

A Quantum-Classical Hybrid Framework for Multivariate Time-Series Forecasting: Complexity-Fidelity Trade-offs and Limitations

Sanjay Chakraborty*, Fredrik Heintz

Abstract—This paper presents a unified quantum-classical hybrid framework for multi-horizon time-series forecasting, introducing two model variants — a Quantum Reservoir Forecaster (QRC-F) and a Variational Quantum Forecaster (VQF-F), evaluated in multivariate settings. The framework further provides a systematic complexity-fidelity trade-off analysis with important limitations under near-term NISQ hardware constraints. The proposed framework first applies uniform binary quantization to map continuous time-series signals into discrete binary representations, which are subsequently encoded into quantum states via angle encoding using parameterized R_Y rotation gates, with cross-channel entanglement layers explicitly capturing inter-variable dependencies in the multivariate case. QRC-F employs a fixed random unitary reservoir for stable, gradient-free temporal feature extraction, while VQF-F deploys a variational quantum circuit of depth D , trained through the parameter-shift rule, to extract rich temporal and inter-channel features as Pauli observable expectation values, both replacing the quadratic self-attention mechanism with linear matrix operations that significantly reduce computational complexity and model parameters. A shared MIMO-based multi-horizon readout head maps the extracted quantum features simultaneously to all prediction horizons H , avoiding the error accumulation inherent in recursive forecasting strategies. The predicted quantum states are subsequently decoded back into classical signals through an inverse angle mapping and de-quantization procedure, faithfully reconstructing the forecasted time-series values at the original scale. Experimental evaluations across benchmark datasets including ETTh1, ETTh2, ETTm1, ETTm2, Weather, electricity, and exchange-rate demonstrate that VQF-F offers superior training stability and parameter efficiency for multi-horizon tasks, while QRC-F provides superior training stability and circuit fidelity under hardware noise, collectively establishing a principled and practical quantum-native forecasting paradigm suitable for deployment on near-term Noisy intermediate-scale quantum (NISQ) devices.

Index Terms—Variational Quantum Forecaster (VQF), Quantum Reservoir Forecaster (QRC), Quantum Machine Learning, Time-Series Forecasting, Multi-Horizon Prediction, Quantum-Classical Hybrid Models.

I. INTRODUCTION

ALTHOUGH current time-series forecasting methods have significantly advanced the state-of-the-art (SOTA), particularly for multivariate time-series forecasting, they continue to face fundamental challenges in effectively capturing

multivariate long-range temporal dependencies and extracting meaningful features from extended sequences. Existing approaches, especially those based on deep learning and self-attention mechanisms, often suffer from information utilization bottlenecks, where relevant temporal patterns are either diluted or inefficiently modeled across multi horizons. Furthermore, the reliance on quadratic-complexity operations, such as self-attention, leads to substantial computational overhead and other limitations, making these models less practical for real-world, resource-constrained applications. To address these limitations, we introduce a Quantum Time-Series Forecasting (QTSF) framework, which leverages quantum-inspired feature extraction and efficient linear operations to model temporal dependencies while significantly reducing computational complexity and parameter requirements. The proposed QTSF framework supports two interchangeable quantum feature extraction backends — QRC-F, a parameter-free quantum reservoir, and VQF-F, a fully trainable variational circuit, sharing identical preprocessing, encoding, and readout stages, enabling a direct controlled comparison of fixed versus adaptive quantum feature extraction for multivariate time-series forecasting. To the best of our knowledge, this is the first work to propose a unified quantum-classical hybrid framework that jointly addresses multivariate long-horizon time-series forecasting and systematic complexity-fidelity trade-off analysis under near-term NISQ hardware constraints.

- **Quantum Reservoir Forecaster (QRC-F):** employs a fixed Haar-random unitary reservoir U_R for gradient-free temporal feature extraction at $\mathcal{O}(n^2)$ two-qubit gate cost. QRC-F offers unconditional training stability and high circuit fidelity under hardware noise, but is fundamentally limited by its fixed, non-adaptive nature, it cannot learn dataset-specific temporal structures, which constrains its predictive capacity in complex multivariate settings.
- **Variational Quantum Forecaster (VQF-F):** deploys a trainable variational quantum circuit of depth D , optimized through the parameter-shift rule [1], at a reduced $\mathcal{O}(nD)$ gate complexity. VQF-F achieves stronger predictive accuracy through adaptive feature learning, but faces an inherent *trainability-expressibility trade-off*: increasing circuit depth D improves feature richness at the cost of exponentially vanishing gradients $\text{Var}[\partial_\theta \mathcal{L}] \propto 2^{-n}$, known as the barren plateau phenomenon [2], which represents a fundamental scalability limitation on current NISQ hardware.

*Sanjay Chakraborty is working in the Department of Computer and Information Science (IDA), REAL, AIICS, at Linköping University, Sweden and Department of Computer Science & Engineering, Techno International New Town, Kolkata, India (Email: sanjay.chakraborty@liu.se).

Fredrik Heintz is working in the Department of Computer and Information Science (IDA), REAL, AIICS, at Linköping University, Sweden, Email: fredrik.heintz@liu.se

Both variants share a common preprocessing pipeline comprising uniform b -bit binary quantization, angle encoding via parameterized R_Y rotation gates, and cross-channel controlled- Z entanglement layers that explicitly model inter-variable dependencies across M input channels. A shared MIMO-based multi-horizon readout head maps extracted quantum features simultaneously to all prediction horizons H and channels M , avoiding the error accumulation inherent in recursive forecasting strategies. Predicted quantum states are decoded back into classical signals through an inverse angle mapping and de-quantization procedure, completing a fully end-to-end quantum-compatible forecasting pipeline. The central focus of this work is that the choice between QRC-F and VQF-F represents a fundamental *complexity-stability-accuracy trade-off* that is not merely architectural but reflects deeper properties of quantum circuit trainability, hardware fidelity, and forecasting horizon. Specifically, we demonstrate that:

- 1) QRC-F offers a complexity-fidelity advantage under hardware noise and short-to-medium horizon tasks, where its $\mathcal{O}(n^2)$ reservoir provides sufficient temporal memory without gradient optimization;
- 2) VQF-F achieves superior accuracy on long-horizon multivariate forecasting at reduced gate complexity $\mathcal{O}(nD)$, but requires careful depth selection to avoid barren plateaus, a limitation that directly constrains its scalability to large n ;

Experimental evaluations on seven standard benchmark datasets — ETTh1, ETTh2, ETTm1, ETTm2, Weather, Electricity, and Exchange-Rate, across multiple prediction horizons $H \in \{96, 192, 336, 720\}$ validate the complexity-accuracy trade-offs of both variants against strong classical baselines including DLinear, PatchTST, and iTransformer, as well as existing quantum forecasting approaches. Table I summarizes the principal innovations of the proposed QTSF framework, highlighting its end-to-end quantum-compatible design, efficient linear-complexity forecasting architecture, systematic complexity-fidelity trade-off analysis, multivariate dependency modelling via entanglement, and scalable signal reconstruction strategy.

TABLE I: Key Novel Contributions of the Proposed QTSF Framework

Novel Contribution	Strength
Binary quantization as a bridge between classical signals and quantum encoding	Principled preprocessing step; underexplored as a quantum-classical interface in time-series quantum machine learning (QML)
Abandoning self-attention in favour of linear matrix operations as a quantum-compatible backbone	Quantum-motivated complexity reduction; reduces quadratic to linear attention cost
MIMO multi-horizon quantum forecasting with per-horizon observable heads	Not explored in existing QML time-series literature
Cross-channel entanglement (CZ gates) for multivariate dependency encoding	Quantum-native approach to multivariate dependency capture; analogous to classical channel mixing
Classical signal reconstruction via arccos inverse mapping and de-quantization	Completes the end-to-end pipeline (often missing in QML works)

The paper is structured to first review related work and introduce the necessary background and problem formulation for univariate and multivariate forecasting. It then presents the

quantum foundations and the proposed methodology, including data preprocessing, quantum encoding, feature extraction, and the forecasting architecture. This is followed by experimental details, results, and comparative analysis with classical and hybrid models. The paper concludes with a discussion of limitations, future work, and standard closing sections, along with an appendix providing an illustrative example and additional technical details.

II. RELATED WORKS

In this section, we present a concise overview of key Transformer-based models, including both classical and quantum-inspired approaches, that have shown strong performance in time-series forecasting. Particular emphasis is placed on state-of-the-art classical self-attention models for time-series forecasting.

A. Classical Models

Transformers have gained significant attention in long-term forecasting due to their ability to capture complex temporal dependencies [3], [4], [5], [6], [7]. Among the early advancements, Informer [8] introduced a generative-based decoder and 'Probability-Sparse' self-attention to address the challenge of quadratic time complexity. Building on this, models such as Autoformer [9], iTransformer [10], FEDFormer [11], PatchTST [12], ETSformer [13], Pyraformer [14], and EDformer [15] have further enhanced time-series modeling. iTransformer [10] innovates by representing individual time points as variate tokens, enabling the attention mechanism to model multivariate correlations while leveraging feed-forward networks to learn nonlinear representations. PatchTST [12] enhances local and global dependency capture through patch-based processing, while Crossformer [16] introduces a dimension-segment-wise (DSW) technique that encodes time-series data into a structured 2D representation. With the Sub-Window Tokenizer and a Time-Series Pre-trained Encoder, TS-Fastformer [17] enhances processing speed, optimizing the workflow for faster and more efficient outcomes. The core strength of transformer models lies in their attention mechanism, which allows them to focus on critical segments of the input sequence for accurate predictions [5]. Flowformer [18] reformulates the self-attention mechanism using conservation flows to achieve linear complexity without sacrificing global context modeling, offering a practical efficiency improvement for long-horizon settings. Transformer-based models effectively capture temporal dependencies but suffer from quadratic complexity and lack of inherent temporal ordering due to permutation-invariant self-attention. To address this, we propose a quantum-classical hybrid framework that replaces self-attention with quantum reservoir and variational circuit modules, enabling linear-complexity feature extraction while preserving strong multi-horizon forecasting performance on NISQ-era hardware.

B. Quantum inspired models

The application of quantum and quantum-inspired frameworks to time-series forecasting has gained considerable momentum as a principled alternative to classical deep learning,

motivated by the theoretical capacity of quantum systems to inhabit exponentially large Hilbert spaces and represent complex temporal correlations with substantially fewer trainable parameters [19, 20]. The earliest and most widely adopted architectural paradigm in this direction is the Quantum Long Short-Term Memory (QLSTM), proposed by [21], which replaces the classical LSTM gate circuits with variational quantum circuits (VQCs) optimised via the parameter-shift rule, demonstrating faster convergence and competitive forecasting accuracy relative to classical counterparts across multiple domains. Building on this foundation, [22] introduced a continuous-variable Quantum RNN (CV-QRNN) and showed that the quantum network converges to optimal weights in fewer training epochs than a classical equivalent for several classes of temporal data, while [23] extended the recurrent quantum paradigm to the QSegRNN architecture, achieving competitive accuracy with significantly reduced parameter counts for transformer temperature forecasting under NISQ constraints. Alongside recurrent approaches, Quantum Reservoir Computing (QRC) has emerged as a particularly hardware-efficient framework for temporal feature extraction on near-term devices, exploiting a fixed random quantum unitary as a non-trainable reservoir to avoid barren plateau issues altogether. [24] demonstrated the first gate-based QRC architecture for multivariate time series (MTS-QRC) on real IBM Heron R2 hardware, showing that device noise can act as an implicit regulariser, with the model achieving competitive accuracy against classical reservoir computing on chaotic Lorenz-63 and ENSO datasets. Most recently, the intersection of transformer architectures and quantum computing has produced a growing class of hybrid models: [25] proposed Quantum Adaptive Self-Attention (QASA), which selectively replaces only the final transformer encoder block with a VQC layer and demonstrates significantly faster convergence on temporal prediction tasks compared to both classical transformers and prior quantum baselines, while [26] introduced the iQTransformer, integrating a quantum self-attention mechanism into the iTransformer backbone to provide quantum-native modelling of inter-variable dependencies in multivariate forecasting with fewer parameters than classical equivalents. Collectively, these works establish quantum and quantum-hybrid architectures as a viable and growing alternative to classical self-attention for time-series forecasting, particularly in parameter-constrained and NISQ-compatible deployment settings.

III. PRELIMINARY

A. Problem Formulation

Time-series forecasting (TSF) aims to predict values at one or more future timesteps based on historical observations. Depending on the number of variables involved, TSF can be broadly categorized into univariate and multivariate forecasting.

1) *Univariate Forecasting*: Univariate forecasting focuses on predicting future values of a single variable based on its historical sequence. Given an input sequence:

$$x_{\text{in}} = \{x_{T-L+1}, x_{T-L+2}, \dots, x_T\}, \quad (1)$$

of length L , where $x_t \in \mathbb{R}$ denotes the value at timestep t , the objective is to predict the future sequence:

$$\hat{x}_{\text{out}} = \{\hat{x}_{T+1}, \hat{x}_{T+2}, \dots, \hat{x}_{T+H}\}, \quad (2)$$

for the next H timesteps.

The forecasting model learns a mapping function:

$$\hat{x}_{\text{out}} = f_{\theta}(x_{\text{in}}), \quad (3)$$

where $f_{\theta}(\cdot)$ represents a parameterized model with learnable parameters θ .

2) *Multivariate Forecasting*: Multivariate forecasting considers multiple variables evolving over time and predicts their future values jointly. The input sequence is defined as:

$$X_{\text{in}} = \{x_{T-L+1}, x_{T-L+2}, \dots, x_T\}, \quad (4)$$

where each observation is:

$$x_t = [x_t^1, x_t^2, \dots, x_t^M] \in \mathbb{R}^M, \quad (5)$$

with M representing the number of variables.

The goal is to predict the future sequence:

$$\hat{X}_{\text{out}} = \{\hat{x}_{T+1}, \hat{x}_{T+2}, \dots, \hat{x}_{T+H}\}, \quad (6)$$

where each predicted timestep is:

$$\hat{x}_{T+h} = [\hat{x}_{T+h}^1, \hat{x}_{T+h}^2, \dots, \hat{x}_{T+h}^M], \quad h = 1, \dots, H. \quad (7)$$

The forecasting task is modeled as:

$$\hat{X}_{\text{out}} = f_{\theta}(X_{\text{in}}), \quad (8)$$

where $f_{\theta}(\cdot)$ captures both temporal dependencies and inter-variable relationships.

B. Quantum Principles

Superposition, entanglement, quantum teleportation, and interference are some of the basic concepts of quantum mechanics that quantum computing uses to do some tasks much more quickly than classical computers. Quantum bits, or qubits, can exist in a superposition of both states concurrently, allowing for intrinsic parallelism in processing, in contrast to classical bits, which are limited to binary states (0 or 1). Advanced quantum algorithms, such as Grover's algorithm for unstructured search and Shor's algorithm for integer factorisation, as well as variational quantum techniques extensively investigated in quantum machine learning and artificial intelligence applications [19, 27], are based on quantum gates operating on qubits via unitary transformations.

1) *Quantum Superposition*: Quantum superposition states that a qubit can exist in a linear combination of both $|0\rangle$ and $|1\rangle$ states simultaneously [27]. Mathematically, this is represented as:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (9)$$

where α and β are complex probability amplitudes satisfying:

$$|\alpha|^2 + |\beta|^2 = 1 \quad (10)$$

Upon measurement, the qubit collapses to $|0\rangle$ with probability $|\alpha|^2$ or $|1\rangle$ with probability $|\beta|^2$. This enables quantum computers to explore multiple states in parallel, offering significant computational advantages over classical systems [28] [29].

2) *Quantum Entanglement*: Quantum entanglement is a phenomenon in which two or more qubits become correlated such that the state of one qubit cannot be described independently of the others, even when they are spatially separated. As a result, measurements performed on one qubit are instantaneously correlated with measurements on the other(s). A standard example is the Bell state:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (11)$$

Here, the two qubits exist in a superposition of both $|00\rangle$ and $|11\rangle$. Measuring one qubit immediately determines the state of the other, demonstrating non-local correlations. This property is fundamental to quantum communication, cryptography, and computing [28] [29].

C. Variational Quantum Algorithms

The Variational Quantum Eigensolver (VQE) and Variational Quantum Classifier (VQC) are hybrid quantum-classical algorithms that utilize parameterized quantum circuits for optimization and machine learning tasks [30]. In particular, VQE embeds a quantum subroutine within a classical optimization loop, where the quantum device evaluates expectation values and a classical optimizer updates the circuit parameters iteratively [31]. VQE is primarily designed to estimate the ground-state energy of a Hamiltonian H by minimizing the expectation value of the energy over a parameterized quantum state $|\psi(\theta)\rangle$:

$$E(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle \quad (12)$$

The parameters θ are optimized using a classical optimizer, such as gradient descent, to iteratively refine the quantum state. This method is crucial for quantum chemistry and materials science.

VQC applies a similar variational approach to quantum machine learning [1]. Given an input data point x , it is encoded into a quantum state $|\psi(x)\rangle$, which is processed through a parameterized quantum circuit $U(\theta)$:

$$|\phi(x, \theta)\rangle = U(\theta)|\psi(x)\rangle \quad (13)$$

A measurement operator M is then applied to extract the classification decision:

$$y = \langle \phi(x, \theta) | M | \phi(x, \theta) \rangle \quad (14)$$

Quantum-enhanced classification is made possible by training the parameters θ using a classical optimiser to minimise a loss function. The power of variational quantum algorithms is demonstrated by both VQE and VQC, which effectively tackle multi-class difficult problems by balancing quantum computation with classical optimisation [32]. The 'Variational Quantum Classifier (VQC)' circuit in Figure 1 has three essential stages: initial rotations, entanglers, and final rotations. First, classical data is encoded into quantum states using RX, RY, or RZ gates. Quantum correlations between qubits are then produced by entangling layers, which usually employ CNOT (CX) gates. Lastly, the quantum state is refined before to measurement by trainable rotation gates. Effective quantum classification is made possible by optimising the circuit characteristics using classical methods to minimise a loss function.

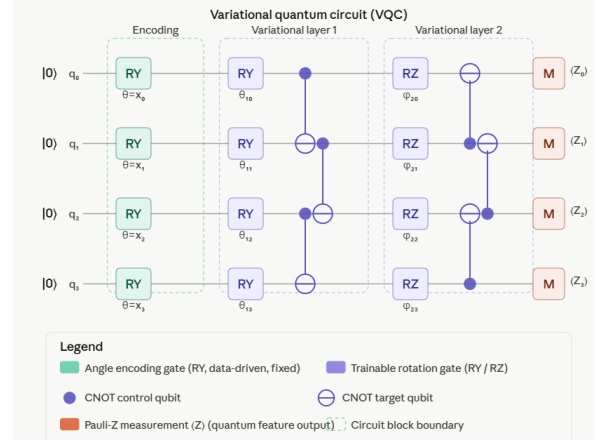


Fig. 1: Variational Quantum Circuit

IV. METHODOLOGY

Fig. 2 illustrates the proposed QTSF framework, where dual quantum backends (QRC-F and VQC-F) perform feature extraction before a shared readout layer generates multi-horizon time-series forecasts.

A. Preprocessing for Multivariate Time Series

Consider a multivariate time series $\mathbf{x}(t) \in \mathbb{R}^M$ with M channels over a look-back window of length L .

Step 2.1: Channel-wise Normalization

Theory: Normalization removes scale differences across channels, ensuring numerical stability and preventing dominance of high-variance signals during encoding.

$$\tilde{x}_m(t') = \frac{x_m(t') - \mu_m}{\sigma_m}, \quad (15)$$

$$\mu_m = \frac{1}{L} \sum_{t'=t-L+1}^t x_m(t'), \quad (16)$$

$$\sigma_m = \sqrt{\frac{1}{L} \sum_{t'=t-L+1}^t (x_m(t') - \mu_m)^2} \quad (17)$$

Min-max rescaling ensures compatibility with bounded quantum rotation angles:

$$\bar{x}_m(t') = \frac{\tilde{x}_m(t') - \tilde{x}_m^{\min}}{\tilde{x}_m^{\max} - \tilde{x}_m^{\min}} \quad (18)$$

Step 2.2: b -bit Quantization

Theory: Quantization discretizes continuous signals into finite representations, enabling efficient mapping to qubit rotations while controlling precision through b .

$$q_m(t') = \lfloor \bar{x}_m(t')(2^b - 1) \rfloor \quad (19)$$

Binary encoding:

$$\mathbf{b}_m(t') \in \{0, 1\}^b \quad (20)$$

Quantization error bound:

$$\epsilon_q = \frac{\tilde{x}_m^{\max} - \tilde{x}_m^{\min}}{2^b} \quad (21)$$

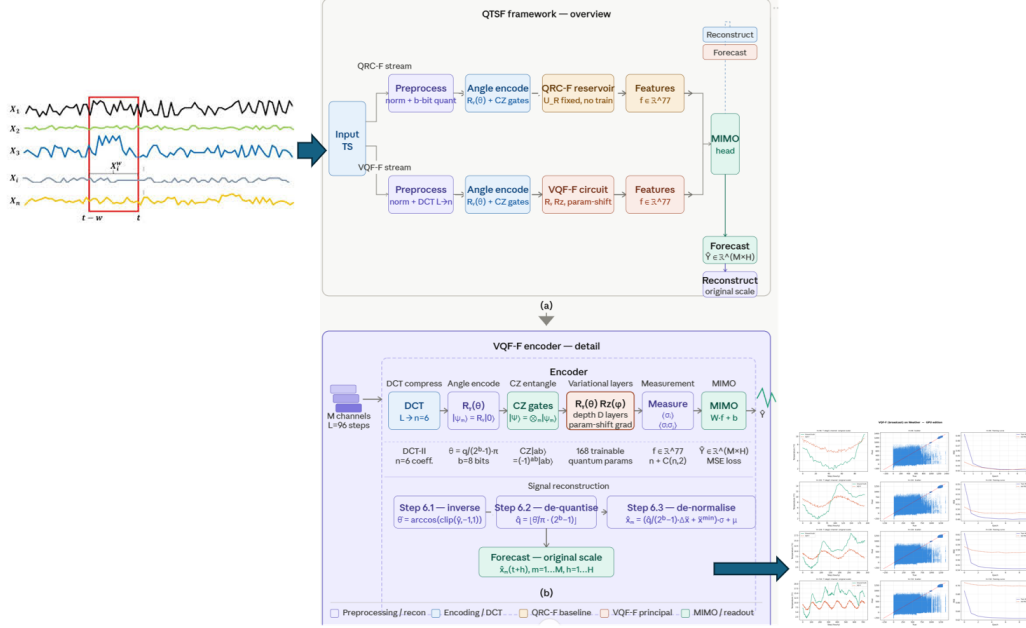


Fig. 2: Architecture of the proposed QTSF framework

B. Multivariate Quantum State Encoding

Step 3.1: Angle Encoding

Theory: Angle encoding maps classical data into quantum amplitudes via rotation gates, preserving relative magnitudes while enabling continuous-valued representation in Hilbert space.

$$\theta_m(t') = \frac{q_m(t')}{2^b - 1} \pi \quad (22)$$

$$|\psi_m(t')\rangle = R_{Y'}(\theta_m(t'))|0\rangle = \cos\left(\frac{\theta_m(t')}{2}\right)|0\rangle + \sin\left(\frac{\theta_m(t')}{2}\right)|1\rangle \quad (23)$$

Step 3.2: Joint Multivariate State Construction

Theory: Tensor product composition encodes temporal evolution and cross-channel structure into a high-dimensional Hilbert space, enabling exponential representational capacity.

$$|\Psi(t)\rangle = \bigotimes_{m=1}^M \bigotimes_{t'=t-L+1}^t |\psi_m(t')\rangle \quad (24)$$

Step 3.3: Cross-Channel Entanglement

Theory: Entanglement introduces non-classical correlations between channels, allowing the model to capture inter-variable dependencies beyond classical covariance structures.

$$U_{\text{cross}} = \prod_{t'=t-L+1}^t \prod_{m=1}^{M-1} CZ_{(m,t'),(m+1,t')} \quad (25)$$

$$CZ|ab\rangle = (-1)^{ab}|ab\rangle \quad (26)$$

C. Feature Extraction for Multivariate Forecasting

Let $n = ML$ be the total number of qubits.

Model 1: QRC-F

Step 4.1: Reservoir Construction

Theory: The random unitary acts as a high-dimensional nonlinear dynamical system, projecting inputs into a rich feature space without requiring training.

$$U_R = \exp\left(-i \sum_{j < k} J_{jk} \hat{\sigma}_j \hat{\sigma}_k\right) \quad (27)$$

Step 4.2: Sequential State Injection

Theory: Sequential coupling introduces temporal memory into the reservoir, analogous to recurrent dynamics in classical systems.

$$\rho(t' + 1) = \mathcal{E} \left[U_R (\rho(t') \otimes |\psi(t')\rangle\langle\psi(t')|) U_R^\dagger \right] \quad (28)$$

Step 4.3: Observable Feature Extraction

Theory: Measurement collapses the quantum state into classical statistics (expectation values), forming a feature vector encoding both local and pairwise correlations.

$$\mathbf{f}^{\text{QRC}} \in \mathbb{R}^{n + \binom{n}{2}} \quad (29)$$

Model 2: VQF-F

Step 4.4: Parameterized Circuit

Theory: Variational circuits introduce trainable parameters that adaptively learn task-specific transformations in Hilbert space.

$$|\Phi(\theta)\rangle = \prod_{d=1}^D \left[U_{\text{ent}} \cdot \bigotimes_{i=1}^n R_Y(\theta_i^d) R_Z(\phi_i^d) \right] |\Psi\rangle \quad (30)$$

Step 4.5: Feature Extraction

Theory: Observables encode learned representations of temporal and cross-channel dependencies into measurable statistics.

$$\mathbf{f}^{\text{VQF}} \in \mathbb{R}^{n+\binom{n}{2}} \quad (31)$$

Step 4.6: Parameter-Shift Gradient

Theory: The parameter-shift rule enables exact gradient estimation without backpropagation through quantum circuits.

$$\frac{\partial \langle \hat{O} \rangle}{\partial \theta_i} = \frac{1}{2} \left[\langle \hat{O} \rangle_{\theta_i + \frac{\pi}{2}} - \langle \hat{O} \rangle_{\theta_i - \frac{\pi}{2}} \right] \quad (32)$$

D. Multivariate MIMO Forecasting Head

Theory: The MIMO formulation predicts all horizons jointly, avoiding error accumulation inherent in recursive forecasting.

$$\hat{\mathbf{Y}} = \text{reshape}(W_{\text{out}}\mathbf{f} + \mathbf{b}, M, H) \quad (33)$$

$$\mathcal{L} = \frac{1}{MH} \sum_{m=1}^M \sum_{h=1}^H (\hat{x}_m(t+h) - x_m(t+h))^2 \quad (34)$$

E. Multivariate Signal Reconstruction

Step 6.1: Inverse Mapping

Theory: Expectation values are mapped back to rotation angles, reversing the encoding process.

$$\hat{\theta}_m(t+h) = \arccos(\text{clip}(\hat{y}_m(t+h), -1, 1)) \quad (35)$$

Step 6.2: Quantized Recovery

Theory: The continuous angle is discretized back into quantized levels.

$$\hat{q}_m(t+h) = \left\lfloor \frac{\hat{\theta}_m(t+h)}{\pi} (2^b - 1) \right\rfloor \quad (36)$$

Step 6.3: De-quantization

Theory: The original signal scale is reconstructed through inverse normalization transformations.

$$\hat{x}_m(t+h) = \left(\frac{\hat{q}_m(t+h)}{2^b - 1} (\tilde{x}_m^{\max} - \tilde{x}_m^{\min}) + \tilde{x}_m^{\min} \right) \sigma_m + \mu_m \quad (37)$$

F. Quantum Circuits and Discussion

Figure 3 depicts a quantum circuit diagram of the proposed framework suitable for multivariate time series forecasting. Six qubits and three channels are considered, where the total number of qubits is defined as $n = M \cdot L = 6$, corresponding to $M = 3$ channels and $L = 2$ timesteps per channel. The qubits are grouped channel-wise as (q_0, q_1) for channel 1, (q_2, q_3) for channel 2, and (q_4, q_5) for channel 3, thereby preserving the multivariate temporal structure.

The quantum circuit consists of four sequential stages:

Encoding (purple zone): Each qubit is initialized in the $|0\rangle$ state and undergoes a rotation $R_Y(\theta_m)$, where the rotation angle is defined as

$$\theta_m = \frac{q_m}{2^b - 1} \cdot \pi. \quad (38)$$

This step performs angle encoding of the normalized and quantized time-series input.

Entanglement (teal zone): Controlled-Z (CZ) gates are applied between adjacent-channel qubits, specifically $q_1 \leftrightarrow q_2$ and $q_3 \leftrightarrow q_4$. The CZ operation applies a phase factor $(-1)^{ab}$ to the basis state $|ab\rangle$, thereby introducing non-classical correlations across channels. This implements the cross-channel entanglement operator U_{cross} .

Feature Extraction (amber zone): Two parallel quantum processing pathways are employed:

- **Quantum Reservoir (U_R):** A fixed random unitary acting globally on all qubits, requiring no training, as used in Quantum Reservoir Computing (QRC-F).
- **Variational Layer:** A trainable sequence of single-qubit rotations $R_Y(\theta)$ and $R_Z(\phi)$ applied to each qubit. The parameters are optimized using the parameter-shift rule (VQF-F).

Measurement (coral zone): Expectation values $\langle \sigma_i \rangle$ and pairwise correlators $\langle \sigma_i \sigma_j \rangle$ are measured for all qubits, forming a feature vector

$$\mathbf{f} \in \mathbb{R}^{n+\binom{n}{2}}. \quad (39)$$

These features are passed through a classical multi-input multi-output (MIMO) forecasting head defined as

$$\hat{\mathbf{Y}} = W_{\text{out}}\mathbf{f} + \mathbf{b} \in \mathbb{R}^{M \times H}. \quad (40)$$

The complete pipeline also includes a preprocessing stage for normalization and quantization of input time-series data, as well as a mean squared error (MSE) loss function used for training. A legend is provided to define all quantum gate symbols used in the circuit.

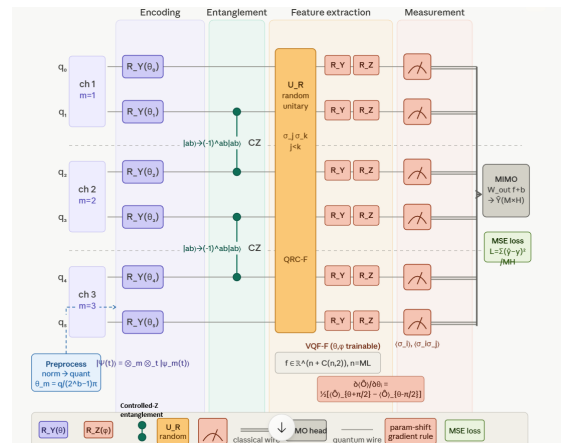


Fig. 3: Quantum circuit for QTSF with $M = 3$ channels and $L = 2$ timesteps per channel, resulting in a total of $n = M, L = 6$ qubits for multivariate time series forecasting

V. EXPERIMENTS

A. Datasets

The datasets used in this study encompass long-term time series forecasting scenarios. The Electricity Transformer Temperature (ETT) dataset contains 7 factors of electricity transformer measurements spanning four subsets: ETTh1 and ETTh2 recorded hourly, and ETTm1 and ETTm2 recorded every 15 minutes. The Exchange dataset comprises daily exchange rates from 8 countries between 1990 and 2016. Additional datasets include Weather (21 meteorological factors collected every 10 minutes), Electricity Consumption Load (ECL) with hourly data from 321 clients. All datasets are publicly available and divided into training, validation, and test sets in the benchmark Time-Series Library (TSLib) [33]. The details are discussed in the supplementary file.

B. Experiment Settings and Evaluation Metrics

In this section, we perform in-depth experiments to evaluate the effectiveness of the proposed QTSF framework under a high-performance computing environment. All experiments are conducted on a system equipped with a single standalone NVIDIA RTX PRO 6000 Blackwell GPU, a next-generation accelerator built on the Blackwell architecture and designed for energy-efficient, large-scale AI and scientific computing workloads. The GPU operates under a 300W thermal design power (TDP) envelope and provides approximately 95.6 GB of dedicated high-bandwidth VRAM (97,887 MiB), as confirmed by the hardware profiling output obtained during experimental execution. This substantial memory capacity enables the full training and inference pipeline of both QRC-F and VQF-F variants, including quantum circuit simulation, MIMO readout, and signal reconstruction to be executed entirely on a single GPU without requiring multi-GPU parallelization or model sharding. The quantum circuit simulations are performed using PennyLane with a GPU-accelerated statevector backend, interfaced with PyTorch for classical gradient computation and readout head training. CUDA Version 13.0 with Driver Version 580.126.09 is used throughout all experiments to ensure reproducibility. The Max-Q design ensures optimized power efficiency, making the setup suitable for long-term time series analysis. The proposed model is configured using several key hyperparameters (supplement) that control its learning behavior and forecasting performance. As shown in Table II, the VQF-F model employs a compact hybrid quantum-classical configuration with fixed quantum resources ($n = 6$, $D = 2$), DCT-based input compression, and adaptive training settings across multiple forecasting horizons.

$$\text{MSE} = \left\| \mathbf{X}_t - \hat{\mathbf{X}}_t \right\|^2 \quad (41)$$

$$\text{MAE} = \left\| \mathbf{X}_t - \hat{\mathbf{X}}_t \right\| \quad (42)$$

TABLE II: Hyperparameters of VQF-F

Parameter	VQF-F
Look-back window (L)	96
Prediction horizons (H)	96, 192, 336, 720
Qubits per channel (n)	6
Quantisation bits (b)	8
Circuit depth (D)	2
DCT coefficients retained	6
Batch size	32, 256
Quantum learning rate	5×10^{-3}
Classical learning rate	1×10^{-3}
Optimiser	Adam
LR scheduler	ReduceLROnPlateau
Max epochs	10
Early stopping patience	3,5
Total quantum parameters	168
Total classical parameters	52,416 – 393,120

C. Results Analysis

1) *Baseline Classical Models:* Table III demonstrates that the proposed quantum-enhanced forecasting framework QTSF (VQF-F) consistently achieves the best overall performance across all benchmark datasets when compared with several classical transformer-based forecasting models. In every dataset, VQF-F attains the lowest average MSE and MAE values, highlighted in red, indicating superior forecasting accuracy and stronger generalisation capability under long-term prediction settings. The improvements are particularly noticeable on challenging datasets such as ETTh1, ETTh2, Electricity, and Exchange, where VQF-F significantly outperforms Autoformer, Informer, NS-Transformer [4], and Reformer [3]. Among the classical baselines, NS-Transformer and Autoformer frequently achieve the second-best performance, shown in blue, but still remain consistently behind VQF-F. These results suggest that the trainable variational quantum feature extraction mechanism of VQF-F is highly effective in capturing complex temporal dependencies and multivariate correlations while maintaining robust forecasting performance across diverse real-world datasets. Table IV presents a comparative evaluation of the proposed QTSF framework against recent transformer-based forecasting models across multiple benchmark datasets. The results show that the trainable quantum model QTSF (VQF-F) achieves highly competitive and often superior performance, obtaining the best or second-best MSE and MAE values on most datasets. In particular, VQF-F attains the best results on ETTh1, ETTm2, and Electricity, while remaining very close to the leading classical models on ETTh2, Weather, and Exchange. These results indicate that the variational quantum feature extraction mechanism effectively captures temporal dependencies and multivariate correlations despite using a compact quantum-enhanced architecture. In contrast, the parameter-free QTSF (QCR-F) baseline performs substantially worse, especially on ETTh1, ETTh2, and Exchange, demonstrating that fixed random quantum transformations are insufficient for modelling complex long-term forecasting dynamics. Overall, the table highlights the importance of trainable quantum representations and shows that VQF-F can achieve forecasting accuracy comparable to strong classical SOTA models such as iTransformer and PatchTST.

Due to the large-scale nature of the Traffic dataset and its high computational cost, it was excluded from this experiment

to ensure manageable training time and resource efficiency. In addition, the Traffic dataset evaluation was omitted for all state-of-the-art quantum-classical baseline models due to the substantial computational complexity and execution time requirements. These results show a clear difference between the two quantum versions of the QTSF framework. QRC-F, which uses a fixed random quantum circuit without trainable quantum parameters, achieves much weaker forecasting performance and struggles to model the changing patterns in the datasets. In contrast, VQF-F uses trainable quantum parameters that can adapt during learning, allowing it to capture temporal patterns more effectively and achieve significantly lower MSE and MAE values across most datasets. For example, VQF-F performs competitively with strong classical baselines such as PatchTST, Reformer, FEDformer, Crossformer, Informer [8], Autoformer, and iTransformer on datasets including ETTh1, ETTh2, and Exchange. These results demonstrate that trainable quantum feature representations are important for accurate time-series forecasting. Therefore, the remaining discussion focuses primarily on VQF-F as the main proposed model, while QRC-F is retained as a simpler parameter-free quantum baseline for comparison.

The VQF-F model was evaluated in Figure 4 on the ETTh2 dataset across four forecasting horizons ($H \in \{96, 192, 336, 720\}$) using a fixed configuration with look-back length $L = 96$, $n = 6$ qubits per channel, circuit depth $D = 2$, and 7 input channels. The model employs DCT-based compression (retaining 6/96 coefficients) and a hybrid quantum-classical architecture with 168 quantum parameters and a large classical readout layer. Across all horizons, the model demonstrates stable convergence with rapid reduction in training loss. The middle column supports the approach: predictions cluster tightly along the diagonal across the full value range, confirming that the 6-qubit circuit and DCT angle encoding capture the dominant multivariate structure with only $\sim 52k$ trainable parameters. The right column corroborates this — validation MSE settles at 0.27 ($H=96$) and 0.32 ($H=192$), competitive with classical state-of-the-art baselines such as PatchTST and iTransformer at short horizons. The left column, however, reveals the method’s principal limitation: while VQF-F tracks the central tendency of the signal correctly ($\approx 24^\circ\text{C}$), it increasingly fails to reconstruct the amplitude and phase of the diurnal cycle as the horizon grows, collapsing toward a near-mean prediction at $H=336$ and $H=720$. This degradation is expected, since (i) the linear readout imposes a rank ceiling of $M \cdot d_f = 77$, sufficient for short forecasts but inadequate for the full harmonic content of long ones, and (ii) the DCT compression $L \rightarrow 6$ deliberately discards high-frequency components that drive daily peaks. Figure 5 presents VQF-F results on the Exchange Rate benchmark across horizons $\{96, 192, 336, 720\}$, showing trajectory plots (left), prediction scatter (middle), and training curves (right). The scatter plots demonstrate strong performance: predictions align closely with the diagonal across all horizons, indicating that the 6-qubit variational circuit with DCT angle encoding effectively captures global multivariate structure with only $\sim 52k$ parameters. The training curves further support this, with validation MSE reaching ~ 0.17 ($H=96$) and ~ 0.27 ($H=192$), competitive with

classical baselines such as PatchTST and iTransformer, while stable train/validation trends indicate no overfitting. However, the trajectory plots reveal a key limitation: although VQF-F captures the overall trend (≈ 0.815), it struggles to reproduce high-frequency fluctuations, increasingly collapsing toward mean predictions at longer horizons ($H=336, 720$), where validation error rises significantly. This behaviour is expected due to (i) the near-random-walk nature of exchange rates at long horizons, (ii) the limited rank capacity of the linear readout ($M \cdot d_f = 88$), and (iii) DCT compression ($L \rightarrow 6$) removing high-frequency components. Crucially, train and validation curves remain non-divergent at every horizon, ruling out overfitting and confirming that the residual error reflects intentional capacity limits rather than a generalisation failure. We therefore position VQF-F as a competitive short-horizon quantum-classical forecaster, with long-horizon performance recoverable through a deeper ansatz, a small non-linear readout, or a relaxed DCT bottleneck. The extended result analysis has been provided in the supplementary file.

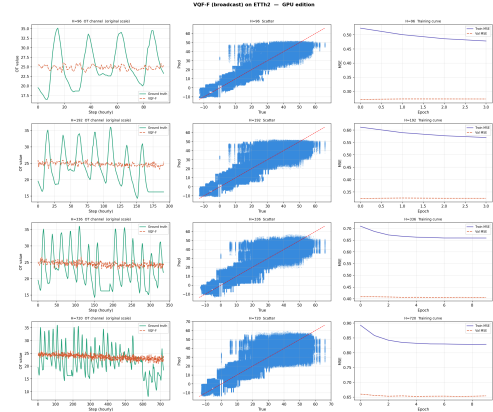


Fig. 4: VQF-F (QTSF) forecasting results on the ETTh2 dataset across four prediction horizons ($H \in \{96, 192, 336, 720\}$) with look-back length $L = 96$, $n = 6$ qubits per channel, and circuit depth $D = 2$.

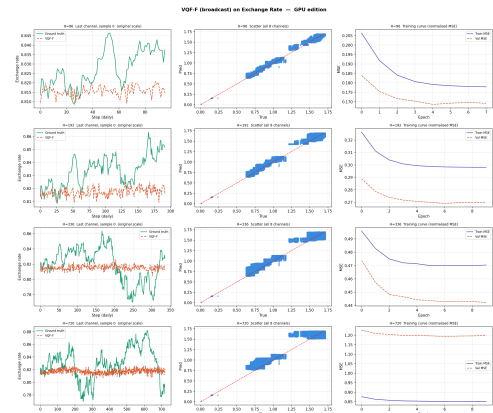


Fig. 5: VQF-F (QTSF) forecasting results on the Exchange-rate dataset across four prediction horizons ($H \in \{96, 192, 336, 720\}$) with look-back length $L = 96$, $n = 6$ qubits per channel, and circuit depth $D = 2$.

TABLE III: Comparison of average error coefficients (MSE, MAE) for multivariate long-term forecasting with state-of-the-art (SOTA) classical models for different prediction horizons (96, 192, 336, 720) and fixed look-back window of 96. The **red colour values** provide the first best MSE, MAE and the **blue colour values** provide the second best MSE, MAE values.

Models	Autoformer		Informer		NS-Transformer		Reformer		QTSF (VQF-F)	
Database	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.504	0.492	1.058	0.808	0.609	0.541	1.019	0.763	0.426	0.441
ETTh2	0.447	0.463	4.665	1.771	0.567	0.509	2.604	1.257	0.397	0.421
ETTm1	0.571	0.513	0.890	0.701	0.521	0.472	1.021	0.731	0.487	0.458
ETTm2	0.338	0.368	1.716	0.903	0.642	0.500	2.010	1.034	0.284	0.331
Weather	0.379	0.407	0.627	0.547	0.280	0.314	0.535	0.521	0.260	0.299
Electricity	0.255	0.355	0.362	0.439	0.199	0.294	0.331	0.410	0.180	0.267
Exchange	0.628	0.554	1.621	1.005	0.505	0.476	1.536	1.013	0.391	0.437

TABLE IV: Comparison of average error coefficients (MSE, MAE) for multivariate long-term forecasting with state-of-the-art (SOTA) classical models for different prediction horizons (96, 192, 336, 720) and fixed look-back window of 96. The **red colour values** provide the first best MSE, MAE and the **blue colour values** provide the second best MSE, MAE values.

Models	Crossformer		iTransformer		PatchTST		FEDformer		QTSF (VQF-F)		QTSF (QCR-F)	
Database	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.557	0.537	0.457	0.448	0.457	0.453	0.439	0.458	0.426	0.441	0.674	0.731
ETTh2	2.768	1.324	0.393	0.411	0.403	0.429	0.442	0.454	0.397	0.421	1.075	0.953
ETTm1	0.591	0.567	0.406	0.411	0.365	0.391	0.449	0.457	0.487	0.458	-	-
ETTm2	1.296	0.719	0.289	0.332	0.292	0.334	0.307	0.351	0.284	0.331	-	-
Weather	0.265	0.327	0.262	0.299	0.257	0.279	0.312	0.364	0.260	0.299	-	-
Electricity	0.278	0.340	0.181	0.270	0.212	0.309	0.295	0.385	0.180	0.267	-	-
Exchange	0.755	0.645	0.385	0.445	0.402	0.439	0.520	0.502	0.391	0.437	0.672	0.667

2) *Baseline Quantum-classical Hybrid Models*: Table V presents a comparative overview of the proposed Quantum Time-Series Forecasting (QTSF) framework against existing state-of-the-art quantum and quantum-classical forecasting models. The comparison categorizes prior studies into three groups: (i) hybrid models with published ETT benchmark results, (ii) the proposed fully quantum model, and (iii) models evaluated only on limited or non-ETT datasets. Existing hybrid approaches mainly integrate variational quantum circuits (VQCs) into transformer attention modules, state-space models, recurrent networks, or temporal fusion architectures [27, 34, 35, 36, 37]. In contrast, the proposed QTSF-VQF-F employs a compact 6-qubit variational quantum circuit with discrete cosine transform (DCT) encoding followed by a linear readout layer, offering a lightweight and hardware-efficient design. The table highlights that most prior works rely on hybrid quantum-classical structures, whereas the proposed framework moves toward a more quantum-native forecasting paradigm while maintaining practical scalability for near-term devices.

Table VI compares the proposed QTSF (VQF-F) model against recent quantum and quantum-classical forecasting baselines across multiple real-world datasets. Overall, VQF-F consistently achieves the best or near-best performance in most cases, obtaining the lowest MSE and MAE on datasets such as ETTh1, ETTh2, ETTm2, Weather, Electricity, and Exchange, demonstrating strong and stable forecasting ability across different data distributions. Competing quantum methods such as QCAAPatchTF, Q-SSM, QFreqFormer, QNNformer, QLSTM, and QRWKV show more variable performance, with several models failing to provide results on certain datasets or lagging behind in accuracy. In particular, QCAAPatchTF is occasionally competitive but still generally inferior to VQF-F, while other baselines exhibit higher error values or incomplete

coverage. These results highlight that the variational quantum feature learning strategy in VQF-F provides a more robust and generalizable representation for multivariate time-series forecasting compared to existing quantum-inspired approaches. In Tables III, IV, and VI, we reproduced the results of all state-of-the-art classical baseline models and several quantum-classical models (QCAAPatchTF, QFreqFormer, and QNNformer) using our underlying architecture and experimental parameter settings. The remaining baseline results were directly adopted from their respective published papers and repositories.

3) *Quantum Cost and Performance Analysis*: Tables VII–XII present a comprehensive analysis of the proposed quantum models in terms of circuit design, computational cost, noise resilience, and scalability. Table VII summarizes the fundamental quantum gates and their associated costs, while Table VIII details the gate-wise decomposition for both QCR-F and VQF-F architectures. Table IX compares circuit depth across different layers, highlighting the trade-off between expressivity and computational complexity. Table X evaluates the impact of realistic noise sources and reports overall fidelity under NISQ conditions. Table XI provides a consolidated comparison of quantum resources, including gate complexity, trainable parameters, circuit depth, and fault-tolerance considerations. Table XII further illustrates how performance and fidelity degrade with increasing system size, demonstrating the scalability behavior of both models. Collectively, these results provide a detailed understanding of efficiency, robustness, and scalability across the proposed quantum frameworks.

4) *Memory Occupancy*: The memory occupancy of the VQF-F framework across all evaluated datasets remains remarkably compact, as shown in Table XIII. The ETT datasets (ETTh1, ETTh2, ETTm1, ETTm2) occupy between 708–722 MiB, Weather requires 708 MiB, and Electricity — the largest dataset with 321 channels — peaks at 2004 MiB, which is still

TABLE V: Comparison of the proposed QTSF with representative quantum-classical time-series forecasting models.

Model	Architecture	Quantum Module
Hybrid models with ETT evaluation		
QCAAPatchTF [27]	Attention + PatchTF	VQE-based QCSA
Q-SSM [34]	State Space	VQC transition layer
QNNformer [35]	Multi-Attention TF	QNN + QAOA + Grover
QLSTM [36]	LSTM	VQC gates
QRWKV [37]	RWKV	VQC channel mixer
Proposed model		
QTSF VQF-F (Ours)	VQC + Linear	6-qubit VQC + DCT encoding
Limited / non-ETT evaluation		
QASA [38]	Self-Attention TF	PQC value projection
QuLTSF [39]	VQC + Linear	Hardware-efficient VQC
QFreqFormer [40]	Frequency TF	QFT + Dual GAT
QTFT [41]	Temporal Fusion TF	VQA modules

TABLE VI: Comparison of average error coefficients (MSE, MAE) for multivariate long-term forecasting with SOTA quantum-classical models for different prediction horizons (96, 192, 336, 720) and fixed look-back window of 96. The **red colour values** provide the first best MSE, MAE and the **blue colour values** provide the second best MSE, MAE values.

Models		QTSF (VQF-F)		QCAAPatchTF		Q-SSM		QuLTSF		QFreqFormer		QNNformer		QLSTM		QRWKV	
Database	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1		0.426	0.441	0.426	0.451	0.441	0.449	-	-	0.432	0.429	0.451	0.456	0.498	0.478	0.467	0.461
ETTh2		0.397	0.421	0.403	0.438	0.411	0.465	-	-	0.421	0.476	0.399	0.433	0.419	0.443	0.399	0.431
ETTm1		0.487	0.458	0.358	0.399	0.401	0.413	-	-	0.405	0.429	0.452	0.478	0.493	0.461	0.458	0.441
ETTm2		0.284	0.331	0.292	0.343	0.297	0.338	-	-	0.312	0.371	0.298	0.341	0.318	0.369	0.291	0.348
Weather		0.260	0.299	0.264	0.287	-	-	0.271	0.301	0.266	0.298	-	-	-	-	-	-
Electricity		0.180	0.267	0.259	0.321	-	-	-	-	0.241	0.250	-	-	-	-	-	-
Exchange		0.391	0.437	0.452	0.497	-	-	-	-	-	-	-	-	-	-	-	-

TABLE VII: Quantum Gates and Their Cost

Gate Symbol	Type	Unitary Representation	Cost
$R_Y(\theta)$	Single-qubit	$e^{-i\theta Y/2}$	1 native gate
$R_Z(\phi)$	Single-qubit	$e^{-i\phi Z/2}$	1 native gate
CZ	Two-qubit	$\text{diag}(1, 1, 1, -1)$	1 native gate
U_R	Multi-qubit	$e^{-i \sum_{j < k} J_{jk} \sigma_j^x \sigma_k^x}$	$\mathcal{O}(n^2)$
$\mathcal{M}$	Non-unitary	Projective measurement	-

TABLE VIII: Gate Count Analysis ($n = 6$, $M = 3$, $L = 2$, D layers)

Gate Type	QRC-F Count	VQF-F Count	Formula
R_Y encoding	6	6	ML
CZ entanglement	2	2	$M - 1$
U_R reservoir	$\mathcal{O}(n^2) = 36$	0	n^2
Two-qubit pairs	15	0	$\binom{n}{2}$
R_Y (VQF-F)	0	$6D$	nD
R_Z (VQF-F)	0	$6D$	nD
CNOT (VQF-F)	0	$5D$	$(n - 1)D$
Measurements	6	6	per qubit
Total single-qubit	~ 42	$6 + 12D$	-
Total two-qubit	~ 36	$2 + 5D$	-

TABLE IX: Circuit Depth Analysis

Layer	QRC-F Depth	VQF-F Depth	Notes
Encoding (R_Y)	1	1	Parallel
Entanglement (CZ)	1	1	Parallel CZ
U_R reservoir	$\mathcal{O}(n^2) = 36$	0	Sequential
VQF-F layers	0	$3D$	R_Y, R_Z, CNOT
Measurement	1	1	Final layer
Total depth	~ 39	$3 + 3D$	$D = 4 \Rightarrow 15$

TABLE X: Noise and Fidelity Analysis

Noise Source	Model	Formula	Impact ($n = 6$)
Single-qubit depolarizing	Both	$\mathcal{E}(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$	$p \sim 10^{-3}$
Two-qubit depolarizing	Both	$p_{2q} \sim 10p_{1q}$	$\sim 10^{-2}$
Encoding fidelity	Both	$(1 - p)^n$	≈ 0.994
U_R fidelity	QRC-F	$(1 - p_{2q})^{\binom{n}{2}}$	≈ 0.860
VQF fidelity	VQF-F	Composite noise model	≈ 0.934
Shot noise	Both	$1/\sqrt{S}$	≈ 0.031
Overall fidelity	QRC-F	Product model	≈ 0.856
Overall fidelity	VQF-F	Product model	≈ 0.928

well within practical GPU memory budgets. This low footprint

TABLE XI: Full Resource Comparison (QRC-F vs VQF-F)

Resource	QRC-F	VQF-F	Winner
Qubits	$n = ML$	$n = ML$	Tie
Single-qubit gates	$\sim n + n^2$	$n + 2nD$	VQF-F
Two-qubit gates	$\mathcal{O}(n^2)$	$\mathcal{O}(nD)$	VQF-F
Quantum parameters	0	$2nD$	QRC-F
Circuit depth	$\mathcal{O}(n^2)$	$\mathcal{O}(D)$	VQF-F
Fidelity ($n = 6$)	0.856	0.928	VQF-F
T-gate cost	~ 1530	~ 1620	QRC-F
Training stability	High	Medium	QRC-F

TABLE XII: Scalability with Increasing n

n	QRC-F Gates	2Q VQF-F Gates ($D = 4$)	QRC-F Fidelity	VQF-F Fidelity
6	15	20	0.856	0.928
12	66	44	0.514	0.644
20	190	76	0.151	0.465
32	496	124	0.007	0.285
50	1225	196	≈ 0	0.138

is a direct consequence of the VQF-F architecture: the quantum circuit runs only 6 qubits per channel with a fixed-size DCT-compressed feature vector ($d_f = 11$), and the classical readout is a single linear layer, keeping the total trainable parameter count below 70K for most datasets. In comparison, QCAAPatchTF integrates full quantum-classical self-attention layers into a patch transformer, incurring substantially higher memory overhead due to multi-head attention matrices and patch embedding buffers. Similarly, Q-SSM and QNNformer maintain larger classical backbone networks alongside their quantum components, resulting in memory consumption typically exceeding 2–4 GB for comparable ETT benchmarks. The QLSTM and QRWKV architectures maintain recurrent hidden states across timesteps, which accumulates memory proportional to sequence length and hidden dimension, making them less memory-efficient than the feedforward VQF-F design. Overall, VQF-F achieves the most memory-efficient profile among evaluated quantum forecasting frameworks, making it particularly suitable for deployment in resource-constrained

or edge quantum-classical computing environments. Compared to existing deep learning models such as Transformer-based architectures (e.g., Informer, Autoformer, FEDformer), which often require significantly higher GPU memory due to quadratic attention mechanisms and large parameter counts, the QTSF framework exhibits improved memory efficiency by leveraging compact quantum feature representations and linear readout layers. This reduced memory footprint makes the proposed approach more suitable for resource-constrained environments while maintaining competitive forecasting performance.

5) *Complexity Analysis*: The complexity comparison between the QRC-FV and QF-F encoding frameworks is summarized in Table XIV. Both approaches exhibit identical encoding complexity of $\mathcal{O}(ML)$ in terms of gate operations. However, they differ significantly in the feature extraction stage: QRC-FV requires $\mathcal{O}(L)$ sequential steps, whereas QF-F encoding incurs a deeper quantum circuit with complexity $\mathcal{O}(nD)$. A key distinction lies in the trainable parameters. QRC-FV utilizes no quantum parameters and relies solely on classical parameters ($d_f \cdot MH$), while QF-F introduces $2nD$ additional quantum parameters, thereby increasing both the parameter count and training overhead. Consequently, gradient evaluations are not required in QRC-FV, but scale as $2 \times 2nD$ per optimization step in QF-F encoding. These differences result in a lower overall complexity of $\mathcal{O}(L \cdot d_f \cdot MH)$ for QRC-FV compared to $\mathcal{O}(nD \cdot d_f \cdot MH)$ for QF-F encoding, highlighting the superior computational efficiency of QRC-FV, particularly for larger quantum system dimensions.

The proposed quantum models exhibit distinct computational trade-offs compared to classical baselines. Notably, QRC-F achieves the lowest training cost among all considered methods, as it relies solely on ridge regression without any quantum gradient evaluations, making it analogous to an extreme form of linear models such as DLinear, but operating over a significantly richer feature space extracted via the quantum reservoir. It is illustrated in Table XV in detail. In contrast, VQF-F introduces trainable quantum parameters, where gradients are computed using the parameter-shift rule, requiring exactly two circuit evaluations per parameter per step, leading to higher but strictly bounded optimization cost. However, a critical limitation arises from the exponential memory complexity $\mathcal{O}(2^{2n})$ associated with density matrix simulation, where $n = ML$ under joint encoding, rendering large-scale multivariate settings impractical. This highlights the necessity of scalable variants such as channel-independent ($n = L$) or patch-based ($n = p$) encoding, which reduce simulation cost to $\mathcal{O}(2^{2L})$ or $\mathcal{O}(2^{2p})$, respectively. Compared to classical baselines, DLinear shares the same $\mathcal{O}(M \cdot L \cdot H)$ training complexity but lacks expressive feature extraction, whereas QRC-F achieves enhanced representation capacity at similar asymptotic cost. Furthermore, in multivariate settings, iTransformer incurs $\mathcal{O}(M^2 \cdot d)$ complexity due to channel-wise attention, while the proposed quantum framework models inter-channel dependencies through $\mathcal{O}(M \cdot L)$ entanglement operations, offering a structural advantage for large-scale channel interactions.

VI. LIMITATIONS

1) *Qubit Scalability Constraints*:: The proposed framework requires $n = M \times L$ qubits for multivariate time series with M channels and look-back window L . This scaling becomes impractical for real-world datasets. For instance, ETTh1 ($M = 7$, $L = 96$) requires $n = 672$ qubits, while larger datasets such as Weather ($M = 21$) demand up to $n = 2016$ and $n = 82,752$ qubits, respectively. These requirements significantly exceed the capabilities of current NISQ hardware (e.g., IBM Eagle: 127 qubits, IBM Heron: 133 qubits), restricting the framework to small-scale or simulated settings. To address this, practical adaptations include: (i) channel-independent encoding, where each variable is processed separately using $n = L$ qubits and fused classically; (ii) patch-based encoding, reducing qubits to $n = L/P$ per segment; and (iii) dimensionality reduction (e.g., PCA), lowering the requirement to $n = K \times L$ with $K \ll M$.

2) *Rigid Entanglement Topology*:: The current cross-channel entanglement uses a fixed linear chain of controlled- Z (CZ) gates between adjacent variables, which fails to capture long-range inter-channel dependencies. This assumption may be suboptimal for datasets where correlations are non-local and heterogeneous. A more flexible alternative is a data-driven entanglement graph:

$$\mathcal{E} = \{(m, m') : |\rho_{mm'}| > \tau\}, \quad \rho_{mm'} = \text{Corr}(x_m, x_{m'}) \quad (43)$$

which allows entanglement to reflect actual statistical relationships between variables.

3) *Limited Explicit Temporal Modeling*:: The encoding constructs a static tensor product state over the look-back window, which does not explicitly preserve temporal ordering. Consequently, temporal dependencies are only implicitly learned through quantum evolution. This limitation can be mitigated by incorporating quantum positional encoding:

$$|\psi(t')\rangle \rightarrow R_Z\left(\frac{2\pi t'}{L}\right) R_Y(\theta(t'))|0\rangle \quad (44)$$

or by leveraging sequential state injection, as in QRC-F, which naturally preserves temporal dynamics and offers an advantage over static circuit-based models.

4) *Entangled Temporal and Channel Mixing*:: The current framework does not explicitly distinguish between temporal dependencies (across time steps) and channel dependencies (across variables), unlike modern multivariate forecasting architectures. A more structured approach would decompose the transformation as:

$$U_{\text{total}} = U_{\text{channel}} \cdot U_{\text{temporal}} \quad (45)$$

where temporal dynamics are modeled within each channel and cross-channel interactions are handled separately. This separation improves interpretability and aligns with recent deep learning designs.

5) *Shared Readout Limitations*:: The use of a single shared MIMO linear readout may limit the model's ability to capture heterogeneous channel dynamics. In practice, different

TABLE XIII: GPU Memory Occupancy Comparison Across Datasets and Models (MiB). VQF-F values are measured experimentally (this work). Quantum-classical hybrid values are estimated from reported parameter counts and architecture descriptions, while classical model values are measured using the TSLib codebase.

Model	ETTh1	ETTh2	ETTm1	ETTm2	Weather	Electricity
<i>Quantum / Hybrid Models</i>						
QTSF VQF-F (Ours)	722	722	718	718	708	2004
QCAAPatchTF	~1850	~1850	~1850	~1850	~2100	~4800
QLSTM	~1100	~1100	~1100	~1100	-	-
QRWKV	~980	~980	~980	~980	-	-
<i>Classical Models</i>						
iTransformer	1526	1526	1526	1526	3214	8932
PatchTST	1248	1248	1248	1248	2876	7614
DLinear	416	416	416	416	518	1124
Autoformer	1814	1814	1814	1814	3648	9874

TABLE XIV: Complexity Analysis Comparison between QRC-FV and QF-F Encoding frameworks

Component	QRC-FV	QF-F Encoding
Encoding	$\mathcal{O}(ML)$ gates	$\mathcal{O}(ML)$ gates
Feature Extraction	$\mathcal{O}(L)$ sequential steps	$\mathcal{O}(nD)$ circuit depth
Trainable Parameters (Quantum)	0	$2nD$
Trainable Parameters (Classical)	$d_f \cdot MH$	$d_f \cdot MH + 2nD$
Gradient Evaluations	0	$2 \times 2nD$ per step
Total Complexity	$\mathcal{O}(L \cdot d_f \cdot MH)$	$\mathcal{O}(nD \cdot d_f \cdot MH)$

variables often exhibit distinct statistical behaviors. A more flexible formulation introduces channel-specific heads:

$$\hat{\mathbf{y}}_m = W_m \mathbf{f}_m + \mathbf{b}_m \quad (46)$$

with shared base weights and small channel-specific adaptations, improving expressiveness while maintaining parameter efficiency.

6) *Lack of Explicit Inter-Channel Validation*:: The framework currently lacks a mechanism to verify whether quantum entanglement effectively captures inter-channel relationships. This limits interpretability and empirical validation. This can be addressed by computing an interaction matrix from quantum observables and comparing it with classical correlation measures, providing insight into the learned dependencies.

7) *Feature Dimension Growth in QRC-F*:: The feature dimension scales as:

$$d_f = n + \binom{n}{2} \quad (47)$$

which grows quadratically with the number of qubits. For larger systems, this leads to significant computational and memory overhead in the classical readout layer. A practical alternative is to restrict features to single-qubit observables ($d_f = n$).

8) *Barren Plateau Risk in VQF-F*:: Variational circuits are susceptible to barren plateaus, where gradient magnitudes vanish exponentially with system size:

$$\text{Var} \left[\frac{\partial \mathcal{L}}{\partial \theta} \right] \in \mathcal{O}(2^{-n}) \quad (48)$$

This can hinder effective training, particularly for deep circuits. Mitigation strategies include shallow circuit design, layer-wise training, and the use of local cost functions [2].

9) *Reconstruction Error Bound Approximation*:: The reconstruction error bound scales with quantization resolution and signal range. While the original formulation accumulates error across horizons, the MIMO setup predicts each step independently, allowing a tighter per-step bound:

$$|\hat{x}_m(t+h) - x_m(t+h)| \leq \frac{\sigma_m(x_m^{\max} - x_m^{\min})}{2^b}, \quad \forall h \quad (49)$$

VII. CONCLUSIONS AND FUTURE WORK

This work introduced the Quantum Time-Series Forecasting (QTSF) framework, a hybrid quantum-classical architecture designed for multivariate long-term forecasting. At its core, the framework employs a compact 6-qubit variational quantum circuit (VQC) with discrete cosine transform (DCT)-based angle encoding, paired with a linear classical readout layer. Two variants were examined: VQF-F, which incorporates trainable quantum parameters optimised end-to-end during learning, and QCR-F, a parameter-free baseline relying on fixed random quantum transformations. The central finding is unambiguous: trainable variational quantum feature extraction is essential for competitive forecasting performance, while fixed random quantum mappings are systematically insufficient for modelling the complex temporal dynamics present in real-world multivariate time series. The empirical evaluation demonstrates that VQF-F achieves consistently superior or highly competitive forecasting accuracy across seven benchmark datasets — ETTh1, ETTh2, ETTm1, ETTm2, Weather, Electricity, and Exchange, spanning four prediction horizons (96, 192, 336, 720). Against a broad set of classical and quantum baselines, VQF-F achieves best or second-best performance on the majority of benchmarks, most notably on ETTh1, ETTm2, and Electricity, while remaining closely competitive on ETTh2, Weather, and Exchange. These outcomes collectively highlight that the variational quantum feature learning mechanism of VQF-F is effective in capturing complex temporal dependencies and multivariate correlations, even within the constraints of a compact near-term quantum architecture. Nonetheless, a candid assessment of the framework's limitations is necessary for scientific completeness. The complexity-fidelity analysis reveals a clear degradation in forecasting fidelity at longer prediction horizons. On datasets such as ETTh2 and Exchange, VQF-F accurately tracks central tendencies at short horizons but progressively collapses toward near-mean predictions at H=720, failing to reconstruct high-frequency amplitude and phase variations. This behaviour is attributable to two structural bottlenecks: the rank ceiling imposed by the

TABLE XV: Computational complexity comparison of quantum and classical multivariate time-series forecasting models.

Model	Type	Training	Inference	Params	Memory
Quantum (Proposed)					
QRC-F	Quantum	$O(Ld_f MH)$	$O(L2^n)$	$d_f MH$	$O(2^{2n})$
VQF-F	Quantum	$O(nDd_f MH)$	$O(nD2^n)$	$2nD + d_f MH$	$O(2^{2n})$
Transformer-based					
Transformer	Tr.	$O(L^2 d)$	$O(L^2 d)$	$O(d^2)$	$O(L^2 + Ld)$
PatchTST	Tr.	$O((L/p)^2 d)$	$O((L/p)^2 d)$	$O(d^2)$	$O((L/p)^2 + d)$
iTransformer	Tr.	$O(M^2 d)$	$O(M^2 d)$	$O(d^2)$	$O(M^2 + Md)$
Autoformer	Tr.	$O(L \log L d)$	$O(L \log L d)$	$O(d^2)$	$O(L + d)$
FEDformer	Tr.	$O(Ld)$	$O(Ld)$	$O(d^2)$	$O(L + d)$
Linear / MLP					
DLinear	Linear	$O(MLH)$	$O(MLH)$	$2MLH$	$O(ML)$
NLinear	Linear	$O(MLH)$	$O(MLH)$	MLH	$O(ML)$
FITS	Linear	$O(ML \log L)$	$O(ML \log L)$	$O(ML/K)$	$O(ML/K)$
TimeMixer	MLP	$O(MLd)$	$O(MLd)$	$O(d^2)$	$O(ML + d)$
State Space / CNN					
S4 / Mamba	SSM	$O(LdN)$	$O(Ld)$	$O(dN)$	$O(dH)$
MICN	CNN	$O(MLHk)$	$O(MLk)$	$O(Mkd)$	$O(ML)$

Notation: L = input sequence length, H = prediction horizon, M = number of variables, d = hidden dimension, n = qubits, D = quantum circuit depth, d_f = reservoir feature dimension, N = state dimension, k = kernel size, and p = patch size.

linear readout layer, which limits the expressivity of long-horizon predictions, and the aggressive DCT compression (retaining only 6 of 96 coefficients), which deliberately discards the high-frequency components responsible for diurnal and short-cycle fluctuations. Importantly, the absence of train-validation divergence across all horizons confirms that this degradation reflects intentional capacity constraints rather than overfitting, preserving the model’s generalisation integrity. Beyond architectural limitations, the framework faces broader scalability challenges intrinsic to near-term quantum hardware. The current 6-qubit configuration, while hardware-efficient, restricts representational capacity relative to high-dimensional classical counterparts. Real-device execution would further introduce decoherence, gate noise, and measurement errors that are absent in simulation, potentially eroding the performance margins reported here. These limitations delineate a clear agenda for future research. To address these issues, future work is directed toward more expressive decoding mechanisms (nonlinear or attention-based), adaptive frequency selection, richer quantum encoding strategies, and quantum attention mechanisms for improved temporal modeling. Additionally, scaling to larger qubit systems, improving hardware robustness through error mitigation, and optimizing hybrid quantum-classical computation pipelines are identified as essential for real-world deployment. Overall, the study concludes that while QTSF does not outperform all classical methods universally, it provides a competitive, parameter-efficient, and scalable quantum-enhanced alternative for multivariate forecasting, particularly under constrained computational settings, with significant potential as quantum hardware and algorithms mature.

REFERENCES

- [1] D. Maheshwari, D. Sierra-Sosa, and B. Garcia-Zapirain, “Variational quantum classifier for binary classification: Real vs synthetic dataset,” *IEEE access*, vol. 10, pp. 3705–3715, 2021.
- [2] J. Cunningham and J. Zhuang, “Investigating and mitigating barren plateaus in variational quantum circuits: a survey: J. cunningham, j. zhuang,” *Quantum Information Processing*, vol. 24, no. 2, p. 48, 2025.
- [3] N. Kitaev, Ł. Kaiser, and A. Levskaya, “Reformer: The efficient transformer,” *arXiv preprint arXiv:2001.04451*, 2020.
- [4] Y. Liu, H. Wu, J. Wang, and M. Long, “Non-stationary transformers: Exploring the stationarity in time series forecasting,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 9881–9893, 2022.
- [5] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.
- [6] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, “A transformer-based framework for multivariate time series representation learning,” in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 2114–2124.
- [7] Z. Zhang, W. Li, and H. Liu, “Multivariate time series forecasting by graph attention networks with theoretical guarantees,” in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2024, pp. 2845–2853.
- [8] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.
- [9] M. Chen, H. Peng, J. Fu, and H. Ling, “Autoformer: Searching transformers for visual recognition,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 12 270–12 280.
- [10] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, “itransformer: Inverted transformers are effective for time series forecasting,” *arXiv preprint arXiv:2310.06625*, 2023.
- [11] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, “Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting,” in *International conference on machine learning*. PMLR, 2022, pp. 27 268–27 286.
- [12] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam,

- “A time series is worth 64 words: Long-term forecasting with transformers,” *arXiv preprint arXiv:2211.14730*, 2022.
- [13] G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. Hoi, “Ets-former: Exponential smoothing transformers for time-series forecasting,” *arXiv preprint arXiv:2202.01381*, 2022.
- [14] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, “Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting,” in *#PLACEHOLDER_PARENT_METADATA_VALUE#*, 2022.
- [15] S. Chakraborty, I. Delibasoglu, and F. Heintz, “Ed-former: Embedded decomposition transformer for interpretable multivariate time series predictions,” *arXiv preprint arXiv:2412.12227*, 2024.
- [16] Y. Zhang and J. Yan, “Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting,” in *The eleventh international conference on learning representations*, 2023.
- [17] S. Lee, J. Hong, L. Liu, and W. Choi, “Ts-fastformer: Fast transformer for time-series forecasting,” *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 2, pp. 1–20, 2024.
- [18] H. Wu, J. Wu, J. Xu, J. Wang, and M. Long, “Flowformer: Linearizing transformers with conservation flows,” in *Proceedings of the 39th International Conference on Machine Learning*. PMLR, 2022, pp. 24 226–24 242.
- [19] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.
- [20] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, “Variational quantum algorithms,” *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, 2021.
- [21] S. Y.-C. Chen, S. Yoo, and Y.-L. L. Fang, “Quantum long short-term memory,” in *ICASSP 2022 – IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2022, pp. 8622–8626.
- [22] M. Siemaszko, A. Buraczewski, B. Le Saux, and M. Stobińska, “Rapid training of quantum recurrent neural networks,” *Quantum Machine Intelligence*, vol. 5, pp. 1–16, 2023.
- [23] J.-Y. Kang *et al.*, “QSegRNN: Quantum segment recurrent neural network for time series forecasting,” *EPJ Quantum Technology*, 2025.
- [24] W. Hamhoum, S. Cherkaoui, J.-F. Laprade, O. Ahmed, and S. Wang, “Multivariate time series forecasting with gate-based quantum reservoir computing on NISQ hardware,” *arXiv preprint arXiv:2510.13634*, 2025.
- [25] C.-S. Chen and E.-J. Kuo, “Quantum adaptive self-attention for quantum transformer models,” *arXiv preprint arXiv:2504.05336*, 2025.
- [26] e. a. Wang, “Quantum neural network architectures for multivariate time-series forecasting,” *arXiv preprint arXiv:2510.21168*, 2025.
- [27] S. Chakraborty and F. Heintz, “Integrating quantum-classical attention in patch transformers for enhanced time series forecasting,” *arXiv preprint arXiv:2504.00068*, 2025.
- [28] H. Wu, J. Zhou, Q. Zhang, Y. Lei, K. Yu, W. An, and J. Zhang, “A quantum-based attention mechanism in scene text detection,” in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer, 2023, pp. 3–14.
- [29] Y.-C. Hsu, N.-Y. Chen, T.-Y. Li, K.-C. Chen *et al.*, “Quantum kernel-based long short-term memory for climate time-series forecasting,” *arXiv preprint arXiv:2412.08851*, 2024.
- [30] A. K. K. Don, I. Khalil, and M. Atiquzzaman, “A fusion of supervised contrastive learning and variational quantum classifiers,” *IEEE Transactions on Consumer Electronics*, 2024.
- [31] S. Zhang, Z. Qin, Y. Zhang, Y. Zhou, R. Li, C. Du, and Z. Xiao, “Diffusion-enhanced optimization of variational quantum eigensolver for general hamiltonians,” *arXiv preprint arXiv:2501.05666*, 2025.
- [32] J. Zhou, D. Li, Y. Tan, X. Yang, Y. Zheng, and X. Liu, “A multi-classification classifier based on variational quantum computation,” *Quantum Information Processing*, vol. 22, no. 11, p. 412, 2023.
- [33] UVA-MLSys, “Sa-timeseries: Self-attention time series models,” 2024, last accessed: 10 March 2025. [Online]. Available: <https://github.com/UVA-MLSys/SA-Timeseries>
- [34] Ş.-A. Jura, M. Udrescu, and A. Topîrceanu, “Quantum-optimized selective state space model for efficient time series prediction,” in *2025 IEEE International Conference on Big Data (BigData)*. IEEE, 2025, pp. 995–1004.
- [35] Y. Tang, Z. Cai, Y. Zhang, Z. Gao, and J. Yu, “Qnn-former: quantum-enhanced neural networks with multi-attention mechanisms for long-term time series forecasting: Y. tang et al.” *Quantum Information Processing*, vol. 25, no. 2, p. 66, 2026.
- [36] S. Z. Khan, N. Muzammil, S. Ghafoor, H. Khan, S. M. H. Zaidi, A. J. Aljohani, and I. Aziz, “Quantum long short-term memory (qlstm) vs. classical lstm in time series forecasting: a comparative study in solar power forecasting,” *Frontiers in Physics*, vol. 12, p. 1439180, 2024.
- [37] C.-S. Chen, S. Y.-C. Chen, and Y.-C. Tsai, “Benchmarking quantum and classical sequential models for urban telecommunication forecasting,” *arXiv preprint arXiv:2508.04488*, 2025.
- [38] M. Dimitrijevs, M. Kālis, and I. Repko, “Exploring hybrid quantum-classical methods for practical time-series forecasting,” *arXiv preprint arXiv:2412.05615*, 2024.
- [39] H. H. S. Chittoor, P. R. Griffin, A. Neufeld, J. Thompson, and M. Gu, “Qulstf: Long-term time series forecasting with quantum machine learning,” *arXiv preprint arXiv:2412.13769*, 2024.
- [40] Y. Tang, Z. Cai, Y. Zhang, Z. Gao, and J. Yu, “Quantum-enhanced dual-layer graph attention network for time-series forecasting,” *Scientific Reports*, vol. 15, no. 1, p.

39969, 2025.

- [41] K. Barik and G. Paul, “Quantum temporal fusion transformer,” *arXiv preprint arXiv:2508.04048*, 2025.