

CometNet: Contextual Motif-guided Long-term Time Series Forecasting

Weixu Wang, Xiaobo Zhou*, Xin Qiao, Lei Wang*, Tie Qiu

School of Computer Science and Technology, Tianjin University, China
 {weixuwang, xiaobo.zhou, xinqiao, wanglei2019}@tju.edu.cn, qiutie@ieee.org

Abstract

Long-term Time Series Forecasting is crucial across numerous critical domains, yet its accuracy remains fundamentally constrained by the receptive field bottleneck in existing models. Mainstream Transformer- and Multi-layer Perceptron (MLP)-based methods mainly rely on finite look-back windows, limiting their ability to model long-term dependencies and hurting forecasting performance. Naively extending the look-back window proves ineffective, as it not only introduces prohibitive computational complexity, but also drowns vital long-term dependencies in historical noise. To address these challenges, we propose CometNet, a novel Contextual Motif-guided Long-term Time Series Forecasting framework. CometNet first introduces a Contextual Motif Extraction module that identifies recurrent, dominant contextual motifs from complex historical sequences, providing extensive temporal dependencies far exceeding limited look-back windows; Subsequently, a Motif-guided Forecasting module is proposed, which integrates the extracted dominant motifs into forecasting. By dynamically mapping the look-back window to its relevant motifs, CometNet effectively harnesses their contextual information to strengthen long-term forecasting capability. Extensive experimental results on eight real-world datasets have demonstrated that CometNet significantly outperforms current state-of-the-art (SOTA) methods, particularly on extended forecast horizons.

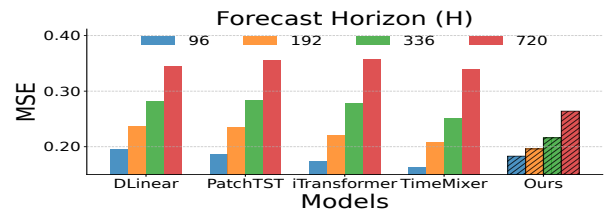
Extended version — <http://arxiv.org/abs/2511.08049>

Introduction

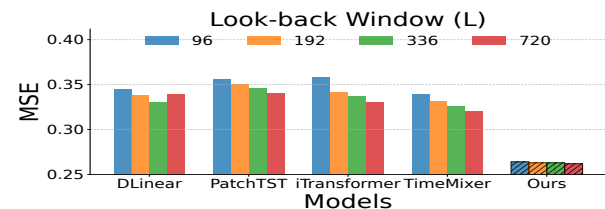
Time series forecasting (TSF) is a fundamental task in modern data science, driving critical decisions in domains ranging from meteorology (Bi et al. 2023), energy management (Balkanski et al. 2023) to transportation (Fang et al. 2025) and finance (Arsenault, Wang, and Patenaude 2025). With technological advancements, TSF methodologies have evolved from traditional statistical models (Ariyo, Adewumi, and Ayo 2014; Bahdanau, Cho, and Bengio 2015) to sophisticated deep learning architectures, including Transformers (Zhang and Yan 2023; Nie et al. 2023) and MLPs (Das et al. 2023; Tang and Zhang 2025). However,

*Corresponding Author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



(a) MSE across various forecast horizons (L = 96)



(b) MSE across various look-back windows (H = 720)

Figure 1: The Receptive Field Bottleneck in LTSF.

achieving accurate and robust long-term time series forecasting (LTSF) remains a persistent and formidable challenge.

While recent advances have introduced powerful models such as PatchTST (Nie et al. 2023), iTransformer (Liu et al. 2024) and TimeMixer++ (Wang et al. 2025a), they suffer from a fundamental architectural constraint: **the Receptive Field Bottleneck**. Specifically, contemporary models typically operate within finite-length look-back windows, attempting to predict future sequences from this limited temporal context (Wu et al. 2023; Chen et al. 2025). This design inherently fragments continuous time series into discrete, context-independent segments, thereby limiting the models' ability to learn long-term dependencies beyond the restricted look-back window (Zeng et al. 2023). Although sliding windows theoretically traverse the entire time series during training, gradient backpropagation remains confined within individual windows (Kang et al. 2024), preventing differentiable learning of long-term dependencies. Consequently, this bottleneck leads to severe performance degradation, as Fig. 1a shows, with a limited look-back window

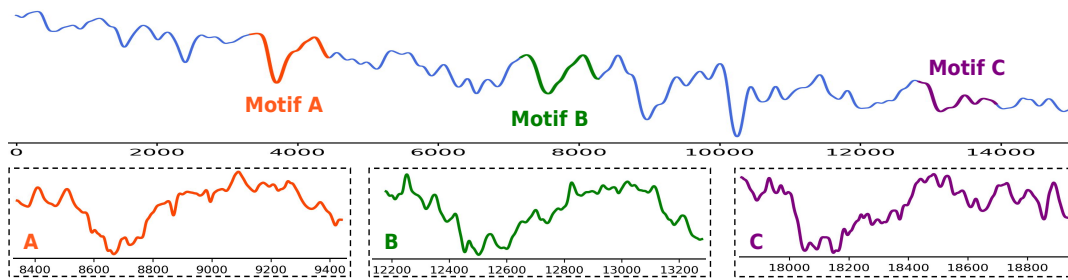


Figure 2: An example of a contextual motif and its recurring instances in ETTm1 dataset.

($L = 96$), the models exhibit substantially growing prediction errors as the forecast horizon extends.

An intuitive approach is to extend the look-back window, enabling models to capture longer-range dependencies. While Fig. 1b confirms that enlarged windows yield improved prediction performance, it also suffers from exponential growth in computational complexity and training time (Tan et al. 2024). Moreover, overly extended look-back windows can drown out meaningful temporal dependencies within irrelevant historical noise, leading to diminishing returns or even performance degradation (Zeng et al. 2023). Although the recent approach like Batched Spectral Attention (BSA) (Kang et al. 2024) enhances long-term dependency modeling by preserving cross-sample temporal correlations, it still struggles to capture meaningful temporal dependencies across thousands of time steps.

Fortunately, the evolution of real-world time series is not disordered. Instead, it is often governed by some recurring, representative long-context patterns (Huang, Chen, and Qiao 2024) that we term contextual motifs. As illustrated in Fig. 2, these motifs can span thousands of time steps and reappear at distant points in the series, while maintaining similar temporal patterns. Their extended temporal horizons encode rich contextual information, like factory production cycles or seasonal climate shifts, which is vital for understanding long-term dynamics. Building on this insight, we argue that these contextual motifs are key to transcending the limitations of local look-back windows. By discovering and leveraging these motifs, we can provide the long-term temporal dependencies needed for long-term forecasting.

To operationalize this insight, we propose **CometNet**, a novel **C**ontextual **M**otif-guided **N**ETwork for Long-term Time Series Forecasting. CometNet introduces a new forecasting paradigm that integrates contextual motifs into the forecasting process, enabling accurate long-term predictions. The framework operates through two core modules. First, a **Contextual Motif Extraction** module is designed to analyze the entire historical time series and establish a comprehensive library of dominant contextual motifs. Second, we further develop a **Motif-guided Forecasting** module that dynamically identify look-back window’s most relevant motif from the pre-established library and uses it to inform the prediction of the distant future. By doing so, our model effectively circumvents the receptive field bottleneck, grounding its predictions not just on the limited look-back window, but also on rich long-term contextual information

covered by the dominant motifs. The main contributions of this paper are summarized as follows:

- We propose CometNet, a novel contextual motif-guided network for long-term time series forecasting that leverages contextual motifs to overcome the receptive field bottleneck.
- We design an efficient contextual motif extraction module that systematically analyzes historical time series to establish a comprehensive library comprising dominant contextual motifs.
- We develop a motif-guided forecasting module that leverages our pre-established motif library to inform future predictions.
- Comprehensive experiments have been conducted across a broad spectrum of LTSF benchmarks, demonstrating the superiority of CometNet over current SOTA methods.

Related Work

In recent years, significant progress in LTSF has been predominantly driven by advancements in deep learning, with Transformer-based architectures and MLP-based models emerging as the two predominant paradigms in the field.

Transformer-based Methods

The Transformer architecture, with its powerful attention mechanism, has been extensively adapted for modeling temporal dependencies. With early variants like Reformer (Kitaev, Kaiser, and Levskaya 2020) and Informer (Zhou et al. 2021) focused on tackling the critical issue of quadratic complexity. To better capture complex dependencies, recent prominent works have moved beyond standard attention, exploring alternative structures to model dependencies. For example, Autoformer (Chen et al. 2021) captures more robust temporal patterns by embedding series decomposition within its attention mechanism, while Fedformer (Zhou et al. 2022) leverage frequency-domain analysis for capturing periodic characteristics. A pivotal advancement came with PatchTST (Nie et al. 2023), which introduced a channel-independent patching mechanism. This approach segments each time series variable into patches and focuses on capturing temporal relationships across these patches, significantly improving performance. This paradigm was further evolved by iTransformer (Liu et al. 2024), which inverted the typical role of attention and feed-forward networks, applying

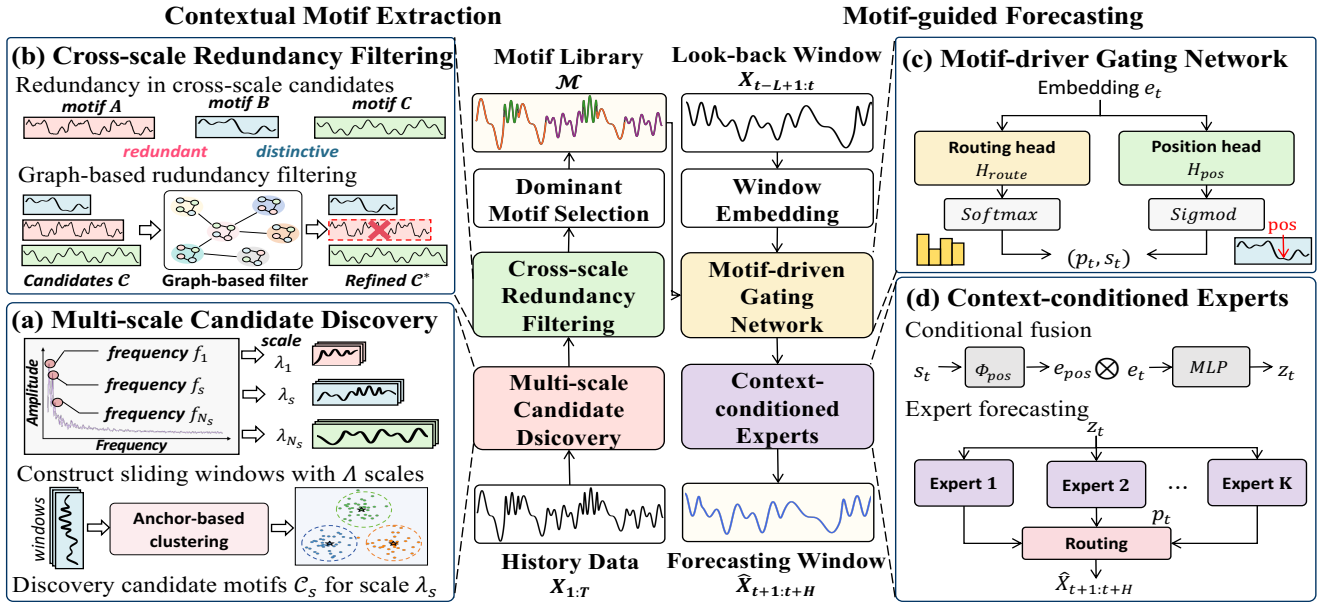


Figure 3: The architecture of CometNet.

attention to model correlations across variables while using MLPs to capture temporal dynamics within each channel.

MLP-based Methods

In response to the high complexity and potential training instability of Transformers, a competing paradigm centered on simpler MLP and linear structures has gained significant traction. DLinear (Zeng et al. 2023) pioneered this trend by demonstrating that a simple linear model applied to decomposed trend and seasonal components can achieve competitive forecasting performance. Subsequent works further advanced this paradigm. TSMixer (Ekambaram et al. 2023) introduced a pure MLP-based architecture that iteratively mixes temporal and feature dimensions, while FreTS (Yi et al. 2023) and FilterTS (Wang et al. 2025b) incorporated frequency-domain filtering to enhance MLP-based predictions. More recently, models like TimeMixer (Wang et al. 2024) and TimeMixer++ (Wang et al. 2025a) explored more sophisticated mixing strategies to further improve performance.

Despite their notable progress, these methods primarily concentrate on optimizing information extraction within localized temporal windows, which lack a dedicated mechanism to explicitly leverage the rich contextual dependencies embedded in the vast history beyond this finite scope. Though recent approach like BSA (Kang et al. 2024) attempts to alleviate this by preserving cross-sample temporal correlations through spectral attention, it still challenged in identifying meaningful temporal dependencies from complex historical sequences.

Methodology

Problem Statement. The objective of LTSF is to learn a mapping function $f : \mathbb{R}^{L \times N} \rightarrow \mathbb{R}^{H \times N}$ that predicts

future sequence $X_{t:t+H} \in \mathbb{R}^{H \times N}$ based on the observed look-back window $X_{t-L+1:t} \in \mathbb{R}^{L \times N}$, where N denotes the number of channels, L and H represent the observation window length and the prediction horizon, respectively. The function f is typically learned from the historical time series $X_{1:T}$ to capture temporal dependencies and generalize to unseen sequences.

Overall Architecture

To overcome the receptive field bottleneck inherent in limited look-back windows, we propose **CometNet**, a novel framework that leverages dominant contextual motifs to capture temporal dependencies far beyond the look-back window, significantly enhancing long-term forecasting performance. As illustrated in Fig. 3, CometNet introduces a two-module paradigm: first, a Contextual Motif Extraction module establishes a dominant contextual motif library from historical data, and second, a Motif-guided Forecasting module utilizes this library to make accurate long-term predictions.

The Contextual Motif Extraction module is designed to analyze the entire historical time series $X_{1:T} \in \mathbb{R}^{T \times N}$, to extract dominant contextual motifs. To handle channel-wise heterogeneity in multivariate time series (MTS), we adopt a channel-independent strategy (Nie et al. 2023; Han, Ye, and Zhan 2024), decomposing $X_{1:T}$ into N univariate series:

$$X_{1:T} \rightarrow \{X_{1:T}^1, \dots, X_{1:T}^n, \dots, X_{1:T}^N\}, \quad (1)$$

where $X_{1:T}^n \in \mathbb{R}^{T \times 1}$ represents the n -th channel. For brevity, we omit the channel superscript n hereafter. A cascaded motif extraction process is then applied to each univariate series to establish a library of its dominant motifs:

$$\mathcal{M} = CME_{extractor}(X_{1:T}), \quad (2)$$

where $\mathcal{M} = \{m_1, m_2, \dots, m_K\}$ denotes the extracted library of K dominant contextual motifs, and

$CMExtractor(\cdot)$ represents the contextual motif extraction function.

The Motif-guided Forecasting module is engineered to leverage this pre-established motif library \mathcal{M} for long-term prediction. We implement this using a Mixture-of-Experts (MoE) architecture, where each dominant motif $m_k \in \mathcal{M}$ is conceptually associated with a dedicated expert network E_k , which specializes in capturing the temporal dynamics inherent to that motif. Specifically, given a look-back window $X_{t-L+1:t}$, the forecasting module employs a motif-guided gating mechanism to dynamically associate the input with the most relevant dominant motif m_k in the library \mathcal{M} . This association not only selects the appropriate expert E_k , but also provides crucial contextual information about the window's position within the motif. The activated expert then generates the pattern-specific forecasts. This forecasting process is formally defined as:

$$\hat{X}_{t+1:t+H} = MGForecaster(X_{t-L+1:t} | \mathcal{M}), \quad (3)$$

where $MGForecaster(\cdot)$ represents the motif-guided forecasting function, which makes predictions conditioned on the pre-established motif library \mathcal{M} .

This design effectively circumvents the receptive field bottleneck, enabling the model to ground its predictions in both the local context of the look-back window and the long-term dependencies embedded in the dominant motifs. Next, we will introduce each module in detail.

Contextual Motif Extraction

The primary goal of this module is to automatically establish a library of dominant contextual motifs from the entire historical time series. The key challenge lies in the unknown scales of these motifs. A naive multi-scale exploration would not only create an exponentially large candidate space but also introduce significant cross-scale redundancy. To address this, we propose a cascaded motif extraction framework that combines multi-scale candidate discovery, cross-scale redundancy filtering, and benefit-driven selection to efficiently construct a high-quality library of dominant motifs.

Multi-scale Candidate Discovery. This initial step aims to efficiently discover a diverse set of candidate motifs across multiple time scales. Given a historical univariate time series $X_{1:T}$, we first apply moving average smoothing (Chen et al. 2021) to obtain a detrended series $X_{1:T}^{det}$ with more stable periodic characteristics. Subsequently, the Fast Fourier Transform (FFT) (Duhamel and Vetterli 1990) is employed to identify the dominant frequencies in the data:

$$A_f = Amp(FFT(X_{1:T}^{det})), \quad (4)$$

where $Amp(\cdot)$ calculates the amplitude spectrum. The corresponding period λ_f for each frequency component f is derived as $\lambda_f = \lceil \frac{T}{f} \rceil$. Based on the amplitude intensity, we select the top- N_s periods $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{N_s}\}$ with the highest spectral magnitudes to serve as the set of relevant scales. For each identified scale $\lambda_s \in \Lambda$, we first downsample the sequence to improve computational efficiency. On

this downsampled series, we employ an anchor-based clustering approach to efficiently group similar subsequences. Specifically, to avoid exhaustive pairwise comparisons, we randomly sample N_r subsequences of length λ_s as candidate anchors $\mathcal{A}_s = \{a_1, \dots, a_{N_r}\}$ and construct their correlation matrix $R^s \in \mathbb{R}^{N_r \times N_r}$, where each element R_{ij}^s is the Pearson correlation coefficient between anchors a_i and a_j . The density score for each anchor a_i at scale λ_s is then calculated as:

$$\rho_i^s = \sum_{j=1}^{N_r} R_{ij}^s \cdot \mathbb{I}(R_{ij}^s > \tau_s), \quad (5)$$

where τ_s is a similarity threshold. After selecting the top N_c anchors with the highest density scores as cluster centroids, all subsequences of length λ_s are assigned to their nearest centroid. From each resulting cluster, we select the most representative subsequence (the medoid) as a candidate motif. This process yields a comprehensive set of candidate motifs from all scales, $\mathcal{C} = \bigcup_{s=1}^{N_s} \mathcal{C}_s$, for the subsequent refinement stages.

Cross-scale Redundancy Filtering. To refine the candidate set \mathcal{C} while preserving its diversity, we develop a cross-scale filtering strategy that eliminates redundant motifs using graph-based analysis. We first model the relationships between all candidate motifs in \mathcal{C} as a weighted, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each vertex $v_i \in \mathcal{V}$ corresponds to a candidate motif $c_i \in \mathcal{C}$. Edge weights are determined by a normalized Dynamic Time Warping (DTW) (Sakoe and Chiba 2003) similarity score, with a sparsity constraint:

$$w_{ij} = S_{DTW}(c_i, c_j) \cdot \mathbb{I}(S_{DTW}(c_i, c_j) > \tau_g), \quad (6)$$

where $S_{DTW}(\cdot)$ is the normalized DTW similarity, and the threshold τ_g controls the granularity of redundancy filtering.

The graph \mathcal{G} is then decomposed into a set of connected components $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_{N_g}\}$. Each component $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k)$ represents a cluster of semantically similar motifs that may span different scales. We obtain a refined candidate set \mathcal{C}^* by selecting a single prototype from each component. The prototype c_k^* for component \mathcal{G}_k is identified as the motif with the highest intra-cluster affinity:

$$c_k^* = \operatorname{argmax}_{c \in \mathcal{V}_k} \frac{1}{|\mathcal{V}_k|} \sum_{c' \in \mathcal{V}_k} S_{DTW}(c, c'). \quad (7)$$

This process effectively filters redundancy while preserving a diverse set of representative candidates.

Dominant Motif Selection. An ideal library of dominant motifs must be representative, comprehensive, and diverse, as these properties fundamentally determine its capacity to provide precise contextual guidance. To this end, we propose a benefit-driven selection strategy to select the final dominant motifs from the refined set \mathcal{C}^* . First, each candidate $c_i^* \in \mathcal{C}^*$ is evaluated for its intrinsic quality using a composite metric:

$$Q(c_i^*) = \alpha_S \cdot Q_S(c_i^*) + \alpha_P \cdot Q_P(c_i^*) + \alpha_A \cdot Q_A(c_i^*), \quad (8)$$

where $Q_S(\cdot)$, $Q_P(\cdot)$, $Q_A(\cdot)$ quantify the saliency, prevalence and atomicity of patten c_i^* , respectively. While α_S , α_P , and α_A are their corresponding weighting coefficients.

Next, we formulate a benefit score that jointly optimizes for pattern quality, overall temporal coverage, and library diversity. The benefit of adding a candidate c_i^* to our current selection \mathcal{S} is calculated as:

$$B(c_i^* | \mathcal{S}) = Q(c_i^*) \cdot Cov(c_i^* | \mathcal{S}) \cdot Div(c_i^* | \mathcal{S}), \quad (9)$$

where $Cov(\cdot)$ measures the marginal temporal coverage gained by adding c_i^* , and $Div(\cdot)$ measures its marginal contribution to the library’s diversity. By iteratively selecting the top- K candidates with the highest benefit scores, we construct the final library of dominant contextual motifs, $\mathcal{M} = \{m_1, m_2, \dots, m_K\}$, that optimally balances these crucial properties. The detailed formulations of these metrics are provided in Appendix.

Motif-guided Forecasting

After establishing the library of dominant contextual motifs \mathcal{M} , we introduce the Motif-guided Forecasting module. This module is designed to leverage the rich, long-term contextual information embedded in the motif library to guide its predictions. Its core is a motif-aware MoE architecture, where the number of experts is equal to the number of dominant motifs, $K = |\mathcal{M}|$. The forecasting process consists of three sequential stages: window embedding, motif-driven gating, and context-conditioned prediction by the experts.

Window Embedding. Given any look-back window $X_{t-L+1:t}$, we first map it into a compact latent representation $e_t \in \mathbb{R}^{d_e}$ via a window embedding module (implemented as a MLP), preserving key temporal information while reducing dimensionality,

$$e_t = LayerNorm(MLP(X_{t-L+1:t})). \quad (10)$$

Motif-driven Gating Network. We then develop a motif-driven gating network to dynamically integrate our pre-established motif library \mathcal{M} into the forecasting process. This network acts as an intelligent router, receiving the window embedding e_t as input and dynamically associating the current local observation with the relevant contextual motif. Unlike standard MoE gates, which only selects which expert to use, our gate also provides window position information in its relevant motif, offering fine-grained prediction guidance.

Specifically, our gating network processes the embedding e_t through two parallel heads, both implemented as lightweight MLPs. The first is a routing head H_{route} , which maps e_t to a logit vector. This vector is then passed through a Softmax function to generate a probability distribution $p_t \in \mathbb{R}^K$, representing the selection probabilities for the K experts,

$$p_t = Softmax(H_{route}(e_t)), \quad (11)$$

the k -th element $p_{t,k}$ represents the confidence in selecting the k -th expert, which corresponds to the k -th dominant motif $m_k \in \mathcal{M}$.

Simultaneously, a parallel position head H_{pos} maps the same embedding e_t into a scalar value $s_t \in [0, 1]$ via a Sigmoid activation function. This scalar encodes the relative position of the current observation within the life cycle of its associated motif.

$$s_t = \sigma(H_{pos}(e_t)). \quad (12)$$

Together, p_t and s_t form a dual instructional signal that specifies both the associated motif and the temporal progression within it. This provides rich, dual-faceted guidance for the downstream experts.

Context-conditioned Experts. Our framework contains K independent context-conditioned expert networks $\{E_1, E_2, \dots, E_K\}$, each designed to specialize in the dynamics of a corresponding dominant motif $m_k \in \mathcal{M}$. The core contribution of these experts lies in their ability to perform context-conditioned forecasting by leveraging the dual instructional signal from the gating network.

For each expert E_k , the shared inputs are the window embedding e_t and its positional score s_t . The process begins with a position encoder Φ_{pos} , implemented as a MLP. It elevates the low-dimensional positional score s_t into a high-dimensional position embedding $e_{pos} \in \mathbb{R}^{d_e}$. Subsequently, a conditional fusion layer concatenates the window embedding e_t with the position embedding e_{pos} and processes them through another MLP to learn their non-linear interactions. This produces a unified, contextually-aware conditional representation $z_t \in \mathbb{R}^{d_e}$.

$$z_t = F_{fuse}(concat(e_t, \Phi_{pos}(s_t))). \quad (13)$$

Then, the highly conditional representation z_t is fed into K parallel, expert-specific prediction heads $\{P_1, P_2, \dots, P_K\}$, to generate a set of expert-specific forecasts $\{\hat{X}_{1,t+1:t+H}, \hat{X}_{2,t+1:t+H}, \dots, \hat{X}_{K,t+1:t+H}\}$, where $\hat{X}_{k,t+1:t+H} \in \mathbb{R}^{H \times 1}$. The final prediction of the entire model $\hat{X}_{t+1:t+H}$ is the weighted sum of all expert outputs, with the weights provided by the gating network’s probability distribution p_t .

$$\hat{X}_{t+1:t+H} = \sum_{k=1}^K p_{t,k} \cdot \hat{X}_{k,t+1:t+H} = \sum_{k=1}^K p_{t,k} \cdot P_k(z_t). \quad (14)$$

This design enables our model to dynamically and granularly adjust its predictive strategy based on both the high-level contextual motif and the subtle variations of the stage within it.

Experiments

Experimental Details

Datasets. We evaluate our proposed model on eight real-world benchmark datasets from various domains including energy, climate, economics and transportation for LTSF task. These datasets encompasses four ETT datasets (ETTth1, ETTth2, ETTm1, ETTm2), Weather, Exchange Rate, Traffic, and Electricity, as previously employed in Autoformer (Chen et al. 2021).

Baselines and Setup. We select 7 state-of-the-art LTSF models as baselines, including TimeMixer++ (Wang et al. 2025a), FilterTS (Wang et al. 2025b), PatchMLP (Tang and Zhang 2025), BSA (Kang et al. 2024), iTransformer (Liu et al. 2024), PatchTST (Nie et al. 2023) and DLinear (Zeng et al. 2023). To ensure fair comparison, all models adopt identical experimental setup, utilizing a finite look-back

Models	CometNet (Ours)		TimeMixer++ (2025)		FilterTS (2025)		PatchMLP (2025)		BSA (2024)		iTransformer (2024)		PatchTST (2023)		DLinear (2023)		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	0.345 <u>0.398</u>	0.361	0.403	0.374	0.391	0.391	0.403	0.427	0.442	0.386	0.405	0.460	0.447	0.397	0.412	
	192	0.368 0.412	<u>0.416</u>	0.441	0.424	<u>0.421</u>	0.444	0.431	0.480	0.481	0.441	0.512	0.477	0.429	0.446	0.441	
	336	0.387 0.423	<u>0.430</u> <u>0.434</u>	0.464	0.441	0.464	0.441	0.490	0.456	0.537	0.521	0.487	0.458	0.546	0.496	0.489	0.467
	720	0.391 0.434	<u>0.467</u> <u>0.451</u>	0.470	0.466	0.470	0.466	0.528	0.496	0.697	0.620	0.503	0.491	0.544	0.517	0.513	0.510
	Avg	0.373 0.417	<u>0.419</u>	0.432	0.433	<u>0.430</u>	0.463	0.447	0.536	0.516	0.454	0.447	0.516	0.484	0.461	0.457	
ETTh2	96	0.205 0.281	0.276	0.328	0.290	0.338	0.305	0.353	<u>0.234</u> <u>0.323</u>	0.297	0.349	0.308	0.355	0.340	0.394		
	192	0.212 0.305	0.342	0.379	0.374	0.390	0.402	0.410	<u>0.290</u> <u>0.362</u>	0.380	0.400	0.393	0.405	0.482	0.479		
	336	0.219 0.318	0.346	0.398	0.406	0.420	0.436	0.437	<u>0.327</u> <u>0.388</u>	0.428	0.432	0.427	0.436	0.591	0.541		
	720	0.228 0.332	<u>0.392</u> <u>0.415</u>	0.418	0.437	0.437	0.449	0.414	0.439	0.427	0.445	0.436	0.450	0.839	0.661		
	Avg	0.216 0.309	0.339	0.380	0.372	0.396	0.395	0.412	<u>0.316</u> <u>0.378</u>	0.383	0.407	0.391	0.411	0.563	0.519		
ETTh1	96	0.210 0.315	<u>0.310</u> <u>0.334</u>	0.321	0.360	0.324	0.362	0.382	0.398	0.334	0.368	0.352	0.374	0.346	0.374		
	192	0.211 0.322	<u>0.348</u> <u>0.362</u>	0.363	0.382	0.369	0.385	0.423	0.428	0.390	0.393	0.374	0.387	0.391	0.391		
	336	0.282 0.368	<u>0.376</u> <u>0.391</u>	0.395	0.403	0.402	0.407	0.492	0.466	0.426	0.420	0.421	0.414	0.415	0.415		
	720	0.303 0.387	<u>0.440</u> <u>0.423</u>	0.462	0.438	0.474	0.446	0.567	0.516	0.491	0.459	0.462	0.449	0.473	0.451		
	Avg	0.252 0.348	<u>0.369</u> <u>0.378</u>	0.385	0.396	0.392	0.400	0.466	0.452	0.407	0.410	0.406	0.407	0.404	0.408		
ETTh2	96	0.117 0.244	0.170	<u>0.245</u>	0.172	0.255	0.180	0.263	<u>0.153</u>	0.258	0.180	0.264	0.183	0.270	0.193	0.293	
	192	0.148 0.280	0.229	0.291	0.237	0.299	0.244	0.306	<u>0.189</u> <u>0.290</u>	0.250	0.309	0.255	0.314	0.284	0.361		
	336	0.156 0.285	0.303	0.343	0.299	0.398	0.312	0.349	<u>0.230</u> <u>0.321</u>	0.311	0.348	0.309	0.347	0.382	0.329		
	720	0.171 0.299	0.373	0.399	0.397	0.394	0.411	0.407	<u>0.303</u> <u>0.369</u>	0.412	0.407	0.412	0.404	0.558	0.525		
	Avg	0.148 0.277	0.269	0.320	0.276	0.321	0.287	0.331	<u>0.218</u> <u>0.309</u>	0.288	0.332	0.290	0.334	0.354	0.402		
Weather	96	0.183	0.209	0.155 <u>0.205</u>	0.162	0.207	0.164	0.210	<u>0.159</u> 0.203	0.174	0.214	0.186	0.227	0.195	0.252		
	192	0.196 0.223	<u>0.201</u> <u>0.245</u>	0.209	0.252	0.211	0.251	0.205	0.246	0.221	0.254	0.234	0.265	0.237	0.295		
	336	0.216 0.241	<u>0.237</u> <u>0.265</u>	0.263	0.292	0.269	0.294	0.252	0.285	0.278	0.296	0.284	0.301	0.282	0.331		
	720	0.264 0.280	<u>0.312</u> <u>0.334</u>	0.344	0.344	0.349	0.345	0.325	0.337	0.358	0.347	0.356	0.349	0.345	0.382		
	Avg	0.215 0.238	<u>0.226</u> <u>0.262</u>	0.244	0.274	0.248	0.275	0.235	0.268	0.258	0.278	0.265	0.285	0.265	0.315		
Exchange	96	0.086	0.216	<u>0.085</u>	0.214	0.081 0.199	0.094	0.215	0.089	0.210	0.086	0.206	0.088	<u>0.205</u>	0.088	0.218	
	192	0.137 0.269	0.175	0.313	<u>0.171</u>	0.294	0.182	0.306	0.183	<u>0.289</u>	0.177	0.299	0.176	0.299	0.176	0.315	
	336	0.176 0.293	0.316	0.420	0.321	0.409	0.339	0.423	0.317	0.403	0.331	0.417	<u>0.301</u> <u>0.397</u>	0.313	0.427		
	720	0.423 0.437	0.851	0.689	0.837	0.688	0.873	0.701	<u>0.647</u> <u>0.536</u>	0.847	0.691	0.901	0.714	0.839	0.695		
	Avg	0.206 0.304	0.357	0.391	0.352	0.397	0.372	0.411	<u>0.309</u> <u>0.360</u>	0.360	0.403	0.367	0.404	0.354	0.414		
Electricity	96	<u>0.139</u> <u>0.228</u>	0.135 0.222	0.151	0.245	0.160	0.257	0.142	0.239	0.148	0.240	0.190	0.296	0.210	0.302		
	192	<u>0.154</u> 0.233	0.147 <u>0.235</u>	0.163	0.256	0.176	0.271	0.163	0.257	0.162	0.253	0.199	0.304	0.210	0.305		
	336	<u>0.167</u> <u>0.247</u>	0.164 0.245	0.180	0.274	0.197	0.292	0.176	0.274	0.178	0.269	0.217	0.319	0.223	0.319		
	720	0.172 0.256	<u>0.212</u>	0.310	0.224	0.311	0.249	0.333	0.221	<u>0.308</u>	0.225	0.317	0.258	0.352	0.258	0.350	
	Avg	0.158 0.241	<u>0.165</u> <u>0.253</u>	0.180	0.271	0.196	0.288	0.176	0.269	0.178	0.270	0.216	0.318	0.225	0.319		
Traffic	96	0.408	0.271	0.392 0.253	0.448	0.309	0.475	0.329	<u>0.392</u>	0.272	0.395	<u>0.268</u>	0.526	0.347	0.650	0.396	
	192	0.419	<u>0.275</u>	0.402 0.258	0.455	0.307	0.483	0.330	<u>0.417</u>	0.281	0.417	0.276	0.522	0.332	0.598	0.370	
	336	0.426 <u>0.279</u>	<u>0.428</u> 0.263	0.472	0.313	0.498	0.336	0.433	0.289	0.433	0.283	0.517	0.334	0.605	0.373		
	720	0.436 <u>0.283</u>	<u>0.441</u> <u>0.282</u>	0.508	0.332	0.542	0.359	0.470	0.310	0.467	0.302	0.552	0.352	0.645	0.394		
	Avg	<u>0.422</u> <u>0.277</u>	0.416 0.264	0.471	0.315	0.500	0.339	0.428	0.288	0.428	0.282	0.529	0.341	0.625	0.383		

Table 1: Full results for the long-term forecasting task. We compare extensive competitive models under different prediction lengths. Avg is averaged from all four prediction lengths, that is $\{96, 192, 336, 720\}$. Best and second best results are highlighted in **red** and blue.

window of $L = 96$ to generate predictions across multiple horizons $H \in \{96, 192, 336, 720\}$. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are selected as evaluation metrics.

For high-dimensional datasets (e.g., Traffic and Electricity), we employ k-dominant frequency hashing (k-DFH) (Kang, Shin, and Lee 2025) to cluster variables into multiple groups, enabling efficient parallel training of variable groups. More details are provided in Appendix.

Main Results

The comprehensive quantitative results of our proposed CometNet against 7 baseline models are presented in Table 1. The results unequivocally demonstrate the superiority

of our approach across a wide array of benchmarks and forecasting horizons. Specifically, across all datasets, CometNet attained either first (7/8) or second (1/8) rank in averaging MSE and MAE performance, establishing robust forecasting superiority.

A key observation from the results is that our approach’s competitive advantage becomes more pronounced as the prediction horizon increases. For instance, on the ETTh2 dataset, while CometNet already outperforms baselines at horizon $H = 96$, (MSE 0.205 vs 0.276), its margin of victory grows substantially at $H = 720$, (MSE 0.228 vs 0.392). This trend holds consistently across nearly all datasets, including ETTh1, ETTh2, ETTm1, ETTm2, Weather, and Exchange, directly verify the validity of our design. Unlike

CometNet Components					ETTh1		ETTh2		ETTm1		ETTm2	
Multi-scale	Filtering	Selection	Gating	Position	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
×	×	×	×	×	0.544	0.531	0.653	0.572	0.483	0.464	0.526	0.539
✓	×	×	✓	✓	0.447	0.466	0.492	0.461	0.395	0.439	0.392	0.361
✓	✓	×	✓	✓	0.430	0.470	0.331	0.376	0.345	0.401	0.384	0.332
✓	✓	✓	×	✓	0.476	0.481	0.292	0.414	0.364	0.422	0.311	0.326
✓	✓	✓	✓	×	0.423	0.442	0.235	0.340	0.326	0.408	0.262	0.308
✓	✓	✓	✓	✓	0.391	0.434	0.228	0.332	0.303	0.387	0.171	0.299

Table 2: Ablation study of CometNet’s key components on ETT datasets ($H = 720$). ✓ denotes that the component is enabled, while × denotes it is ablated. The best results are highlighted in red.

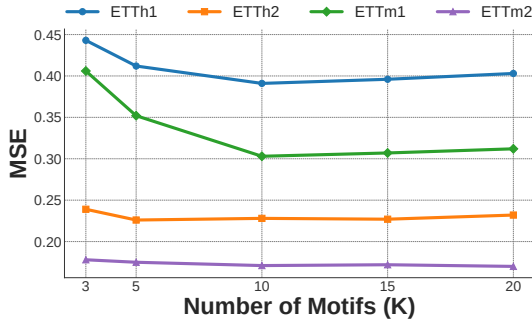


Figure 4: Impact of the number of motifs (K) on the forecasting performance.

baseline models constrained by receptive field limitations, CometNet leverages learned long-context motifs to guide long-term time series forecasting, thus preventing the error accumulation that plagues long-horizon forecasting.

The slightly weaker performance of CometNet on the Electricity and Traffic datasets can be attributed to the high number of variables in these datasets. This is an expected result since our group-based motif discovery trick trains each group rather than each variable to improve training efficiency. Notably, as prediction horizons extend, CometNet’s rank consistently improves, ultimately achieving top performance at the longest horizons. This further confirms the value of our motif-guided approach for long-term forecasting.

Ablation Study

To validate the individual contributions of CometNet’s key components, we conduct a comprehensive ablation study on the ETT datasets with a forecast horizon $H = 720$. The results, presented in Table 2, unequivocally demonstrate the effectiveness and synergistic cooperation of each designed module. We start with a baseline where all motif-related components are disabled, which yields the poorest performance and confirms the critical need for long-range contextual information. Then, the stepwise reintroduction of Multi-scale discovery, Cross-scale filtering, and motif selection components demonstrates consistent and significant improvements, confirming that each step in our Contextual Motif Extraction process is vital for distilling a high-quality motif library. For the Motif-guided Forecasting module, we

verified that removing either the expert routing gate mechanism or the refined position signal leads to a marked drop increase in prediction loss. This highlights the necessity of both dynamically selecting the correct motif-based expert and properly localizing the window’s relative position within its motif’s lifecycle. Ultimately, the full CometNet model, with all components integrated, achieves the best performance across all metrics, effectively validating our design and proving that each component plays an indispensable role in overcoming the receptive field bottleneck.

Sensitive Study

We conduct a sensitivity analysis to evaluate the impact of the number of dominant motifs K in our library. As illustrated in Fig. 4, when K is too small (e.g., $K=3$), our model’s performance is suboptimal, as an insufficient number of motifs limits the library’s expressive power. As K increases, the performance consistently improves, most datasets achieving performance peaks around $K = 10$. This suggests that a library of this size strikes an optimal balance, offering a rich set of diverse and representative patterns. Interestingly, further increasing K beyond this point (e.g., to 15 or 20) leads to either stabilization or a slight degradation in performance. This indicates that an overly large library may introduce redundancy or noise, slightly hindering the model’s ability to pinpoint the most relevant context.

Conclusion

This paper introduces CometNet, a Contextual Motif-guided framework for LTSF that leverages recurrent contextual motifs to capture long-term dependencies, thereby significantly enhancing prediction accuracy. Extensive experiments across eight real-world datasets show that CometNet significantly outperforms SOTA methods, especially for extended forecast horizons. These results highlight the benefits of leveraging contextual motifs to guide the modeling of long-term temporal dependencies.

Acknowledgments

This work is supported in part by the National Science Fund for Distinguished Young Scholars (No. 62325208), in part by the National Natural Science Foundation of China (No. 62572342, 62272339), and in part by the National Science Foundation of Tianjin (No. 23ZGZNGX00020).

References

- Ariyo, A. A.; Adewumi, A. O.; and Ayo, C. K. 2014. Stock Price Prediction Using the ARIMA Model. In Al-Dabass, D.; Orsoni, A.; Cant, R. J.; Yunus, J.; Ibrahim, Z.; and Saad, I., eds., *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014, Cambridge, United Kingdom, March 26-28, 2014*, 106–112. IEEE.
- Arsenault, P.; Wang, S.; and Patenaude, J. 2025. A Survey of Explainable Artificial Intelligence (XAI) in Financial Time Series Forecasting. *ACM Comput. Surv.*, 57(10): 265:1–265:37.
- Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Balkanski, E.; Périer, N.; Stein, C.; and Wei, H. 2023. Energy-Efficient Scheduling with Predictions. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.
- Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; and Tian, Q. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nat.*, 619(7970): 533–538.
- Chen, H.; Rossi, R. A.; Kim, S.; Mahadik, K.; and Eldardiry, H. 2025. Probabilistic hypergraph recurrent neural networks for time-series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, 82–93.
- Chen, M.; Peng, H.; Fu, J.; and Ling, H. 2021. AutoFormer: Searching Transformers for Visual Recognition. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, 12250–12260. IEEE.
- Das, A.; Kong, W.; Leach, A.; Mathur, S.; Sen, R.; and Yu, R. 2023. Long-term Forecasting with TiDE: Time-series Dense Encoder. *Trans. Mach. Learn. Res.*, 2023.
- Duhamel, P.; and Vetterli, M. 1990. Fast Fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4): 259–299.
- Ekambaram, V.; Jati, A.; Nguyen, N.; Sinthong, P.; and Kalagnanam, J. 2023. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting. In Singh, A. K.; Sun, Y.; Akoglu, L.; Gunopulos, D.; Yan, X.; Kumar, R.; Ozcan, F.; and Ye, J., eds., *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, 459–469. ACM.
- Fang, Y.; Liang, Y.; Hui, B.; Shao, Z.; Deng, L.; Liu, X.; Jiang, X.; and Zheng, K. 2025. Efficient Large-Scale Traffic Forecasting with Transformers: A Spatial Data Management Perspective. In Sun, Y.; Chierichetti, F.; Lauw, H. W.; Perlich, C.; Tok, W. H.; and Tomkins, A., eds., *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, V.1, KDD 2025, Toronto, ON, Canada, August 3-7, 2025*, 307–317. ACM.
- Han, L.; Ye, H.; and Zhan, D. 2024. The Capacity and Robustness Trade-Off: Revisiting the Channel Independent Strategy for Multivariate Time Series Forecasting. *IEEE Trans. Knowl. Data Eng.*, 36(11): 7129–7142.
- Huang, H.; Chen, M.; and Qiao, X. 2024. Generative Learning for Financial Time Series with Irregular and Scale-Invariant Patterns. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.
- Kang, B. G.; Lee, D.; Kim, H.; Chung, D.; and Yoon, S. 2024. Introducing spectral attention for long-range dependency in time series forecasting. *Advances in Neural Information Processing Systems*, 37: 136509–136544.
- Kang, J.; Shin, Y.; and Lee, J. 2025. VarDrop: Enhancing Training Efficiency by Reducing Variate Redundancy in Periodic Time Series Forecasting. In Walsh, T.; Shah, J.; and Kolter, Z., eds., *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA, 17742–17749*. AAAI Press.
- Kitaev, N.; Kaiser, L.; and Levskaya, A. 2020. Reformer: The Efficient Transformer. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.
- Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2024. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.
- 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, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Sakoe, H.; and Chiba, S. 2003. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1): 43–49.
- Tan, M.; Merrill, M. A.; Gupta, V.; Althoff, T.; and Hartvigsen, T. 2024. Are Language Models Actually Useful for Time Series Forecasting? In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J. M.; and Zhang, C., eds., *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*.
- Tang, P.; and Zhang, W. 2025. Unlocking the Power of Patch: Patch-Based MLP for Long-Term Time Series Forecasting. In Walsh, T.; Shah, J.; and Kolter, Z., eds., *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, 12640–12648. AAAI Press.
- Wang, S.; Li, J.; Shi, X.; Ye, Z.; Mo, B.; Lin, W.; Ju, S.; Chu, Z.; and Jin, M. 2025a. TimeMixer++: A General Time Series

Pattern Machine for Universal Predictive Analysis. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net.

Wang, S.; Wu, H.; Shi, X.; Hu, T.; Luo, H.; Ma, L.; Zhang, J. Y.; and Zhou, J. 2024. TimeMixer: Decomposable Multi-scale Mixing for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Wang, Y.; Liu, Y.; Duan, X.; and Wang, K. 2025b. FilterTS: Comprehensive Frequency Filtering for Multivariate Time Series Forecasting. In Walsh, T.; Shah, J.; and Kolter, Z., eds., *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA, 21375–21383*. AAAI Press.

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, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; An, N.; Lian, D.; Cao, L.; and Niu, Z. 2023. Frequency-domain MLPs are More Effective Learners in Time Series Forecasting. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.

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, Y.; and Yan, J. 2023. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, 11106–11115. AAAI Press.

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 Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvári, C.; Niu, G.; and Sabato, S., eds., *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, 27268–27286. PMLR.