MPTSNet: Integrating Multiscale Periodic Local Patterns and Global Dependencies for Multivariate Time Series Classification

Yang Mu[1], **Muhammad Shahzad**[1], **Xiao Xiang Zhu**[1, 2*]
[1]Technical University of Munich, Germany
[2]Munich Center for Machine Learning (MCML), Germany
{yang.mu, muhammad.shahzad, xiaoxiang.zhu}@tum.de

Abstract

Multivariate Time Series Classification (MTSC) is crucial in extensive practical applications, such as environmental monitoring, medical EEG analysis, and action recognition. Real-world time series datasets typically exhibit complex dynamics. To capture this complexity, RNN-based, CNN-based, Transformer-based, and hybrid models have been proposed. Unfortunately, current deep learning-based methods often neglect the simultaneous construction of local features and global dependencies at different time scales, lacking sufficient feature extraction capabilities to achieve satisfactory classification accuracy. To address these challenges, we propose a novel Multiscale Periodic Time Series Network (MPTSNet), which integrates multiscale local patterns and global correlations to fully exploit the inherent information in time series. Recognizing the multi-periodicity and complex variable correlations in time series, we use the Fourier transform to extract primary periods, enabling us to decompose data into multiscale periodic segments. Leveraging the inherent strengths of CNN and attention mechanism, we introduce the PeriodicBlock, which adaptively captures local patterns and global dependencies while offering enhanced interpretability through attention integration across different periodic scales. The experiments on UEA benchmark datasets demonstrate that the proposed MPTSNet outperforms 21 existing advanced baselines in the MTSC tasks.

Code — https://github.com/MUYang99/MPTSNet
Datasets — https://timeseriesclassification.com/dataset.php

Introduction

Multivariate time series classification (MTSC) has garnered significant attention due to its widespread applications across various domains. This classification technique plays a crucial role in analyzing complex, multi-dimensional temporal data, offering insights that are invaluable in fields ranging from healthcare (Wang et al. 2022; An et al. 2023), human activity recognition (Yang et al. 2015; Li et al. 2023a), traffic (Zhao et al. 2024) to environmental monitoring (Ienco and Gaetano 2007; Rußwurm et al. 2023) and industrial processes (Farahani et al. 2023).

*Corresponding author

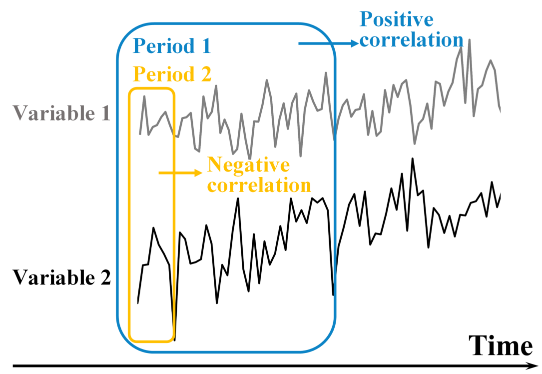

Figure 1: Varying correlations between the subsequences of two variables across different periodic scales in MTS data.

The challenges in solving MTSC tasks stem from two primary factors. Firstly, time series data inherently possesses periodic characteristics, presenting similar trends across different periods, which leads to data redundancy. Simultaneously, hidden information overlaps and interacts across multiple periods (Wu et al. 2023), complicating the exploration of time series data. Secondly, time series subsequences among different variables exhibit varying degrees of correlation or even opposite correlations at different periodic scales (Cai et al. 2024) in MTS data, as shown in Fig. 1. Thus, the extraction of local intra-period features and the capture of global inter-period dependencies at multiple periodic scales become paramount for effective MTSC.

Traditional approaches to time series classification, such as bag-of-patterns (Baydogan, Runger, and Tuv 2013) or shapelet-based methods (Lines et al. 2012), typically require a preprocessing step to transform time series into a wide range of subsequences or patterns as candidate features. Subsequently, discriminative subsequences are selected from these candidates for classification purposes. However, this approach often results in an expansive feature space, which not only complicates the feature selection process but may also lead to decreased accuracy (Schäfer and Leser 2017), particularly in multivariate settings where the complexity of data is inherently higher.

Recently, deep neural networks have demonstrated re-

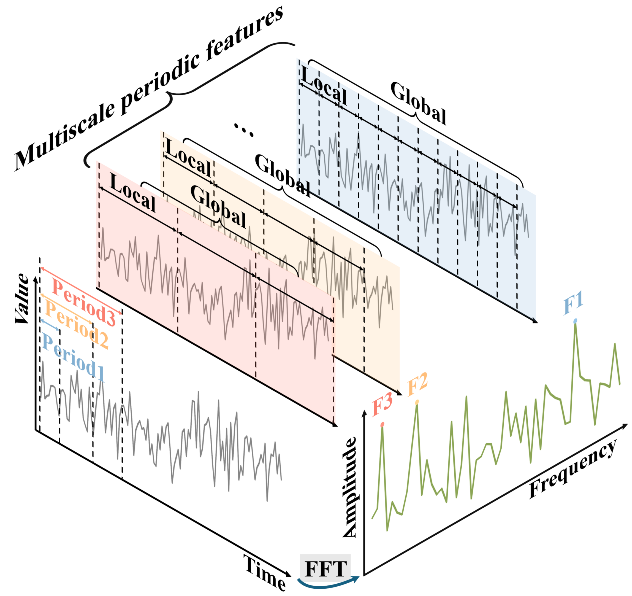

Figure 2: Multiscale periodic feature analysis of time series data using FFT. By decomposing time series into multiple frequency components, it reveals both local patterns and global dependencies across different periodic scales.

markable potential in various time series tasks, outperforming traditional methods in many aspects. RNN-based methods (Karim et al. 2017; Yu, Kim, and Mechefske 2021) utilize loops within their architecture to maintain and propagate information across time steps, but limit the ability to capture long-term dependencies due to the vanishing gradient problem. CNN-based methods (Cui, Chen, and Chen 2016; Zhao et al. 2017) excel at learning spatial hierarchies of features through convolutional filters, the inductive bias allows them to capture local patterns effectively, but they are less adept at modeling global features comprehensively. Transformer-based methods (Wen et al. 2022; Zuo et al. 2023) leverage self-attention mechanisms to model dependencies between different positions in the input sequence, adeptly handling long-range dependencies. However, they struggle to effectively extract features from adjacent time points that exhibit localized pattern characteristics. While these models have shown promise, they often fall short in simultaneously constructing local features and global dependencies within time series data across different periodic scales, resulting in suboptimal feature extraction.

To address these limitations, we propose a general framework called Multiscale Periodic Time Series Network (MPTSNet). The core concept, illustrated in Fig. 2, involves transforming time-domain data into frequency-domain representations using the Fast Fourier Transform (FFT). This approach decomposes complex time series into multiple frequency components, revealing both local intra-period patterns and global inter-period dependencies across various periodic scales. The architecture of MPTSNet is supported by the PeriodicBlock, which decomposes time series into multiscale periodic segments. These segments are then processed by two key components sequentially: an Inception-based Local Extractor, which captures corresponding local features, and an Attention-based Global Capturer, which extracts global dependencies. This modular architecture enables MPTSNet to effectively analyze time series data at multiple periodic scales. Our contributions can be summarized in three key aspects:

- We propose an end-to-end general framework MPTSNet that addresses the multi-periodicity of time series and complex correlations among variable subsequences at different time scales. This novel architecture is capable of extracting multiscale periodic features effectively.
- Inspired by the strengths of CNN in capturing local features and attention mechanism in modeling global dependencies, we design a PeriodicBlock comprising an Inception-based Local Extractor and an Attention-based Global Capturer to extract both local features and global dependencies from time series data.
- Our extensive experiments on real-world datasets demonstrate that the proposed MPTSNet outperforms existing advanced baselines in MTSC tasks. Furthermore, MPTSNet offers enhanced interpretability through its ability to localize temporal patterns with integrated global attention across different periodic scales.

Related Work

Multivariate Time Series Classification

Recent years have witnessed the emergence of various deep learning models for MTSC. These can be broadly categorized into three main groups: 1) CNN-based models. The Multichannel Deep Convolutional Neural Network (MCDCNN) (Zheng et al. 2014) pioneered the application of CNNs to MTSC by capturing intra-variable features through one-dimensional convolutions and combining them with fully connected layers. More recently, OS-CNN (Tang et al. 2022) has been proposed, featuring an innovative Omni-Scale block that uses a set of prime numbers as convolution kernel sizes to effectively cover receptive fields across all scales, optimizing performance across different datasets. 2) RNN/CNN Hybrid models. Models such as LSTM-FCN (Karim et al. 2017) and MLSTM-FCN (Karim et al. 2019) have been introduced, combining LSTM layers to capture short- and long-term dependencies with stacked CNN layers to extract features from the time series. 3) Transformer-based models. These models have gained prominence in *General Time Series Analysis*, addressing multiple tasks including classification, imputation, short-term and long-term forecasting, and anomaly detection (Chen et al. 2024). Notable examples include FEDformer (Zhou et al. 2022), Flowformer (Wu et al. 2022), PatchTST (Nie et al. 2023), and Crossformer (Zhang and Yan 2023), which have been developed and refined in recent years, leveraging their excellent scaling behaviors. Additionally, MLP-based (Zhang et al. 2022; Li et al. 2023b), GNN-based (Liu et al. 2024) and MIL-based (Chen et al. 2024) models have also achieved impressive performance. Since MIL-based models employ a binary paradigm for multi-class problems, unlike other approaches, they are excluded from quantitative comparisons.

However, a key assumption in many of these models is that the correlations between different variables remain constant across various time resolutions. This assumption often leads to inadequate representation of pairwise relationships between variables at different periodic scales.

Local and Global Modeling for Sequence

Modeling local patterns and global dependencies has proven effective for correlation learning and feature extraction in sequence modeling (Wang et al. 2023). In speech recognition, the SOTA model Interformer (Lai et al. 2023) employs convolution to extract local information and self-attention to capture long-term dependencies in parallel branches. However, this parallel processing of local and global features might introduce redundancy, potentially reducing efficiency. In general time series analysis, TimesNet (Wu et al. 2023) transforms 1D time series into various 2D tensors using CNN-based layers to capture both local and global variations. Directly capturing local and global features with CNNs may lead to a loss in precision. Additionally, while TimesNet provides visualizations of the 2D transformations, interpreting these in the context of the original 1D time series can be challenging. In time series forecasting, MSGNet (Cai et al. 2024) employs a graph convolution module for inter-series correlation learning and a multi-head attention module for intra-series correlation learning. However, it introduces increased complexity and computational cost due to its adaptive graph convolution module; MICN (Wang et al. 2023) combines local feature extraction and global correlation modeling using a convolution-based structure; however, its performance can be highly sensitive to the choice of hyperparameters, such as scale sizes in the multi-scale convolution, requiring meticulous tuning for optimal results. Moreover, the use of full CNN layers in MICN sacrifices the interpretability.

To address these existing limitations, we propose MPTSNet, which sequentially applies convolutional and attention modules to capture local and global features. By requiring only the number of extraction scales, other parameters are adaptively configured based on time series characteristics, reducing sensitivity to parameter tuning and avoiding redundant computations. Moreover, this approach enhances the interpretability of MTSC by integrating multi-scale attention maps.

Methodology

Problem Formulation

MTSC involves the analysis of a set of time series data, where each series is composed of multiple variables observed over time. A MTS data is formally defined as $X = \{x_1, x_2, \ldots, x_d\} \in \mathbb{R}^{d \times l}$, where d is the number of variables and l is the length of the series. In the context of MTSC, the goal is to classify these time series into predefined categories. Given a dataset $\mathcal{X} = \{X_1, X_2, \ldots, X_m\} \in \mathbb{R}^{m \times d \times l}$ with corresponding labels $\eta = \{y_1, y_2, \ldots, y_m\}$, where m denotes the number of time series, the task is to train a classifier $f(\cdot)$ capable of predicting the label y for new, unlabeled MTS data.

Model Architecture Overview

The overview of our proposed MPTSNet is shown in Fig. 3, which is designed to extract local patterns and global dependencies at various periodic scales. Initially, the model identifies the main periods from the entire training dataset using FFT. The core component, PeriodicBlock, processes time series data in parallel. It begins by segmenting batch data into multiple scales based on the identified main periods. Subsequently, the Local Extractor and Global Capturer are employed within the PeriodicBlock to extract intra-period local and inter-period global features sequentially, as depicted in the right panel of Fig. 3. To account for potential discrepancies in frequency-amplitude relationships between batch data and the overall training set, FFT is also applied to each batch. This process queries the amplitudes corresponding to the main periods, which are then used as weights calculated by Softmax to aggregate multiscale periodic features. The PeriodicBlock outputs a weighted sum of features extracted from different periodic scales. Residual connections between stacked PeriodicBlocks enhance information flow. Finally, the extracted features are fed into the Classification head to generate predicted labels.

Main Periods Identification

Grounded in the inherent characteristics of time series data, we aim to enhance MTSC accuracy by exploiting two fundamental properties: multi-periodicity and complex correlations among variables across different periodic scales, as mentioned before. The identification of main periods is a critical step in our methodology, as it serves as the foundation for subsequent multiscale analysis.

Inspired by the work (Wu et al. 2023), we employ the FFT to identify the principal periods. This technique effectively converts time-domain data into the frequency domain, allowing for a comprehensive analysis of the underlying periodic structures. The process can be formalized as follows:

$$\mathbf{AMP} = \mathrm{Avg}(\mathrm{Amplitudes}(\mathrm{FFT}(X \in \mathbb{R}^{d \times l}))), \tag{1}$$

where $X \in \mathbb{R}^{d \times l}$ represents the input time series, d denotes the number of variables, and l is the sequence length. The function $\mathrm{FFT}(\cdot)$ computes the Fast Fourier Transform, while $\mathrm{Amplitudes}(\cdot)$ calculates the amplitude values. $\mathrm{Avg}(\cdot)$ averages the amplitudes across all variables, resulting in $\mathbf{AMP} \in \mathbb{R}^l$, which represents the mean amplitude for each frequency.

$$f_1, \ldots, f_k = \underset{f_* \in \{1, \ldots, \lfloor \frac{T}{2} \rfloor\}}{\arg\mathrm{Topk}} (\mathbf{AMP}), \tag{2}$$

$$p_i = l/f_i, \quad i \in \{1, \ldots, k\}, \tag{3}$$

To focus on the most significant periodic components, we select the top k amplitudes using the $\arg\mathrm{Topk}$ function. This operation yields the frequencies $f_1, \ldots, f_k$, which correspond to the k most prominent periods $p_1, \ldots, p_k$. The parameter k is a hyperparameter that determines the number of periodic scales to consider in subsequent analyses.

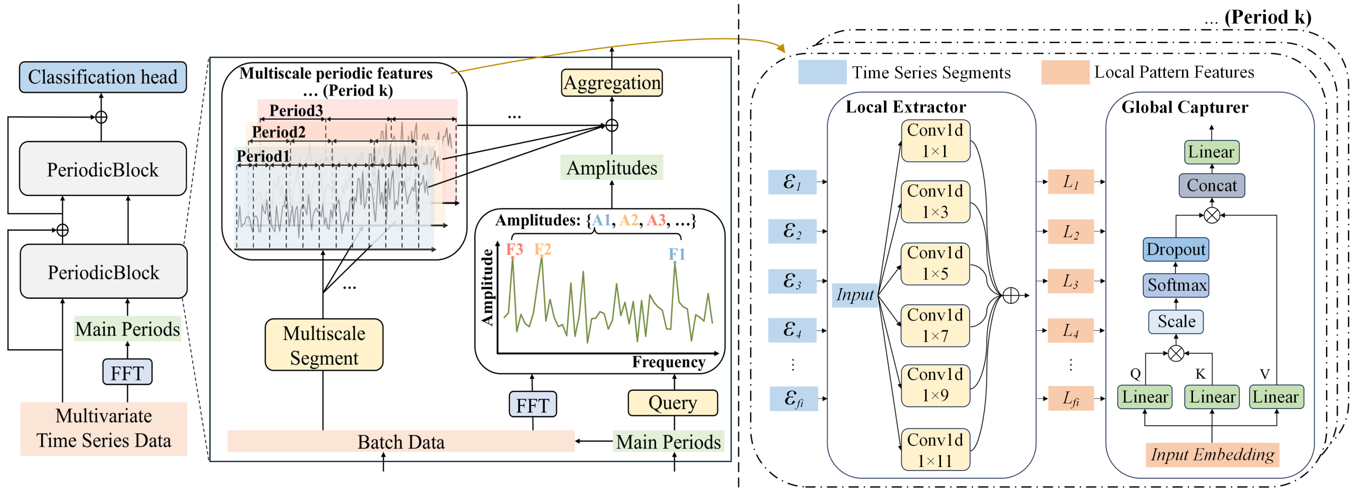

Figure 3: Overview of the Multiscale Periodic Time Series Network (MPTSNet). The model processes multivariate time series data through FFT to identify the main periods. The data is then segmented into multiscale periodic components, where Local Extractor and Global Capturer in PeriodicBlock extract features sequentially at different scales. The features are aggregated by corresponding amplitudes and passed through the Classification head for final prediction.

The identification of these main periods establishes the groundwork for multiscale analysis. This approach allows us to decompose the original time series into multiple periodic components, and such decomposition is crucial for understanding the complex temporal dynamics present in MTS data.

The extracted periods serve as the basis for reshaping the input data into multiple representations, each corresponding to a different time scale. This transformation can be expressed as:

$$X'^{i} = \text{Reshape}_{p_i, f_i}(\text{Pad}(X)), \quad i \in \{1, \dots, k\}. \tag{4}$$

where $\text{Pad}(\cdot)$ extends the time series with zeros to ensure compatibility with the reshaping operation, and $X'^{i} \in \mathbb{R}^{d \times p_i \times f_i}$ represents the i-th reshaped time series based on the i-th periodic scale.

PeriodicBlock

The PeriodicBlock serves as the core component of our MPTSNet architecture, designed to extract both local and global features from the reshaped time series data at different periodic scales. Prior to processing by the PeriodicBlock, each reshaped tensor $X'^{i} \in \mathbb{R}^{d \times p_i \times f_i}$ is embedded into a higher-dimensional space:

$$\mathcal{E}'^{i} = \text{Embedding}(X'^{i}), \quad i \in \{1, \dots, k\}. \tag{5}$$

where $\mathcal{E}'^{i} \in \mathbb{R}^{d_{\text{embed}} \times p_i \times f_i}$ and d_{embed} is the embedding dimension.

Local Extractor To capture specific and comprehensive intra-period local patterns, we propose a multiscale adaptive convolution module inspired by the Inception architecture. For each of the f_i embeddings $\mathcal{E}^{i} \in \mathbb{R}^{d_{\text{embed}} \times p_i}$, this module comprises parallel 1D convolutional layers with varying kernel sizes:

$$L_j^{i,ker} = \text{Conv1D}_{ker}(\mathcal{E}^{i}), \quad ker \in \{1, 3, 5, 7, 9, 11\}, \tag{6}$$

where Conv1D_{ker} represents a 1D convolutional layer with kernel size ker. The outputs of these convolutions are concatenated along the channel dimension:

$$L_j^{i} = \text{Mean}(L_j^{i,ker}), \quad j \in \{1, \dots, f_i\}. \tag{7}$$

This multi-kernel size Inception-based approach allows the model to capture local patterns at various granularities, enhancing its ability to detect relevant features.

Global Capturer To model inter-period dependencies and adapt to different time steps for multiscale periods, we designed a Global Capturer based on the multi-head attention mechanism. The Global Capturer transforms each local feature L^{i} into query, key, and value representations:

$$Q^{i} = W_Q L^{i}, \quad K^{i} = W_K L^{i}, \quad V^{i} = W_V L^{i}, \tag{8}$$

where W_Q, W_K, and W_V are learnable weight matrices. The multi-head attention is then computed as:

$$\text{Attention}(Q^{i}, K^{i}, V^{i}) = \text{Softmax}\left(\frac{Q^{i}(K^{i})^{T}}{\sqrt{d_k}}\right) V^{i}, \tag{9}$$

$$G^{i} = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^{O}. \tag{10}$$

where h is the number of attention heads, d_k is the dimension of each head, and W^{O} is a learnable output projection matrix. This mechanism allows the model to capture complex global dependencies across all time series segments.

Adaptive Aggregation and Classification To fuse the k multiscale periodic features, we employ an adaptive aggregation using weights calculated from the queried amplitudes:

Data/Model	LSTNet *SIG.'18*	LSSL *ICLR'22*	FEDf. *ICLR'22*	Flowf. *ICLR'22*	SCINet *NIPS'22*	Dlinear *AAAI'23*	PatchTST *ICLR'23*	MICN *ICLR'23*	TimesNet *ICLR'23*	Crossf. *ICLR'23*	M.TCN *ICLR'24*	Ours
EthanolConcentration	<u>39.9</u>	31.1	31.2	33.8	34.4	36.2	32.8	35.3	35.7	38.0	36.3	**43.3**
FaceDetection	65.7	66.7	66.0	67.6	68.9	68.0	68.3	65.2	68.6	68.7	**70.8**	<u>69.8</u>
Handwriting	25.8	24.6	28.0	<u>33.8</u>	23.6	27.0	29.6	25.5	32.1	28.8	30.6	**34.4**
Heartbeat	77.1	72.7	73.7	<u>77.6</u>	77.5	75.1	74.9	74.7	**78.0**	<u>77.6</u>	77.2	75.6
JapaneseVowels	98.1	98.4	98.4	<u>98.9</u>	96.0	96.2	97.5	94.6	98.4	**99.1**	98.8	98.6
PEMS-SF	86.7	86.1	80.9	86.0	83.8	75.1	89.3	85.5	<u>89.6</u>	85.9	89.1	**94.2**
SelfRegulationSCP1	84.0	90.8	88.7	92.5	92.5	87.3	90.7	86.0	91.8	92.1	**93.4**	<u>92.8</u>
SelfRegulationSCP2	52.8	52.2	54.4	56.1	57.2	50.5	57.8	53.6	57.2	<u>58.3</u>	**60.3**	57.2
SpokenArabicDigits	**100**	**100**	**100**	98.8	98.1	81.4	98.3	97.1	99.0	97.9	98.7	<u>99.5</u>
UWaveGestureLibrary	<u>87.8</u>	85.9	85.3	86.6	85.1	82.1	85.8	82.8	85.3	85.3	86.7	**88.1**
Ours 1-to-1-Wins	3	4	9	3	3	8	1	10	5	7	5	-
Ours 1-to-1-Draws	1	2	0	1	0	0	0	0	1	0	0	-
Ours 1-to-1-Losses	6	4	1	6	7	2	9	0	4	3	5	-
Avg. accuracy (↑)	71.8	70.9	70.7	73.2	71.7	67.9	72.5	70.0	73.6	73.2	<u>74.2</u>	**75.4**
Avg. rank (↓)	6	7.5	7.6	4.6	6.7	8.6	6.1	9.2	4.1	4.5	<u>3</u>	**2.4**

Table 1: Setting 1 experimental results. Performance comparison with the recent advanced *General Time Series Analysis* frameworks on 10 UEA datasets. The best results are highlighted in *bold*, while the second best are *underlined*.

$$\alpha_i = \text{Softmax}(\mathbf{AMP}_i), \quad i \in \{1, \ldots, k\}, \tag{11}$$

where $\mathbf{AMP}_i$ represents the amplitude corresponding to the i-th selected frequency. The aggregated feature representation is then computed as:

$$Z = \sum_{i=1}^{k} \alpha_i G^i. \tag{12}$$

This adaptive weighting scheme ensures that the model emphasizes the most relevant periodic components in the final representation.

The aggregated features Z are then passed through the Classification head, typically consisting of fully connected layers followed by a softmax activation:

$$y = \text{Softmax}(W_c Z + b_c), \tag{13}$$

where W_c and b_c are the weight matrix and bias vector of the classification layer, respectively, and y is the predicted class probability distribution.

By combining the Local Extractor for intra-period local pattern extraction, the Global Capturer for modeling inter-period global dependencies, and the adaptive aggregation mechanism, the PeriodicBlock effectively captures the complex temporal dynamics present in multivariate time series data across multiple periodic scales.

Experiments

Datasets

The public UEA benchmark datasets (Bagnall et al. 2018), collected from various real-world applications, constitute a comprehensive archive for multivariate time series classification across several domains, including human activity, speech, medical EEG, and audio data, etc. We utilize the UEA benchmark datasets to evaluate the performance of the proposed MPTSNet. These datasets vary in length, dimensions, and the size of their training/testing sets.

Baselines

Setting 1 Experiments Based on the experimental results of papers (Luo and Wang 2024; Wu et al. 2023), this set of experiments is conducted on 10 UEA datasets and compared with several recent advanced *General Time Series Analysis* frameworks, which address tasks including classification, imputation, forecasting, and anomaly detection. The comparison includes ModernTCN (Luo and Wang 2024), Crossformer (Zhang and Yan 2023), TimesNet (Wu et al. 2023), MICN (Wang et al. 2023), PatchTST (Nie et al. 2023), Dlinear (Zeng et al. 2023), SCINet (Liu et al. 2022), Flowformer (Wu et al. 2022), FEDformer (Zhou et al. 2022), LSSL (Gu, Goel, and Ré 2021), and LSTNet (Lai et al. 2018).

Setting 2 Experiments Following the results of papers (Liu et al. 2024; Li et al. 2021), this set of experiments is conducted on 25 UEA datasets and compared with recent advanced MTSC-dedicated models, including TodyNet (Liu et al. 2024), OS-CNN and MOS-CNN (Tang et al. 2022), ShapeNet (Li et al. 2021), TapNet (Zhang et al. 2020), MLSTM-FCN (Karim et al. 2019), and WEASEL+MUSE (Schäfer and Leser 2017). The traditional methods EDI, DTWI, and DTWD based on Euclidean Distance, dynamic time warping, and the nearest neighbor classifier are also included.

Implementation Details

The benchmark results for all baseline methods are sourced from their respective publications, ensuring consistent training parameters across comparisons. Our proposed model is implemented using a computational infrastructure consisting of a server running Ubuntu 20.04.3 LTS, equipped with 8 NVIDIA GeForce RTX 3090 GPUs. The performance is evaluated by computing the accuracy, 1-to-1 Wins/Draws/Losses, average accuracy, and average rank.

Data/Model	EDI	DTWI	DTWD	W.+MUSE *arxiv'17*	M.-FCN *Neur.'19*	TapNet *AAAI'20*	ShapeNet *AAAI'21*	OS-CNN *ICLR'22*	MOS-CNN *ICLR'22*	TodyNet *Info.'24*	Ours
ArticularyWordRecognition	97.0	98.0	98.7	99.0	97.3	98.7	98.7	98.8	**99.1**	98.7	97.7
AtrialFibrillation	26.7	26.7	20.0	33.3	26.7	33.3	40.0	23.3	18.3	46.7	**53.3**
BasicMotions	67.5	**100**	97.5	**100**	95.0	**100**	**100**	**100**	**100**	**100**	**100**
Cricket	94.4	98.6	**100**	**100**	91.7	95.8	98.6	99.3	99.0	**100**	94.4
DuckDuckGeese	27.5	55.0	60.0	57.5	67.5	57.5	**72.5**	54.0	61.5	58.0	68.0
Epilepsy	66.7	97.8	96.4	**100**	76.1	97.1	98.7	98.0	99.6	97.1	97.1
EthanolConcentration	29.3	30.4	32.3	13.3	37.3	32.3	31.2	24.0	41.5	35.0	**43.3**
ERing	13.3	13.3	13.3	43.0	13.3	13.3	13.3	88.1	91.5	91.5	**94.4**
FaceDetection	51.9	51.3	52.9	54.5	54.5	55.6	60.2	57.5	59.7	62.7	**69.8**
FingerMovements	55.0	52.0	53.0	49.0	58.0	53.0	58.9	56.8	56.8	**67.6**	64.0
HandMovementDirection	27.9	30.6	23.1	36.5	36.5	37.8	33.8	44.3	36.1	**64.9**	63.5
Handwriting	37.1	50.9	60.7	60.5	28.6	35.7	45.1	**66.8**	**67.7**	43.6	34.4
Heartbeat	62.0	65.9	71.7	72.7	66.3	75.1	**75.6**	48.9	60.4	**75.6**	**75.6**
Libras	83.3	89.4	87.2	87.8	85.6	85.0	85.6	95.0	**96.5**	85.0	87.2
LSST	45.6	57.5	55.1	59.0	37.3	56.8	59.0	41.3	52.1	**61.5**	60.4
MotorImagery	51.0	39.0	50.0	50.0	51.0	59.0	61.0	53.5	51.5	64.0	**65.0**
NATOPS	86.0	85.0	88.3	87.0	88.9	93.9	88.3	96.8	95.1	**97.2**	94.4
PenDigits	97.3	93.9	97.7	94.8	97.8	98.0	97.7	98.5	98.3	98.7	**98.9**
PEMS-SF	70.5	73.4	71.1	N/A	69.9	75.1	75.1	76.0	76.4	78.0	**94.2**
PhonemeSpectra	10.4	15.1	15.1	19.0	11.0	17.5	29.8	29.9	29.5	**30.9**	14.4
RacketSports	86.8	84.2	80.3	**93.4**	80.3	86.8	88.2	87.7	92.9	80.3	87.5
SelfRegulationSCP1	77.1	76.5	77.5	71.0	87.4	65.2	78.2	83.5	82.9	89.8	**92.8**
SelfRegulationSCP2	48.3	53.3	53.9	46.0	47.2	55.0	**57.8**	53.2	51.0	55.0	57.2
StandWalkJump	20.0	33.3	20.0	33.3	6.7	40.0	**53.3**	38.3	38.3	46.7	**53.3**
UWaveGestureLibrary	88.1	86.9	90.3	91.6	89.1	89.4	90.6	**92.7**	92.6	85.0	88.1
Ours 1-to-1-Wins	15	13	14	10	19	18	12	12	15	14	-
Ours 1-to-1-Draws	2	2	2	2	1	4	1	3	2	3	-
Ours 1-to-1-Losses	8	10	9	13	5	3	12	10	8	8	-
Avg. accuracy (↑)	56.8	62.3	62.6	62.1	60.0	64.3	67.6	68.2	69.9	72.6	**74.0**
Avg. rank (↓)	7.52	6.48	5.8	5.36	6.4	5.24	3.84	4.28	3.8	3.16	**3.12**

Table 2: Setting 2 experimental results. Performance comparison with the recent advanced MTSC-dedicated models on 25 UEA datasets. In the table, 'N/A' indicates that the results for the corresponding method could not be obtained due to memory or computational limitations (Liu et al. 2024).

Experimental Results

Setting 1 Results Table 1 presents the performance comparison of our proposed method with 11 advanced *General Time Series Analysis* frameworks across 10 diverse MTSC datasets. Our approach demonstrates superior performance, achieving an average accuracy of 75.4%, and ranks first in terms of average rank, outperforming all competitors. A detailed analysis reveals that our method consistently excels on challenging datasets. For instance, on the EthanolConcentration and PEMS-SF datasets, known for their complex patterns and high dimensionality, our method significantly outperforms other approaches. In direct comparison to the recently state-of-the-art ModernTCN, our method achieves a substantial improvement of 1.2% in average accuracy and 0.6 in average rank, underscoring its effectiveness in handling complex multivariate time series data.

Setting 2 Results Our method consistently demonstrates superior performance across the 25 UEA datasets compared to 10 advanced MTSC-dedicated models in Table 2. It secures a solid win-loss record, achieving the highest average accuracy of 74% and the lowest average rank of 3.12. Notably, our approach surpasses competitors in eleven of the twenty-five datasets. Moreover, it demonstrates strong generalization capabilities by yielding competitive results across diverse domains, including human activity recognition, healthcare, and speech recognition.

Ablation Studies

Model Design Variants We conducted ablation studies on MPTSNet with three variants across 10 datasets in Setting 1:

- **w/o-LocalExt**: Removes the Local Extractor, capturing only global dependencies at each time point.
- **w/o-GlobalCap**: Removes the Global Capturer, relying solely on local patterns for each segment.
- **w/o-MP**: Omitted the multi-scale FFT decomposition, applying Local Extractor and Global Capturer to the entire time series instead.

The results of our ablation studies shown in Table 3 provide significant insights into the contributions of different components of the MPTSNet model. The full MPTSNet model achieved the highest average accuracy of 0.754, demonstrating the effectiveness of integrating local pattern

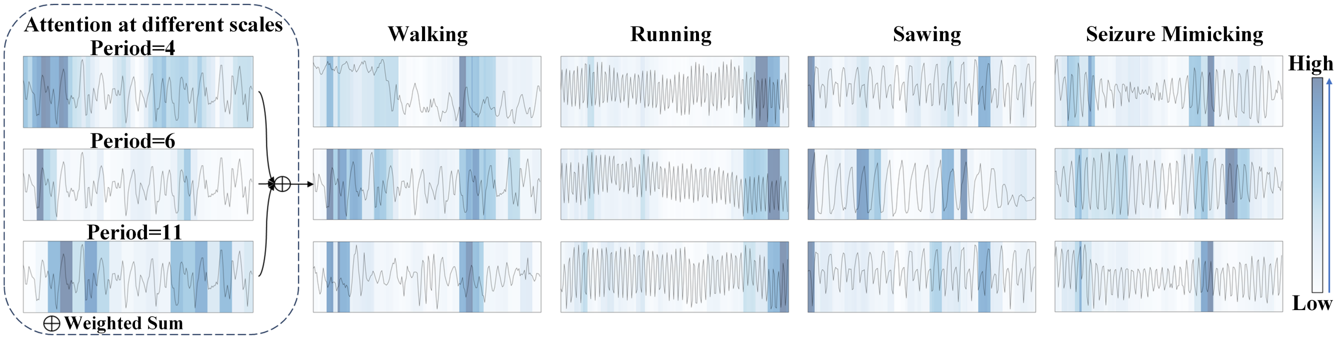

Figure 4: Interpretability visualization of the MPTSNet model on the Epilepsy dataset. Left: attention maps at different periodic scales (single example); Right: composite attention maps by weighted summing the individual scale maps. MPTSNet captures varying local patterns across periodic scales while achieving enhanced interpretability by the integration of multi-scale attention.

extraction, global dependency modeling, and multiscale periodic decomposition. Removing the Local Extractor (w/o-LocalExt) and the Global Capturer (w/o-GlobalCap) resulted in reduced accuracies of 0.737 and 0.731, respectively, highlighting their critical roles in capturing local and global features. The most significant performance drop was observed without Multiscale Periodic decomposition (w/o-MP), with accuracy falling to 0.709, underscoring the importance of multiscale periodic feature analysis.

Model Design Variants	Accuracy	Setting of k	Accuracy
MPTSNet	**0.754**	k=7	0.751
w/o-LocalExt	0.737	k=5	**0.754**
w/o-GlobalCap	0.731	k=3	0.742
w/o-MP	0.709	k=1	0.734

Table 3: The ablation study was conducted on 10 UEA datasets within Setting 1 experiments. Average accuracies are reported, with the best performance indicated in *bold*.

Setting of *k* The hyperparameter k is the sole parameter that needs to be set in MPTSNet to determine the number of periodic scales. Experimental results in Table 3 show a consistent improvement in average accuracy as k increases from 1 (0.734) to 5 (0.754), followed by a slight decline at k=7 (0.751). This trend suggests that incorporating multiple periodic scales enhances the model capacity to capture complex temporal patterns, with k=5 providing an optimal balance between information extraction and noise avoidance. The marginal performance decrease at k=7 indicates that excessive periodic scale extraction may introduce noise, particularly for datasets with less pronounced periodicity. These findings underscore the importance of judicious k selection, balancing the trade-off between capturing sufficient periodic information and mitigating the risk of overfitting. While k=5 emerges as the optimal value for the diverse datasets examined, further research into adaptive k-selection methods could potentially enhance the generalization capabilities of MPTSNet across varied time series data.

Enhanced Interpretability

Our proposed MPTSNet model offers a unique insight into the decision-making process through its interpretability visualization. By integrating global attention across various periodic scales and weighting them accordingly, we can pinpoint the specific time intervals that contribute most significantly to the classification of each time series. As an example, we illustrate the visualization of the attention maps using one of the datasets, the Epilepsy dataset in Fig. 4. This dataset, sourced from the UEA archive, encompasses wrist motion recordings from multiple participants engaged in four distinct activities: Walking, Running, Sawing, and Seizure Mimicking. Our analysis is conducted on three randomly selected samples from each class.

The left portion of the figure presents the attention maps from different periodic scales, which clearly demonstrate that the model attention is drawn to varying temporal locations across different periodic scales. To obtain a holistic view, we computed composite attention maps by combining individual scale attention maps, weighted by their corresponding amplitudes. These composite maps, presented in the right panel, are juxtaposed with the original signals. The resulting visualizations reveal distinct patterns for different classes while maintaining similarities within the same class. This observation underscores the ability of MPTSNet to differentiate between various time series classes by identifying and exploiting key temporal patterns and extracting meaningful representations.

Conclusion

To improve the performance of multivariate time series classification, we introduce MPTSNet, a novel deep learning framework. Our approach incorporates periodicity as a fundamental time scale and leverages the complementary strengths of CNN and attention mechanisms to thoroughly learn multiscale local and global features within time series data. Extensive experiments demonstrate the effectiveness of MPTSNet in various real-world applications. By visualizing attention maps, we provide insights into how MPTSNet identifies crucial temporal patterns, enabling more interpretable and reliable classification results.

Acknowledgments

The work is jointly supported by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (grant number: 01DD20001), by German Federal Ministry for Economic Affairs and Climate Action in the framework of the "national center of excellence ML4Earth" (grant number: 50EE2201C), by the European Union's Horizon Europe research and innovations actions programme in the framework of the "Multi-source and Multi-scale Earth observation and Novel Machine Learning Methods for Mineral Exploration and Mine Site Monitoring (MultiMiner)" (grant number: 101091374), by the Excellence Strategy of the Federal Government and the Länder through the TUM Innovation Network EarthCare and by Munich Center for Machine Learning.

References

An, Q.; Rahman, S.; Zhou, J.; and Kang, J. J. 2023. A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. *Sensors*, 23(9): 4178.

Bagnall, A.; Dau, H. A.; Lines, J.; Flynn, M.; Large, J.; Bostrom, A.; Southam, P.; and Keogh, E. 2018. The UEA multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*.

Baydogan, M. G.; Runger, G.; and Tuv, E. 2013. A bag-of-features framework to classify time series. *IEEE transactions on pattern analysis and machine intelligence*, 35(11): 2796–2802.

Cai, W.; Liang, Y.; Liu, X.; Feng, J.; and Wu, Y. 2024. Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 11141–11149.

Chen, X.; Qiu, P.; Zhu, W.; Li, H.; Wang, H.; Sotiras, A.; Wang, Y.; and Razi, A. 2024. TimeMIL: Advancing Multivariate Time Series Classification via a Time-aware Multiple Instance Learning. *arXiv preprint arXiv:2405.03140*.

Cui, Z.; Chen, W.; and Chen, Y. 2016. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*.

Farahani, M. A.; McCormick, M.; Gianinny, R.; Hudacheck, F.; Harik, R.; Liu, Z.; and Wuest, T. 2023. Time-series pattern recognition in Smart Manufacturing Systems: A literature review and ontology. *Journal of Manufacturing Systems*, 69: 208–241.

Gu, A.; Goel, K.; and Ré, C. 2021. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*.

Ienco, D.; and Gaetano, R. 2007. Tiselac: time series land cover classification challenge. *TiSeLaC: Time Series Land Cover Classification Challenge*, 2.

Karim, F.; Majumdar, S.; Darabi, H.; and Chen, S. 2017. LSTM fully convolutional networks for time series classification. *IEEE access*, 6: 1662–1669.

Karim, F.; Majumdar, S.; Darabi, H.; and Harford, S. 2019. Multivariate LSTM-FCNs for time series classification. *Neural networks*, 116: 237–245.

Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, 95–104.

Lai, Z.-H.; Zhang, T.-H.; Liu, Q.; Qian, X.; Wei, L.-F.; Chen, S.-L.; Chen, F.; and Yin, X.-C. 2023. InterFormer: Interactive Local and Global Features Fusion for Automatic Speech Recognition. *arXiv preprint arXiv:2305.16342*.

Li, G.; Choi, B.; Xu, J.; Bhowmick, S. S.; Chun, K.-P.; and Wong, G. L.-H. 2021. Shapenet: A shapelet-neural network approach for multivariate time series classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 8375–8383.

Li, Y.; Yang, G.; Su, Z.; Li, S.; and Wang, Y. 2023a. Human activity recognition based on multienvironment sensor data. *Information Fusion*, 91: 47–63.

Li, Z.; Rao, Z.; Pan, L.; and Xu, Z. 2023b. Mts-mixers: Multivariate time series forecasting via factorized temporal and channel mixing. *arXiv preprint arXiv:2302.04501*.

Lines, J.; Davis, L. M.; Hills, J.; and Bagnall, A. 2012. A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 289–297.

Liu, H.; Yang, D.; Liu, X.; Chen, X.; Liang, Z.; Wang, H.; Cui, Y.; and Gu, J. 2024. Todynet: temporal dynamic graph neural network for multivariate time series classification. *Information Sciences*, 120914.

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

Luo, D.; and Wang, X. 2024. Moderntcn: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.

Rußwurm, M.; Courty, N.; Emonet, R.; Lefèvre, S.; Tuia, D.; and Tavenard, R. 2023. End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196: 445–456.

Schäfer, P.; and Leser, U. 2017. Multivariate time series classification with WEASEL+ MUSE. *arXiv preprint arXiv:1711.11343*.

Tang, W.; Long, G.; Liu, L.; Zhou, T.; Blumenstein, M.; and Jiang, J. 2022. Omni-Scale CNNs: a simple and effective kernel size configuration for time series classification. In *International Conference on Learning Representations*.

Wang, H.; Peng, J.; Huang, F.; Wang, J.; Chen, J.; and Xiao, Y. 2023. Micn: Multi-scale local and global context modeling for long-term series forecasting. In *The eleventh international conference on learning representations*.

Wang, W. K.; Chen, I.; Hershkovich, L.; Yang, J.; Shetty, A.; Singh, G.; Jiang, Y.; Kotla, A.; Shang, J. Z.; Yerrabelli, R.; et al. 2022. A systematic review of time series classification techniques used in biomedical applications. *Sensors*, 22(20): 8016.

Wen, Q.; Zhou, T.; Zhang, C.; Chen, W.; Ma, Z.; Yan, J.; and Sun, L. 2022. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2023. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *The Eleventh International Conference on Learning Representations*.

Wu, H.; Wu, J.; Xu, J.; Wang, J.; and Long, M. 2022. Flowformer: Linearizing transformers with conservation flows. *arXiv preprint arXiv:2202.06258*.

Yang, J.; Nguyen, M. N.; San, P. P.; Li, X.; and Krishnaswamy, S. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Ijcai*, volume 15, 3995–4001. Buenos Aires, Argentina.

Yu, W.; Kim, I. Y.; and Mechefske, C. 2021. Analysis of different RNN autoencoder variants for time series classification and machine prognostics. *Mechanical Systems and Signal Processing*, 149: 107322.

Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11121–11128.

Zhang, T.; Zhang, Y.; Cao, W.; Bian, J.; Yi, X.; Zheng, S.; and Li, J. 2022. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures. *arXiv preprint arXiv:2207.01186*.

Zhang, X.; Gao, Y.; Lin, J.; and Lu, C.-T. 2020. Tapnet: Multivariate time series classification with attentional prototypical network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 6845–6852.

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

Zhao, B.; Lu, H.; Chen, S.; Liu, J.; and Wu, D. 2017. Convolutional neural networks for time series classification. *Journal of systems engineering and electronics*, 28(1): 162–169.

Zhao, J.; Li, Q.; Hong, Y.; and Shen, M. 2024. MetaRock-ETC: Adaptive Encrypted Traffic Classification in Complex Network Environments via Time Series Analysis and Meta-Learning. *IEEE Transactions on Network and Service Management*.

Zheng, Y.; Liu, Q.; Chen, E.; Ge, Y.; and Zhao, J. L. 2014. Time series classification using multi-channels deep convolutional neural networks. In *International conference on web-age information management*, 298–310. Springer.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, 27268–27286. PMLR.

Zuo, R.; Li, G.; Choi, B.; Bhowmick, S. S.; Mah, D. N.-y.; and Wong, G. L. 2023. SVP-T: a shape-level variable-position transformer for multivariate time series classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 11497–11505.