaction patterns are clearly non-uniform and differ substantially across datasets, indicating that DACE Net does not impose a fixed dependency structure among channels. Instead, it adaptively emphasizes different channel pairs according to the characteristics of the underlying signals. This dataset-specific structure suggests that effective multichannel physiological modeling requires selective interaction rather than treating all channel relationships equally, further supporting the channel refinement mechanism in DACE Net.

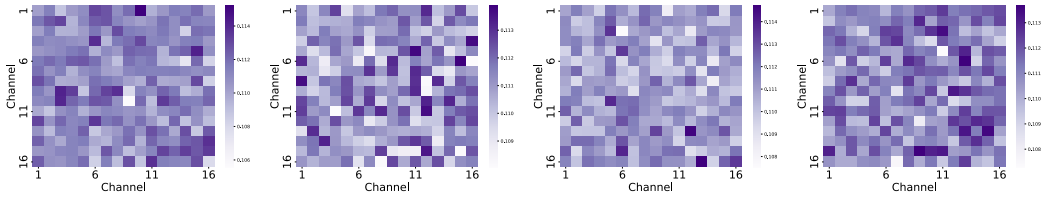


Figure 8: Visualization of the learned delay weights $\mathbf{W}^{\text{delay}} \in \mathbb{R}^{C \times C \times (\tau+1)}$ in DACE Net on the APAVA dataset (16 channels). Each subfigure shows the channel-to-channel weight matrix at a specific delay d. For clarity, only a subset of delay slices is shown.

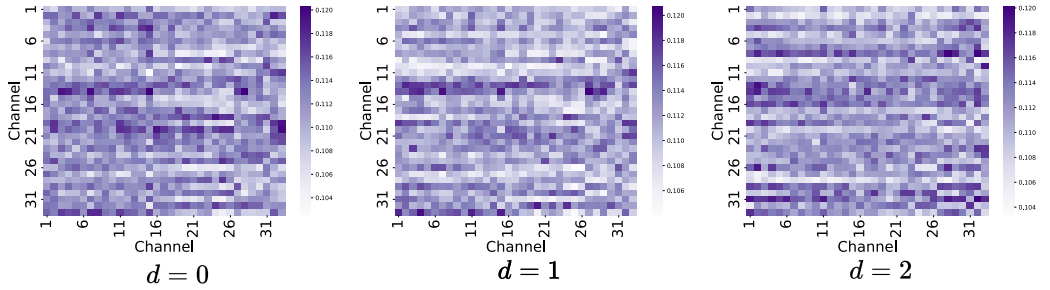
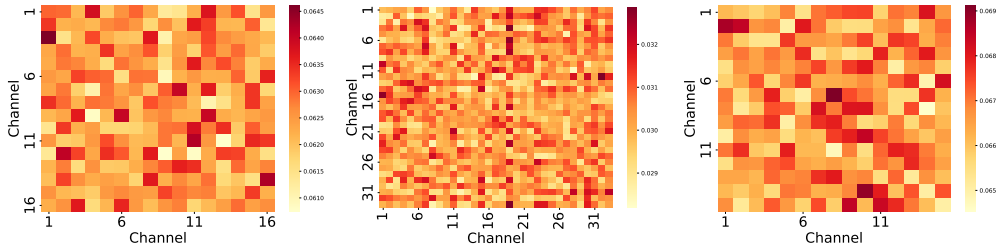


Figure 9: Visualization of the learned delay weights $\mathbf{W}^{\text{delay}} \in \mathbb{R}^{C \times C \times (\tau+1)}$ in DACE Net on the TDBRAIN dataset (33 channels). Each subfigure shows the channel-to-channel weight matrix at a specific delay d. For clarity, only a subset of delay slices is shown.



(a) APAVA (b) TDBRAIN (c) PTB

Figure 10: Visualization of the learned channel weights $\mathbf{W}^{\text{channel}} \in \mathbb{R}^{C \times C}$ in DACE Net on the APAVA, TDBRAIN, and PTB datasets. Each element represents the learned interaction strength between a pair of channels.