Wavelet-Based MLP for Efficient Sensor Signal Measurement and Forecasting

Zheng Zhou, Yu-Jie Xiong, and Chun-Ming Xia

***Abstract*—Measuring and predicting in real-world time series is a crucial task. However, the existing approaches often suffer from high computational complexity, are highly sensitive to noise, and lack interpretability, which hinders their effectiveness in practical applications. In this article, we propose WaveMLP, a novel lightweight framework with wavelet decomposition and multilayer perceptrons (MLPs). Unlike conventional models, WaveMLP fundamentally enhances measurement accuracy by decomposing raw sensor signals into low-frequency physical quantities and high-frequency measurement noise through the discrete wavelet transform (DWT). WaveMLP comprises three key components. First, it employs a computationally efficient perfect reconstruction Haar wavelet to capture abrupt changes in signals and decompose them into low- and high-frequency components. Next, a parameter-efficient MLP module processes these components: separate low- and high-frequency MLPs extract features from their respective frequency bands, ensuring high inference speed. Finally, an adaptive fusion mechanism nonlinearly combines the low- and high-frequency components, enhancing model robustness and preventing overfitting. The approach achieves two key advances in Instrumentation & Measurement (I&M): 1) signal feature capture capability. WaveMLP separates the low- and high-frequency components of sensor signals, effectively capturing the main and nonlinear features of the signal to improve prediction performance and 2) real-time deployment capability for edge instrumentation. WaveMLP reduces training time by 20%–40%. Validated on eight real-world datasets, WaveMLP matches SOTA while reducing parameters by orders of magnitude. This work establishes wavelet-MLP as a new paradigm for resource-constrained measurement systems.**

***Index Terms*—Time–frequency analysis, time-series forecasting, wavelet transform.**

I. Introduction

INSTRUMENTATION and measurement systems increasingly rely on advanced signal processing techniques to address the challenges in real-time sensor data analysis. The rise of deep learning introduces novel approaches for time-series forecasting, but many models struggle to balance computational efficiency with accuracy in industrial measurement contexts. For instance, the transformer architecture, while widely adopted for its sequence modeling capabilities, suffers from high computational demands and prolonged training times. As the input sequence grows, the computational burden of the transformer method gradually makes it difficult to balance performance and efficiency. These limitations are particularly critical in Instrumentation & Measurement (I&M) applications, such as vibration monitoring [1] and environmental sensor networks [2], [3], where real-time performance and resource constraints are paramount. In industrial measurement systems, traditional linear models [4], [5], [6] often rely on complex frequency-domain transformations (e.g., Fourier transforms) to capture temporal patterns. However, these models face two key limitations: 1) they assume stationarity in sensor signals, which is rarely valid for nonstationary industrial environments (e.g., machinery degradation or fluctuating environmental conditions) and 2) their reliance on channel independence increases computational overhead, making them unsuitable for edge deployment in IoT-based measurement systems.

To tackle these challenges, we propose WaveMLP, a lightweight time-series forecasting model tailored for I&M applications. Unlike the existing models, WaveMLP extends time-series analysis into the wavelet domain, enabling efficient decomposition of sensor signals into approximate representations (low-frequency component) and detailed features (high-frequency noise). The proposed model directly addresses the critical issues in industrial measurement systems: first, it leverages wavelet decomposition to isolate high-frequency noise, a common requirement in sensor signal denoising. Second, by separating low- and high-frequency components, WaveMLP captures both long-term trends and transient anomalies. Overall, our contributions are summarized as follows.

1) We propose WaveMLP, a low-parameter and highly efficient forecasting model. In the majority of instances, it diminishes the training time of the lightweight linear model by 20%–40%. Compared to large-scale models, WaveMLP reduces the parameter count by several orders of magnitude.
2) WaveMLP provides a new direction for decomposing time series, which is different from the method of decomposing time series into trend and seasonal terms. This will provide more references for the field of time series and indicate a promising direction.
3) WaveMLP achieves SOTA performance on several real-world and artificially synthesized speech datasets with minimal parameter count and extremely fast speed using simple and low complexity models.

Received 26 August 2025; revised 20 November 2025; accepted 10 December 2025. Date of publication 19 January 2026; date of current version 29 January 2026. This work was supported in part by the Science and Technology Commission of Shanghai Municipality under Grant 21DZ2203100 and in part by the National Natural Science Foundation of China under Grant 62006150. The Associate Editor coordinating the review process was Dr. Grazia Iadarola. *(Corresponding author: Yu-Jie Xiong.)*

The authors are with the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China (e-mail: xiong@sues.edu.cn).

Digital Object Identifier 10.1109/TIM.2026.3652755

II. Related Work

We depict various preceding models, categorizing them according to their distinctive characteristics. According to whether the signal is decomposed or not and whether the model runs in the time or frequency domain, it can be divided into the following four categories.

A. Nondecomposition in Time Domain

Many deep learning models are formulated for the predictive analysis of time-series data, each designed with specialized architectures to overcome the particular challenges of the domain. While RNN-based models, exemplified by LSTNet [7], are traditionally preferred for their adeptness at handling sequential data, they struggle with high computational demands and error propagation in long-range predictions. In contrast, CNN-based frameworks, such as SCINet [8], excel at capturing local dependencies but often struggle to model the long-range, global interactions critical for accurate and comprehensive forecasting. Advances that combine causal convolution with expansive convolutional kernels ameliorate these issues by incorporating broader contextual information more effectively [9].

Transformers, with architectures informer [10] and reformer [11], are posited as groundbreaking for their ability to handle extensive sequences through self-attention mechanisms. Yet, studies show that these complex models are occasionally outperformed by simpler linear models, challenging the assumption that greater complexity yields better results. PatchTST [12] applies principles from vision transformers [13] to time series. iTransformer [14] reimagines the linkage between data form and token representation. Moreover, MTGNN [15] utilizes graph neural networks (GNNs) to explore intricate spatiotemporal dependencies, providing a nuanced understanding of interconnected datasets and surpassing more complex architectures with simpler, well-designed linear models RLinear [16].

B. Decomposition in Time Domain

In new applications to time series, decomposition into simpler, more interpretable components, such as seasonal patterns and underlying trends, is pivotal. RobustSTL [17] proposes robust algorithms for extracting trends using sparse regularization coupled with minimum absolute deviation loss. A comprehensive survey [18] critiques the robustness, scalability, and efficiency of these models, suggesting future directions for more adaptable and effective strategies. Autoformer [19] represents a significant leap forward by replacing traditional self-attention with the fast Fourier transform (FFT) to more adeptly manage sequence connections and delay accumulations, facilitating the processing of longer sequences with reduced computational overhead.

C. Nondecomposition in Frequency Domain

Shifting focus to frequency-domain analysis represents a notable trend, as researchers strive to transform complex time-domain data into a more tractable frequency-domain format. This leads to the development of models, such as FITS [5] and FiLM [20], which employ techniques like the Fourier transform to improve signal clarity and understand temporal patterns within the frequency domain. Similarly, a time–frequency-domain transformation method [21] is proposed to overcome harmonic suppression issues through parameter optimization. These models underscore the increasing appreciation for frequency-domain analysis in enhancing the precision and efficiency of forecasting models, particularly where conventional time-domain approaches falter.

D. Decomposition in Wavelet Domains

Integrating wavelet analysis with machine learning, approaches FEDformer [22] and CoST [23] utilize both Fourier and wavelet transforms to achieve robust performance across various datasets by detecting global periodic patterns and local specifics. The wavelet domain presents a unique avenue for enhancing interpretability and predictive accuracy in time series analysis [24], as evidenced by the development of sophisticated network structures, the multilevel wavelet decomposition network [25], [26]. In addition, there are some advances in the application of wavelet transform in the fields of speech and biological signals [27], [28]. Our methodology capitalizes on these advancements by employing wavelet transforms to effectively decompose signals into their fundamental frequencies and details, leveraging the inherent strengths of both low- and high-frequency signals to improve both the accuracy and interpretability of forecasts.

III. Method

In this section, we provide a detailed explanation of our novel approach. WaveMLP entails transforming the time domain into the wavelet domain, thereby enabling a comprehensive examination of the time-series distribution from diverse frequency perspectives. The complete process of our model and the corresponding pseudocode of WaveMLP are presented in Fig. 1 and Algorithm 1, respectively.

Problem Definition: Typically, time-series data manifest as multivariate variables, prompting research into multivariate time-series forecasting. Given a historical observation data $X = \{x_1, \ldots, x_l\} \in \mathbb{R}^{L\times N}$, predicting future $\hat{L}$ time steps $Y = \{y_1, \ldots, y_{\hat{l}}\} \in \mathbb{R}^{\hat{L}\times N}$, where $X_{l,:}$ is the multivariate data recorded at time l and $X_{:,n}$ is the entire sequence corresponding to the variable with index n.

A. Wavelet Domain for Sensor Signal Analysis

Wavelet transform facilitates the decomposition of time series into low-frequency components and high-frequency components, effectively capturing both frequency and time information. This capability enables wavelet transform to discern the distribution characteristics of time series, thereby exhibiting significant potential in industrial measurement systems. When applying the discrete wavelet transform (DWT) to real data, a time series of length L is typically decomposed into two equally long feature sequences of length $L/2 + C$, where C accounts for any additional coefficients introduced by the transform, based on the selected wavelet function.

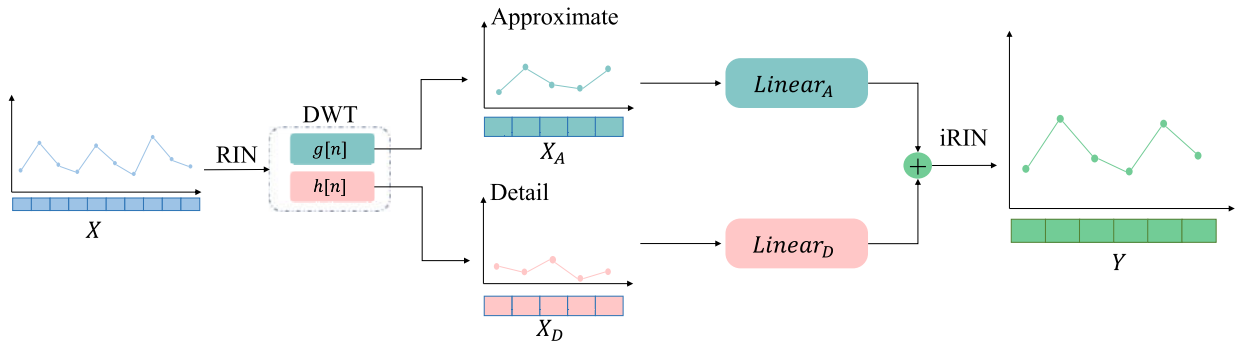

Fig. 1. Process of WaveMLP. The input time series is first normalized to zero mean using reversible instance normalization (RIN). It is then decomposed via Haar wavelet transform (DWT) into low- and high-frequency components. Each component is processed by an independent linear projection layer. The transformed features are fused through a GLU. Finally, the output is denormalized via inverse RIN (iRIN) to produce the prediction in the original scale.

Algorithm 1 WaveMLP

Require: historical observation data $X = \{x_1, \ldots, x_l\} \in \mathbb{R}^{L\times N}$; input length L; predicted length S; variates number N; LF (Wavelet low-pass filter), HF (Wavelet high-pass filter).

▷ Reversible instance normalization.
$X \leftarrow RevIN(X)$ ▷ $X \in \mathbb{R}^{L\times N}$
▷ Low-pass filter coefficients $= \begin{bmatrix}1 & 1\end{bmatrix}$.
$X_A \leftarrow LF(X)$ ▷ $X_A \in \mathbb{R}^{\frac{L}{2}\times N}$
▷ High-pass filter coefficients $= \begin{bmatrix}1 & -1\end{bmatrix}$.
$X_D \leftarrow HF(X)$ ▷ $X_D \in \mathbb{R}^{\frac{L}{2}\times N}$
▷ Multi-layer Perceptron works on the last dimension.
$X_A, X_D \leftarrow X_A^\top, X_D^\top$ ▷ $X_A \in \mathbb{R}^{N\times\frac{L}{2}}$, $X_D \in \mathbb{R}^{N\times\frac{L}{2}}$
▷ Optional gating mechanisms
if $isGLU == True$ **then**:
    ▷ Applying linear projection $\frac{L}{2} \to 2S$.
    $Y_A, Y_D = Linear(X_A), Linear(X_D)$ ▷ $Y_A \in \mathbb{R}^{N\times 2S}$, $Y_D \in \mathbb{R}^{N\times S}$
    $Y \leftarrow GLU(Y_A, Y_D)$ ▷ $Y \in \mathbb{R}^{N\times S}$
else
    ▷ Applying linear projection $\frac{L}{2} \to S$.
    $Y_A, Y_D = Linear(X_A), Linear(X_D)$ ▷ $Y_A \in \mathbb{R}^{N\times S}$, $Y_D \in \mathbb{R}^{N\times S}$
    $Y \leftarrow Y_A + Y_D$ ▷ $Y \in \mathbb{R}^{N\times S}$
end if
$Y \leftarrow iRevIN(Y^\top)$ ▷ $Y \in \mathbb{R}^{S\times N}$
return Y ▷ Return the prediction result Y

Importantly, this decomposition is reversible and lossless. The decomposition is represented as follows:

$$S(t) = \sum_k A_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j>j_0}\sum_k D_{j,k}\psi_{j,k}(t) \tag{1}$$

where $S(t)$ is a time series, and $A_{j,k}$ and $D_{j,k}$ are the wavelet coefficients representing the low- (approximation) and high-frequency (detail) components, respectively. These coefficients are obtained by taking the inner product of the original signal with the scaling function φ and wavelet functions ψ as follows:

$$A_{j_0,k} = \left\langle S(t), \varphi_{j_0,k}(t)\right\rangle \tag{2}$$

$$D_{j,k} = \left\langle S(t), \psi_{j,k}(t)\right\rangle \tag{3}$$

where the scale function $\varphi_{j,k}(t)$ and wavelet function $\psi_{j,k}(t)$. It can be expressed as follows:

$$\varphi_{j,k}(t) = 2^{\frac{j}{2}}\varphi\left(2^j t - k\right) \tag{4}$$

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^j t - k\right). \tag{5}$$

Therefore, by continuously compressing and translating the coverage signal, the signal representations of various frequency bands are obtained. So, DWT is a series of high- and low-pass filter banks. When the filter step size is selected as 2, the length of each iteration of the filter is halved.

The lossless nature of the DWT is guaranteed by the properties of perfect reconstruction filter banks [29]. The decomposition process defined by (2) and (3) is mathematically invertible. The original signal $S(t)$ can be exactly reconstructed from its wavelet coefficients $A_{j,k}$ and $D_{j,k}$ through the inverse discrete wavelet transform (iDWT)

$$S(t) = \sum_k A_{j_0,k}\tilde{\varphi}_{j_0,k}(t) + \sum_{j>j_0}\sum_k D_{j,k}\tilde{\psi}_{j,k}(t). \tag{6}$$

$\tilde{\varphi}$ and $\tilde{\psi}$ are the synthesis scaling and wavelet functions, dual to the analysis functions φ and ψ. For orthogonal wavelets, the Haar wavelet used in this work, the analysis and synthesis filters are identical. This reversibility ensures that no information is lost during the decomposition process, which is a critical foundation for WaveMLP. In practical numerical implementations, the discrete wavelet transform (DWT) and its inverse (iDWT) are reversible to within machine precision, making them effectively lossless for applications, such as signal processing and forecasting. Let $\tilde{X}$ denote the finite-precision representation of the original signal X. The numerical DWT and iDWT operations can be modeled

$$\tilde{X}' = \text{iDWT}\left(\text{DWT}\left(\tilde{X} + \delta_1\right) + \delta_2\right) + \delta_3 \tag{7}$$

where δ_1 is the initial quantization error of $\tilde{X}$, δ_2 is the round-off error introduced during DWT coefficient computation, and δ_3 is the round-off error from the iDWT computation. For orthogonal wavelets (e.g., Haar), the DWT and iDWT are orthogonal transformations, represented by matrices W and W^T, respectively, satisfying $W^T W = I$. Thus, the numerical reconstruction becomes

$$\tilde{X}' = W^T\left(W\left(\tilde{X} + \delta_1\right) + \delta_2\right) + \delta_3$$

$$= \tilde{X} + \delta_1 + W^T \delta_2 + \delta_3. \tag{8}$$

The total reconstruction error E is then

$$E = \tilde{X}' - \tilde{X} = \delta_1 + W^T \delta_2 + \delta_3. \tag{9}$$

Due to the orthogonality of W^T, the L2 norm is preserved ($\|W^T \delta_2\| = \|\delta_2\|$), and the norm of the total error is bounded by the following equation:

$$\|E\| \leq \|\delta_1\| + \|\delta_2\| + \|\delta_3\|. \tag{10}$$

This analysis demonstrates two key properties: 1) the orthogonal transform structure ensures that errors (e.g., δ_2) are not amplified during reconstruction and 2) the total error is a linear combination of the initial and computational errors. The magnitudes of δ_2 and δ_3 are determined by the unit round-off error u of the floating-point format (e.g., $u \approx 2.22 \times 10^{-16}$ for double precision). For the Haar transform, which involves minimal arithmetic operations, the accumulated error is of order $O(u\|X\|)$. Therefore, when using double-precision arithmetic, the theoretical reconstruction error is negligible in practice.

B. WaveMLP

WaveMLP consists of three components: wavelet transform, multilayer perceptron (MLP), and GLU. The wavelet transform effectively analyzes various signals, captures nonlinear changes, and has low resource consumption, making it easy to deploy. Compared to attention-based models, MLP has fewer parameters and faster inference speed, which meets the requirements of lightweight design. The GLU is used to promote nonlinear fusion of low- and high-frequency components and prevent overfitting [30]. The wavelet transform decomposes the time series j times to obtain 2^j feature sequences. The first one represents the approximation of the signal in the low-frequency part, while the remaining $2^j - 1$ represents the detailed features of the signal in different frequency bands in the high-frequency part. We use Haar wavelet as the wavelet function to construct the filter, but since the vanishing moment of Haar wavelet is 1, we only perform one decomposition. The transformation matrix is represented as follows:

$$\text{Haar} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{11}$$

Following the wavelet transform, two feature sequences are generated: an approximate representation of the time series and detailed features. These sequences are then processed by two separate MLPs: a low-frequency MLP and a high-frequency MLP. The low-frequency MLP learns a projection transformation, mapping the approximate representation from the preceding L time steps to the predicted $\hat{L}$ time steps, thereby capturing the contour features of the predicted sequence. Conversely, the high-frequency MLP learns an analogous projection transformation, mapping the detailed features from the past L time steps to the predicted $\hat{L}$ time steps, thereby capturing the detailed features of the predicted sequence. The fusion of these two transformed feature sequences yields the final prediction sequence. The overall process can be expressed as the following formula:

$$\begin{aligned} \text{DWT}(X) &= \sum_{k=0}^{K-1} \left[X[2n-k]\, g[k], X[2n-k]\, h[k] \right] \\ X_A, X_D &= \sum_{k=0}^{K-1} X[2n-k]\, g[k], \sum_{k=0}^{K-1} X[2n-k]\, h[k] \\ Y_A, Y_D &= \text{Linear}(X_A), \text{Linear}(X_D) \\ \hat{Y} &= \text{GLU}(Y_A, Y_D) \end{aligned} \tag{12}$$

where $X_A \in \mathbb{R}^{L/2 \times N}$ is the low-frequency component of X, $X_D \in \mathbb{R}^{L/2 \times N}$ is the detailed feature of X, DWT is a Haar transformation performed along the time direction, and the two MLPs project X_A and X_D as $Y_A \in \mathbb{R}^{\hat{L} \times N}$, $Y_D \in \mathbb{R}^{\hat{L} \times N}$, and $\hat{Y}$ are the outputs.

C. GLU

To achieve an adaptive and nonlinear fusion of the low- and high-frequency components processed by their respective MLPs, we employ a GLU [30]. The GLU mechanism acts as a learned, dynamic gate that controls the information flow from each component, enhancing the model's representational capacity and robustness while mitigating the risk of overfitting. The outputs from the low- and high-frequency MLP branches, denoted as Y_A and Y_D, are combined as an input to the GLU. The fusion process is defined as follows:

$$\text{GLU}(Y_A, Y_D) = Y_A + Y_D \otimes \sigma(Y_A) \tag{13}$$

where $\otimes$ represents the elementwise Hadamard product.

The choice of a gated linear unit for fusion is motivated by its ability to perform *adaptive* and *nonlinear* integration of the low- and high-frequency components. Unlike simple additive ($Y_A + Y_D$) or multiplicative ($Y_A \odot Y_D$) fusion, which applies a fixed operation, the GLU employs a learned gating mechanism. The sigmoid gate $\sigma(Y_A)$ generates a dynamic weight between 0 and 1 for each element in the low-frequency component. This gate effectively acts as a filter that decides "how much" of the detailed, high-frequency information Y_D should be retained at each temporal point, conditioned on the current context provided by the low-frequency trends. Therefore, the GLU provides a more sophisticated and expressive fusion strategy compared to static baselines, enabling WaveMLP to robustly handle the complex dynamics present in real-world sensor signals.

IV. Experiments

Experiments on eight real-world time-series benchmarks compare them with SOTA models to evaluate the performance and efficiency of WaveMLP.

Datasets: All datasets are publicly available and widely used real-world datasets from different fields, mainly including electricity, traffic, weather, exchange rates, and ETT [19].

Baselines: We compare WaveMLP with SOTA time-series forecasting models to evaluate performance and efficiency, including: Dinear [4], FITS [5], TPATCN [9], informer [10],

TABLE I
COMPARISON OF HYPERPARAMETERS IN DIFFERENT MODELS ON THE ELECTRICITY DATASET. AMONG THEM, THE FLAG "-" INDICATES THAT THIS HYPERPARAMETER IS NOT REQUIRED

Models	learning_rate	patience	e_layers	n_heads	train_epochs
WaveMLP (ours)	0.0005	3	-	-	50
iTransformer	0.0001	3	2	8	10
PatchTST	0.0001	10	3	16	100
TPATCN	0.0001	3	1	8	50

reformer [11], PatchTST [12], iTransformer [14], FreTS [31], and CrossGNN [32].

Implementation details: According to FITS settings, we set the length of the input sequence to $T = 720$ and the length of the prediction sequence to $S \in \{96, 192, 336, 720\}$. At the same time, in order to avoid information leakage, we apply normalization to each time slot rather than to the overall data. Use mean squared errors (MSEs) and mean absolute errors (MAEs) as evaluation metrics. All experiments were conducted on a single NVIDIA RTX 4090 24-GB GPU in PyTorch. All experiments were conducted using the same settings to ensure fairness. Parameter configuration is displayed in Table I. A pivotal advantage of WaveMLP is its minimal reliance on hyperparameter tuning, necessitating only three common parameters: a standard learning rate (0.0005), conventional early stopping patience (3), and a maximum epoch limit (50). Unlike complex models requiring extensive optimization of architecture-specific parameters, WaveMLP's lightweight design eliminates this overhead. Its performance remains robust across standard hyperparameter values, as validated in our ablation study, underscoring its efficiency and ease of deployment.

A. Comparison With SOTA Models

In this section, we first show a comparison of WaveMLP and SOTA models in efficiency and then present a comparison between WaveMLP and SOTA models in long-term predictions, in order to evaluate the performance of our model in prediction tasks.

1) Comparison of Efficiency With SOTA Models: We evaluate the efficiency of WaveMLP in terms of model parameters, inference time, and training time. Tables II and III compare the parameter counts, multiply-accumulate operations (MACs) [33], inference time, and training time across various models on multiple datasets, with a fixed input length of $T = 720$ and prediction lengths $S \in \{96, 192, 336, 720\}$. In terms of model complexity and deployment efficiency, WaveMLP exhibits significantly fewer parameters and lower computational cost (measured in MACs) compared to other models. This lightweight design translates into faster inference: the inference time of WaveMLP is approximately half that of FITS, demonstrating its advantage in real-time or resource-constrained applications. Regarding training efficiency, WaveMLP also achieves superior performance. The training time is reported as the average per epoch, and across most datasets, WaveMLP converges notably faster than competing models. Compared to linear models, WaveMLP trains over 40% faster, and its training time is only about 10% of that required by large-parameter models. This high training efficiency is attributed to the use of filter banks in the DWT, which decomposes the signal into multiple frequency bands through simple addition and subtraction operations—avoiding the need for complex, computationally intensive transformations. Overall, WaveMLP achieves an excellent balance between accuracy and efficiency, with low parameter count, minimal inference latency, and rapid training convergence, making it highly suitable for practical industrial forecasting applications. The effectiveness of edge measurement systems depends on end-to-end latency, which includes sensor acquisition, interface transmission, computational processing, and actuator response. While measuring interface-specific latencies (such as PCIe or USB) requires specialized hardware instrumentation, the computational efficiency of WaveMLP fundamentally alters the latency profile in edge deployment settings. WaveMLP achieves inference latencies of 3 ms in CPU. When compared with typical interface latency characteristics—where USB 2.0/3.0 usually shows 1–10-ms round-trip latency and PCIe operates in the microsecond range—WaveMLP's computational overhead becomes secondary to interface transmission delays. This reversal of the conventional latency hierarchy is especially significant: when model inference latency falls below or matches the minimum interface latency, computational efficiency becomes the dominant factor in achieving real-time performance. The exceptional efficiency of WaveMLP, evidenced by its minimal parameter count, MACs, and inference latency, strongly suggests its suitability for real-time operating systems (RTOSs). A fundamental requirement for RTOS is a predictable and short worst case execution time (WCET) to ensure a deterministic task loop time (T_{loop}). The consistently low and stable computational load of WaveMLP directly translates to a favorable WCET profile. This makes it a promising candidate for deployment as a high-priority real-time task, capable of meeting stringent timing constraints in edge instrumentation. In contrast, models with large and variable computational graphs would struggle to provide such timing guarantees, even on an RTOS.

2) Comparison of Performance With SOTA Models: The results of long-term time-series prediction are shown in Table IV, with a fixed input length of $T = 720$ and prediction length of $S \in \{96, 192, 336, 720\}$. By comparing WaveMLP with the highly competitive SOTA model, our model achieved satisfactory performance. Our model achieves consistently superior performance. In terms of prediction tasks, WaveMLP performs satisfactorily and reaches the current SOTA level. Across 57 tasks (MSE/MAE metrics), WaveMLP achieves first or second place in all cases. Notably, WaveMLP exhibited exceptional robustness in high-frequency financial data due to its dynamic weight adaptation, which mitigates overfitting to transient market fluctuations. For seasonal meteorological data (Weather), it achieved superior MAE compared to linear models, highlighting its capacity to model periodic patterns. These results validate its lightweight architecture and dynamic fusion strategy, which balances local-detail extraction with global-trend modeling.

TABLE II
PARAMETER QUANTITY, MACS, AND AVERAGE INFERENCE TIME REQUIRED FOR WAVEMLP AND OTHER MAINSTREAM MODELS

Attribute	**WaveMLP**	FITS	DLinear	iTransformer	Informer	CrossGNN	Reformer	PatchTST	FreTS	TPATCN
Parameters (K)	**40**	43	138	304	704	730	1100	1300	2300	1500
MACs (G)	**0.17**	0.44	1.42	3.11	6.42	9.56	8.64	9.8	30.32	11.2
Infer time (ms)	**0.30**	0.79	0.43	1.2	2.5	3.4	2.7	1.3	2.8	2.3

TABLE III
AVERAGE TRAINING TIME PER EPOCH ON DATASETS (IN SECONDS). THE BEST RESULT IS HIGHLIGHTED IN BOLD (INPUT LENGTH $L = 720$ AND PREDICTION HORIZON $T \in \{96, 192, 336, 720\}$)

Dataset	Weather				Electricity				Traffic				Exchange			
Horizon	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720
FreT	59.44	59.39	59.37	59.18	363	352	358	359	582	578	588	584	4.16	4.11	3.94	3.75
TPATCN	30.82	31.46	36.57	38.91	44.54	47.55	56.27	66.21	52.72	58.45	67.13	82.74	3.93	3.76	3.86	3.98
iTranformer	13.28	18.13	13.27	14.24	19.12	20.33	22.58	28.62	35.79	37.03	39.84	49.23	2.79	2.74	2.70	2.70
DLinear	7.35	7.44	8.30	9.19	15.43	17.37	19.92	25.48	24.78	28.26	33.02	40.77	1.56	1.63	**1.53**	1.58
FIT	8.02	8.31	8.56	9.68	19.93	21.74	24.08	32.09	32.49	35.56	43.44	59.93	1.57	**1.52**	**1.53**	**1.56**
WaveMLP (Ours)	**7.23**	**7.35**	**7.78**	**8.75**	**15.05**	**16.61**	**18.89**	**25.28**	**24.27**	**26.58**	**30.20**	**40.01**	**1.53**	1.58	1.59	1.62

Dataset	ETT-h1				ETT-h2				ETT-m1				ETT-m2			
Horizon	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720
FreT	5.85	5.80	5.75	5.46	5.88	5.88	5.70	5.53	22.03	22.02	22.08	22.03	21.95	21.96	22.01	22.07
TPATCN	6.33	5.93	6.61	6.89	6.24	6.17	6.15	7.07	23.91	24.53	25.32	27.12	24.08	24.84	26.83	27.57
iTranformer	3.75	3.80	3.66	3.78	3.74	3.68	3.71	3.81	11.16	11.36	11.72	11.93	10.83	10.92	10.97	11.81
DLinear	1.93	**1.96**	2.02	2.15	1.99	1.99	2.00	2.18	5.27	5.45	5.83	6.22	5.28	**5.42**	5.68	6.12
FIT	2.03	2.05	2.18	2.21	2.00	1.98	2.04	2.08	5.70	5.92	6.23	6.47	5.72	5.85	5.91	6.47
WaveMLP (Ours)	**1.91**	1.99	**2.01**	**2.04**	**1.90**	**1.96**	**1.95**	**2.01**	**5.32**	**5.39**	**5.52**	**6.10**	**5.27**	5.46	**5.54**	**5.97**

B. Analysis

1) Different Decomposition Levels Analysis: We investigate the impact of multilevel wavelet decomposition on model performance, where decomposition depths are denoted as level-1, level-2, and level-3. As shown in Table V, increasing the decomposition level enables the model to better capture the hierarchical and essential features of the time series. However, deeper decomposition requires additional linear layers to process the resulting subbands, leading to a growing number of model parameters and increased computational overhead. Consequently, while higher levels may improve representational capacity, they also tradeoff efficiency, highlighting the need for a balanced choice in practice.

2) Different Wavelet Function Analyses: We evaluate several higher order wavelet bases—including db4, db8, sym2, and sym4 [34], [35]—to assess their impact on forecasting performance compared to the Haar wavelet. Although these wavelets offer improved regularity and multiresolution properties, they do not yield consistent gains in accuracy and introduce additional computational overhead. In contrast, the Haar wavelet maintains high predictive performance while remaining computationally lightweight, making it a practical and effective choice for temporal modeling. These findings are summarized in Table VI.

3) Visualization Analysis: We visualize the prediction results on different datasets in Fig. 2 and display the maximum and average errors. It can be seen that WaveMLP can effectively model signals and make accurate predictions in different scene datasets.

Furthermore, the weights of the linear layers are visualized to demonstrate that features, such as periodicity and trends, are effectively captured by the model, as shown in Fig. 3. The heatmaps depicting the linear layer weights of the WaveMLP model across eight diverse time-series datasets (exchange, traffic, weather, electricity, ETTh1, ETTh2, ETTm1, and ETTm2) reveal critical insights into its temporal dependency modeling and adaptive mechanism. A predominant diagonal dominance observed in datasets: exchange, traffic, and ETTm1 underscores the model's prioritization of recent input–output correspondences, aligning with short-term temporal dynamics, whereas broader gradient patterns in weather and electricity highlight WaveMLP's capacity to capture extended dependencies in systems with prolonged memory effects or non-Markovian behavior. The heterogeneous weight distributions—ranging from concentrated blue/yellow clusters in ETTh1 to diffuse gradients in ETTm2—demonstrate the model's multiscale sensitivity, dynamically emphasizing dataset-specific temporal structures through its hierarchical architecture. Notably, the amplified weight magnitudes near sequence boundaries (e.g., upper-right corners in traffic and lower-left in weather) suggest that WaveMLP strategically reinforces boundary-adjacent temporal regions, which often encode pivotal contextual information for accurate forecasting.

TABLE IV

LONG-TERM FORECASTING COMPARISON. THE BEST RESULTS ARE IN BOLD AND THE SECOND BEST ARE UNDERLINED. USE THE MEAN TO BALANCE THE DIFFERENCES IN DIFFERENT PREDICTION LENGTHS AND COUNT THE BEST AND SECOND RESULTS

Models		**WaveMLP(ours)**		DLinear		FITS		FreTS		iTransformer		Reformer		Informer		PatchTST		CrossGNN		TPATCN	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exchange	96	**0.083**	0.203	0.087	0.213	0.088	0.208	0.525	0.531	0.118	0.253	1.117	0.902	1.104	0.866	0.124	0.251	**0.082**	**0.198**	1.299	0.918
	192	**0.174**	**0.297**	0.196	0.337	0.181	0.302	0.958	0.731	0.232	0.355	1.158	0.913	1.175	0.875	0.287	0.390	0.183	0.300	1.132	0.815
	336	0.338	0.424	**0.269**	**0.387**	0.338	0.418	1.100	0.765	0.434	0.489	1.291	0.959	1.297	0.960	0.702	0.586	0.348	0.425	1.954	1.128
	720	1.025	0.762	**0.946**	0.738	0.964	**0.733**	2.518	1.188	1.050	0.750	1.530	1.051	1.165	0.908	1.371	0.867	1.221	0.828	1.493	0.972
	Avg	0.405	0.420	**0.374**	0.418	0.392	**0.415**	1.275	0.803	0.459	0.463	1.274	0.956	1.185	0.902	0.621	0.523	0.458	0.437	1.469	0.958
weather	96	0.168	0.221	0.170	0.230	0.168	0.223	0.154	0.214	0.169	0.222	0.398	0.421	0.217	0.294	**0.149**	**0.205**	0.162	0.218	0.468	0.490
	192	0.212	0.259	0.220	0.280	0.211	0.258	0.199	0.261	0.211	0.258	0.651	0.562	0.303	0.353	0.199	**0.248**	**0.197**	0.250	0.480	0.471
	336	0.260	0.297	0.258	0.310	0.257	0.294	0.253	0.309	0.273	0.302	0.641	0.554	0.485	0.477	**0.248**	**0.291**	0.249	0.294	0.510	0.498
	720	0.322	0.340	0.321	0.364	0.322	0.341	0.329	0.366	0.325	0.343	0.713	0.606	0.732	0.614	0.317	**0.335**	**0.313**	0.337	0.666	0.574
	Avg	0.245	0.279	0.242	0.296	0.239	0.279	0.233	0.287	0.244	0.281	0.600	0.535	0.434	0.434	**0.228**	**0.269**	0.230	0.274	0.531	0.508
Electricity	96	**0.133**	**0.230**	0.135	0.234	0.136	0.235	0.134	0.234	0.135	0.232	0.294	0.377	0.529	0.531	0.147	0.244	0.134	0.231	0.382	0.434
	192	**0.149**	**0.243**	0.150	0.249	0.151	0.248	0.151	0.249	0.153	0.251	0.298	0.377	0.567	0.541	0.152	0.245	**0.149**	0.245	0.385	0.441
	336	0.164	0.259	0.163	0.262	0.167	0.264	0.422	0.292	0.167	0.266	0.348	0.418	0.542	0.554	**0.162**	**0.258**	0.165	0.261	0.383	0.444
	720	0.204	**0.292**	0.199	0.297	0.205	0.296	0.455	0.312	**0.196**	**0.292**	0.334	0.406	0.547	0.556	0.168	0.261	0.204	0.294	0.388	0.445
	Avg	0.162	**0.256**	**0.161**	0.260	0.164	0.260	0.419	0.321	0.162	0.260	0.318	0.394	0.451	0.485	0.168	0.261	0.163	0.257	0.385	0.441
Traffic	96	0.386	**0.270**	0.387	0.274	0.390	0.274	0.388	0.390	**0.371**	0.275	0.663	0.362	0.845	0.487	0.382	0.278	0.394	0.283	0.823	0.476
	192	0.398	**0.274**	0.400	0.281	0.400	0.277	0.412	0.292	**0.382**	0.278	0.696	0.378	0.897	0.500	0.396	0.282	0.407	0.288	0.817	0.468
	336	0.411	**0.280**	0.412	0.287	0.412	0.281	0.422	0.299	**0.398**	0.285	0.698	0.376	1.463	0.819	0.404	0.288	0.441	0.295	1.487	0.804
	720	0.449	0.294	0.452	0.295	0.450	0.301	0.455	0.312	0.428	**0.285**	0.722	0.98	0.618	0.934	**0.412**	**0.292**	0.474	0.311	0.99	1.25
	Avg	0.411	**0.279**	0.412	0.284	0.413	0.283	0.419	0.323	**0.394**	0.280	0.694	0.378	1.22	0.685	0.398	0.285	0.429	0.294	1.029	0.749
ETT-h1	96	0.380	0.403	0.385	0.410	**0.377**	**0.400**	0.480	0.480	0.404	0.432	1.006	0.773	1.248	0.874	0.431	0.441	0.38	0.409	1.271	0.931
	192	0.421	0.431	0.427	0.437	0.412	**0.421**	0.552	0.530	0.465	0.473	1.025	0.779	1.246	0.870	0.497	0.486	**0.408**	0.427	2.018	1.178
	336	0.452	0.450	0.479	0.478	**0.430**	**0.436**	0.586	0.549	0.501	0.499	1.181	0.821	1.355	0.864	0.610	0.553	0.434	0.442	1.410	0.991
	720	0.470	0.480	0.527	0.533	**0.428**	**0.452**	0.645	0.577	0.601	0.563	1.126	0.818	1.297	0.885	0.720	0.597	0.454	0.471	1.428	0.999
	Avg	0.430	0.441	0.454	0.464	**0.411**	**0.427**	0.565	0.534	0.492	0.491	1.084	0.797	1.286	0.873	0.564	0.519	0.419	0.437	1.532	1.025
ETT-h2	96	**0.270**	**0.335**	0.296	0.362	0.271	0.336	0.428	0.454	0.368	0.393	2.774	1.306	3.023	1.44	0.318	0.384	0.280	0.352	7.562	2.506
	192	0.336	0.378	0.345	0.394	**0.331**	**0.374**	0.603	0.540	0.437	0.438	4.740	1.683	4.546	1.654	0.398	0.428	0.342	0.389	5.063	2.009
	336	0.358	0.399	0.455	0.460	**0.354**	**0.395**	0.532	0.508	0.453	0.455	4.275	1.656	3.780	1.536	0.394	0.432	0.364	0.407	3.886	1.758
	720	0.385	0.430	0.782	0.621	**0.378**	**0.422**	1.050	0.733	0.440	0.466	3.335	1.437	4.495	1.825	0.519	0.506	0.406	0.444	3.437	1.630
	Avg	0.337	0.385	0.469	0.459	**0.333**	**0.381**	0.654	0.558	0.424	0.438	3.781	1.520	3.961	1.613	0.407	0.437	0.348	0.398	4.987	1.976
ETT-m1	96	0.309	0.352	0.318	0.366	0.312	0.354	0.334	0.382	0.332	0.382	0.652	0.593	0.836	0.678	0.325	0.375	**0.306**	**0.349**	1.264	0.913
	192	0.343	0.372	0.350	0.383	**0.338**	**0.368**	0.365	0.397	0.364	0.400	0.780	0.644	0.833	0.670	0.360	0.400	0.342	0.370	1.617	1.034
	336	0.371	0.388	0.375	0.396	**0.367**	**0.385**	0.415	0.431	0.394	0.415	0.950	0.723	1.067	0.801	0.402	0.428	0.373	0.390	1.174	0.859
	720	0.421	0.416	0.427	0.427	**0.415**	**0.412**	0.497	0.488	0.447	0.446	1.114	0.803	1.413	0.964	0.571	0.493	0.419	0.417	0.964	0.730
	Avg	0.361	0.382	0.367	0.393	**0.358**	**0.379**	0.402	0.424	0.384	0.410	0.874	0.690	1.037	0.778	0.414	0.424	0.360	0.381	0.939	0.656
ETT-m2	96	**0.162**	**0.253**	0.167	0.259	**0.162**	**0.253**	0.188	0.278	0.185	0.268	0.753	0.678	0.512	0.547	0.179	0.267	0.165	0.257	5.795	1.941
	192	**0.216**	**0.290**	0.237	0.316	**0.216**	**0.290**	0.250	0.316	0.269	0.328	1.114	0.826	1.535	0.956	0.256	0.318	0.221	0.296	1.315	0.948
	336	0.270	0.327	0.282	0.342	**0.268**	**0.325**	0.312	0.353	0.334	0.362	2.218	1.158	2.171	1.117	0.320	0.366	0.272	0.334	1.939	1.090
	720	**0.350**	**0.378**	0.389	0.416	**0.350**	**0.378**	0.391	0.415	0.413	0.415	2.766	1.254	6.218	1.946	0.415	0.427	0.358	0.396	4.732	1.704
	Avg	**0.249**	0.312	0.268	0.333	**0.249**	**0.311**	0.285	0.340	0.300	0.343	1.712	0.979	2.609	1.141	0.292	0.344	0.254	0.320	3.445	1.421
$1^{st} + 2^{nd}$Count		**55**		19		52		7		13		0		0		18		12		0	
training time		**× 1**		× 1.1		× 1.4		× 8.2		× 2.1		> 9		> 9		> 9		> 9		× 4.1	
infer time		**× 1**		× 1.4		× 2.6		× 9.3		× 4		× 9		× 8.3		× 4.3		× 10.1		× 7.6	

TABLE V

METRICS OF MSE/MAE FOR DIFFERENT LEVELS IN WAVELET DECOMPOSITION. THE BEST RESULT IS HIGHLIGHTED IN BOLD, INPUT LENGTH $T = 720$, AND PREDICTION LENGTH $S = 96$

Decomposition level	Exchange	Weather	Electricity	Traffic	ETT-h1	ETT-h2	ETT-m1	ETT-m2
level-1	**0.089** / 0.211	**0.168** / **0.221**	**0.133** / 0.230	0.386 / 0.270	0.380 / 0.403	**0.270** / **0.335**	**0.309** / 0.352	**0.162** / 0.253
level-2	0.091 / 0.212	**0.168** / 0.222	**0.133** / **0.229**	0.386 / 0.277	0.377 / 0.401	0.276 / 0.337	**0.309** / **0.351**	0.163 / 0.253
level-3	**0.089** / **0.210**	0.168 / 0.223	**0.133** / **0.229**	**0.385** / **0.267**	**0.375** / **0.400**	0.275 / 0.337	**0.309** / **0.351**	**0.162** / **0.252**

Furthermore, the consistent diagonal patterns across ETT subseries validate the model's robustness in multivariate industrial settings, while stark contrasts in weight sparsity (e.g., sparse activations in exchange versus dense gradients in electricity) underscore its flexibility in adapting to diverse temporal modalities.

4) Hyperparameter Analysis: To quantitatively assess the robustness of WaveMLP to hyperparameter variations, we conduct an ablation study on the learning rate. The model exhibits consistent and superior performance across a range of commonly used values—0.0001, 0.0005, and 0.0009—indicating low sensitivity to this key hyperparameter. This stability reduces the need for extensive manual tuning and highlights the practicality of WaveMLP in real-world deployments, as further shown in Table VII.

5) Model Generalization Analysis: To demonstrate the scalability of WaveMLP, we integrate it into the baseline DLinear model and evaluate its performance on standard time-series

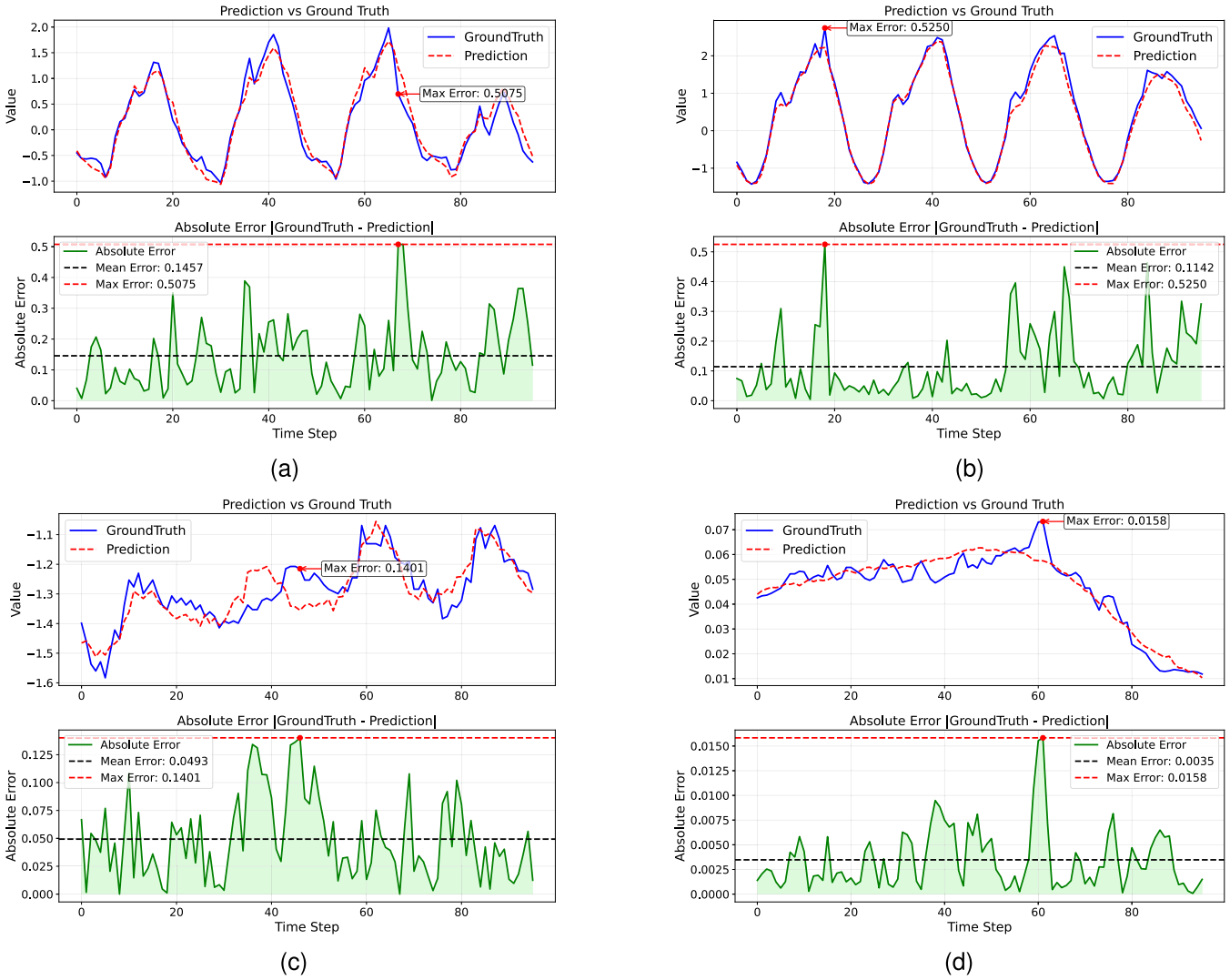

Fig. 2. Visualize the results and observe the differences between the predicted and actual values. (a) Electricity. (b) Traffic. (c) ETT-H1. (d) Weather.

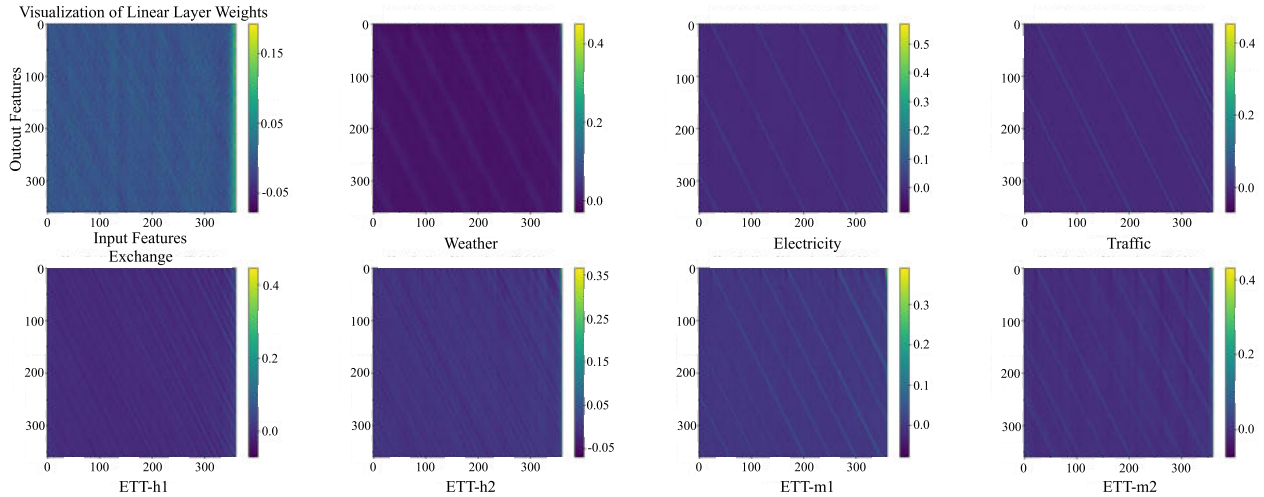

Fig. 3. Visualize the weights of the MLP for low-frequency components and further analyze the impact of low-frequency components.

forecasting benchmarks. The results, presented in Fig. 4, show consistent improvements, confirming that WaveMLP can effectively enhance the existing architectures with minimal modification. Furthermore, to validate the generalization capa-

TABLE VI

METRICS OF MSE/MAE FOR DIFFERENT WAVELET FUNCTIONS ANALYSIS. THE BEST RESULT IS HIGHLIGHTED IN BOLD, INPUT LENGTH $T = 720$, AND PREDICTION LENGTH $S = 96$

Wavelet function	Exchange	Weather	Electricity	Traffic	ETT-h1	ETT-h2	ETT-m1	ETT-m2
haar	**0.089 / 0.211**	**0.168 / 0.221**	**0.133 / 0.230**	**0.386 / 0.270**	0.380 / 0.403	**0.270 / 0.335**	**0.309 / 0.352**	**0.162 / 0.253**
db4	0.094 / 0.218	0.172 / 0.226	0.137 / 0.234	0.390 / 0.274	0.395 / 0.415	0.290 / 0.349	0.319 / 0.361	0.168 / 0.256
db8	0.100 / 0.226	0.174 / 0.228	0.138 / 0.233	0.391 / 0.274	0.397 / 0.415	0.297 / 0.354	0.323 / 0.364	0.171 / 0.258
sym2	0.099 / 0.223	0.169 / 0.223	0.135 / 0.231	0.388 / 0.271	**0.377 / 0.402**	0.282 / 0.343	0.311 / 0.353	0.164 / **0.253**
sym4	0.091 / 0.213	0.169 / 0.223	0.135 / 0.232	0.390 / 0.275	0.390 / 0.412	0.283 / 0.344	0.314 / 0.357	0.164 / **0.253**

TABLE VII

MSE/MAE ACROSS HYPERPARAMETER SETTINGS (BEST IN BOLD)

learning_rate	Exchange	Weather	Electricity	Traffic	ETT-h1	ETT-h2	ETT-m1	ETT-m2
0.0001	0.091 / **0.209**	0.170 / **0.221**	**0.132** / 0.232	**0.382** / 0.268	0.378 / **0.398**	0.273 / 0.340	0.310 / **0.349**	**0.162** / 0.254
0.0005	**0.089** / 0.210	0.168 / 0.223	0.133 / **0.229**	0.385 / **0.267**	0.375 / 0.400	**0.275 / 0.337**	**0.309** / 0.351	**0.162** / 0.252
0.0009	**0.089** / 0.212	**0.165** / 0.225	0.136 / 0.230	0.387 / 0.268	**0.374** / 0.403	0.277 / 0.344	0.311 / 0.353	0.164 / **0.251**

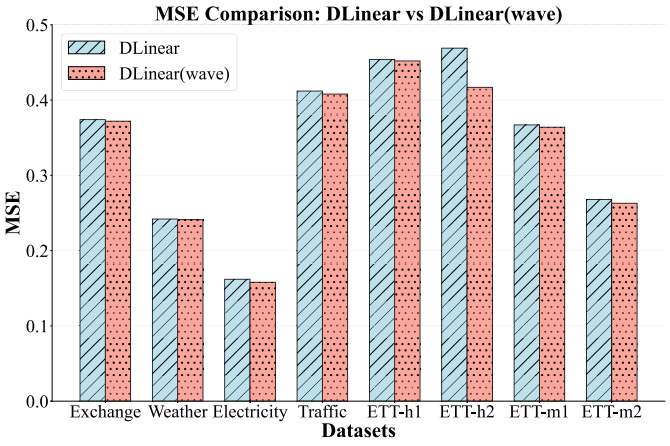

Fig. 4. Performance comparison of DLinear with WaveMLP integration.

TABLE VIII

PERFORMANCE COMPARISON IN SPEECH SIGNAL MODELING

models	WaveMLP(ours)	ReLU	GELU	PAU	APL	GR-KAN	KAN
MSE	**0.079**	0.154	0.117	0.090	0.121	0.085	0.081

TABLE IX

RECONSTRUCTION ERROR UNDER SINGLE-PRECISION (FP32) ARITHMETIC

Dataset	ETT-h1	ETT-h2	ETT-m1	ETT-m2
MSE	1.03e-12	3.27e-12	1.01e-12	2.95e-12
RMSE	1.02e-06	1.81e-06	1.00e-06	1.72e-06

bility of WaveMLP across domains, we apply it to a speech processing task under the experimental setup described in [36]. The performance gains achieved by WaveMLP, summarized in Table VIII, demonstrate its adaptability to different data modalities and model paradigms beyond time-series forecasting. Together, these experiments illustrate that WaveMLP is not only scalable to mainstream forecasting models but also generalizes well to diverse application scenarios.

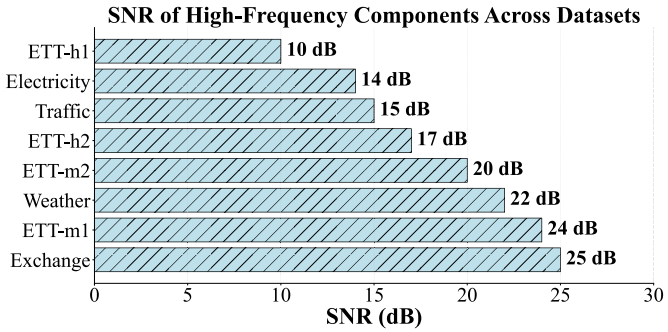

Fig. 5. SNR of high-frequency components across datasets.

6) Low- and High-Frequency Components Analysis: We analyze the impact of low- and high-frequency components on signal modeling in WaveMLP, while explaining the statement that high-frequency components are dominated by noise. To quantitatively validate the assumption that high-frequency components are dominated by low-energy details, we compute the signal-to-noise ratio (SNR) between low- and high-frequency components as follows:

$$\text{SNR (dB)} = 10\log_{10}\left(\frac{\|X_A\|_2^2}{\|X_D\|_2^2}\right) \tag{14}$$

where X_A and X_D are the low- and high-frequency coefficients from Haar DWT. As shown in Fig. 5, SNR values range from 10 to 25 dB. This implies that the high-frequency components account for only 0.3%–9% of the total signal power (since $P_{\text{HF}}/P_{\text{total}} = 1/(1 + 10^{\text{SNR}/10})$). We further isolate the low- and high-frequency components for experimentation, and the results are reported in Fig. 6. The outcomes show that using only low frequencies yields performance similar to WaveMLP, whereas using only high frequencies results in significantly inferior performance. This further indicates that the high-frequency components are dominated by noise.

7) Reconstruction Error Analysis: To empirically validate the practical losslessness of the DWT, we quantify the reconstruction error under single-precision floating-point arithmetic

TABLE X
NOISE ROBUSTNESS COMPARISON ACROSS DATASETS (AVERAGE MSE). BEST RESULTS ARE IN BOLD

Model	Exchange		Weather		ETTh1		ETTh2	
	Gaussian	Rayleigh	Gaussian	Rayleigh	Gaussian	Rayleigh	Gaussian	Rayleigh
DLinear	0.206	0.215	0.172	0.172	0.395	0.395	0.328	0.329
iTransformer	0.254	0.265	0.170	0.171	0.473	0.472	0.411	0.417
WaveMLP (Ours)	**0.150**	**0.156**	**0.169**	**0.169**	**0.389**	**0.388**	**0.298**	**0.299**

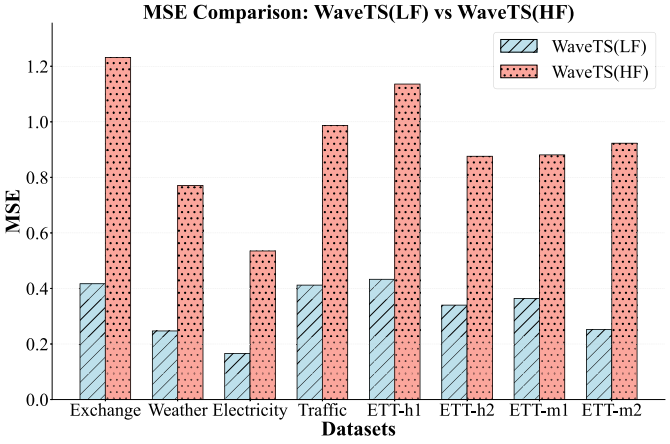

Fig. 6. Performance comparison between only low-frequency components—WaveMLP (LF) and only high-frequency components—WaveMLP (HF).

(FP32), reflecting various deployment scenarios. We perform a one-level Haar DWT followed by an immediate iDWT on each dataset and compute the MSE and RMSE between the original and reconstructed signals. The results across all datasets are summarized in Table IX, confirming that the reconstruction errors remain at the level of machine epsilon, which verifies the practical losslessness of the transformation in real-world implementations.

The results demonstrate exceptionally low reconstruction errors. The largest RMSE observed is 1.81×10^{-6} for the electricity dataset, which is negligible compared to the signal magnitude and inherent sensor noise. This provides conclusive empirical evidence that the wavelet decomposition and reconstruction process in WaveMLP is functionally lossless for sensor signal measurement.

8) Noise Robustness Analysis: The resilience of a model to noise is critical for its deployment in practical industrial settings, where sensor signals are often contaminated. To quantitatively assess this, we conduct a comprehensive noise robustness evaluation. We introduce two common types of noise—additive White Gaussian noise (AWGN) and multiplicative noise (Rayleigh distribution)—into the input sequences of the test sets from the exchange, weather, ETTh1, and ETTh2 datasets. The noise is applied at three severity levels, defined by input SNRs of 20, 15, and 10 dB. We compare WaveMLP against DLinear and iTransformer. For a concise and clear comparison, Table X reports the average MSE across the three noise levels for each model and noise type. The results lead to two primary conclusions. First, WaveMLP demonstrates superior robustness, achieving the lowest average MSE in all scenarios. This performance gap is especially pronounced under multiplicative noise, which is more disruptive to the temporal structure of the data. Second, the lightweight WaveMLP even surpasses DLinear, another lightweight model, indicating that its robustness stems not merely from model simplicity but from its principled wavelet-based architecture.

TABLE XI
ABLATION STUDY ON FEATURE FUSION STRATEGIES (MSE). BEST RESULTS ARE IN BOLD

Fusion Method	ETTh1	ETTh2	Traffic	Exchange
Addition	0.382	0.273	0.140	0.090
Multiplication	0.386	0.275	0.149	0.092
WaveMLP	**0.380**	**0.270**	**0.133**	**0.089**

9) Different Fusion Methods Analysis: To quantitatively validate the effectiveness of the GLU for feature fusion, we compare it against several common fusion strategies.

1) *Addition:* Output = $Y_A + Y_D$.
2) *Multiplication:* Output = $Y_A \odot Y_D$.
3) *Proposed GLU:* Output = $Y_A + Y_D \otimes \sigma(Y_A)$.

We evaluate these fusion variants on a subset of datasets (ETTh1, ETTh2, traffic, and exchange) under a prediction length of 96. The results, measured in MSE, are summarized in Table XI. The results clearly demonstrate the superiority of the proposed GLU fusion. It consistently achieves the lowest MSE across all tested datasets. The GLU's dynamic gating mechanism allows it to selectively integrate the most relevant information from both components, leading to more robust and accurate forecasts.

V. CONCLUSION

This article presents WaveMLP, a lightweight time-series forecasting model designed for instrument transformation measurement. The method introduces a wavelet-based approach that converts time-domain data into the wavelet domain, achieving strong performance in forecasting tasks. WaveMLP advances I&M through three contributions: it enables effective signal decomposition into low-frequency trends and high-frequency noise via wavelet analysis, facilitating real-time denoising and multiscale feature extraction; it employs a lightweight structure that reduces computational cost; and it demonstrates superior forecasting accuracy in comprehensive benchmarks.

REFERENCES

[1] L. Yang et al., "Submillimeter acoustic vibration measurement and monitoring using a single smartphone," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–15, 2025.

[2] R. Wang, Y. Fan, H. Yun, R. Rayhana, L. Bai, and Z. Liu, "Multiserver federated learning for building energy prediction with wind speed," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–14, 2025.

[3] S. Qiao et al., "Investigation and practical applications of seismo-electromagnetic joint exploration," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025.

[4] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 9, pp. 11121–11128.

[5] Z. Xu, A. Zeng, and Q. Xu, "FITS: Modeling time series with 10 k parameters," in *Proc. Int. Conf. Learn. Represent.*, 2023, pp. 1–17. [Online]. Available: https://openreview.net/forum?id=bWcnvZ3qMb

[6] Z. Zhou, X. Guo, Y.-J. Xiong, and C.-M. Xia, "Kalman-SSM: Modeling long-term time series with Kalman filter structured state spaces," *IEEE Signal Process. Lett.*, vol. 31, pp. 2470–2474, 2024.

[7] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long-and short-term temporal patterns with deep neural networks," in *Proc. Int. ACM Conf. Res. Devel. Inf. Retrieval (SIGIR)*, 2018, pp. 95–104.

[8] M. Liu et al., "SCINet: Time series modeling and forecasting with sample convolution and interaction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 5816–5828.

[9] T. Sun, S. Wang, H. Yang, J. Yang, and Z.-G. Hou, "Multisource information fusion for continuous prediction of joint angles using TCN combined with temporal pattern attention mechanism," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025.

[10] H. Zhou et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, May 2021, pp. 11106–11115.

[11] N. Kitaev, Ł. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–13. [Online]. Available: https://openreview.net/forum?id=rkgNKkHtvB

[12] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *Proc. Int. Conf. Learn. Represent.*, 2023, pp. 1–15. [Online]. Available: https://openreview.net/pdf?id=Jbdc0vTOcol

[13] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–22. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[14] Y. Liu et al., "ITransformer: Inverted transformers are effective for time series forecasting," in *Proc. Int. Conf. Learn. Represent.*, 2024, pp. 1–22. [Online]. Available: https://openreview.net/forum?id=JePfAI8fah

[15] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 753–763.

[16] T. Kim, J. Kim, Y. Tae, C. Park, J.-H. Choi, and J. Choo, "Reversible instance normalization for accurate time-series forecasting against distribution shift," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–11. [Online]. Available:https://openreview.net/forum?id=cGDAkQo1C0p

[17] Q. Wen, J. Gao, X. Song, L. Sun, H. Xu, and S. Zhu, "RobustSTL: A robust seasonal-trend decomposition algorithm for long time series," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 5409–5416.

[18] Q. Wen et al., "Transformers in time series: A survey," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, Aug. 2023, pp. 6778–6786.

[19] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, Mar. 2021, pp. 22419–22430.

[20] T. Zhou et al., "FiLM: Frequency improved legendre memory model for long-term time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 12677–12690.

[21] C. Shao, Y. Liu, C. Sun, R. Li, and J. Tan, "A high-precision roundness error separation method based on time-frequency-domain transformation for parameter optimization," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–10, 2025.

[22] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FedFormer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 27268–27286.

[23] G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. C. H. Hoi, "CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–22. [Online]. Available: https://openreview.net/forum?id=PilZY3omXV2

[24] J. Liu, W. Zhao, and S. Li, "Compressive sensing empirical wavelet transform for frequency-banded power measurement considering interharmonics," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025.

[25] J. Wang, Z. Wang, J. Li, and J. Wu, "Multilevel wavelet decomposition network for interpretable time series analysis," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 2437–2446.

[26] L. Sasal, T. Chakraborty, and A. Hadid, "W-transformers: A wavelet-based transformer framework for univariate time series forecasting," in *Proc. 21st IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2022, pp. 671–676.

[27] V. Lostanlen, D. Haider, H. Han, M. Lagrange, P. Balazs, and M. Ehler, "Fitting auditory filterbanks with multiresolution neural networks," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust. (WASPAA)*, Oct. 2023, pp. 1–5.

[28] A. Patwa, M. M. U. Rahman, and T. Y. Al-Naffouri, "Heart murmur and abnormal PCG detection via wavelet scattering transform and 1D-CNN," *IEEE Sensors J.*, vol. 25, no. 7, pp. 12430–12443, Apr. 2025.

[29] S. Mallat, *A Wavelet Tour of Signal Processing*. Amsterdam, The Netherlands: Elsevier, 1999.

[30] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 933–941.

[31] K. Yi et al., "Frequency-domain MLPs are more effective learners in time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 76656–76679.

[32] Q. Huang et al., "CrossGNN: Confronting noisy multivariate time series via cross interaction refinement," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 46885–46902.

[33] M. A. Nahmias, T. F. de Lima, A. N. Tait, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Photonic multiply-accumulate operations for neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–18, Jan. 2020.

[34] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, Oct. 1988.

[35] E. L. Lema-Condo, F. L. Bueno-Palomeque, S. E. Castro-Villalobos, E. F. Ordoñez-Morales, and L. J. Serpa-Andrade, "Comparison of wavelet transform symlets (2–10) and Daubechies (2–10) for an electroencephalographic signal analysis," in *Proc. IEEE 24th Int. Conf. Electron., Electr. Eng. Comput. (INTERCON)*, Aug. 2017, pp. 1–4.

[36] H. Li, Y. Hu, C. Chen, S. M. Siniscalchi, S. Liu, and E. S. Chng, "From KAN to GR-KAN: Advancing speech enhancement with KAN-based methodology," 2024, *arXiv:2412.17778*.