Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

M-S4DLite: Orthogonal Decoupling for Lightweight Time-Series Classification via Diagonal State-Space Models

jian chen

chenjian24@mails.ucas.ac.cn

yunfei han
yi wang
shiwei guo
yupeng ma

Article

Keywords: Time-Series Classification, State-Space Models, Orthogonal Decoupling, Spectral Oversmoothing, Inductive Bias Calibration, Lightweight Deep Learning

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8919991/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

M-S4DLite: Orthogonal Decoupling for Lightweight Time-Series Classification via Diagonal State-Space Models

Jian Chen[1,2,3*], Yunfei Han[1,3], Yi Wang[1,3], Shiwei Guo[1,2,3], Yupeng Ma[1,3*]

[1]Xinjiang Institute of Physical and Chemical Technology, Chinese Academy of Sciences, Urumqi, 830011, Xinjiang, China.
[2]University of Chinese Academy of Sciences, Beijing, 100049, China.
[3]Key Laboratory of Minority Language Resources and Information Processing of Xinjiang, Chinese Academy of Sciences, Urumqi, 830011, Xinjiang, China.

*Corresponding author(s). E-mail(s): chenjian24@mails.ucas.ac.cn; ypma@ms.xjb.ac.cn;

Abstract

Emerging state space models have revolutionized time series forecasting by efficiently capturing long range dependencies, yet their application to classification tasks reveals a fundamental semantic misalignment. Rooted in continuous control theory, standard models suffer from spectral oversmoothing where continuous inductive bias tends to erode the sharp and discrete discriminative motifs critical for classification. To resolve this theoretical conflict, we propose M-S4DLite, a lightweight framework built on the principle of Orthogonal Decoupling. By structurally separating local feature extraction from global evolution, we employ a discrete patching module to preserve high frequency transients and synergize it with a diagonalized S4D backbone for global reasoning. This architecture effectively recalibrates the inductive bias to reconcile discrete local precision with continuous global context while maintaining quasi linear complexity. Extensive experiments demonstrate that M-S4DLite establishes a new Pareto frontier in the balance between efficiency and accuracy, consistently achieving superior performance on diverse benchmarks and exhibiting robustness against resolution degradation compared to pure continuous or convolutional baselines.

Keywords: Time-Series Classification,State-Space Models,Orthogonal Decoupling,Spectral Oversmoothing,Inductive Bias Calibration,Lightweight Deep Learning

1 Introduction

Time-Series Classification (TSC) constitutes the foundational layer for diverse ubiquitous applications, from real-time arrhythmia detection in wearable healthcare to fault diagnosis in industrial IoT [1]. Unlike forecasting tasks that focus on trend extrapolation, classification hinges on the precise identification of *discriminative shapelets*—discrete, often high-frequency subsequences buried within noisy signals. While traditional ensemble methods like the Canonical Interval Forest (CIF) [2] and TS-CHIEF [3] established strong accuracy baselines, they struggle with scalability on high-dimensional data, prompting a shift toward deep learning paradigms.

However, the quest for optimal TSC architectures currently faces a fundamental "Granularity-Context Conflict." On one hand, discriminative features (e.g., the onset of a gesture or a heartbeat spike) are inherently local and transient. Convolutional Neural Networks (CNNs), such as InceptionTime [1] and the recent ModernTCN [4], excel at capturing these local motifs but suffer from limited receptive fields. Capturing global dependencies requires stacking excessively deep layers, leading to optimization instability and parameter redundancy.

On the other hand, the field has witnessed a surge in Large Time Series Models (LTSMs) adapting Transformer architectures [5, 6]. While approaches like ReasonTSC [7] leverage global attention to achieve state-of-the-art capabilities, they are plagued by the "accuracy-efficiency dilemma." The quadratic complexity ($\mathcal{O}(L^2)$) and billion-parameter scales render them prohibitively expensive for resource-constrained edge devices. Even efficient variants like TimeMixer [8] pose a significant computational burden compared to linear baselines [9].

To bridge this gap, Structured State-Space Models (SSMs) have emerged as a promising third paradigm, theoretically offering global modeling with quasi-linear complexity via Fast Fourier Transform (FFT) [10, 11]. Yet, a blind migration of these continuous models to classification reveals a critical "Semantic Misalignment." Standard SSMs, rooted in continuous control theory [12], implicitly prioritize spectral smoothness. This continuous inductive bias is beneficial for reconstructing smooth trajectories but detrimental for classification, as it tends to "oversmooth" the sharp, discrete transients that define class boundaries. Consequently, while recent SSM variants like DecMamba [13] succeed in forecasting, general-purpose SSMs often fail to match the granular precision of specialized lightweight modules like LITE [14] or ROCKET [15].

Addressing this theoretical conflict, we propose M-S4DLite, a lightweight framework built on the principle of *Orthogonal Decoupling*. We identify that local feature extraction and global dependency modeling require conflicting inductive biases. Therefore, we synergize a discrete patching module (to strictly preserve high-frequency locality and prevent smoothing) with a diagonalized S4D backbone (to interpolate

low-frequency global trends). This methodology essentially re-calibrates the inductive bias of SSMs, achieving state-of-the-art accuracy with quasi-linear complexity ($\mathcal{O}(L \log L)$). Our contributions are:

1. **Orthogonal Decoupling Strategy:** We propose a novel architectural design that structurally separates local tokenization from global reasoning. This effectively resolves the inductive bias mismatch, allowing continuous SSMs to model discrete classification tasks without feature oversmoothing.
2. **Linear Complexity Backbone:** Leveraging the strictly diagonal structure of S4D [11] and FFT-based spectral interactions, we construct a global receptive field with efficiency comparable to simple linear models, far surpassing quadratic Transformers in throughput and memory usage.
3. **Pareto Optimality and Robustness:** Extensive experiments confirm that M-S4DLite establishes a new Pareto frontier in efficiency-accuracy trade-offs. Crucially, the model demonstrates superior cross-scale robustness under varying sequence lengths, outperforming recent baselines like ModernTCN [4] and TimeMIL [16].

2 Related Work

2.1 Deep Convolutional Paradigms

Convolutional Neural Networks (CNNs) have historically constituted the backbone of Time-Series Classification (TSC), evolving from simple discriminators to complex hierarchical feature extractors. Early foundations were laid by the Fully Convolutional Network (FCN) [17] and ResNet variants, which introduced residual connections to prevent gradient vanishing. The landmark InceptionTime model [1] marked a significant milestone by utilizing parallel convolutional kernels of varying lengths to capture diverse temporal patterns. Subsequent hybrid architectures, such as LSTM-FCN [18], attempted to augment convolutional features with recurrent dependencies, often at the cost of parallelization efficiency.

Recently, the focus has shifted towards modernizing pure CNN architectures to rival the Transformer paradigm. ModernTCN [4] rejuvenated CNNs by introducing large-kernel depthwise convolutions to expand the effective receptive field while retaining local inductive biases. Furthermore, lightweight designs like LITE [14] have pushed efficiency limits specifically for edge deployment. However, despite these structural innovations, CNN-based methods fundamentally suffer from the "locality bottleneck": expanding the receptive field to cover global dependencies requires stacking excessively deep layers. This inevitably introduces parameter redundancy and optimization difficulties, limiting their ability to model long-range trajectories efficiently.

2.2 Transformers and Foundation Models

To transcend the locality limitations of CNNs, Transformer-based architectures have been extensively adapted for time-series analysis. PatchTST [19] pioneered channel-independent patching, proving that segmenting signals into sub-series tokens preserves

local semantics better than point-wise attention. Following this, Pathformer [20] introduced adaptive pathways for multi-scale resolutions, while Shapeformer [21] integrated shapelet discovery with self-attention.

The landscape has recently witnessed a shift towards Time-Series Foundation Models. Leveraging Large Language Models (LLMs), approaches like Time-LLM [5] and UniTS [6] utilize cross-modality reprogramming. Similarly, ReasonTSC [7] employs LLMs to enhance interpretability. While powerful, these models face the "accuracy-efficiency dilemma." The quadratic complexity ($\mathcal{O}(L^2)$) and billion-scale parameters render them prohibitively expensive for real-time edge deployment. Although variants like TimeMixer [8] attempt to mitigate this via multiscale mixing, the computational overhead remains significant compared to linear alternatives [9], necessitating a lighter global modeling paradigm.

2.3 Structured State-Space Models

Structured State-Space Models (SSMs) have emerged as a compelling third paradigm, theoretically bridging the gap between CNN efficiency and Transformer global modeling. Grounded in the HiPPO theory [12] of optimal polynomial projection, SSMs map inputs to latent states via continuous-time evolution. The pioneering S4 model [10] enabled modeling extremely long sequences with linear complexity ($\mathcal{O}(L)$), which was further refined by the diagonalized S4D [11] for enhanced parallelization.

However, we argue that the current SSM literature suffers from a "Forecasting-Classification Gap." Most advancements, such as DecMamba [13] and Mamba-2 [22], are optimized for signal reconstruction, benefiting from the implicit spectral smoothness of continuous integration. When applied to classification, this continuous inductive bias becomes a liability, leading to the "oversmoothing" of discrete, high-frequency discriminative motifs (e.g., sudden anomalies or gesture onsets). Unlike forecasting, where smoothness aids generalization, classification demands the preservation of sharp transients. Our M-S4DLite framework specifically addresses this by introducing an orthogonal decoupling strategy, using discrete patching to re-calibrate the inductive bias before the continuous S4D backbone.

2.4 Frequency Domain Learning and Robustness

Complementary to time-domain modeling, frequency-domain analysis offers a global perspective on signal periodicity. Methods like SciNet [23] and Ftmixer [24] explicitly model interactions in the frequency domain. M-S4DLite internalizes this insight: by leveraging the convolution theorem, we treat global reasoning as a spectral interaction problem computed via Fast Fourier Transform (FFT). This naturally integrates long-range dependencies with minimal overhead, avoiding the computational cost of self-attention.

Crucially, robust generalization requires tackling non-stationarity—where statistical properties shift over time. While RevIN [25] is standard in forecasting, its role in lightweight classification is less explored. We adopt a tailored reversible normalization strategy to stabilize input distributions for the S4D backbone. This ensures robustness

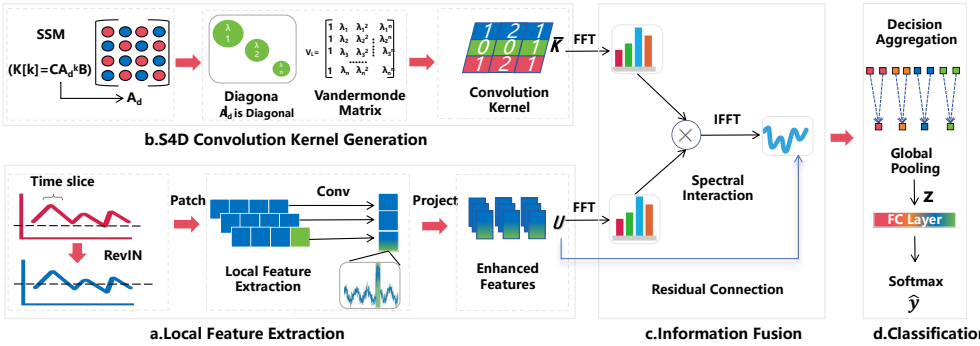


Fig. 1 The overall architecture of M-S4DLite. (a) **Local Tokenization**: The input series is normalized via RevIN and unfolded into patches to form the discrete driving signal $\mathbf{U}$. (b) **Continuous S4D Evolution**: A diagonal state matrix $\mathbf{A}$ is parameterized in the complex domain to model system dynamics. (c) **Spectral Interaction**: The global convolution kernel $\overline{\mathbf{K}}$ is computed and applied via FFT, achieving quasi-linear complexity. (d) **Decision Head**: Aggregated features are mapped to class probabilities.

against sensor noise and domain shifts, a vital capability for bridging the gap between theoretical modeling and practical edge deployment.

3 Methodology

The overall architecture of our proposed M-S4DLite framework is illustrated in Fig. 1.

3.1 Problem Formulation

Let $\mathcal{D} = \{(\mathbf{X}^{(i)}, y^{(i)})\}_{i=1}^{M}$ denote the dataset containing M samples. Each input $\mathbf{X}^{(i)} \in \mathbb{R}^{C \times L}$ represents a multivariate time series with C variates (channels) and temporal length L. The corresponding label is $y^{(i)} \in \{1, \dots, K\}$, where K is the number of classes. Our objective is to learn a parameterized mapping $\mathcal{F}_\Theta : \mathbb{R}^{C \times L} \to \mathbb{R}^K$ to minimize the empirical cross-entropy loss.

3.2 Local Tokenization via Discrete Patching

Real-world time series often exhibit non-stationarity. To mitigate distribution shifts, we first apply Reversible Instance Normalization (RevIN). Given an input instance $\mathbf{X}$ (superscript omitted for brevity), the normalized series $\tilde{\mathbf{X}} \in \mathbb{R}^{C \times L}$ is computed as:

$$\tilde{\mathbf{X}} = \gamma \odot \frac{\mathbf{X} - \mu_{\mathbf{X}}}{\sqrt{\sigma_{\mathbf{X}}^2 + \epsilon}} + \beta \tag{1}$$

where $\mu_{\mathbf{X}}, \sigma_{\mathbf{X}}^2 \in \mathbb{R}^{C \times 1}$ are the instance-specific mean and variance, $\gamma, \beta \in \mathbb{R}^C$ are learnable affine parameters, and $\odot$ denotes the element-wise Hadamard product.

Subsequently, we employ an Orthogonal Decoupling strategy. To prevent the continuous backbone from smoothing out high-frequency local variations, we explicitly

discretize the signal into patches. We define a patching operator $\mathcal{P}(\cdot)$ with patch length P and stride S, transforming $\tilde{\mathbf{X}}$ into a sequence of N_p patches, where $N_p = \lfloor \frac{L-P}{S} \rfloor + 1$. These patches are projected into a latent space of dimension D_{model} via a linear transformation $\mathbf{W}_{emb}$:

$$\mathbf{U} = \mathcal{P}(\tilde{\mathbf{X}})\mathbf{W}_{emb} \in \mathbb{R}^{N_p \times D_{model}} \tag{2}$$

Crucially, the tensor $\mathbf{U}$ serves as the discrete *driving signal* $u(t)$ for the subsequent state-space model, acting as a high-frequency preservation buffer.

3.3 Continuous Dynamics via Diagonal S4

The core of M-S4DLite is the Structured State Space (S4) model, which views the discrete feature sequence as the sampled response of a continuous-time system. We model the latent state evolution $\mathbf{h}(t) \in \mathbb{C}^N$ and output signal $y(t)$ via the Ordinary Differential Equation (ODE):

$$\begin{cases} \mathbf{h}'(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}u(t) \\ y(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}u(t) \end{cases} \tag{3}$$

Here, N denotes the state dimension, and the state matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ governs the system dynamics. To ensure linear scalability $\mathcal{O}(L)$ and efficient parameterization, we strictly constrain $\mathbf{A}$ to be Diagonal (S4D).

Stable Parameterization: We parameterize the diagonal matrix $\mathbf{A}$ in the complex domain to guarantee stability. For the n-th diagonal element ($n \in \{1, \dots, N\}$):

$$\mathbf{A}_{nn} = -\exp(\theta_{real}^{(n)}) + j \cdot \theta_{imag}^{(n)} \tag{4}$$

where $\theta_{real}, \theta_{imag} \in \mathbb{R}^N$ are learnable parameters, and j is the imaginary unit. This formulation enforces the real part of eigenvalues to be negative ($\mathrm{Re}(\mathbf{A}_{nn}) < 0$), ensuring system stability over long sequences.

3.4 Discretization and Spectral Interaction

To process the discrete input $\mathbf{U}$ (sampled at intervals Δ), the continuous ODE in Eq. (3) must be discretized. We employ the Zero-Order Hold (ZOH) principle. The discrete state matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ are derived as:

$$\overline{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad \overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B} \tag{5}$$

where Δ is a learnable timescale parameter and $\mathbf{I}$ is the identity matrix.

Global Convolution Kernel: The system response can be equivalently formulated as a global convolution $y = \overline{\mathbf{k}} * u$. The global kernel $\overline{\mathbf{K}} \in \mathbb{C}^{N_p}$ represents the discretized impulse response:

$$\overline{\mathbf{K}} = \left(\mathbf{C}\overline{\mathbf{B}}, \mathbf{C}\overline{\mathbf{A}\mathbf{B}}, \dots, \mathbf{C}\overline{\mathbf{A}}^{N_p-1}\overline{\mathbf{B}} \right) \tag{6}$$

Table 1 Quantitative comparison on three benchmarks. We report $F_1 \uparrow$ scores. M-S4DLite demonstrates superior performance, particularly on long-sequence tasks (UWave), validating the efficacy of the global S4D backbone. Best results are in **bold**.

	MIT-BIH ECG ($F_1 \uparrow$)			**UCI-HAR** ($F_1 \uparrow$)			**UWave-All** ($F_1 \uparrow$)		
Model	**H=16**	**H=32**	**H=64**	**H=16**	**H=32**	**H=64**	**H=16**	**H=32**	**H=64**
DLinear	0.9513	0.9531	0.9575	0.9789	0.9782	0.9817	0.9456	0.9497	0.9481
Pathformer	0.9953	0.9961	0.9970	0.9812	0.9884	0.9905	0.9651	0.9742	0.9749
PatchTST	0.9861	0.9852	0.9883	0.9789	0.9769	0.9816	0.9282	0.9609	0.9429
ModernTCN	0.9693	0.9720	0.9903	0.6806	0.9064	0.9796	0.6143	0.7869	0.8074
TimeMixer	0.9698	0.9731	0.9757	0.9811	0.9816	0.9832	0.9439	0.9347	0.9274
TimeMIL	0.9732	0.9842	0.9894	0.9775	0.9706	0.9831	0.8110	0.8536	0.8956
M-S4DLite	**0.9962**	**0.9969**	**0.9972**	**0.9817**	**0.9889**	**0.9909**	**0.9707**	**0.9778**	**0.9817**

Spectral Interaction via FFT: To bypass the $\mathcal{O}(N_p^2)$ complexity of time-domain convolution, we leverage the Convolution Theorem. The operation is performed in the frequency domain:

$$\mathbf{Y} = \mathcal{F}^{-1}\left(\mathcal{F}(\mathbf{U}) \odot \mathcal{F}(\overline{\mathbf{K}})\right) + \mathbf{DU} \tag{7}$$

where $\mathcal{F}$ denotes the Fast Fourier Transform (FFT), $\mathcal{F}^{-1}$ is its inverse, and $\odot$ is the Hadamard product. This reduces the computational complexity to $\mathcal{O}(N_p \log N_p)$, enabling the model to capture global dependencies across the entire sequence.

3.5 Feature Aggregation and Classification

The output $\mathbf{Y} \in \mathbb{R}^{N_p \times D_{model}}$ contains context-aware representations. To obtain a sequence-level representation vector $\mathbf{z} \in \mathbb{R}^{D_{model}}$, we apply Global Average Pooling (GAP):

$$\mathbf{z} = \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{Y}_i \tag{8}$$

Finally, the class probability distribution is computed via a linear classifier:

$$\hat{\mathbf{y}} = \text{Softmax}(\mathbf{W}_{cls}\mathbf{z} + \mathbf{b}) \tag{9}$$

where $\mathbf{W}_{cls} \in \mathbb{R}^{K \times D_{model}}$ and $\mathbf{b} \in \mathbb{R}^K$.

3.6 Experimental Settings

To ensure a comprehensive evaluation, we benchmark M-S4DLite against state-of-the-art models on three public datasets selected to represent distinct structural challenges: MIT-BIH ECG (medical diagnostics containing sharp morphological transients, 5 classes, length 187), UCI-HAR (human activity with complex periodicity, 6 activities, length 561), and UWave Gesture Library (long-range motion trajectories, 8 gestures, length 945). All experiments follow a strict 7:1:2 train/val/test split.

We compare our model against representative baselines from three categories: (1) CNN-based: ModernTCN [4] and TimeMixer [8]; (2) Transformer-based: PatchTST [19], Pathformer [20], and TimeMIL [16]; (3) MLP-based: DLinear [9].

Implementation Details: We sweep hidden sizes $H \in \{16, 32, 64\}$, fixing other hyperparameters (Adam optimizer, LR 10^{-3}, batch size 32) to ensure a fair comparison. To ensure statistical stability, we report the mean performance over 5 independent runs with early stopping (patience 5). The primary metric is the class-averaged F_1-score. Efficiency latency measurements are performed under consistent hardware constraints.

3.7 Main Results

Table 1 presents the comprehensive comparison across three diverse benchmarks. M-S4DLite achieves state-of-the-art performance, securing the top rank in all 9 experimental settings. This consistent dominance strongly validates the superiority of the "Orthogonal Decoupling" architecture in resolving the granularity-context conflict.

Dominance on Long Sequences (Global Reasoning). On the UWave dataset, where sequences are long ($L = 945$) and capturing global trajectory is critical, M-S4DLite consistently outperforms all baselines. Notably, at $H = 64$, it achieves an F_1 of 0.9817, significantly surpassing ModernTCN (0.8074) and PatchTST (0.9429). This empirically proves that our Global S4D Kernel effectively captures long-range dependencies via spectral interaction, overcoming the receptive field limitations of CNNs while avoiding the quadratic complexity of Transformers.

Precision on Local Motifs (Anti-Oversmoothing). On datasets requiring fine-grained local feature extraction, M-S4DLite demonstrates exceptional precision. On UCI-HAR, it achieves an F_1 of 0.9909, outperforming the strong Pathformer baseline (0.9905). Crucially, on the medical diagnostics task (MIT-BIH ECG), which relies heavily on identifying specific local morphological anomalies (e.g., QRS complex spikes), our model achieves an F_1 of 0.9962 even with a compact hidden dimension ($H = 16$), surpassing the heavy Transformer baseline Pathformer (0.9953). This result is vital: it empirically validates that our Discrete Patching mechanism effectively shields high-frequency motifs from the spectral smoothing bias of the continuous S4D backbone, successfully reconciling local precision with global modeling.

3.8 Analytical Experiments

Resolution Robustness and Structural Stability. Real-world IoT sensors often face unstable sampling rates or bandwidth constraints, necessitating models that are robust to resolution degradation. To evaluate this, we conducted a stress test by progressively downsampling the UWave sequences from the original length L=945 to coarser granularities $L \in \{470, 230, 120\}$.

As illustrated in Fig. 2, M-S4DLite exhibits exceptional stability, maintaining an essentially flat performance curve (black line) with $F_1 > 0.98$ even at the lowest resolution. In sharp contrast, baseline architectures struggle with structural volatility. Specifically, CNN-based models like ModernTCN display significant instability (dropping to $\sim$ 0.85 at L=470) because their fixed-size kernels rely on rigid geometric priors; when the temporal resolution changes, the semantic meaning of the receptive field shifts, leading to feature misalignment. Similarly, Transformer variants (e.g.,

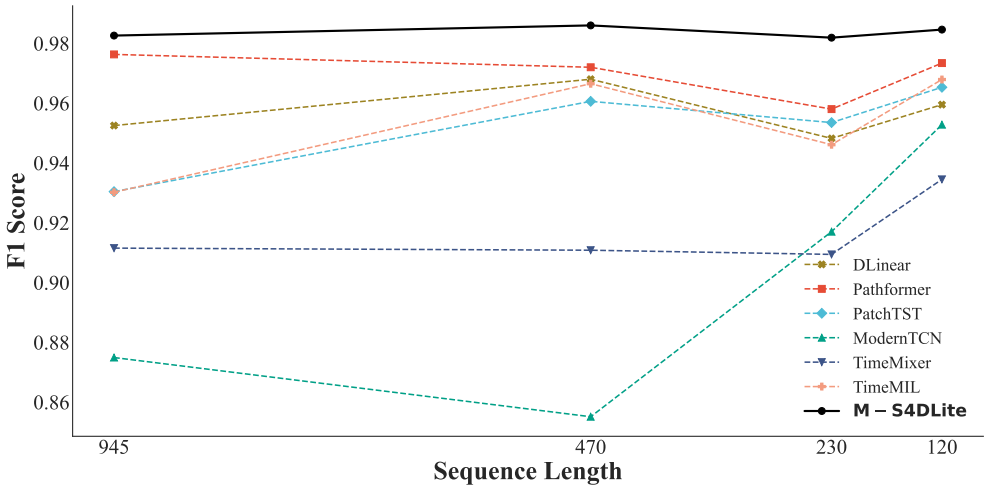


Fig. 2 Robustness across sequence lengths. M-S4DLite exhibits the minimal performance drop as L decreases, validating that the global S4D backbone effectively interpolates information even from coarse-grained inputs.

PatchTST) suffer from token aliasing artifacts as the patch semantics change with downsampling.

We attribute M-S4DLite's superior robustness to the intrinsic continuous nature of our Orthogonal Decoupling architecture. Unlike discrete convolutions that operate on fixed grids, the S4D backbone models the temporal dynamics as a continuous Ordinary Differential Equation (ODE). This allows the model to learn the underlying function of the signal trajectory rather than memorizing discrete points. Consequently, even when the input grid becomes sparse (L=120), the S4D backbone effectively interpolates the global context, while our discrete patching module preserves the essential local shapelets. This resolution-invariant characteristic confirms M-S4DLite's suitability for dynamic edge environments where data fidelity cannot be guaranteed.

Pareto Optimality in the Accuracy-Efficiency Landscape. The holy grail of edge-deployed time-series classification is to maximize accuracy while minimizing computational overhead. Fig. 3 visualizes this trade-off, positioning M-S4DLite on the optimal Pareto frontier (Top-Left corner of the right subplot).

Inference Latency Analysis: As detailed in Fig. 3 (Left), M-S4DLite records a total inference time of **1040.5 ms**, which is competitively close to the simplistic linear baseline (DLinear, 839.4 ms) and significantly faster than Transformer-based counterparts like TimeMIL (>10,000 ms) and Pathformer (∼2174 ms). This acceleration is theoretically grounded in our architectural choices: unlike the $\mathcal{O}(L^2)$ complexity inherent in the self-attention mechanisms of Transformers, our FFT-based spectral interaction operates with quasi-linear $\mathcal{O}(L \log L)$ complexity. Furthermore, by diagonalizing the state matrix, we eliminate the heavy matrix multiplications typical of standard RNNs, ensuring that throughput remains high even for long sequences.

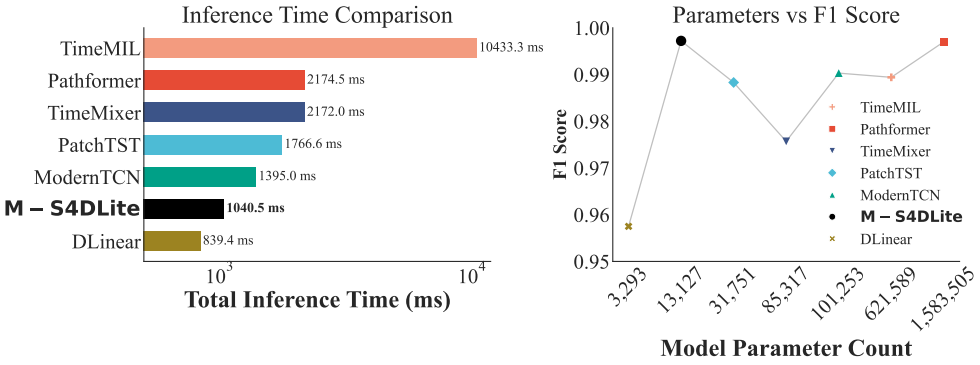


Fig. 3 Efficiency Analysis. Left: Inference latency. Right: F_1 vs. Parameters. M-S4DLite achieves the optimal Pareto frontier (Top-Left corner), balancing lightweight design with high accuracy.

Parameter Efficiency and Information Density: The scatter plot in Fig. 3 (Right) reveals a stark contrast in model capacity usage. M-S4DLite achieves the highest F1 score (≈ 0.997) with merely $\sim$**13k parameters**. In comparison, heavy foundation models like Pathformer require over **1.5 million parameters**—two orders of magnitude more—to achieve comparable or slightly inferior performance. This suggests that the "brute-force" approach of scaling up model size yields diminishing returns in classification tasks. Instead, our results validate the "Lite" philosophy: by structurally decoupling local feature extraction (Patching) from global dynamics (S4D), M-S4DLite achieves superior *information density*, capturing the essential discriminative physics of the signal without redundant parametrization. This makes it uniquely suited for deployment on resource-constrained microcontrollers where memory footprints are strictly capped.

Structural Sensitivity and Inductive Bias Calibration. To understand the interplay between discrete tokenization and continuous modeling, we performed a fine-grained sensitivity analysis on the Patch Length P and State Dimension N, as visualized in Fig. 4.

The Resonance of Patching Granularity: The top row of Fig. 4 reveals a consistent convex optimization curve across all datasets, identifying a critical "semantic resonance zone" around $P \approx 12$. We attribute this to two competing failure modes:

1. *Semantic Fragmentation ($P < 8$):* When patches are too small, the local morphological features (e.g., a single heartbeat spike) are sliced into fragmented tokens. This forces the S4D backbone to waste capacity reconstructing local continuity from incoherent pieces rather than modeling global dynamics.
2. *Intra-Patch Oversmoothing ($P > 16$):* Conversely, excessively large patches aggressively compress information before it enters the model. High-frequency discriminative transients are averaged out within the projection layer (intra-patch aliasing), causing irreversible information loss that the subsequent backbone cannot recover.

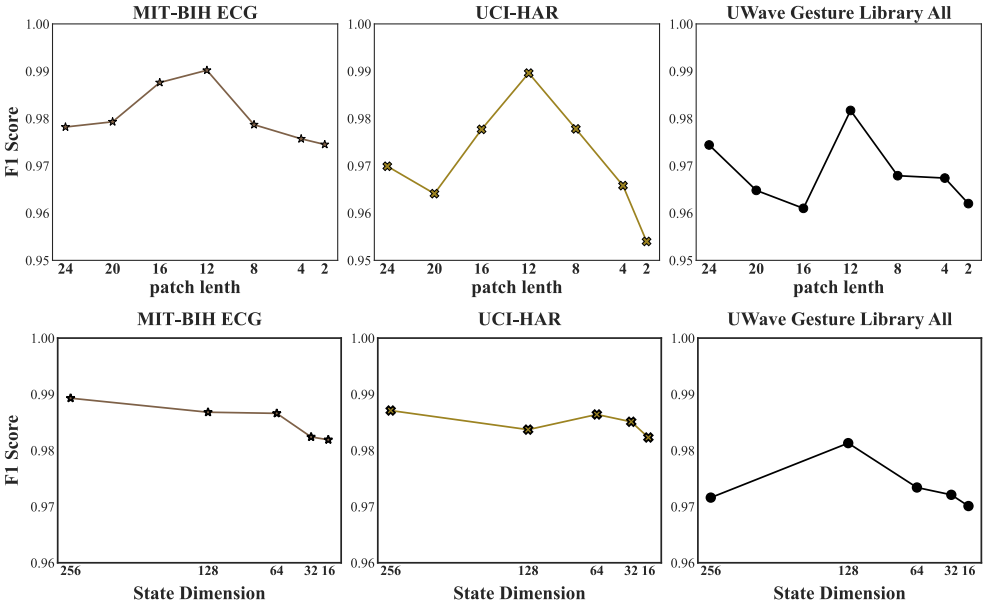


Fig. 4 Sensitivity Analysis. Top: Impact of Patch Length P. Performance drops at extreme values, confirming the need for a calibrated discretization scale. Bottom: Impact of State Dimension N.

The peak performance at $P \approx 12$ confirms that aligning the tokenization scale with the *characteristic timescale* of the signal's physical events is crucial for optimal performance.

Spectral Sufficiency of the State Space: The bottom row of Fig. 4 demonstrates that performance improves with State Dimension N but rapidly saturates beyond $N = 64$. This saturation is theoretically significant. Unlike dense Recurrent Neural Networks which often require hidden states of size 512+ to prevent forgetting, the diagonal S4D matrix effectively decomposes the signal into independent orthogonal frequency modes. The plateau indicates that the underlying dynamics of these physiological and kinematic signals are inherently *low-rank*; a state dimension of 64 provides sufficient "spectral capacity" to memorize all relevant history. Increasing N further yields diminishing returns, validating our lightweight design choice to cap model complexity without sacrificing expressivity.

3.9 Component-wise Ablation and Mechanism Verification

To rigorously verify the internal validity of the proposed Orthogonal Decoupling architecture, we conducted a component-wise ablation study across all three datasets. The results, visualized in Fig. 5, reveal the distinct role each module plays in reconciling the granularity-context conflict.

The Imperative of Global S4D Reasoning. The most striking observation occurs in the long-sequence regime (UWave, $L = 945$). Removing the S4D backbone

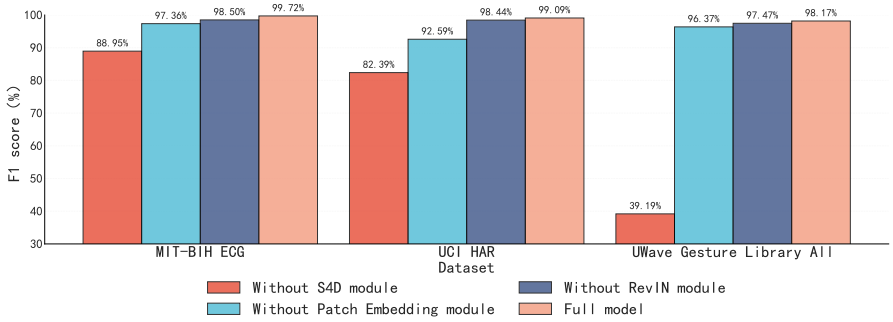


Fig. 5 Ablation Study. Removing S4D leads to the most significant drop, confirming its role as the core global reasoner. Removing Patching also degrades performance, validating the necessity of discrete tokenization.

(Red Bar) causes a catastrophic performance collapse, with the F1 score plummeting to **39.19%**. This empirical evidence underscores that for complex, long-duration trajectories, local features alone are insufficient. The S4D module is not merely an enhancer but the fundamental engine for *temporal integration*. Without it, the model degenerates into a shallow, local-only processor, completely losing the capacity to track the global evolution of gestures, effectively reducing classification to random guessing.

Discrete Patching as a Spectral Barrier. Removing the Patch Embedding module (Light Blue Bar) leads to a consistent degradation across all tasks, particularly on MIT-BIH (dropping from 99.72% to 97.36%). This validates our core hypothesis regarding *Spectral Oversmoothing*. Standard SSMs define a continuous-time integration process, which inherently acts as a low-pass filter. Without the discretization provided by Patching, the S4D backbone acts directly on raw point-wise signals, aggressively smoothing out the high-frequency transients (such as the sharp R-peaks in ECGs) that are critical for discrimination. The Patching module effectively functions as a "structural barrier," encapsulating these local sharp details into tokens before they are subjected to global integration, thus preserving high-frequency fidelity.

Robustness via RevIN. Finally, removing Reversible Instance Normalization (Dark Blue Bar) results in a moderate performance drop. While not as structurally critical as S4D or Patching, its presence is vital for mitigating non-stationarity. By normalizing the input distribution, RevIN ensures that the S4D backbone learns the intrinsic shape dynamics rather than overfitting to the shifting statistics of signal amplitude, thereby enhancing generalization across diverse subjects.

Visual Interpretability and Manifold Separability. We further corroborate the theoretical advantages of M-S4DLite by visualizing the internal feature dynamics and the resulting latent space topology.

Microscopic Dynamics (Activation Heatmaps): Fig. 6 provides a comparative view of the feature activations across channels over time. The "With S4D" panel exhibits distinct, continuous horizontal activation bands. This implies that the S4D backbone successfully captures *temporal consistency*, effectively "locking onto" specific latent frequency modes and sustaining them across the sequence. In sharp contrast, the "Without S4D" panel displays a sparse, scattered, and stochastic activation pattern. This "salt-and-pepper" noise indicates that without the global state-space integration,

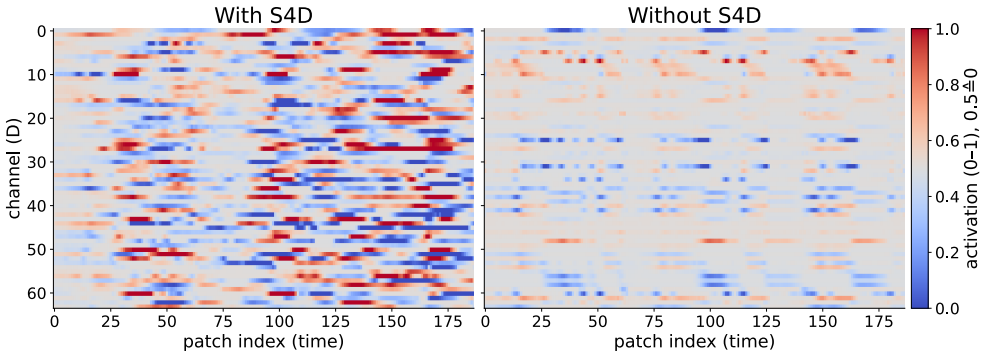


Fig. 6 Visualization of Activation Heatmaps. Top: M-S4DLite (with S4D) shows coherent temporal focus. Bottom: w/o S4D results in scattered noise, lacking global guidance.

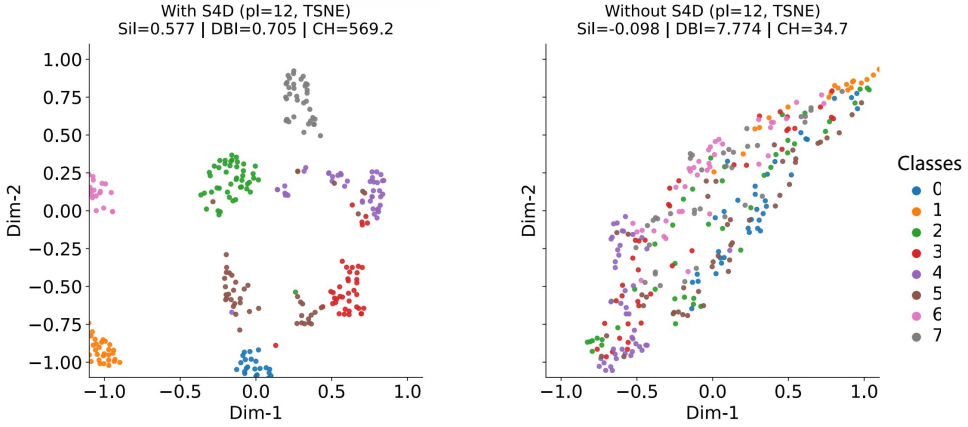


Fig. 7 t-SNE Embeddings on UWave. Left: M-S4DLite shows clear class separation. Right: Removing S4D leads to severe mixing, indicating a loss of global discriminative power.

the model fails to model inter-patch dependencies, reacting merely to instantaneous local variations without forming a coherent temporal representation.

Macroscopic Topology (t-SNE Embeddings): This difference in feature quality directly impacts the decision boundary, as evidenced by the t-SNE projections in Fig. 7.

- **Manifold Structure:** M-S4DLite produces well-defined, compact clusters with significant inter-class margins. The high Silhouette Score (0.577) quantitatively confirms this structural integrity.
- **Manifold Collapse:** Conversely, removing the S4D backbone leads to a severe "manifold collapse," indicated by the negative Silhouette Score (-0.098). As shown in the right panel of Fig. 7, data points degenerate into a diagonal smear where distinct classes are inextricably mixed.

This visual dichotomy proves that while Patching extracts the necessary local semantics, it is the S4D backbone that acts as a *topological organizer*. It maps these local tokens onto a linearly separable manifold by integrating them into global trajectories, thereby preventing the model from degenerating into a simple "Bag-of-Words" discriminator.

4 Conclusion

In this work, we addressed the fundamental "Semantic Misalignment" between continuous state-space models and discrete classification tasks. We identified that the spectral smoothness inherent in standard SSMs tends to erode high-frequency discriminative motifs, creating a barrier to high-precision classification. To resolve this, we proposed M-S4DLite, a framework grounded in the principle of Orthogonal Decoupling. By strategically synergizing discrete patching—which acts as a structural barrier against oversmoothing—with a strictly diagonal S4D backbone for global reasoning, we effectively re-calibrated the inductive bias of continuous models.

Our empirical results confirm that M-S4DLite establishes a new Pareto frontier in the accuracy-efficiency landscape. It demonstrates that high-performance classification does not require quadratic complexity or billion-scale parameters; rather, it requires the correct structural alignment between local granularity and global context. Moreover, the model exhibits exceptional cross-scale robustness, validating its ability to interpolate global trends from coarse-grained inputs without losing local precision. This makes M-S4DLite a viable solution for high-precision deployment on resource-constrained edge devices, bridging the gap between advanced theoretical modeling and practical industrial applications.

Future research will focus on extending this lightweight backbone to handle irregularly sampled data via continuous spectral interactions, and exploring its potential in self-supervised pre-training frameworks to leverage the growing availability of unlabeled time-series data.

Acknowledgments. We would like to express our sincere gratitude for the financial support provided by the projects listed in the Funding section, which enabled the successful completion of this research.

Declarations

- **Funding**
  This work was supported by the following projects: the Xinjiang Talent Development Fund (Grant No. XJRC-2025-KJ-PY-KJLJ-047); the Xinjiang Uygur Autonomous Region "Tianshan Talents" Leading Talent Project (No. 2023TSYCLJ0022); the Xinjiang Uygur Autonomous Region "Tianshan Talents" Innovation Team Project (No. 2023TSYCTD0011); the Distinguished Young Scientists Program of the Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2025D01E53); and the Science and Technology Development Program of the Silk Road Economic Belt Innovation-Driven Development Pilot Zone

and Urumqi-Changji-Shihezi National Independent Innovation Demonstration Zone (No. 2024LQ03002).

- **Conflict of interest**
  We confirm that there are no known conflicts of interest associated with this publication, and no significant financial support has been received that could have influenced the results. All authors have read and approved the manuscript, and no other individuals who meet the criteria for authorship are omitted. All authors agree on the author order of the manuscript.
- **Ethics approval**
  Not applicable.
- **Consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Data availability**
  The data used in this research are from publicly available datasets. The links to these datasets are provided below:
  - MIT-BIH ECG: https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data
  - UCI HAR: https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones
  - UWaveGestureLibraryAll: https://www.timeseriesclassification.com/description.php?Dataset=UWaveGestureLibraryAll
- **Code availability**
  The code implementation will be made publicly available upon acceptance.
- **Author contribution**
  Jian Chen: Writing – original draft, model design and implementation, data preprocessing, experimental design, code implementation, data analysis. **Yunfei Han**: Experimental design, conceptual guidance, framework optimization, model training and tuning. **Yi Wang**: Methodology, data curation, visualization, writing – review & editing. **Shiwei Guo**: Algorithm design, model tuning, experimental data analysis. **Yupeng Ma**: Writing – review & editing, supervision, methodology, model validation, research direction planning.
  The corresponding author is Yupeng Ma, who is responsible for communicating with the journal during the editorial process and coordinating with other authors on progress, submission of revisions, and final approval of proofs.

References

[1] Ismail Fawaz, H., Lucas, B., Forestier, G., *et al.*: Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery **34**, 1936–1962 (2020)

[2] Middlehurst, M., Large, J., Bagnall, A.: The canonical interval forest (cif) classifier for time series classification. In: 2020 IEEE International Conference on Big Data (Big Data), pp. 188–195 (2020). IEEE

[3] Shifaz, A., Pelletier, C., Petitjean, F., Webb, G.I.: Ts-chief: a scalable and accurate forest algorithm for time series classification. Data Mining and Knowledge Discovery **34**, 742–775 (2020)

[4] Luo, D., Wang, X.: Moderntcn: A modern pure convolution structure for general time series analysis. In: The Twelfth International Conference on Learning Representations (2024)

[5] Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., et al.: Time-llm: Time series forecasting by reprogramming large language models. arXiv preprint arXiv:2310.01728 (2023)

[6] Gao, S., Koker, T., Queen, O., Hartvigsen, T., Tsiligkaridis, T., Zitnik, M.: Units: Building a unified time series model. CoRR (2024)

[7] Zhou, J., Li, D., Li, L., Chen, Z., Wu, S., Ye, H., Lou, J., Spanos, C.J.: Enhancing llm reasoning for time series classification by tailored thinking and fused decision. arXiv preprint arXiv:2506.00807 (2025)

[8] Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J.Y., Zhou, J.: Timemixer: Decomposable multiscale mixing for time series forecasting. arXiv preprint arXiv:2405.14616 (2024)

[9] Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, pp. 11121–11128 (2023)

[10] Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)

[11] Gu, A., Goel, K., Gupta, A., Ré, C.: On the parameterization and initialization of diagonal state space models. Advances in Neural Information Processing Systems **35**, 35971–35983 (2022)

[12] Gu, A., Dao, T., Ermon, S., Rudra, A., Ré, C.: Hippo: Recurrent memory with optimal polynomial projections. Advances in Neural Information Processing Systems **33**, 1474–1487 (2020)

[13] Feng, J., Zhang, J., Cao, G., Liu, Z., Ding, Y.: Decmamba: Mamba utilizing series decomposition for multivariate time series forecasting. Computers, Materials & Continua **82**(1) (2025)

[14] Ismail-Fawaz, A., Devanne, M., Berretti, S., Weber, J., Forestier, G.: Look into the lite in deep learning for time series classification. International Journal of Data Science and Analytics, 1–21 (2025)

[15] Dempster, A., Petitjean, F., Webb, G.I.: Rocket: exceptionally fast and accurate

time series classification using random convolutional kernels. Data Mining and Knowledge Discovery **34**, 1454–1495 (2020)

[16] Chen, X., Qiu, P., Zhu, W., Li, H., Wang, H., Sotiras, A., Wang, Y., Razi, A.: Timemil: Advancing multivariate time series classification via a time-aware multiple instance learning. arXiv preprint arXiv:2405.03140 (2024)

[17] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)

[18] Karim, F., Majumdar, S., Darabi, H., Chen, S.: Lstm fully convolutional networks for time series classification. IEEE Access **6**, 1662–1669 (2017)

[19] Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. arXiv preprint arXiv:2211.14730 (2022)

[20] Chen, P., Zhang, Y., Cheng, Y., Shu, Y., Wang, Y., Wen, Q., Yang, B., Guo, C.: Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. arXiv preprint arXiv:2402.05956 (2024)

[21] Le, X.-M., Luo, L., Aickelin, U., Tran, M.-T.: Shapeformer: Shapelet transformer for multivariate time series classification. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1484–1494 (2024)

[22] Dao, T., Gu, A.: Transformers are ssms: Generalized models and efficient algorithms through structured state space duality. arXiv preprint arXiv:2405.21060 (2024)

[23] Liu, M., Zeng, A., Chen, M., Xu, Z., Lai, Q., Ma, L., Xu, Q.: Scinet: Time series modeling and forecasting with sample convolution and interaction. Advances in Neural Information Processing Systems **35**, 5816–5828 (2022)

[24] Li, Z., Qin, Y., Cheng, X., Tan, Y.: Ftmixer: Frequency and time domain representations fusion for time series modeling. arXiv preprint arXiv:2405.15256 (2024)

[25] Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.-H., Choo, J.: Reversible instance normalization for accurate time-series forecasting against distribution shift. In: International Conference on Learning Representations (2021)