



OPEN Quantum-enhanced dual-layer graph attention network for time-series forecasting

Yongli Tang¹, Zhongqi Cai¹, Yue Zhang¹✉, Zhenlun Gao¹ & Jinxia Yu²

Time-series forecasting plays a pivotal role in domains such as traffic prediction, financial analysis, and energy consumption monitoring. However, real-world time series often exhibit intertwined patterns—trends, seasonality, and latent structures—that pose significant challenges to forecasting accuracy. This paper proposes a novel forecasting model named QFreqFormer, which stands for Quantum Frequency Transformer. It combines the Quantum Fourier Transform (QFT) with a Dual-Layer Graph Attention Network (D-PAD) to effectively tackle the complexities of time-series forecasting. The QFT module exploits quantum parallelism and superposition to decompose time-series data into frequency components, offering a compact spectral representation. To further enhance the model's ability to capture intricate multi-frequency patterns, we propose the Quantum Frequency Decomposition-Reconstruction (Q-FR-Q) module, which progressively separates high- and low-frequency components using quantum parallel processing. The D-PAD framework integrates Graph Convolutional Networks (GCNs) with attention mechanisms to dynamically model temporal dependencies across frequency layers. Experimental results on benchmark datasets demonstrate that the proposed model consistently outperforms state-of-the-art methods in terms of Mean Squared Error (MSE) and Mean Absolute Error (MAE) across various horizons. In addition, the model demonstrates strong transfer learning capability, underscoring its robustness and generalizability across heterogeneous forecasting scenarios. This study introduces a quantum-enhanced deep learning framework that improves both forecasting accuracy and computational efficiency, offering practical advantages in real-world applications.

Keywords Quantum-enhanced deep learning, Time-series forecasting, Graph attention network, Quantum Fourier transform (QFT), Frequency decomposition-reconstruction

Time-series forecasting plays a vital role in numerous real-world domains, such as traffic¹, finance², and energy³. Real-world time-series data typically involve a combination of trend, seasonality, and latent fluctuations, which are often intricately intertwined and difficult to disentangle, making accurate forecasting a challenging task. This paper proposes a novel forecasting model named QFreqFormer, which stands for Quantum Frequency Transformer. It combines the Quantum Fourier Transform (QFT) with a Dual-Layer Graph Attention Network (D-PAD) to effectively tackle the complexities of time-series forecasting.

Traditional decomposition methods have been widely adopted to extract trend and seasonal components based on predefined statistical assumptions^{4–7}. While such approaches offer interpretability and ease of implementation, they often struggle to capture complex, multi-scale patterns and non-stationary behaviors present in real-world time series⁸.

To enhance frequency-domain representation, researchers have explored quantum-inspired methods. The Quantum Fourier Transform (QFT) has been studied for its potential to decompose time-series signals into frequency components by leveraging quantum parallelism and superposition^{9–11}. The QFT offers a global frequency-domain perspective and has been employed as a tool for extracting broad-spectrum features—providing a shallow-level representation of the temporal structure¹². However, existing QFT-based methods exhibit several limitations. First, QFT emphasizes frequency extraction but lacks the capacity to model nonlinear dependencies or dynamic interactions between frequency components—an essential aspect of complex time-series modeling¹³. Second, reconstruction often relies on linear superposition of isolated frequency bands, limiting expressive power and reducing adaptability in scenarios with multi-scale or entangled temporal dependencies.

¹School of Software, Henan Polytechnic University, Jiaozuo 454000, Henan, China. ²School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, Henan, China. ✉email: y_zhang@hpu.edu.cn

To address these challenges, we introduce a Quantum Frequency Decomposition–Reconstruction–Decomposition (Q-FR-Q) module as the core of the model's deep representation learning. This module refines frequency components iteratively by decomposing raw input via QFT, reconstructing enhanced features, and re-decomposing signals to separate entangled patterns^{14,15}. Such iterative refinement improves disentanglement of multi-frequency dynamics and long-range temporal dependencies.

We further design a Dual-Layer Graph Attention Network (DL-GAN) that integrates QFT with graph-based attention mechanisms. This architecture enables localized feature extraction from decomposed components and adaptively models inter-frequency interactions through graph-relational learning^{16,17}.

The major contributions of this work are summarized as follows:

- We propose a QFT-based framework to capture global frequency structures in time-series data.
- We design a Q-FR-Q module to progressively refine and disentangle frequency components.
- We introduce a dual-layer graph attention module that models hierarchical inter-frequency dependencies.
- The framework unifies frequency-domain analysis with temporal graph learning for robust forecasting.
- Experiments on multiple benchmarks validate the superior performance of our model on both short- and long-term forecasting tasks.

The remainder of this paper is organized as follows. Chapter 1 reviews the related literature. Chapter 2 presents the proposed QFreqFormer model, including the QFT-based frequency encoding and the dual-layer graph attention mechanism. Chapter 3 describes the experimental setup, results, and analysis, covering generalization and ablation studies. Chapter 4 concludes the paper and outlines potential future research directions.

Related work

Transformer-based long-term time series forecasting

In recent years, a significant number of studies have been dedicated to the application of Transformer models in the context of long-term time series forecasting. LogTrans³ has been shown to effectively capture local information and substantially reduce spatial complexity by incorporating a convolutional self-attention layer and LogSparse design, a technique that has proven particularly effective in dealing with long series. Informer¹⁸ proposes the ProbSparse self-attention mechanism, which combines with the distillation technique to efficiently filter out the most important key values, which further reduces the computational complexity and improves the prediction accuracy. Autoformer⁵ utilizes decomposition and autocorrelation concepts from traditional time series analysis methods, employing an automatic approach to capture intrinsic patterns in the time series. This enhances the model's capacity to adapt to periodicity and trend. FEDformer¹⁵ employs Fourier augmentation structure, thereby enabling the model to attain linear complexity in the forecasting process and effectively enhancing the time series modelling capability. Pyraformer¹⁹ utilizes a pyramidal attention module that combines intra- and inter-scale connections to enhance the ability to capture multi-level time-series features while reducing computational complexity.

Despite these innovations, many Transformer-based models rely on local attention and overlook global temporal semantics, which limits their ability to capture long-range dependencies. To address this issue, we propose a novel framework that combines quantum frequency decomposition with a dual-layer graph attention mechanism. This approach captures both global frequency patterns and temporal dependencies, thereby improving long-term forecasting accuracy.

Quantum neural networks in time-series forecasting

Quantum Neural Networks (QNNs) have recently attracted attention for their potential to represent complex patterns and their compatibility with frequency-domain analysis^{20,21}. For example, QCLSTM²² extends LSTM by embedding quantum circuits to enhance memory, showing improved performance in chaotic time-series settings²³. Quantum Convolutional Neural Networks (QCNNs)²⁴ use quantum feature maps and entanglement to extract global periodic signals, demonstrating generalization on frequency-rich datasets²⁵. Hybrid models like Quanformer²⁶ integrate quantum circuits into Transformer blocks to capture amplitude-phase relationships and long-range dependencies²⁷. These methods leverage quantum parallelism and high-dimensional encoding^{28,29}.

However, most existing QNN models focus on static frequency representations or shallow circuits and lack dynamic mechanisms for modeling inter-frequency dependencies at multiple temporal scales.

Decomposition in time series analysis

Classical methods like ARIMA³⁰ and ETS³¹ decompose time series into trend and seasonal components. Modern deep learning models continue this tradition^{5,32,33}. For instance, Autoformer⁵ incorporates decomposition as internal blocks to capture trend and seasonal components progressively. DLinear³² combines decomposition with linear layers, offering a simple yet effective structure. LaST³³ applies variational inference to learn and disentangle trend and seasonal representations.

However, these methods often overlook the complex multi-frequency patterns inherent in real-world time-series data. To overcome the limitations of traditional decomposition techniques and the static nature of Quantum Fourier Transform (QFT), we propose a Quantum Frequency Decomposition–Reconstruction–Decomposition (Q-FR-Q) framework for time-series analysis. By leveraging the frequency extraction capabilities of QFT, the Q-FR-Q module introduces a progressive decomposition approach that refines frequency components. It first decomposes the data using QFT, reconstructs key components by amplifying significant frequency features, and then re-decomposes the signal for finer separation of high- and low-frequency components. This iterative refinement allows the model to handle complex, multi-frequency interactions, improving its ability to model time-varying dependencies and enhancing forecasting accuracy.

Methodology

Problem formulation

Given a multivariate time series comprising a lookback window of length T , denoted as $X = \{x_{t-T+1}, x_{t-T+2}, \dots, x_t\} \in \mathbb{R}^T$, where x_t is the value at time step t . The objective is to forecast future H values $\hat{X} = \{\hat{x}_{t+1}, \hat{x}_{t+2}, \dots, \hat{x}_{t+H}\} \in \mathbb{R}^H$. Therefore, the forecasting task can be formulated as follows.

$$\hat{X} = f(X, \Phi) \quad (1)$$

where f is the deep learning network for the task, and Φ denotes all learnable parameters of f .

Quantum Fourier transform module

In this work, we propose a quantum-enhanced feature extraction method via the Quantum Fourier Transform (QFT), integrated into a graph convolutional neural network (GCN) architecture. The primary function of the QFT module is to transform temporal data into quantum space, thereby capturing frequency-domain features that are beneficial for downstream graph convolutions. The QFT module operates through three fundamental stages: amplitude encoding, quantum Fourier transform, and measurement and post-processing Stage.

Phase 1: Amplitude Encoding :

The temporal features X are embedded into quantum states using amplitude encoding. In this phase, each temporal feature at time step t is mapped to a quantum state $|\psi_x\rangle$, mathematically expressed as:

$$|\psi_x\rangle = \frac{1}{\|X\|} \sum_{t=1}^T x_t |t\rangle \quad (2)$$

Here, x_t is the normalized temporal feature vector, and $|t\rangle$ represents the computational basis state corresponding to time step t . This encoding efficiently represents the time-series data in quantum space, preparing it for subsequent quantum operations.

Phase 2: Quantum Fourier Circuit : The quantum Fourier transform is implemented through a quantum circuit that performs a discrete Fourier transform on the quantum state, mapping time-domain information into the frequency domain. The quantum Fourier transform $\text{QFT}(|j\rangle)$ of a quantum state $|j\rangle$ is given by :

$$\text{QFT}(|j\rangle) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i \frac{jk}{N}} |k\rangle \quad (3)$$

The quantum circuit utilizes a sequence of quantum gates, including Hadamard gates, controlled-phase shift gates, and SWAP gates, to efficiently execute the Fourier transform and extract frequency-domain features from the input time-series data.

Phase 3: Measurement & post-processing :

After the quantum Fourier transform is performed, the quantum state is measured in the Z -basis, generating probability distributions $P(k)$ corresponding to the frequency components of the input data:

$$P(k) = |\langle k | \text{QFT}(\psi_x) \rangle|^2 \quad (4)$$

These quantum-enhanced frequency features $Q \in \mathbb{R}^{B \times N \times T}$ are obtained from the measurement results and concatenated with classical features for further processing.

Quantum Fourier transform in the frequency component region of time series data

The Quantum Fourier Transform (QFT) efficiently transforms time-domain data into the frequency domain, enabling the separation of high-frequency and low-frequency components based on their temporal fluctuations. In a classical Fourier transform, the signal $x(t)$ is decomposed into frequency components $X(f)$ as:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt \quad (5)$$

For discrete data, the Discrete Fourier Transform (DFT) is given by:

$$X_k = \sum_{t=0}^{T-1} x_t e^{-i2\pi \frac{kt}{T}}, \quad (6)$$

where X_k represents the frequency component at index k . Similarly, the Quantum Fourier Transform (QFT) of a quantum state $|\psi_x\rangle$ is expressed as:

$$\text{QFT}(|\psi_x\rangle) = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{jk}{2^n}} |k\rangle, \quad (7)$$

where $|k\rangle$ represents the frequency basis states.

The key advantage of QFT is its ability to distinguish between high-frequency and low-frequency components based on the energy states of the quantum system. High-frequency components correspond to quantum states with higher energy levels, while low-frequency components correspond to quantum states with lower energy levels. This distinction is achieved through the periodicity of the complex exponential terms $e^{2\pi ijk/2^n}$, where higher frequencies produce faster oscillations and lower frequencies produce slower oscillations.

Thus, QFT inherently separates the frequency components of the time-series data into distinct quantum states, with high-frequency information occupying higher energy states and low-frequency information occupying lower energy states. This allows for an efficient extraction of both rapid fluctuations and long-term trends in the data.

Dual-layer graph attention mechanism

A deep learning model is proposed that combines a two-layer graph convolutional neural network (GCN) and a multi-head attention mechanism. The primary function of the model is to propagate node features and to model graph structures in high-dimensional time-series data. The model's principal innovation is the combination of two-layer graph convolution and multi-head attention mechanism. Through multi-layer graph convolution, the model learns the relationship between nodes, and through multi-head attention, it enhances the dependency between them. Specifically, the model captures both local and global information propagation among nodes through two-layer graph convolution operations, while the weighted relationships among nodes are further refined using the multi-head attention mechanism after each layer of convolution, thus enhancing the flexibility and accuracy of the information propagation process.

In this section, we introduce the various modules of the model in detail, especially the information transfer process in the two-layer graph convolution. The analysis will focus on potential space embedding, the attention mechanism, feature embedding, graph convolution operation and its propagation process, as well as the design of the output layer, especially how to deal with hidden information and improve the model's ability to model node features through two-layer graph convolution.

Latent embedding

The input data is first mapped into a high-dimensional space through a potential space embedding layer, allowing each node's features to have a stronger representation in potential space. This step provides the initial feature representation for subsequent graph convolution operations and attention mechanisms. The transformation of the potential space embedding is as follows:

$$Z^{(0)} = \text{LatenEmbed}(X) \quad (8)$$

In this context, $Z^{(0)} \in \mathbb{R}^{B \times N \times K \times \text{latent_dim}}$ denotes the feature tensor extracted from the latent space, where B is the batch size, N is the number of nodes, K is the number of time steps or segments, and latent_dim represents the dimensionality of the latent space. The incorporation of these latent features enables a more intricate representation for the subsequent graph convolutional network, enhancing its ability to model complex spatiotemporal dependencies.

Multi-head attention mechanism

The introduction of a multi-head attention mechanism enables the calculation of distinct weighting relationships for each node. Each attention head generates a distinct representation of the inter-node dependencies through a variety of linear transformations and computes the similarity through a dot product, thereby generating an attention matrix. Specifically, given a feature tensor $Z^{(0)}$ in the latent space, we obtain a feature representation for each attention head as follows:

$$H_i^{(0)} = \text{Linear}_i(Z^{(0)}), \quad i = 1, 2, \dots, \text{num_heads} \quad (9)$$

where $H_i^{(0)} \in \mathbb{R}^{B \times N \times K \times \text{latent_dim}}$ denotes the representation of the i -th attention head. The similarity matrix computed for each attention head is obtained by means of a dot product and SoftMax normalization:

$$A_i = \text{SoftMax} \left(\frac{H_i^{(0)} H_i^{(0)\top}}{\sqrt{d_k}} \right), \quad i = 1, 2, \dots, \text{num_heads} \quad (10)$$

Finally, by averaging the attention matrices of all heads, we obtain the global attention matrix $A^{(0)}$:

$$A^{(0)} = \frac{1}{\text{num_heads}} \sum_{i=1}^{\text{num_heads}} A_i \quad (11)$$

This attention matrix reflects the weighted relationship between nodes, which plays a crucial role in the subsequent graph convolution process, especially when the information is propagated to capture the complex interrelationships between the nodes more effectively.

Feature embedding

Feature embedding is defined as the transformation of the input raw features into a hidden feature representation by linear transformation. This operation facilitates the model’s capacity to discern intricate patterns among nodes. The formula is given below:

$$X' = \text{FeatureEmbed}(X) \tag{12}$$

The original input feature is denoted by $X \in \mathbb{R}^{B \times N \times K \times T}$, where B is the batch size, N is the number of nodes, K is the number of temporal segments, and T is the number of features per segment. The embedded feature is denoted by $X' \in \mathbb{R}^{B \times N \times K \times \text{hidden_features}}$, where *hidden_features* represents the dimensionality of the projected feature space.

Dual-layer graph convolution and feature propagation

The fundamental principle of this model is predicated on the implementation of a two-layer graph convolution, a structural design that facilitates the effective capture of intricate relationships between node features. By employing two layers of graph convolution, the model is capable of sequentially propagating node information through the adjacency matrices $A^{(0)}$ and $A^{(1)}$.

First-layer graph convolution: In the initial layer of graph convolution, the feature information is propagated using the attention-derived adjacency matrix $A^{(0)}$ as follows:

$$X_1 = \sigma(A^{(0)} \cdot X; W^{(0)}) \tag{13}$$

The employment of a graph convolution operation facilitates the capture of relationships between local nodes, concomitant with the updating of node features through learnable weighting. The node features resulting from the first-layer graph convolution are denoted by $X_1 \in \mathbb{R}^{B \times N \times \text{latent_dim} \times \text{hidden_features}}$.

Second layer graph convolution: The second layer of graph convolution further propagates the features of the first layer, a process that captures deeper relationships between nodes on a global scale:

$$X_2 = \text{LayerNorm}(A^{(1)} \cdot X_1 W^{(1)} + X_1) \tag{14}$$

$X_2 \in \mathbb{R}^{B \times N \times \text{latent_dim} \times \text{hidden_features}}$ is the output of the second layer of graph convolution, which further refines the relationships between the nodes, allowing each node to perform feature updates from more levels of information.

The architecture of the proposed bi-layer attention-based graph convolution module is illustrated in Fig. 1. The D-PAD module consists of two cascaded graph attention layers designed to model hierarchical dependencies in the frequency domain. The first layer focuses on intra-frequency interactions by constructing a graph over components within the same frequency band, allowing fine-grained modeling of localized temporal patterns. The second layer captures inter-frequency relationships by enabling message passing between high- and low-frequency subgraphs, thereby integrating global context. Node representations are refined through attention-weighted aggregation and residual connections, allowing the model to effectively encode both local and global dependencies in a frequency-aware manner.

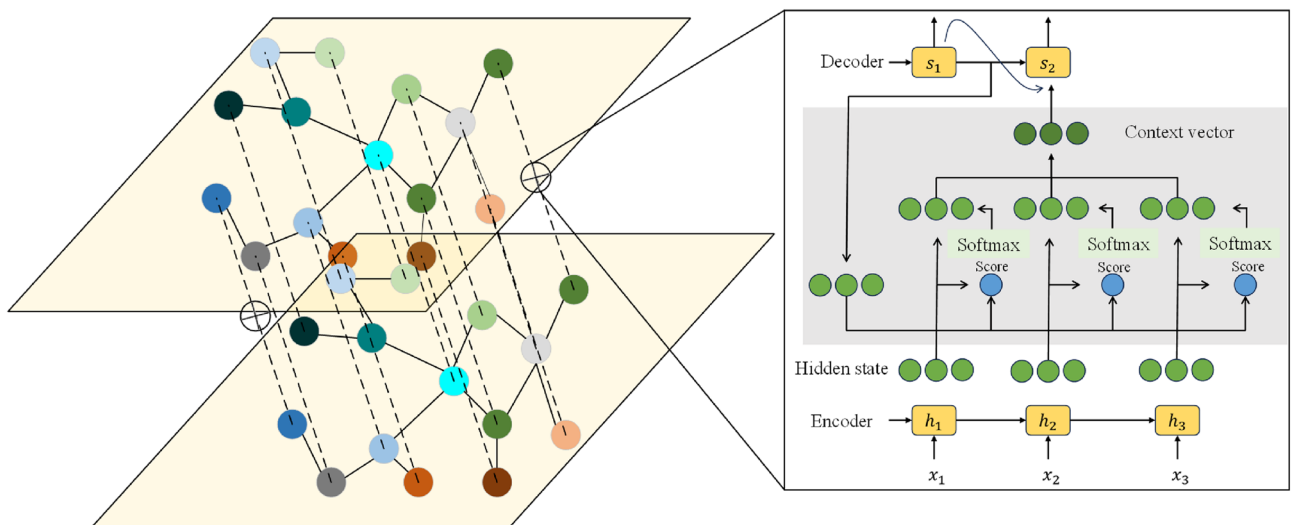


Fig. 1. Illustration of the two-layer graph network with attention-based decoding.

Model

The proposed model integrates Quantum Fourier Transform (QFT) with a Dual-Layer Graph Convolutional Network (GCN) to address key challenges in time series forecasting, particularly in capturing intricate multi-frequency patterns and long-term dependencies. The QFT module transforms temporal signals into quantum states, enabling efficient conversion from the time domain to the frequency domain. This facilitates the separation of high-frequency and low-frequency components, which are subsequently processed by a two-layer graph convolutional network. The first layer employs an attention-guided adjacency matrix for localized feature aggregation, capturing short-range dependencies among nodes, while the second layer refines global relationship modeling by propagating node features across the graph. By combining quantum-enhanced frequency decomposition with hierarchical graph convolution, the model effectively learns both fine-grained local patterns and broader global trends, significantly improving forecasting accuracy. The overall model architecture is illustrated in Fig. 2. The architecture of QFreqFormer comprises three core components. The Quantum Fourier Transform (QFT) module transforms the input time series into a compact frequency-domain representation, enabling the model to effectively capture periodic patterns and harmonic structures. The Quantum Frequency Decomposition-Reconstruction (Q-FR-Q) module further enhances this representation by isolating and refining high- and low-frequency components in parallel, thereby improving the model's ability to distinguish multi-scale variations. Finally, the Dual-layer Graph Attention Network (D-PAD) module captures complex temporal dependencies by modeling interactions between frequency components through hierarchical graph-based attention. The outputs of all components are concatenated and fed into a decoder to produce the final forecasts, achieving joint learning in both the frequency and temporal domains.

Experiments

Multivariate long-term forecasting

Datasets

We evaluate the performance of our proposed QFreqFormer on five commonly used benchmark datasets for long-term prediction: weather, traffic, electricity and ETT datasets (for ETT datasets we choose ETT h1 and ETT m1). Note that the ETT h1 dataset is relatively small-scale compared to others, while ETT m1 and weather are medium-sized datasets. Traffic and electricity, with more than 800 and 300 variables respectively, each containing tens of thousands of time points, are naturally large datasets. In general, smaller datasets contain

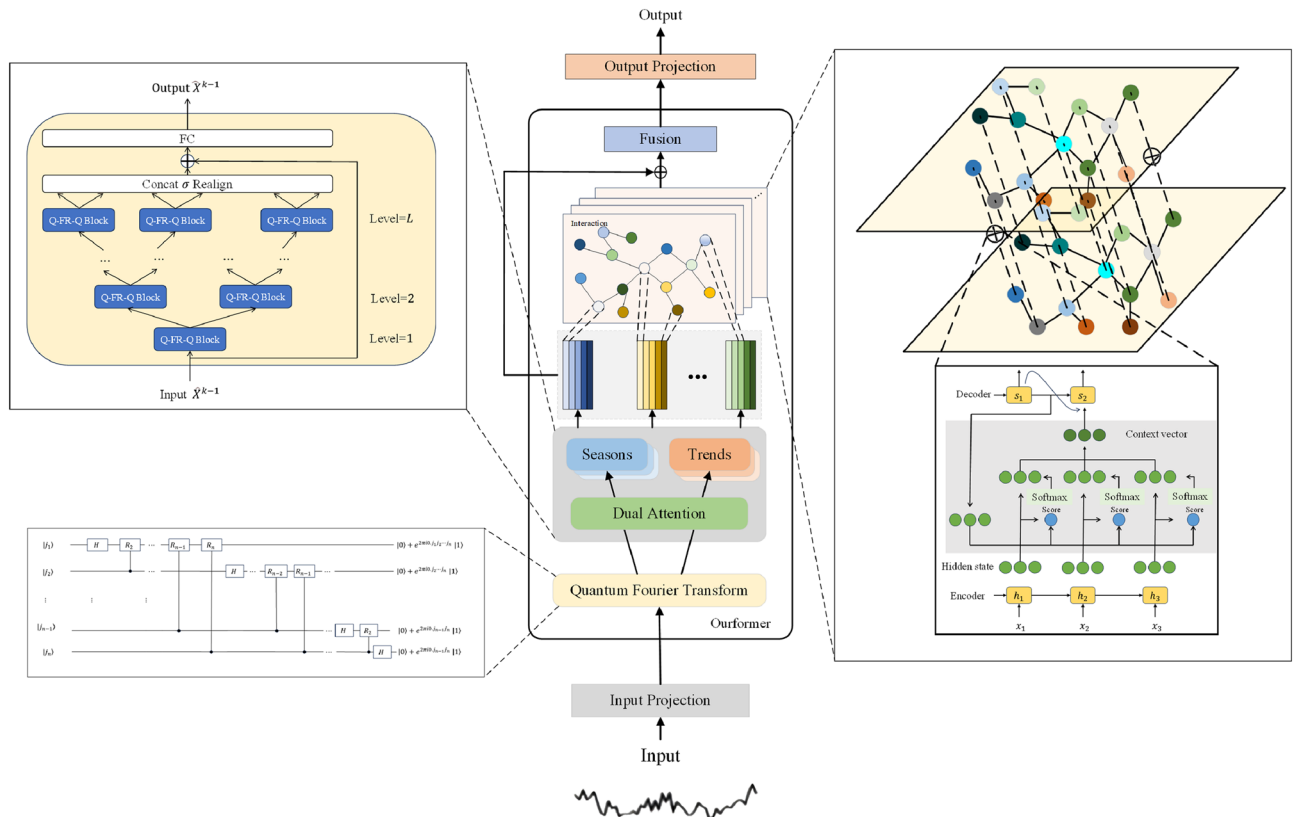


Fig. 2. Overall architecture of the proposed quantum-enhanced forecasting model, which incorporates Quantum Fourier Transform (QFT), dual-path attention for seasonal-trend decomposition, hierarchical Q-FR-Q blocks for multi-scale frequency refinement, and graph-based sequence decoding to comprehensively model temporal dynamics and inter-node dependencies.

more noise, while larger datasets exhibit more stable data distributions. Detailed statistics on the size of the dataset are shown in Table 1.

Implementation Details

To ensure the reproducibility of our results, the source code of QFreqFormer has been made publicly available at <https://github.com/caizhongqi/QNNformer/tree/main/QFreqFormer>. The entire framework is implemented in Python using PyTorch 1.9.0. Model training is conducted on a workstation equipped with three NVIDIA GeForce RTX 3080 Ti GPUs and an Intel Core i9-13900K processor, running Ubuntu 20.04 with CUDA version 11.1. We employ the Adam optimizer with an initial learning rate of 0.0001, a batch size of 32, and a weight decay of $1e-5$. To enhance training stability, gradient clipping with a maximum norm of 1.0 is applied. Early stopping is triggered if the validation loss does not improve for five consecutive epochs.

To address the issue of distributional shift commonly observed in time-series inputs, the Reversible Instance Normalization (RevIN) module is incorporated at both the input and output stages of the model. During the forward pass, RevIN normalizes each time series individually to stabilize the input distribution, and subsequently restores the original scale after prediction. This design enhances robustness under non-stationary conditions.

The Quantum Fourier Transform (QFT) and Quantum Frequency Decomposition–Reconstruction–Decomposition (Q-FR-Q) modules are implemented as modular, fully differentiable components within the PyTorch framework. The QFT module applies a unitary transformation inspired by quantum circuits to project time-domain sequences into the frequency domain. To ensure compatibility with modern GPU hardware, we adopt a real-valued matrix approximation of the complex-valued QFT, which preserves representational fidelity while maintaining computational efficiency.

The Q-FR-Q module adopts a progressive refinement strategy to disentangle complex multi-frequency structures and enhance temporal representation. It first performs a coarse-grained decomposition via the QFT to capture primary spectral components. This is followed by a reconstruction phase, in which salient frequency features are selectively amplified through a residual attention mechanism with learnable, temperature-scaled multi-head attention. Finally, a refined QFT projection is applied to the enhanced signals to further separate entangled frequency bands, enabling hierarchical modeling of frequency interactions.

Both QFT and Q-FR-Q modules are positioned between the input embedding layer and the dual-layer graph attention encoder (D-PAD), forming an integral part of the feature extraction pipeline. Their modular architecture ensures seamless integration with the rest of the model, and their differentiable formulation supports end-to-end optimization within a unified training process.

Baselines and metrics : We choose SOTA and representative LTSF models as our baselines, including Transformer-based models like PatchTST (2023), FEDformer (2022), Autoformer (2021), Informer (2021), in addition to two CNN-based models containing MICN³⁴ and TimesNet³⁵, with the significant MLP-based model DLinear (2023) to serve as our baselines. To assess the performance of these models, we employ widely used evaluation metrics: MSE and MAE.

Results : The experimental results, as shown in the table 2, demonstrate the performance of our model in long-term forecasting across multiple datasets, particularly in major benchmarks such as Traffic, Electricity, and Weather. In these datasets, QFreqFormer outperforms all baseline methods significantly. On other datasets, the model achieves superior performance across most or all prediction horizons. When compared to the current state-of-the-art model, D-PAD, QFreqFormer achieves a relative reduction of 3.7% in MSE and 2.9% in MAE. In comparison to the best-performing MLP-based model, DLinear, our model demonstrates a relative reduction of 11.1% in MSE and 8.7% in MAE. Moreover, when compared to the top CNN-based model, TimesNet, QFreqFormer achieves a significant relative reduction of 21.4% in MSE and 13.3% in MAE.

Comparison with classical spectral baselines : To further isolate and justify the contributions of the quantum-enhanced modules (QFT and Q-FR-Q), we follow the reviewer’s suggestion and introduce additional baseline models based on well-established classical decomposition techniques. Specifically, we incorporate four representative spectral methods: Empirical Mode Decomposition (EMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Variational Mode Decomposition (VMD), and Discrete Wavelet Transform (DWT). These classical modules replace the QFT and Q-FR-Q components in our model pipeline, while keeping the downstream prediction structure (e.g., GCN/GAT) consistent. This ensures a fair evaluation of decomposition capability alone.

As shown in Table 3, QFreqFormer consistently outperforms these spectral-based variants across major datasets such as Weather, Electricity, and Traffic. In particular, the average MSE reduction over the best-performing classical decomposition method (VMD-GAT) is 9.4%, and MAE reduction is 7.1%. These results further validate the advantage of quantum-based spectral reasoning in improving temporal representation for long-term forecasting.

Dataset	Weather	Traffic	Electricity	ETTh1	ETTm1	ETTh2	ETTm2
Variables	21	862	321	7	7	7	7
Timesteps	52,696	17,544	26,304	17,420	69,680	17,420	69,680
Frequency	10 Minutes	1 Hour	1 Hour	1 Hour	15 Minutes	1 Hour	15 Minutes

Table 1. Statistics of popular datasets for benchmark.

Dataset	Steps	EMD + GAT		CEEMDAN + GAT		VMD + GAT		DWT + GAT		QFreqFormer (Ours)	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Weather	96	0.301	0.301	0.296	0.296	<u>0.273</u>	<u>0.273</u>	0.281	0.281	0.153	0.194
	192	0.288	0.288	0.279	0.279	<u>0.268</u>	<u>0.268</u>	0.275	0.275	0.194	0.234
	336	0.264	0.301	0.257	0.296	<u>0.245</u>	<u>0.273</u>	0.250	0.281	0.238	0.272
	720	0.323	0.323	0.315	0.315	<u>0.305</u>	<u>0.305</u>	0.312	0.312	0.329	0.325
Traffic	96	0.334	0.334	0.328	0.328	<u>0.310</u>	<u>0.310</u>	0.319	0.319	0.341	0.204
	192	0.322	0.322	0.314	0.314	<u>0.297</u>	<u>0.297</u>	0.304	0.304	0.377	0.246
	336	0.311	0.287	0.303	0.278	<u>0.294</u>	<u>0.259</u>	0.296	0.267	0.384	0.242
	720	0.344	0.344	0.337	0.337	<u>0.322</u>	<u>0.322</u>	0.333	0.333	0.403	0.278
Electricity	96	0.296	0.296	0.286	0.286	<u>0.263</u>	<u>0.263</u>	0.275	0.275	0.124	0.207
	192	0.288	0.288	0.276	0.276	<u>0.255</u>	<u>0.255</u>	0.263	0.263	0.137	0.229
	336	0.284	0.284	0.271	0.271	<u>0.252</u>	<u>0.252</u>	0.263	0.263	0.156	0.252
	720	0.301	0.301	0.289	0.289	<u>0.268</u>	<u>0.268</u>	0.276	0.276	0.183	0.277

Table 2. Comparison with classical spectral decomposition baselines across different forecast horizons. The best results are highlighted in **bold**, and the second-best in underline.

Models	Steps	QFreqFormer		D-PAD		PatchTST		DLinear		Informer		FEDformer		Autoformer	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Weather	96	0.153	0.194	<u>0.162</u>	<u>0.201</u>	0.174	0.214	0.176	0.237	0.354	0.405	0.238	0.314	0.249	0.329
	192	0.194	0.234	<u>0.217</u>	<u>0.251</u>	0.221	0.254	0.220	0.282	0.419	0.434	0.275	0.329	0.325	0.370
	336	0.238	0.272	<u>0.239</u>	<u>0.273</u>	0.278	0.296	0.265	0.319	0.583	0.543	0.339	0.377	0.351	0.391
	720	0.329	0.325	<u>0.335</u>	<u>0.364</u>	0.343	0.362	0.353	0.362	0.916	0.705	0.389	0.409	0.415	0.426
Traffic	96	0.341	0.204	<u>0.359</u>	<u>0.236</u>	0.492	0.324	0.410	0.282	0.733	0.410	0.576	0.359	0.597	0.371
	192	0.377	0.246	<u>0.377</u>	<u>0.245</u>	0.487	0.303	0.423	0.287	0.777	0.435	0.610	0.380	0.607	0.382
	336	0.384	0.242	<u>0.391</u>	<u>0.253</u>	0.417	0.383	0.436	0.296	0.776	0.434	0.608	0.375	0.623	0.387
	720	0.403	0.278	<u>0.413</u>	<u>0.282</u>	0.542	0.337	0.466	0.315	0.827	0.466	0.621	0.375	0.639	0.395
Electricity	96	0.124	0.207	<u>0.128</u>	<u>0.218</u>	0.180	0.218	0.140	0.237	0.304	0.393	0.186	0.302	0.196	0.313
	192	0.137	0.229	<u>0.142</u>	<u>0.233</u>	0.188	0.275	0.153	0.249	0.327	0.417	0.197	0.311	0.211	0.324
	336	0.156	0.252	<u>0.161</u>	<u>0.254</u>	0.291	0.297	0.169	0.267	0.333	0.422	0.213	0.328	0.214	0.327
	720	0.183	0.277	<u>0.190</u>	<u>0.282</u>	0.328	0.345	0.203	0.301	0.351	0.427	0.233	0.344	0.236	0.342
ETTh1	96	<u>0.405</u>	<u>0.415</u>	0.379	0.402	0.422	0.432	0.452	0.465	0.462	0.475	0.427	0.952	0.374	0.368
	192	<u>0.436</u>	<u>0.430</u>	0.421	0.428	0.379	0.407	0.426	0.430	0.428	0.456	0.436	1.542	0.446	0.434
	336	0.390	0.409	<u>0.428</u>	<u>0.410</u>	0.421	0.400	0.477	0.409	0.410	0.486	0.465	1.642	0.447	0.479
	720	0.486	0.476	<u>0.507</u>	<u>0.605</u>	0.465	0.412	0.453	0.476	0.499	0.515	0.551	1.619	0.469	0.490
ETTh1	96	<u>0.297</u>	<u>0.300</u>	0.294	0.297	0.361	0.324	0.332	0.343	0.348	0.343	0.349	0.462	0.271	0.293
	192	<u>0.364</u>	0.366	0.362	<u>0.374</u>	0.374	0.392	0.378	0.429	0.392	0.423	0.442	0.586	0.458	0.486
	336	0.355	0.356	<u>0.390</u>	<u>0.367</u>	0.403	0.424	0.453	0.485	0.485	0.502	0.415	0.871	0.364	0.379
	720	0.428	0.437	<u>0.461</u>	<u>0.446</u>	0.475	0.480	0.514	0.521	0.525	0.538	0.561	1.267	0.520	0.519
ETTh2	96	0.297	0.343	<u>0.300</u>	<u>0.348</u>	0.294	0.343	0.180	0.251	0.256	0.260	0.264	0.260	0.462	0.271
	192	0.379	0.243	<u>0.248</u>	<u>0.250</u>	0.244	0.378	0.252	0.289	0.298	0.306	0.309	0.303	0.586	0.318
	336	0.377	0.283	<u>0.304</u>	<u>0.311</u>	0.317	0.382	0.324	0.331	0.327	0.342	0.348	0.342	0.871	0.364
	720	0.453	0.442	<u>0.460</u>	<u>0.475</u>	0.467	0.451	0.470	0.480	0.481	0.497	0.507	0.521	1.267	0.520
ETTh2	96	0.170	0.251	<u>0.178</u>	<u>0.256</u>	0.177	0.260	0.180	0.264	0.167	0.260	0.355	0.462	0.180	0.271
	192	0.226	0.289	<u>0.243</u>	<u>0.298</u>	0.248	0.306	0.250	0.309	0.224	0.303	0.595	0.586	0.252	0.318
	336	<u>0.293</u>	<u>0.331</u>	0.283	0.327	0.304	0.342	0.311	0.348	0.281	0.342	1.270	0.871	0.324	0.364
	720	0.389	0.380	0.389	0.381	0.403	0.397	0.412	0.407	0.397	0.421	3.001	1.267	0.410	0.420

Table 3. Multivariate time series forecasting results. The input length H = 336. The best results are highlighted in **bold**, and the second-best results are underlined.

Ablation study

In this section, we conduct a series of ablation experiments to evaluate the contributions of key components within the proposed model. Specifically, we assess the individual impact of the Quantum Fourier Transform (QFT), the Quantum Frequency Decomposition-Reconstruction (Q-FR-Q) module, and the Dual-Layer Graph

Attention Network (referred to as Graph-Attention Network, or GAN). Furthermore, we explore the effectiveness of combining quantum-enhanced feature extraction with graph convolution and attention mechanisms to validate their superiority in time-series forecasting.

To maintain clarity and consistency, we adopt standardized names for all model variants throughout the paper. The full model is referred to as *QFreqFormer* (Quantum Frequency Transformer), which integrates the Quantum Fourier Transform (QFT), the Quantum Frequency Reconstruction module (Q-FR-Q), and a Dual-Layer Graph Attention Network (D-PAD). Three ablated variants are evaluated: (1) *DL-GAN*, which retains D-PAD but replaces quantum components with traditional frequency operations; (2) *GAN-only*, which uses D-PAD with a standard Fourier Transform, excluding both QFT and Q-FR-Q; and (3) *GCN-only*, which substitutes the attention mechanism with standard Graph Convolutional Networks and excludes all quantum modules. These names are used consistently across sections to ensure comprehensibility and ease of comparison.

Experimental setup

The experimental protocol is consistent with the methodology outlined in Section 4.1, with the use of the same datasets and evaluation metrics (MSE and MAE). In the ablation study, we systematically remove or replace critical components of the model and evaluate the impact on forecasting performance, aiming to understand the contribution of each module to the overall prediction accuracy.

- **Full Model:** This is the complete model that integrates Quantum Fourier Transform (QFT), Quantum Frequency Reconstruction (Q-FR-Q), and Generative Adversarial Network (GAN) as described in Section 3. This configuration represents the core contribution of this paper and is used for comparison with the other ablation models.
- **GAN-only Model:** This model utilizes only the Dual-Layer Graph Attention Network (GAN), combined with traditional Fourier Transform, excluding the QFT and Q-FR-Q modules. It serves to evaluate the independent contribution of the GAN in time-series forecasting, disregarding the impact of frequency decomposition.
- **GCN-only Model:** This model employs only the traditional Graph Convolutional Network (GCN) along with traditional Fourier Transform, excluding both the QFT and Q-FR-Q modules. This configuration is used to assess the performance of the traditional graph neural network in time-series forecasting, without the influence of quantum Fourier transform and frequency reconstruction.
- **QFT + GCN Model:** This model combines the QFT and Q-FR-Q modules with the traditional GCN. It is used to evaluate the contribution of quantum frequency decomposition for feature extraction and analyze the role of the traditional graph neural network in this configuration.

Results and analysis

Table 4 presents the results of the ablation study for multivariate long-term forecasting tasks, with evaluation metrics (MSE and MAE) reported for each model across various datasets. The results offer detailed insights into the individual contributions of each model component to overall performance.

As shown in the table, the model using only traditional Fourier Transform performs significantly worse compared to those incorporating quantum-enhanced feature extraction (QFT-only, Q-FR-Q-only, and GAN-only). Specifically, the inclusion of the QFT module (QFT-only) results in substantial improvements in both MSE and MAE, emphasizing the advantages of quantum Fourier transform in extracting frequency-domain features, particularly in capturing the multi-frequency patterns inherent in time-series data. The combination of QFT and Q-FR-Q outperforms QFT-only, particularly in datasets with more complex frequency patterns, such as the Traffic and Electricity datasets. The Q-FR-Q module notably enhances performance by efficiently separating high- and low-frequency components, thereby improving the model's ability to capture intricate temporal patterns.

Model	Metric	Original		GAN-only		GCN-only		QFT + GCN		D-PAD	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2	96	0.297	<u>0.343</u>	0.347	0.361	0.351	0.359	0.344	0.351	<u>0.300</u>	<u>0.348</u>
	192	0.379	0.349	<u>0.352</u>	<u>0.381</u>	0.355	0.387	0.350	0.382	0.384	0.356
	336	0.377	0.334	<u>0.339</u>	<u>0.381</u>	0.342	0.383	0.337	0.381	0.381	0.345
	720	0.453	0.442	<u>0.453</u>	<u>0.461</u>	0.456	0.462	0.449	0.458	0.460	0.475
ETTm2	96	0.170	0.251	<u>0.256</u>	<u>0.179</u>	0.254	0.177	0.256	0.175	0.178	0.256
	192	0.226	0.289	<u>0.231</u>	<u>0.292</u>	0.233	0.295	0.229	0.291	0.243	0.298
	336	0.293	0.331	<u>0.294</u>	<u>0.341</u>	0.297	0.346	0.284	0.329	0.283	0.327
	720	0.389	0.380	<u>0.392</u>	<u>0.384</u>	0.397	0.391	0.389	0.381	0.389	0.381
Weather	96	0.153	0.194	0.199	0.247	0.201	0.252	<u>0.198</u>	<u>0.244</u>	0.258	0.298
	192	0.194	0.234	0.281	0.252	0.286	0.354	0.285	0.351	<u>0.284</u>	<u>0.314</u>
	336	0.238	0.272	0.279	0.237	0.382	0.342	0.298	0.369	<u>0.238</u>	<u>0.272</u>
	720	0.329	0.325	0.357	0.348	0.462	0.451	0.355	0.347	0.329	0.329

Table 4. Ablation study results on ETTh2, ETTm2, and Weather datasets. The best results are highlighted in bold, while the second-best results are underlined. The second-best experimental results have been underlined accordingly

Furthermore, the GAN-only model highlights the importance of graph convolution and attention mechanisms in capturing long-term dependencies and global relationships within the data. While this configuration does not incorporate quantum frequency decomposition, it shows performance improvements over the baseline, likely attributable to the modeling of inter-node relationships through attention mechanisms. Finally, the full model (QFT + Q-FR-Q + GAN) achieves better overall performance compared to other configurations across the evaluated datasets, suggesting that the combination of QFT, Q-FR-Q, and GAN contributes synergistically to forecasting accuracy. By integrating quantum-enhanced frequency decomposition with graph convolution and attention mechanisms, the full model demonstrates enhanced capability in capturing both local and global temporal dependencies, resulting in improved predictive accuracy under standardized conditions.

Experiments conclusion

The results of the ablation study suggest that each component contributes meaningfully to the overall performance of the model. The QFT module supports the extraction of frequency-domain features, while the Q-FR-Q module facilitates the separation of high- and low-frequency components, thereby improving the modeling of temporal patterns. The GAN module, through graph convolutions and attention mechanisms, enhances the model's ability to represent long-term dependencies by capturing complex interdependencies.

Ultimately, the full model (QFT + Q-FR-Q + GAN) provides the best results, validating the substantial advantages of combining quantum-enhanced feature extraction with deep learning techniques for time-series forecasting.

Transfer learning

Experimental Setup : In order to evaluate the portability of the QFreqFormer, the model was benchmarked against three baselines: the D-PAD, the PatchTST and the FEDformer. Two different portability experiments were designed. In the context of evaluating transferability across different datasets, the model is first pre-trained on Electricity and ETTh1. Subsequently, we fine-tuned the model using ETTh2 and ETTh2. To assess its transferability to future data, the models are pre-trained on the first 70% of the training data from the Weather and ETTh1 datasets. Following this preliminary training, the models are then fine-tuned with the objective of optimising the performance on the remaining 30% of the training data for each individual dataset.

Transfer Learning : In this study, we explore the impact of our proposed innovations on transfer learning, particularly evaluating the performance of QFreqFormer when pre-trained on the Electricity dataset and fine-tuned on other datasets. As shown in Table 5, while fine-tuning results in a slightly higher MSE compared to pre-training and fine-tuning on the same dataset, this is expected, as fine-tuning typically involves fewer training epochs and adaptation to new data distributions. Nevertheless, the performance of QFreqFormer remains superior to other models, indicating its ability to effectively transfer learned features across different tasks.

To further evaluate the domain-agnostic generalizability of QFreqFormer, we conduct cross-domain transfer learning experiments. Specifically, we pre-train the model on electricity-related datasets (ETTh2 and ETTh2) and fine-tune it on heterogeneous target datasets, including Weather (meteorological), ETTh1 (electricity), and Traffic (transportation). These experiments demonstrate the model's ability to adapt to diverse temporal dynamics and varying data distributions, highlighting its robustness across different application domains.

The integration of the Quantum Fourier Transform (QFT) module, Quantum Frequency Decomposition-Reconstruction (Q-FR-Q) module, and Dual-Layer Graph Attention Network (D-PAD) significantly enhances the model's performance in transfer learning scenarios. The QFT module leverages the parallelism and superposition properties of quantum computing to decompose time-series data into frequency components, providing a robust foundation for capturing high- and low-frequency patterns. This enhanced frequency-domain representation facilitates better feature transfer, ensuring that the model retains relevant knowledge when

Model	Step	QFreqFormer		D-PAD		PatchTST		FEDformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.153	0.194	0.207	0.251	0.211	0.254	0.238	0.314
	192	0.194	0.234	0.261	0.296	0.274	0.302	0.275	0.329
	336	0.238	0.272	0.214	0.247	0.295	0.317	0.339	0.377
	720	0.329	0.325	0.331	0.342	0.342	0.358	0.389	0.409
ETTh1	96	0.405	0.415	0.411	0.428	0.437	0.395	0.377	0.381
	192	0.436	0.430	0.439	0.435	0.441	0.434	0.417	0.446
	336	0.390	0.409	0.395	0.412	0.405	0.417	0.420	0.453
	720	0.486	0.476	0.491	0.487	0.517	0.502	0.419	0.474
Traffic	96	0.341	0.204	0.411	0.428	0.437	0.395	0.377	0.381
	192	0.377	0.246	0.439	0.435	0.441	0.434	0.417	0.446
	336	0.384	0.242	0.395	0.412	0.405	0.417	0.420	0.453
	720	0.403	0.278	0.491	0.487	0.517	0.502	0.419	0.474

Table 5. Transfer learning results on Weather and ETTh1 show QFreqFormer achieves the best overall performance. The best experimental results are highlighted in bold. The best experimental results have been highlighted in bold as indicated

applied to new datasets. The Q-FR-Q module further refines this by progressively separating mixed-frequency components, improving the model's ability to capture complex temporal dependencies, which is crucial for transfer learning tasks with diverse data characteristics. Moreover, the D-PAD module dynamically models the complex interactions between frequency components using graph convolutions and attention mechanisms, ensuring that long-term dependencies and intricate interdependencies across datasets are effectively captured. The combination of quantum-enhanced frequency decomposition and graph-based attention mechanisms not only improves the model's performance in the source task but also provides significant advantages in fine-tuning across different datasets, thus validating the robustness and wide applicability of QFreqFormer in transfer learning applications.

The results demonstrate that the performance of each model is significantly improved with the addition of these innovations. Specifically, PatchTST, when enhanced with our innovations, performs best on both datasets, significantly improving the model's prediction accuracy. Similarly, FEDformer and Autoformer show substantial improvements, particularly when handling longer time series, with reductions in both MSE and MAE. Additionally, Informer's overall performance improved after incorporating our innovations, though it did not perform as well as the other models in some cases, suggesting that our approach is equally effective in capturing the temporal dependencies in time series data across different models. Overall, the introduction of our innovations notably enhances the performance of these models across various datasets, with the most pronounced advantage observed in long-term time series prediction tasks. This validates the generality and effectiveness of our proposed time series modeling method across different architectures.

Statistical significance validation

To further strengthen the scientific rigor of our evaluation and verify that the observed performance gains of QFreqFormer are not due to random variation, we incorporate statistical significance testing into our analysis.

We adopt two widely used statistical methods: the *paired t-test* and the *Wilcoxon signed-rank test*, which assess whether the differences in performance between QFreqFormer and baseline models are statistically significant. The paired t-test is suitable for normally distributed paired differences, whereas the Wilcoxon test serves as a robust non-parametric alternative when normality cannot be assumed.

We perform 10 independent training runs for each model using different random seeds, focusing on the 336-step forecasting horizon. The MAE and MSE values from QFreqFormer are compared with those from five strong baselines: PatchTST, DLinear, Informer, FEDformer, and Autoformer.

The resulting *p*-values are reported in Table 6. All *p*-values are below the commonly used threshold of 0.05, demonstrating that the performance improvements achieved by QFreqFormer are statistically significant. This confirms the robustness and effectiveness of our proposed quantum-enhanced architecture across multiple trials.

Computational efficiency and resource analysis

To evaluate the practical deployability of QFreqFormer, we provide a detailed assessment of its computational efficiency and memory consumption. While the quantum-inspired modules (QFT and Q-FR-Q) significantly enhance forecasting performance, it is essential to quantify their computational overhead to assess real-world applicability.

We conduct experiments under a unified hardware configuration—a single NVIDIA RTX 3080 Ti (12GB VRAM) and an Intel Core i9-13900K CPU—using input length $H = 336$ and batch size 32. Three key metrics are reported for comparison with five representative baselines (DLinear, PatchTST, Informer, FEDformer, and Autoformer): average training time per epoch (in seconds), inference latency per sample (in milliseconds), and peak GPU memory usage (in GB).

As shown in Table 7, QFreqFormer incurs a moderate computational load (23.4s / 0.68ms / 3.5GB), slightly higher than lightweight models such as DLinear, but substantially more efficient than attention-based baselines including Informer, FEDformer, and Autoformer. This demonstrates that while QFreqFormer introduces additional complexity, it remains computationally tractable.

Importantly, these efficiency metrics should be interpreted in conjunction with QFreqFormer's consistently superior predictive accuracy, which validates that the added cost is well-compensated by performance gains. Moreover, the model is fully implemented using vectorized PyTorch operations, allowing for efficient GPU acceleration and practical scalability.

Comparison method	Paired <i>t</i> -test (<i>p</i> -value)	Wilcoxon Test (<i>p</i> -value)
QFreqFormer vs. PatchTST	0.0021	0.0033
QFreqFormer vs. DLinear	0.0045	0.0079
QFreqFormer vs. Informer	0.0006	0.0004
QFreqFormer vs. FEDformer	0.0018	0.0026
QFreqFormer vs. Autoformer	0.0029	0.0041

Table 6. Statistical significance analysis of QFreqFormer against baseline models (forecast horizon = 336). Values represent *p*-values from 10 paired runs. Statistically significant differences at $p < 0.05$ are shown in blue.

Model	Training (s)	Inference (ms)	GPU (GB)
DLinear	12.1	0.42	2.1
QFreqFormer	<u>23.4</u>	<u>0.68</u>	<u>3.5</u>
Informer	44.7	1.52	6.1
FEDformer	48.9	1.64	6.4
Autoformer	46.2	1.48	5.8
PatchTST	51.6	1.57	6.2

Table 7. Comparison of runtime and GPU memory usage. Input length $H = 336$, batch size = 32. Bold values indicate the best among all models. Underlined values indicate the second-best results.

Conclusion

This work introduces a quantum-enhanced forecasting framework for multivariate time-series analysis, integrating the Quantum Fourier Transform (QFT), a Quantum Frequency Decomposition–Reconstruction–Decomposition (Q-FR-Q) mechanism, and a Dual-Layer Graph Attention Network (DL-GAN). Through the incorporation of frequency-domain transformations and graph-based relational modeling, the framework establishes a structured pipeline for capturing both global periodic patterns and local temporal dependencies. Quantitative evaluations conducted on multiple benchmark datasets demonstrate consistent improvements in predictive accuracy compared to representative baselines under standardized conditions.

Notwithstanding these advantages, the proposed method exhibits several limitations that merit further investigation. First, the QFT and Q-FR-Q modules impose a considerable computational burden due to the repeated manipulation of complex-valued spectral representations. This constraint may hinder deployment in scenarios demanding low-latency inference or high-frequency data processing. Second, the DL-GAN component utilizes a static graph topology, which may be insufficient for capturing dynamic or non-stationary inter-frequency dependencies, especially in environments characterized by evolving temporal structures. Third, the model's performance deteriorates in short-horizon prediction settings, where localized fluctuations dominate and frequency-domain representations alone may fail to adequately reflect abrupt changes.

Future research should consider algorithmic refinements to reduce computational complexity, such as incorporating approximate quantum transformations or sparse spectral filtering. Dynamic graph construction strategies and adaptive relational learning schemes may enhance the model's responsiveness to structural changes in the data. Furthermore, hybrid approaches that combine frequency-aware architectures with local time-domain encoding—such as multi-scale convolutions or attention-based residual modules—could improve short-term prediction reliability. Addressing these aspects is crucial for improving the model's applicability in real-world operational settings, particularly those requiring adaptability, interpretability, and computational efficiency.

Data availability

Data is provided within the manuscript or supplementary information files

Received: 15 May 2025; Accepted: 8 October 2025

Published online: 14 November 2025

References

- Chen, L. et al. Multi-scale adaptive graph neural network for multivariate time series forecasting. *IEEE Trans. Knowledge Data Eng.* **35**(10), 10748–10761 (2023).
- Lai, G., Chang, W.-C., Yang, Y., Liu, H.: Modeling long-and short-term temporal patterns with deep neural networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 95–104 (2018)
- Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., Yan, X.: Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems* **32** (2019)
- Hyndman, R. J. & Athanasopoulos, G. *Forecasting: Principles and Practice* (OTexts, Melbourne, Australia, 2018).
- Wu, H., Xu, J., Wang, J. & Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inform. Process. Syst.* **34**, 22419–22430 (2021).
- Gardner, E. S. Jr. Exponential smoothing: The state of the art. *J. Forecast.* **4**(1), 1–28 (1985).
- Taylor, S. J. & Letham, B. Forecasting at scale. *The Am. Stat.* **72**(1), 37–45 (2018).
- Lim, B. & Zohren, S. Time-series forecasting with deep learning: A survey. *Phil. Trans. Royal Soc. A* **379**(2194), 20200209 (2021).
- Schuld, M., Sinayskiy, I. & Petruccione, F. An introduction to quantum machine learning. *Contemp. Phys.* **56**(2), 172–185 (2015).
- Shor, P.W.: Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th Annual Symposium on Foundations of Computer Science, pp. 124–134 (1994). IEEE
- Osorio, S.L.J., Ruiz, M.A.R., Mendez-Vazquez, A., Rodriguez-Tello, E.: Fourier series guided design of quantum convolutional neural networks for enhanced time series forecasting. arXiv preprint [arXiv:2404.15377](https://arxiv.org/abs/2404.15377) (2024)
- Jones, C. et al. Benchmarking quantum models for time-series forecasting. *IEEE* **2**, 22–27 (2024).
- Wang, J.: Multivariate time series forecasting and classification via gnn and transformer models. *J. Comput. Technol. Software* **3**(9) (2024)
- Zhang, H., Liu, H., Yi, X. & Geng, Z. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. *Adv. Neural Inform. Process. Syst.* **35**, 21045–21056 (2022).
- Zhou, T., Ma, Z., Wen, Q., Sun, L., Jin, R.: Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: International Conference on Machine Learning, pp. 27268–27286 (2022). PMLR
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint [arXiv:1710.10903](https://arxiv.org/abs/1710.10903) (2017)

17. Wu, Z. et al. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Networks Learn. Syst.* **32**(1), 4–24 (2020).
18. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of AAAI, vol. 35, pp. 11106–11115 (2021)
19. Liu, J., Deng, A., Sun, L., Wei, Z., Wang, X.: Pyraformer: Low-complexity pyramidal attention for long-range time series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 7248–7256 (2022)
20. Schuld, M., Sinayskiy, I. & Petruccione, F. The quest for a quantum neural network. *Quantum Inform. Process.* **13**(11), 2567–2586 (2014).
21. Zhou, L., Du, J. & Pan, J.-W. Quantum long short-term memory. *Phys. Rev. A* **102**(5), 052309 (2020).
22. Chen, S.Y.-C., Yoo, S., Fang, Y.-L.L.: Quantum long short-term memory, 8622–8626 (2022). IEEE
23. Chen, Y., Khaliq, A., Furati, K.M.: Quantum recurrent neural networks with encoder-decoder for time-dependent partial differential equations. arXiv preprint [arXiv:2502.13370](https://arxiv.org/abs/2502.13370) (2025)
24. Rivera-Ruiz, M.A., Juárez-Osorio, S.L., Mendez-Vazquez, A., López-Romero, J.M., Rodríguez-Tello, E.: 1d quantum convolutional neural network for time series forecasting and classification. In: Mexican International Conference on Artificial Intelligence, pp. 17–35 (2023). Springer
25. Henderson, M., Shakya, S., Cook, T. & Tenev, T. Quantum convolutional neural networks: Powering image recognition with quantum circuits. *Quantum Mach. Intell.* **2**(1), 1–9 (2020).
26. Chen, C.-S., Tsai, A.H.-W., Huang, S.-C.: Quantum multimodal contrastive learning framework. arXiv preprint [arXiv:2408.13919](https://arxiv.org/abs/2408.13919) (2024)
27. Abbas, A. et al. The power of quantum neural networks. *Nature Computat. Sci.* **1**(6), 403–409 (2021).
28. Du, Y., Wang, X., Guo, N., Yu, Z., Qian, Y., Zhang, K., Hsieh, M.-H., Reberntrost, P., Tao, D.: Quantum machine learning: A hands-on tutorial for machine learning practitioners and researchers. arXiv preprint [arXiv:2502.01146](https://arxiv.org/abs/2502.01146) (2025)
29. Cerezo, M. et al. Variational quantum algorithms. *Nature Rev. Phys.* **3**, 625–644 (2021).
30. Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. *Time Series Analysis: Forecasting and Control* (Wiley, Hoboken, 2015).
31. Hyndman, R., Koehler, A. B., Ord, J. K. & Snyder, R. D. *Forecasting with Exponential Smoothing: the State Space Approach* (Springer, Berlin, Germany, 2008).
32. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37. Washington, DC, USA, pp. 11121–11128 (2023). AAAI Press
33. Sun, L., Liu, X., Wu, H., Xu, Y., Long, G.: Learning disentangled representations for time series with variational temporal decomposition. In: International Conference on Learning Representations (ICLR) (2023)
34. Wang, W., Yin, L., Liu, Y., Sun, H.: Mrfnet: Multi-receptive field network for multivariate time-series prediction. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA, pp. 1234–1243 (2023). ACM
35. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. Proceedings of the 11th International Conference on Learning Representations (2023)

Author contributions

Y.T. and Z.C. wrote the main manuscript text and contributed to the literature review and methodology. Y.Z. supervised the research design and provided critical revisions throughout the manuscript. Z.G. performed the data analysis and implemented the model experiments. J.Y. prepared the figures and assisted in result interpretation. All authors reviewed and approved the final manuscript.

Funding

This research was supported in part by the National Natural Science Foundation of China (Project No. 62472144) under the General Program, titled “Research on privacy preserving techniques for blockchain based on non-interactive zero-knowledge proof from lattice”, covering the period from January 2025 to December 2028. Additional support was provided by the Key Scientific and Technological Research Project of Henan Province (Project No. 252102210100). The authors gratefully acknowledge the financial support from both foundations.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025