QAAR-SIREN: quantum-augmented attention and residual SIREN for time-series forecasting

Information Sciences 754 (2026) 123756

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

QAAR-SIREN: quantum-augmented attention and residual SIREN for time-series forecasting

Abdulkadir Sengur [a], Massimo Salvi [b,*], Prabal Datta Barua [c], Ravinesh Deo [d], Yan Li [d], U.R. Acharya [d]

[a] Electrical-Electronics Engineering Department, Technology Faculty, Firat University, Elazig, Turkey
[b] PolitoBIOMed Lab, Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, Italy
[c] School of Business, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[d] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia

ARTICLE INFO

Keywords:
Time series forecasting
Sinusoidal representation network
Attention pooling
Quantum machine learning

ABSTRACT

Time series forecasting remains challenging in the presence of nonstationarity, regime changes, and observation noise. Many existing machine learning approaches rely on complex architectures that often lead to unstable training and limited robustness. To address these limitations, we propose QAAR-SIREN, a compact forecasting framework that improves stability through residual learning and complementary feature representations. Instead of predicting absolute values, the model forecasts temporal increments, mitigating nonstationarity effects. It integrates three information sources: raw temporal lags, attention-based contextual summarization, and lightweight nonlinear features extracted from a shallow variational quantum circuit applied to the most recent observation. The quantum component functions as a compact nonlinear feature extractor that enriches the input representation without increasing architectural complexity. Experiments on synthetic signals with regime transitions and heterogeneous noise, as well as real-world datasets from climate, energy demand, finance, and transportation, demonstrate that QAAR-SIREN achieves strong and stable predictive performance. The model attains coefficients of determination up to approximately 0.985 with low mean squared error. Ablation studies confirm that observed gains arise from the complementary effects of residual learning, attention-based context aggregation, and quantum feature extraction.

1. Introduction

Time series forecasting is a fundamental challenge in data science, where the goal is to predict future values based on historical temporal patterns [1]. Accurate prediction is particularly difficult for time series due to their complex temporal relationships, which include trends, seasonality, and structural breaks, in contrast to independent data sets [2,3]. Standard assumptions regarding independence and identical distribution are further challenged by real-world data problems such calendar impacts, time-zone changes, missing values, irregular sampling, and measurement drift [4].

Classical forecasting methods, including exponential smoothing, ARIMA, and state-space models, remain widely used because they offer interpretable decompositions and a principled way to quantify uncertainty when the underlying dynamics are reasonably stable

* Corresponding author.
E-mail address: massimo.salvi@polito.it (M. Salvi).

https://doi.org/10.1016/j.ins.2026.123756
Received 14 November 2025; Received in revised form 24 May 2026; Accepted 5 June 2026
Available online 7 June 2026

[5]. For more complex settings, especially when the data exhibit strong nonlinear behavior or long-range temporal dependencies, deep learning models based on convolution and attention have shown clear advantages in learning richer temporal patterns. Their performance can be further improved when exogenous information, such as calendar effects, events, or domain-specific variables, is available [6,7]. Whatever modeling strategy is adopted, evaluation must reflect the sequential nature of forecasting. In practice, the validation involves rolling-origin or walk-forward validation, reporting scale-aware error measures, and checking robustness under distribution shifts so that the reported performance better matches real forecasting conditions, where only past observations are available at each step [8].

Over the last decade, machine learning has expanded the forecasting toolkit with global models that learn across many related series, probabilistic outputs that quantify uncertainty, and architectures tailored to long-horizon dependencies. The benefits of training probabilistic models across several series for better calibration and accuracy were shown by early recurrent techniques like DeepAR [9]. Purely feed-forward designs, exemplified by N-BEATS, achieved strong univariate accuracy through backward/forward residual stacks learned end-to-end without hand-crafted seasonality, often rivaling statistical baselines on public benchmarks [10]. For multi-horizon settings with covariates, the Temporal Fusion Transformer combined static/context gating with attention to deliver both performance and interpretability through variable- and time-dependent importance scores [11]. More recent transformer-based architectures designed for computational efficiency and longer input contexts, such as Informer and patch-based models like PatchTST, have further strengthened scalability and robustness, particularly for high-frequency data and long temporal sequences [12,13]. At the same time, large community benchmarks such as the M4 and M5 competitions have provided a more rigorous basis for comparing forecasting models, highlighting the continued importance of strong statistical baselines, careful rolling-origin evaluation, and probabilistic performance measures [14]. Together, these studies have shown that deep learning methods can be effective across a broad range of time-series applications.

In this study, we propose QAAR-SIREN, a hybrid architecture designed to improve time-series forecasting by combining three complementary sources of information. The first branch retains the raw lagged observations, preserving the original temporal information without additional transformation. The second branch applies a lightweight attention mechanism to the input window, assigning different importance to individual time steps and producing a compact context vector that summarizes the most relevant temporal patterns. The third stream employs a shallow two-qubit variational quantum circuit to encode the most recent observation using rotation gates and entanglement operations, resulting in four expectation values that capture quantum-enhanced features that are potentially unavailable to classical encoding approaches. These three data streams are combined and processed by a Sinusoidal Representation Network (SIREN), which employs sine activation functions built specifically to capture high-frequency temporal patterns and provide stable gradients for coordinate-based neural networks [15]. To improve stability during training, the model is formulated in a residual setting, where it predicts the change with respect to the most recent observation instead of forecasting the next value directly. The quantum circuit parameters are updated through the parameter-shift rule, while the classical components are optimized end to end with Adam, enabling joint training of the full architecture in a stable manner. Experiments were carried out on datasets from several domains, including climate observations, electricity demand, financial market indices, and passenger traffic. Each dataset was split into training, validation, and test segments with a standard sliding window setup. Performance was assessed with mean squared error to capture scale-dependent accuracy and with R^2 to quantify explained variance. We also compared against strong baselines using the same splits and reported results on the held-out test segments to reflect true generalization. The main contributions of this study are:

1. A novel three-stream input that combines raw lags, a short attention summary of the window, and a four-value feature vector from a shallow two-qubit quantum circuit. This joint representation is fed to a frequency-calibrated SIREN regressor for next-step prediction.
2. An end-to-end training framework where the quantum circuit and the neural network are trained together in a single loop. Circuit angles are updated using the parameter shift rule, and network weights are learned through backpropagation, ensuring stable joint learning across quantum and classical components.
3. A lightweight attention pooling module tailored to sliding windows that replaces recurrent states by producing a fixed-size context vector that preserves important temporal cues while keeping the input compact.

1.1. Related works

A comprehensive study of deep learning for healthcare time-series prediction was carried out by Morid et al. [3], who created a unifying taxonomy that arranges modeling options across typical clinical applications. The findings emphasize unresolved issues with irregular sampling, uncertainty measurement, and clinically useful explanations. The authors reported consistent performance gains of deep models over traditional baselines with sufficient data and proper design. In their assessment of deep learning techniques for time-series analysis, Gamboa et al. [16] created a systematic overview of convolutional, recurrent, and autoencoding architectures, highlighting the fact that learnt representations usually perform better than manually constructed pipelines for forecasting and classification. A stock price prediction system that combines machine learning techniques, statistical models, and Long Short-Term Memory (LSTM) in an agglomerative design was created by Mehtab et al. [17]. The authors demonstrated that deep learning and hybrid approaches outperformed classical baselines for both direction prediction and price forecasting on real market data. Parray et al. [18] analyzed stock price movement with machine learning. They built features from historical prices and technical indicators. They compared several classifiers and regressors on real market data. Tree ensembles and support vector machines performed well.

Recurrent networks were competitive with longer input windows. Environmental applications have also benefited from modern machine learning, with Barrera-Animas et al. [19] evaluated rainfall prediction with modern machine learning. They found that gradient boosting and random forests worked well for short horizons while recurrent and attention-based models improved multi-step forecasts. Bontempi et al. [20] surveyed machine learning strategies for time series forecasting, contrasting local and global modeling approaches and discussing feature engineering with lags and sliding windows. Rather than focusing on a single benchmark, that study mainly offered practical guidance on model design. More recent work has examined forecasting problems with domain-specific challenges. For example, Zeroual et al. [21] compared several deep learning approaches for COVID-19 forecasting using country-level time series of cases and deaths, evaluating recurrent and convolutional models against classical baselines. Their results showed that deep models often reduced prediction error, although performance remained sensitive to the choice of input window and to data quality. More recently, Akpinar et al. [22] introduced an uncertainty-aware liquid neural network for noisy time-series data, combining adaptive state dynamics with calibrated uncertainty estimates to improve both accuracy and robustness relative to strong baseline methods.

After the success of deep learning methods, an emerging research direction explores quantum machine learning for time series prediction, leveraging quantum phenomena to potentially capture temporal patterns in fundamentally new ways. Quantum machine learning for time series represents a paradigm shift from classical approaches, as quantum bits can exist in superposition and quantum entanglement can capture complex correlations between different time points. Saghafi and Mili [23] were the first to take this approach by using time series modeling to represent the results of repeated single-qubit circuits. They created models to predict future error rates and error polarity after analyzing error dynamics from a superconducting device over numerous runs. The authors reported clear temporal regularities in the data and showed improvements over naive and static baselines. Rivera Ruiz et al. [24] further investigated the use of quantum machine learning for forecasting by developing parameterized quantum circuits and hybrid quantum–classical models. Their study considered standard forecasting tasks with covariates and multiple prediction horizons, and compared the proposed methods with recurrent and attention-based baselines. The reported results were competitive, and the paper also provided useful details on the training procedure and computational cost, both in simulation and on limited quantum hardware. Alaminos et al. [25] examined GDP growth forecasting using a combination of quantum computing and deep learning, where quantum routines were used for representation learning and neural networks for prediction. Using country-level macroeconomic series over multiple horizons, they observed performance gains in some settings compared with strong classical alternatives, while also discussing the effects of limited data and computational constraints. Habibi et al. [26] framed short-term load prediction as a univariate time series task and proposed quantum artificial intelligence for electrical load forecasting. On two lab loads, they constructed a hybrid quantum–classical network that was trained on utility-style load traces and showed reduced error rates compared to conventional deep models. They also noted improved robustness under noise and discussed scalability.

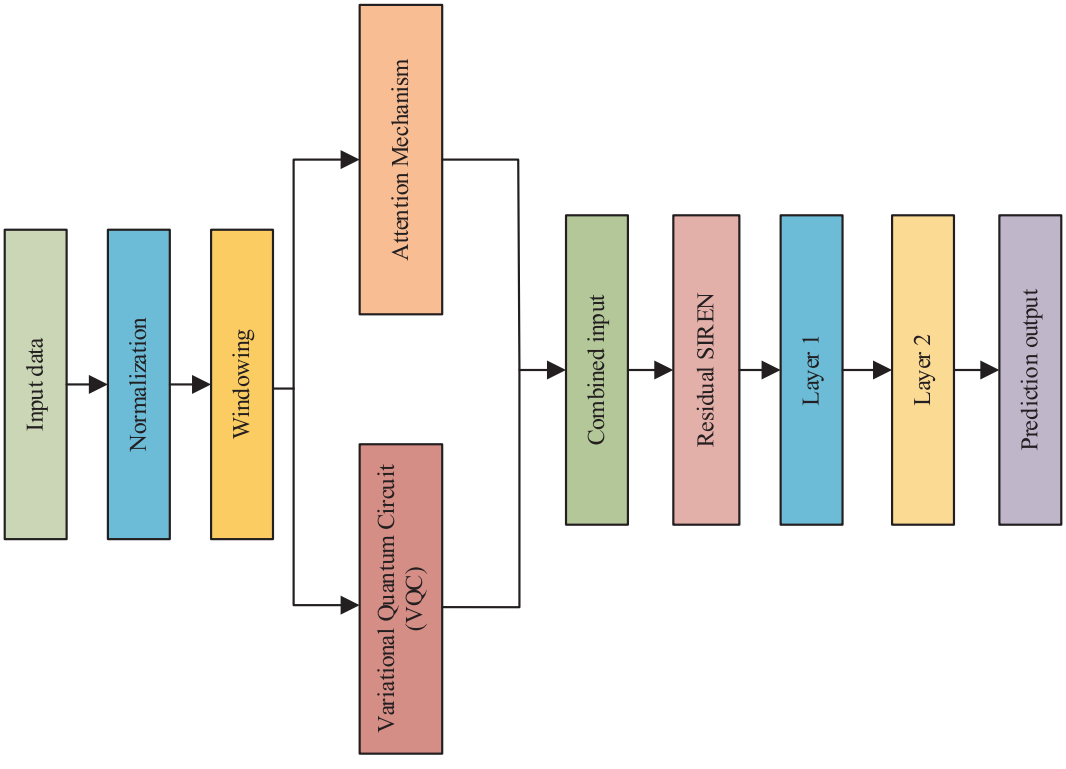

Fig. 1. Methodological flowchart of the proposed model.

However, a number of significant drawbacks restrict the application of current quantum methods for time series forecasting. First of all, most quantum methods operate in isolation without leveraging state-of-the-art techniques such as attention mechanisms, which have redefined sequence modeling in natural language processing and computer vision. Second, many quantum time series methods focus on complex, deep quantum circuits that may be impractical for near-term quantum devices due to noise and decoherence limitations. Third, there has been limited research into hybrid architectures that can combine the representational capacity of quantum encoding with the demonstrated success of specialized neural network architectures created for temporal data.

2. Materials and methods

We present a time series forecasting approach that uses residual learning with a quantum augmented-attention-based SIREN backbone. The input time series data is initially normalized and framed into sliding windows, which are then converted into a supervised learning format [22]. The model is trained to predict the change from the last value in each window, and the final forecast is obtained by adding this predicted change back to the previous value, as summarized in Fig. 1.

In the forward pass, two auxiliary branches are used to enrich the input before it reaches the final regressor. The first is an attention module that compresses the sliding window into a compact context vector, highlighting the most informative temporal patterns. The second is a shallow variational quantum circuit (VQC) [27], which processes only the most recent value in the window and returns four expectation values. These outputs form a small continuous feature vector that introduces an additional nonlinear representation of the input. In our framework, the VQC is not intended to act as a standalone predictor; instead, it serves as a lightweight feature extractor that complements the classical information. From this perspective, the quantum outputs can be viewed as learned nonlinear basis features, similar in spirit to kernel-based mappings in classical machine learning, while preserving a very shallow circuit design.

These quantum features are concatenated with the raw temporal window and the attention-based context vector, and the combined representation is passed to a SIREN network with sinusoidal activations. This design allows the model to capture fine, high-frequency temporal patterns that standard activations often miss. The SIREN outputs a residual, which is added to the last observed value to produce the final prediction. Training minimizes mean squared error on the residuals using Adam optimization with early stopping, and gradients are propagated through all components, including the quantum circuit via the parameter-shift rule. Evaluation reports test MSE and R^2, and the results are visualized by comparing true and predicted trajectories, in accordance with the schematic shown in Fig. 1.

2.1. Sinusoidal neural network (SIREN)

A SIREN is a coordinate-based multilayer perceptron that uses the sine function as its nonlinearity, enabling it to represent signals with rich, high-frequency structure [15]. Given an input coordinate $x \in \mathbb{R}^{d_{in}}$ and desired output $y \in \mathbb{R}^{d_{out}}$, the network defines a function $f_\theta : x \in \mathbb{R}^{d_{in}} \to y \in \mathbb{R}^{d_{out}}$ by the recursion:

$$
\begin{aligned}
z_1(x) &= \omega_0 W_1 x + b_1 \\
a_1(x) &= \sin(z_1(x)) \\
z_\ell(x) &= W_\ell a_{\ell-1}(x) + b_\ell, \ell = 2, \cdots, L-1 \\
a_\ell(x) &= \sin(z_\ell(x)) \\
f_\theta(x) &= W_L a_{L-1}(x) + b_L
\end{aligned}
\tag{1}
$$

The scalar $\omega_0 > 0$ controls the frequency sensitivity of the first layer: larger values make the model more responsive to rapid variations in the input, whereas smaller values lead to smoother functions near initialization. Training is performed on samples $\{(x_i, y_i)\}_{i=1}^{N}$ by minimizing a loss such as mean squared error [28]:

$$
\min_\theta L(\theta) = \left(\frac{1}{N}\right) \sum_{i=1}^{N} \left\| f_\theta(x_i) - y_i \right\|_2^2 \tag{2}
$$

The loss can also be augmented with task-specific regularization (for example, PDE residuals when f is a physical field). Even with a single hidden layer, the model can already be interpreted as a learned superposition of sinusoidal components. In component form [29], the output can be written as:

$$
f_\theta(x) = W_L \sin(\omega_0 W_1 x + b_1) + b_L = \sum_k \alpha_k \sin(\langle \xi_k, x \rangle + \phi_k) \tag{3}
$$

where directions ξ_k are determined by the rows of W_1, the phases ϕ_k by b_1, and the amplitudes α_k by W_L. By stacking layers, the network composes these sinusoidal responses and can represent spatially or temporally varying frequencies instead of a single global Fourier expansion. Another useful property of SIRENs is that their derivatives remain simple and bounded, which is advantageous in tasks where gradients or curvature information is required [30]. For $\sigma(z) = \sin z$:

$$\sigma'(z) = \cos(z),\ \sigma''(z) = -\sin(z), \sigma^{n}(z) = \sin\left(z + \frac{n\pi}{2}\right) \tag{4}$$

and therefore, by the chain rule, the Jacobian and Hessian of f_θ exist everywhere and remain analytic. For example, with respect to input:

$$\nabla x f_\theta(x) = W_L diag(\cos z_{L-1}) W_{L-1} \cdots diag(\cos z_1)(\omega_0 W_1) \tag{5}$$

To keep signals and gradients well-scaled across depth, the weight variances can be chosen so as to preserve unit pre-activation variance. Assume that the inputs are standardized, so that $Var[x_j] \approx 1$ and weights are i.i.d. with $\mathrm{Var}[W_{\ell,ij}] = \sigma_\ell^2$. If z_ℓ is approximated as zero-mean and uses $E[\sin^2(z)] \approx \frac{1}{2}$ and $E[\cos^2(z)] \approx \frac{1}{2}$ when Var[z]≈1, forward variance propagation gives:

$$Var[z_\ell] \approx n_{\ell-1}\sigma_\ell^2 Var[a_{\ell-1}] \approx n_{\ell-1}\sigma_\ell^2 \cdot \left(\frac{1}{2}\right) \tag{6}$$

which suggests, by imposing $Var[z_\ell] \approx 1$:

$$\sigma_\ell^2 \approx \frac{2}{n_{\ell-1}} (\ell \geq 2) \tag{7}$$

For the first layer, $z_1 = \omega_0 W_1 x + b_1$:

$$Var[z_1] \approx \omega_0^2 n_0 \sigma_1^2 Var[x] \approx 1 \Rightarrow \sigma_1^2 \approx \frac{1}{\omega_0^2 n_0} \tag{8}$$

A similar backward analysis for the pre-activation gradient $\delta_\ell = \partial L / \partial z_\ell$ gives:

$$\delta_{\ell-1} = \left(W_\ell^T \delta_\ell\right) \odot \cos(z_{\ell-1}), Var[\delta_{\ell-1}] \approx n_\ell \sigma_\ell^2 \cdot \left(\frac{1}{2}\right) Var[\delta_\ell] \tag{9}$$

which leads to the same condition and thus a depth-stable recipe. Under this view, ω_0 in the first layer becomes the main parameter controlling frequency sensitivity. Intuitively, if coordinates are in the range [−1,1], taking $\omega_0 \approx 30$ means a unit movement in input can sweep several sine periods in the first layer, allowing early representation of high frequencies. If a smoother response is desired, or if the inputs are not properly normalized, it is generally preferable to use a smaller ω_0 or to rescale the inputs accordingly. Due to the periodic nature of the sine activation, SIRENs are able to model both stationary and non-stationary patterns through suitable linear combinations and compositions of sinusoidal responses. As with other universal function approximators, a SIREN can represent any continuous function on a compact domain when given sufficient width, with the argument extending naturally to periodic activations. In practice, the depth and width of the network determine how rich the local frequency representation can become, while the learned parameters adjust directions, phases, and amplitudes to fit the signal of interest.

When additional constraints are required, they can be incorporated directly into the loss function. For example, if $u = f_\theta$ solves a partial differential equation $\mathscr{D}[u] = 0$, a physics loss $\lambda \left\| \mathscr{D}[f_\theta] \right\|^2$ can be introduced during training. Boundary conditions enter as penalty terms, such as $\mu \left\| f_{\theta(x)} - g(x) \right\|^2$ on boundary samples.

From an implementation perspective, a practical SIREN uses sine activations in every hidden layer and a linear output layer, together with an initialization scheme chosen to keep activations and gradients well scaled. Specifically, the first layer is commonly initialized as:

$$W_{1,ij}\ N\left(0, \frac{1}{\omega_0^2 n_0}\right), W_{l,ij}\ N\left(0, \frac{2}{n_{l-1}}\right) for \ell \geq 2 \tag{10}$$

In practice, small or zero biases, standardized coordinates, and an ω_0 selected according to the expected frequency content of the target signal provide a stable starting point for training. This setup allows gradient-based optimization to capture both low- and high-frequency structure while preserving access to derivatives of any order.

2.2. Quantum encoding

Quantum encoding is the process of mapping classical data into quantum states, enabling a circuit to process information using superposition and entanglement [26,30]. Formally, we choose a feature map $U_\phi(x)$ that depends on an input x and prepare $|\psi(x)\rangle = U_\phi(x)|0\rangle$ trainable gates U_θ, then act on $|\psi(x)\rangle$, and measurements turn the state back into numbers via expectation values like $\langle O \rangle = \langle \psi(x) / O / \psi(x) \rangle$. The encoding determines the hypothesis space the model can explore, much like a kernel: different U_ϕ induce different inner products $k(x, x') = |\langle \psi(x) | \psi(x') \rangle|^2$. Common choices include basis encoding (writing bits into computational basis states), amplitude encoding (packing a whole vector into amplitudes, efficient in qubits but costly to prepare), and angle encoding, which rotates qubits by functions of the input (e.g., $R_X(\kappa x), R_Z(\kappa x)$) and is practical on near-term hardware. Entangling layers after the data-dependent rotations let the circuit represent correlations that single-qubit maps cannot. Gradients with respect to circuit

parameters are accessible through the parameter-shift rule, $\partial\langle O\rangle = 1/2\big(\langle O\rangle_{\theta+\pi/2} - \langle O\rangle_{\theta-\pi/2}\big)$, enabling standard gradient-based training. In practice, effective encodings require a balance between expressivity and trainability: inputs are typically scaled so that rotation angles remain in a reasonable range, circuit depth is kept limited to reduce the risk of barren plateaus, and observables are selected to reflect the structure one aims to model.

2.3. Proposed quantum-augmented attention and residual SIREN

The proposed QAAR-SIREN algorithm performs one-step forecasting by learning residuals and fusing three complementary inputs: raw lags, an attention summary, and compact quantum features within a SIREN predictor. Given a window $x_t = [y_{t-w+1}, \cdots, y_t] \in \mathbb{R}^w$, the target is the increment $r_{t+1} = y_{t+1} - y_t$ and the final prediction is reconstructed by $\hat{y}_{t+1} = y_t + \hat{r}_{t+1}$. The attention pathway compresses the window into $c_t \in \mathbb{R}^{d_a}$. In its light linear form, it is:

$$c_t = W_a x_t + b_a \tag{11}$$

while an alternative normalized attention uses scores $s_t = W_s x_t + b_s$, weights $\alpha = \mathrm{softmax}(s_t)$ and

$$c_t = \sum_{i=1}^{w} \alpha_i x_{t-w+i}, \frac{\partial \alpha}{\partial s} = \mathrm{diag}(\alpha) - \alpha\,\alpha^{\mathrm{T}} \tag{12}$$

The quantum pathway maps only the most recent sample $x_t^{(w)}$ into a two-qubit variational circuit with parameters θ. With data-dependent input rotations $R_X\big(\kappa x_t^{(w)}\big)$, followed by an entangling CNOT, the circuit state $|\psi_t(\theta)\rangle$ yields four expectation values:

$$z_{t,1} = \langle Z \otimes I\rangle, z_{t,2} = \langle I \otimes Z\rangle, z_{t,3} = \langle Z \otimes Z\rangle, z_{t,4} = \langle X \otimes X\rangle \tag{13}$$

which form $z_t = \big[z_{t,1}, z_{t,2}, z_{t,3}, z_{t,4}\big]^T \in \mathbb{R}^4$. These features are differentiable with respect to circuit angles via the parameter-shift rule: for any parameter θ_k:

$$\frac{\partial z_{t,j}}{\partial \theta_k} = \frac{1}{2} \cdot \left(z_{t,j}\left(\theta_k + \frac{\pi}{2}\right) - z_{t,j}\left(\theta_k - \frac{\pi}{2}\right) \right) \tag{14}$$

The three streams are concatenated as:

$$h_t = [x_t; c_t; z_t] \in \mathbb{R}^{w+d_a+4} \tag{15}$$

and fed to a residual SIREN, a small MLP with sinusoidal activations suited to high-frequency temporal structure:

$$a_1 = \sin(w_0(W_1 h_t + b_1)), a_2 = \sin(W_2 a_1 + b_2), \hat{r}_{t+1} = W_3 a_2 + b_3 \tag{16}$$

Here $w_0 > 0$ scales the first layer's pre-activation to place the sine nonlinearity in a useful regime. The residual integration is at the model output level, not as an internal skip, so training is done on $\hat{r}_{t+1}$ versus r_{t+1}, then $\hat{y}_{t+1} = y_t + \hat{r}_{t+1}$. Learning minimizes mean squared error on residuals over a batch B:

$$L = \frac{1}{|B|} \cdot \sum_{t \in B} (\hat{r}_{t+1} - r_{t+1})^2 \tag{17}$$

Gradients propagate through SIREN in the usual way. Writing $\delta = \frac{\partial L}{\partial \hat{r}_{t+1}}$, one obtains,

$$\frac{\partial L}{\partial W_3} = \delta \cdot a_2^T, \frac{\partial L}{\partial b_3} = \delta$$

$$\frac{\partial L}{\partial W_2} = \big(\big(W_3^T \delta\big) \odot \cos(W_2 a_1 + b_2)\big) \cdot a_1^T, \frac{\partial L}{\partial b_2} = \big(W_3^T \delta\big) \odot \cos(W_2 a_1 + b_2)$$

$$\frac{\partial L}{\partial W_1} = \big(w_0 \cdot \big(W_2^T\big(\big(W_3^T \delta\big) \odot \cos(W_2 a_1 + b_2)\big)\big) \odot \cos(w_0(W_1 h_t + b_1))\big) \cdot h_t^T$$

$$\frac{\partial L}{\partial b_1} = w_0 \cdot \big(W_2^T\big(\big(W_3^T \delta\big) \odot \cos(W_2 a_1 + b_2)\big)\big) \odot \cos(w_0(W_1 h_t + b_1)) \tag{18}$$

These give the upstream gradient w.r.t. the concatenated input,

$$g_h = \frac{\partial L}{\partial h_t} = W_1^T\big(w_0 \cdot \big(W_2^T\big(\big(W_3^T \delta\big) \odot \cos(W_2 a_1 + b_2)\big)\big) \odot \cos(w_0(W_1 h_t + b_1))\big) \tag{19}$$

Splitting $g_h = [g_x; g_c; g_z]$ cross the three blocks yield direct updates for attention and quantum parts. For linear attention:

$$\frac{\partial L}{\partial W_a} = g_c \cdot x_t^T, \frac{\partial L}{\partial b_a} = g_c \tag{20}$$

and for softmax attention, the additional Jacobian $\partial\alpha/\partial_s$ above applies inside the chain rule before forming C_t. For the quantum angles:

$$\frac{\partial L}{\partial \theta_k} = \sum_{j=1}^{4} g_{z,j} \frac{\partial z_{t,j}}{\partial \theta_k} = \frac{1}{2} \cdot \sum_{j=1}^{4} g_{z,j} \left(z_{t,j}\left(\theta_k + \frac{\pi}{2}\right) - z_{t,j}\left(\theta_k - \frac{\pi}{2}\right) \right) \tag{21}$$

which combines cleanly with any first-order optimizer. Inference simply forms x_t, c_t, z_t, computes $\hat{r}_{t+1}$ through the SIREN and adds back y_t to produce $\hat{y}_{t+1}$.

3. Experimental works

3.1. Simulation data experiments

The first experiment for time-series prediction was conducted with a simulated sinusoidal signal, which is given as follows:

$$y(t) = \sin\left(2\pi \frac{t}{50}\right) + 0.35\sin\left(2\pi \frac{t}{7} + \frac{\pi}{6}\right) + n(t) \tag{22}$$

where n(t) represents Gaussian noise. For our first experiments, we chose a simple waveform that included two sinusoidal components and a Gaussian noise component. The generated signal consists of 1400 samples created by combining two sinusoidal waves. The first component is the dominant one, with a period of 50 samples and an amplitude of 1.0. The second adds a weaker oscillation, with a period of 7 samples, an amplitude of 0.35, and a phase shift of $\pi/6$. To make the series more realistic, we added white Gaussian noise at a signal-to-noise ratio of 30 dB. The signal was then standardized with z-score normalization so that differences in mean and scale were removed, resulting in a zero-mean, unit-variance series suitable for analysis with a 50-sample window and for the forecasting experiments that followed. The network used a three-layer SIREN structure with two hidden layers, each containing 32 neurons, and a sinusoidal activation scaling factor of w_0 set to 8. The window length determined the input dimension, the 16-dimensional attention embedding, and an extra four-dimensional quantum feature vector. The VQC included one quantum layer with two qubits, each controlled by RX and RZ rotation gates, followed by a CNOT entanglement operation. The input amplitude was encoded with a scaling factor of 0.8, while the rotation parameters were initialized with small random values centered around 0.05. Training was carried out for up to 150 epochs using Adam with a learning rate of 5×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a batch size of 16. A validation split of 10% was used from the training set, and early stopping was triggered when the validation mean squared error failed to improve for 15 consecutive epochs. Model performance was then evaluated on a held-out test set using both MSE and the coefficient of determination (R^2).

Fig. 2 compares the predicted and ground-truth signals produced by the proposed QAAR-SIREN model and the conventional SIREN baseline. Even though the waveform was corrupted with additive Gaussian noise at an SNR of 30 dB, both models remained closely aligned with the underlying signal across the training and test portions. The prediction of the QAAR-SIREN model shows high temporal alignment and robustness against noise, capturing both the primary and secondary frequency components of the signal with minimal distortion (Fig. 2a-b). While the general trend of the signal is preserved, the prediction of the SIREN model exhibits a noticeable loss of fine-grained detail, especially in the higher-frequency regions toward the end of the sequence (Fig. 2c-d). Compared to QAAR-SIREN, the classical model shows a lag in phase tracking and slightly reduced amplitude fidelity. This visual comparison highlights the advantage of integrating attention and quantum-inspired encoding for improved temporal feature learning and more accurate time-series reconstruction.

Table 1 reports the quantitative results for the conventional SIREN model and the proposed QAAR-SIREN. Both models achieved very high accuracy, with R^2 values above 0.99, indicating a close match between the predicted and true signal values. Although QAAR-SIREN produced a slightly higher mean squared error than the baseline SIREN, it achieved a marginally better R^2, suggesting a small overall improvement in how well the signal was reconstructed.

Table 2 reports the forecasting performance of QAAR-SIREN and a classical SIREN baseline across different signal-to-noise ratios (SNRs). Reported values correspond to the mean performance over 20 independent runs for each SNR level, using identical random seeds and noise realizations for both models. To determine whether the observed differences were statistically meaningful, we carried out paired tests over repeated runs. For each SNR level, QAAR-SIREN and SIREN were compared using the Wilcoxon signed-rank test on paired outcomes obtained with the same random seeds. Statistical significance was assessed separately for MSE and R^2.

At low SNR levels (5–20 dB), both models face clear reconstruction difficulties because of the stronger noise corruption. In this setting, QAAR-SIREN consistently produces lower MSE and higher R^2 than the classical SIREN, and the differences are statistically significant for both metrics ($p < 0.01$). This suggests that the proposed architecture delivers more reliable gains precisely in the noisy conditions where robustness matters most.

When the SNR increases to intermediate values (25–30 dB), both models improve and their results become more similar. At 25 dB, QAAR-SIREN still shows a statistically significant advantage in MSE, whereas at 30 dB the significant difference remains only for R^2. This suggests that the gap gradually shrinks as the noise level decreases, although QAAR-SIREN still retains a measurable advantage in

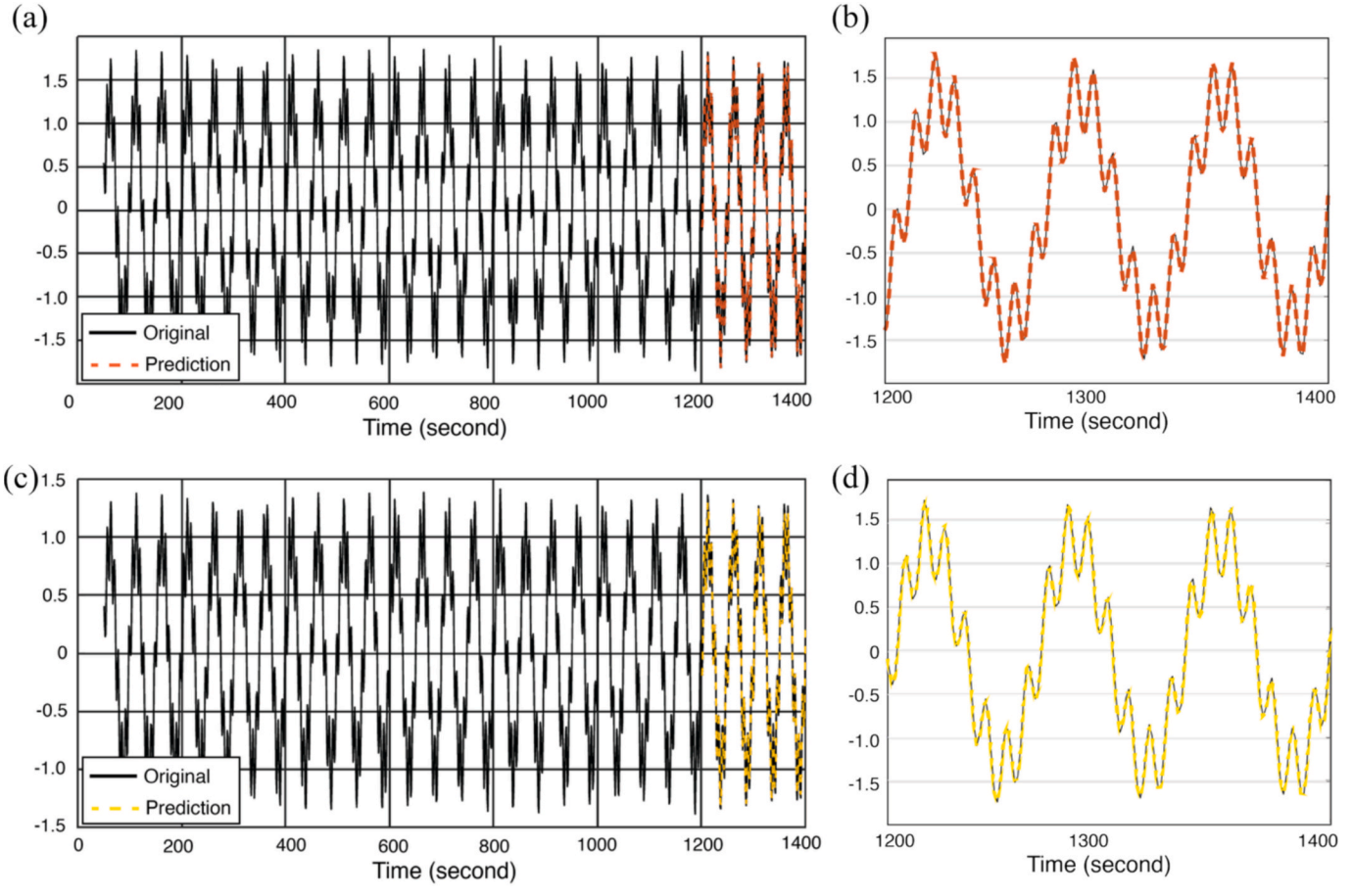

Fig. 2. Comparison of input and predicted signals using QAAR-SIREN and conventional SIREN on simulated sinusoidal data with Gaussian noise (SNR = 30 dB). a) Whole signal with QAAR-SIREN prediction, b) Zoomed view of QAAR-SIREN prediction (red), c) Whole signal with SIREN signal, d) Zoomed view of conventional SIREN prediction (yellow).

Table 1
A quantitative evaluation of the obtained predictions with both QAAR-SIREN and the original SIREN on simulated data.

Method	MSE	R^2
SIREN	0.0020	0.9964
QAAR-SIREN	0.0023	0.9977

Table 2
Quantitative evaluation of the obtained predictions using both QAAR-SIREN and the original SIREN under different SNR levels. * Statistically significant difference with respect to SIREN according to the Wilcoxon signed-rank test ($p < 0.01$).

SNR (dB)	Method	MSE	R^2
5	SIREN	0.4304	0.5810
	QAAR-SIREN	0.4236*	0.5876*
10	SIREN	0.1749	0.8231
	QAAR-SIREN	0.1722*	0.8258*
15	SIREN	0.0641	0.9355
	QAAR-SIREN	0.0626*	0.9370*
20	SIREN	0.0218	0.9780
	QAAR-SIREN	0.0214*	0.9784*
25	SIREN	0.0071	0.9930
	QAAR-SIREN	0.0069*	0.9928
30	SIREN	0.0028	0.9965
	QAAR-SIREN	0.0024	0.9976*
35	SIREN	0.0010	0.9990
	QAAR-SIREN	0.0009	0.9990
40	SIREN	0.0003	0.9996
	QAAR-SIREN	0.0003	0.9996

at least one metric.

At high SNR values (35–40 dB), both models operate on relatively clean signals and approach a saturation regime, with MSE values close to zero and R^2 values close to one. In this regime, performance differences become negligible and are no longer statistically distinguishable, reflecting that both architectures perform near-optimally when noise is minimal.

Overall, these results indicate that the proposed architecture demonstrates its strongest advantage in high-noise conditions (5)–(20) dB), where both metrics show consistent statistical significance, providing robust improvements precisely where noise resilience is most needed. At moderate noise levels (25–30 dB), the gap narrows but remains significant for at least one metric, while at very low noise levels (35–40 dB), both models converge to near-optimal performance. This pattern demonstrates that QAAR-SIREN provides systematic and statistically significant improvements in challenging noisy regimes, without compromising performance on cleaner signals.

In our second simulated signal, we used a more complex nonlinear waveform designed to represent better real-world nonstationary and noisy conditions, which is given as follows:

$$env_1(t) = 0.7 + 0.3\sin(2\pi \cdot 1.3t)$$

$$f_1(t) = 3 + 15t$$

$$y_1(t) = env_1(t)\sin(2\pi f_1(t)t + \pi \cdot u(t - 0.55))$$

$$env_2(t) = 0.4 + 0.2\cos(2\pi \cdot 0.9t)$$

$$\phi_2(t) = 2\pi\left(6t + 8t^2\right)$$

$$y_2(t) = env_2(t)\sin(\phi_2(t))$$

$$y_{lin}(t) = y_1(t) + y_2(t)$$

$$y_{nl}(t) = 0.18tanh\left((h \cdot y_{lin}(t))^2\right), \quad h = [0.2,\ 0.5,\ 0.3]$$

$$y_3(t) = 0.35\sin(2\pi f_3(t)t + 1.2$$

$$y_{\text{lin}}(t) = y_{\text{lin}}(t) + u(t - 0.67) \cdot y_3(t)$$

$$\sigma(t) = 0.15(1 + 0.8\,sin(2\pi \cdot 2.1t))$$

$$a_t = 0.6a_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$$

$$\eta(t) = \sigma(t)a_t$$

$$s(t) = \sum_{k=1}^{12} A_k \delta(t - t_k),\ \ A_k = \alpha_k \beta_k,\ \ \alpha_k \in \{+0.5, -0.5\},\ \ \beta_k \in [0.4, 0.8]$$

$$y(t) = y1(t) + y2(t) + u(t - 0.67)y3(t) + y_{nl}(t) + \eta(t) + s(t) \tag{23}$$

where, u(·) denotes the unit step function, which equals 1 when its argument is non-negative and 0 otherwise. $\Delta_{t,tk}$ represents the Kronecker delta, which equals 1 when $t = t_k$ and 0 otherwise. As seen in Eq. (23), the signal *y(t)* combines multiple components, including an amplitude- and frequency-modulated chirp with a sudden phase shift around the mid-point, a secondary quadratic-phase component with mild amplitude modulation, and a low-order nonlinear feedback term that introduces smooth distortions resembling Volterra-type dynamics. In the later portion of the signal, a regime change is applied, where both frequency and scale increase, simulating a transition to a higher-energy state. To make the task more demanding, we also introduced heteroskedastic autoregressive noise, which creates time-varying variance, together with a small number of sparse impulsive spikes to mimic outliers commonly observed in real sensor data.

Fig. 3 compares the proposed QAAR-SIREN with the original SIREN on the nonlinear and nonstationary signal defined in Equation (23). The QAAR-SIREN prediction remains closely aligned with the underlying waveform across both the training and test portions, capturing amplitude changes and phase transitions even in the noisier, high-frequency regions. This is especially evident near the end of the sequence, where the signal becomes more oscillatory and the heteroskedastic noise is more pronounced: in that part of the series, QAAR-SIREN still shows stable and adaptive behavior. By contrast, although the standard SIREN reproduces the overall waveform reasonably well, its fit becomes less precise in high-frequency segments and around the nonlinear transitions. The mismatch between prediction and ground truth is particularly noticeable toward the end of the sequence, which suggests that the conventional SIREN is less effective at handling abrupt regime changes and heteroskedastic fluctuations than the QAAR-SIREN variant. Table 3 provides a quantitative comparison between the two models in terms of MSE and R^2. The QAAR-SIREN achieved a lower MSE value of 0.6314 and a higher R^2 value of 0.6826 compared to the original SIREN, which obtained an MSE of 0.6972 and an R^2 of 0.6495.

3.2. Simulated healthcare signals

In addition to the synthetic sinusoidal and nonlinear waveforms, we tested the model on simulated physiological signals typically used in healthcare time-series analysis. These included respiratory rhythm, heart rate variability, blood pressure, and electrocardiogram-like patterns. These signals were designed to reflect the noise characteristics and temporal dynamics typically seen in biomedical data. By combining different types of signal sources, the experimental setup was intended to test the model's ability to handle varying degrees of temporal irregularity, generalize across physiological domains, and adapt to both abrupt and gradual changes in amplitude and frequency.

As a first benchmark, we generated a realistic arterial blood pressure (BP) trace for forecasting. The signal covers 24 s and was sampled at 250 Hz, producing a long sequence with physiologically plausible pulse morphology, slow baseline drift, and beat-to-beat variability. Input windows of length $win = 36$ samples were constructed, and the series was standardized using z-score normalization to make training more stable under amplitude variations. The final 500 samples were reserved exclusively for evaluation, so the model was required to extrapolate under regime shifts and noise conditions similar to those encountered in bedside monitoring.

Fig. 4 compares QAAR-SIREN and SIREN models on this simulated blood pressure signal. QAAR-SIREN follows the oscillatory pattern closely, reproducing the typical systolic and diastolic structure while also tracking gradual baseline changes and local outliers. In the predicted segment, the match with the original signal remains close in both timing and amplitude. In contrast, the original SIREN exhibits slight amplitude mismatches and phase delays, particularly during regions of high variability. The prediction tends to smooth out sharp transitions and fails to fully represent occasional spikes that are characteristic of physiological noise or motion artifacts.

Table 4 summarizes the quantitative results for both models using MSE and R^2 metrics. The QAAR-SIREN achieved an MSE of 0.0138 and an R^2 of 0.9859, outperforming the original SIREN, which obtained an MSE of 0.0213 and an R^2 of 0.9781.

We also generated a synthetic heart-rate-variability (HRV) signal to provide a realistic yet controllable benchmark for evaluating time-series prediction models. This signal was designed to reproduce the natural fluctuations in beat-to-beat intervals that arise from the interplay of sympathetic and parasympathetic nervous system activity. To make the simulation more realistic, the HRV signal was built to include both low- and high-frequency oscillations consistent with typical physiological rhythms, together with stochastic variability and occasional abrupt deviations resembling real cardiac behavior. Because the resulting series combined structured temporal patterns with random fluctuations, it provided a useful test case for assessing how well each model could capture nonlinear dynamics, tolerate noise, and maintain stable predictions under conditions closer to those found in physiological monitoring. Similar to the previous experiments, we formed input windows of length win equal to 36 samples and standardized the series with z-score normalization to stabilize training across amplitude changes. The last 200 samples were reserved strictly for evaluation, so the model must extrapolate across regime shifts and noise typical of bedside monitoring.

Fig. 5 illustrates the predicted and original heart-rate-variability signals obtained using the QAAR-SIREN and the original SIREN models. Both models successfully follow the main oscillatory trends of the simulated HRV waveform, including the irregular fluctuations and abrupt variations typical of physiological dynamics. The QAAR-SIREN prediction aligns more closely with the original

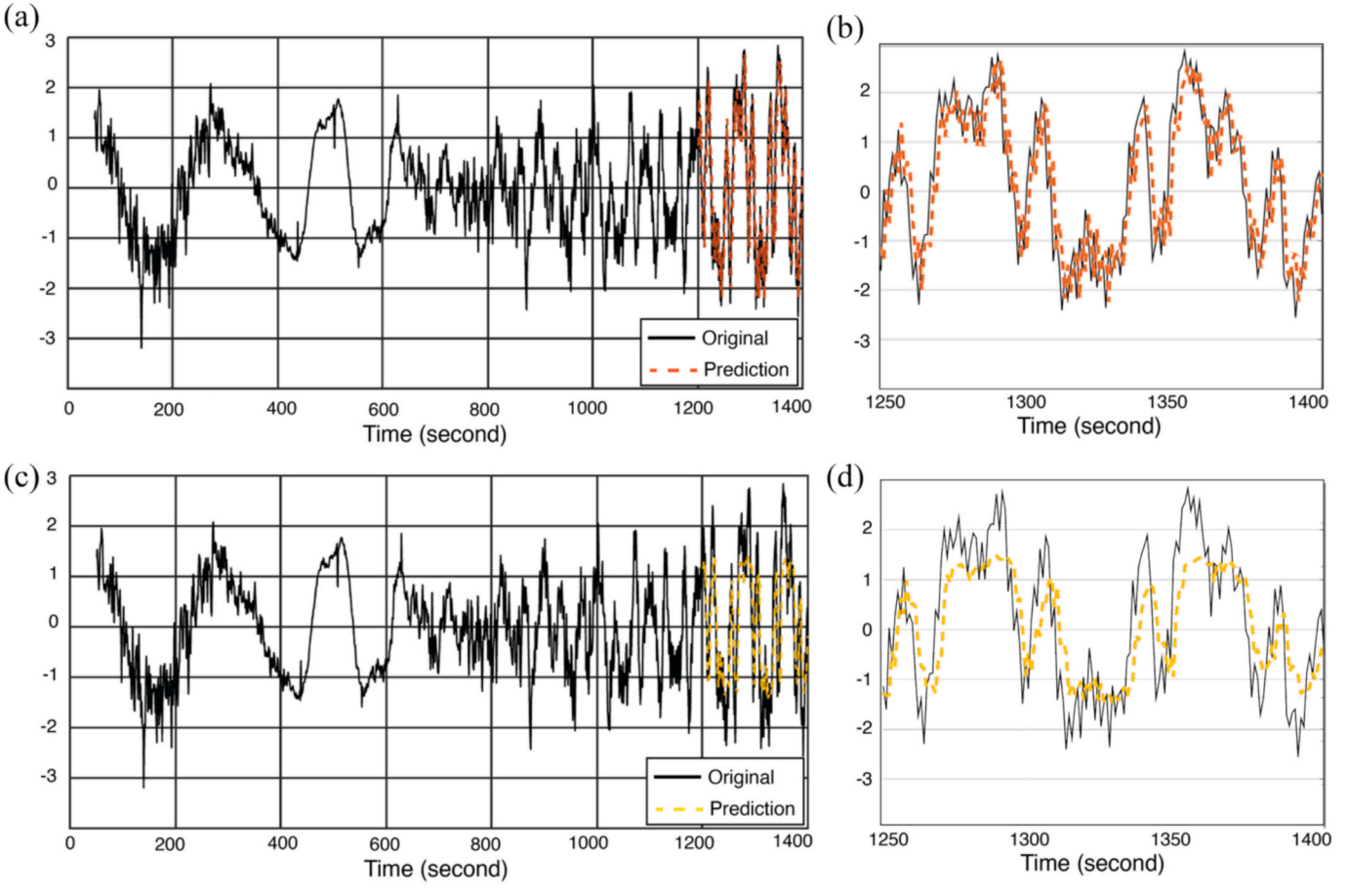

Fig. 3. Comparison of input and predicted simulated blood pressure signals using QAAR-SIREN and conventional SIREN. a) Whole signal with QAAR-SIREN prediction, b) Zoomed view of QAAR-SIREN prediction (red), c) Whole signal with SIREN prediction, d) Zoomed view of conventional SIREN prediction (yellow).

Table 3
A quantitative evaluation of the obtained predictions with both QAAR-SIREN and the original SIREN on simulated data.

Method	MSE	R^2
SIREN	0.6972	0.6495
QAAR-SIREN	0.6314	0.6826

signal during both low- and high-amplitude regions, maintaining consistency even in areas affected by noise and transient peaks.

Table 5 reports the numerical comparison between the two methods. QAAR-SIREN achieved an R^2 of 0.8674, very close to the 0.8690 obtained by the baseline SIREN, showing that both models explained most of the variance in the data. Although the numerical gap is small, QAAR-SIREN generated smoother and more consistent forecasts, with fewer abrupt changes from one prediction to the next, which suggests greater robustness to the irregular fluctuations typical of cardiac rhythm signals.

As a final validation step, we employed a synthetic breathing pattern to examine how well the QAAR-SIREN architecture generalizes across authentic physiological scenarios. This respiratory trace was designed to resemble natural breathing dynamics, including the typical asymmetry between inspiratory and expiratory phases, slow baseline drift, and occasional events such as deep sighs or brief pauses in breathing. Random variability was also introduced to reproduce breath-to-breath changes in timing and intensity, while additive white noise was used to mimic measurement artifacts from monitoring equipment. As a result, the simulated respiratory dataset provided a challenging benchmark: non-stationary, noisy, and intentionally corrupted, yet well suited for testing how each algorithm tracks gradual cyclic patterns, adapts to changing respiratory frequencies, and maintains forecasting accuracy under realistic measurement uncertainty.

Fig. 6 compares QAAR-SIREN and SIREN on the simulated respiratory rhythm signal. The QAAR-SIREN prediction stays very close to the true respiratory waveform, capturing both the amplitude modulation and the slowly varying oscillatory pattern of the breathing cycles. Even in portions of the sequence containing irregular events, such as deeper breaths or small amplitude fluctuations, the model remains smooth and accurate, showing good adaptability to nonlinear changes in both frequency and amplitude. The original SIREN also reconstructs the overall breathing rhythm reasonably well, but it shows small deviations in higher-energy segments and a slight delay around the peaks of the larger oscillations. This points to lower sensitivity to rapid transitions and amplitude irregularities than QAAR-SIREN. Table 6 confirms the same trend in quantitative terms: QAAR-SIREN reached an R^2 of 0.9987, compared with 0.9896 for the standard SIREN.

3.3. Real-world data experiments

We conducted experiments on real-world time series datasets, including the Jena Climate Dataset [31], electric power consumption datasets [32], and stock market price datasets [33]. The Jena Climate dataset was collected at the weather station of the Max Planck Institute for Biogeochemistry in Jena, Germany. It contains 420,551 observations and 15 meteorological variables, including temperature, air pressure, humidity, and wind speed, recorded every 10 min over several years. Because of its multivariate structure and long temporal span, it is widely used in time-series research, especially for climate analysis and forecasting.

The Electric Power Consumption dataset records hourly electricity usage across U.S. regions and includes more than 2.1 million observations, making it suitable for tasks such as demand forecasting and anomaly detection. For financial applications, the S&P 500 equity dataset provides daily records for constituent companies, including opening prices, intraday highs and lows, closing prices, dividend-adjusted values, and trading volume, and is often used for stock analysis and directional prediction. A broader financial dataset extends this setting by including additional indices and market variables, which makes it useful for studying global trends and portfolio allocation strategies. Another finance-related source is the MarketWatch sentiment dataset [34], which combines news-derived sentiment scores with trading data to support the analysis of sentiment-driven price movements and related investment strategies. Finally, the AirPassengers series contains monthly totals of international airline passengers from 1949 to 1960, for a total of 144 observations [35]. It exhibits a clear upward trend together with strong seasonality and is therefore often used as a benchmark for models dealing with multiplicative seasonal structure.

Table 7 shows that QAAR-SIREN consistently outperforms SIREN across all datasets, often by substantial margins. The most dramatic improvement occurs on Climate Temperature data, where R^2 increases from 0.9317 to 0.9995 and MSE decreases by 99.4%. Financial datasets show strong gains: S&P500 MSE drops by $\sim$ 70% (from 0.0430 to 0.0127) with R^2 rising from 0.9114 to 0.9738, while NIKKEI shows a 78% MSE reduction with R^2 increasing from 0.9544 to 0.9898. NASDAQ100 exhibits the largest relative improvement with 97% MSE reduction and R^2 jumping from 0.8755 to 0.9964. Electricity datasets show more modest but consistent gains (4% for PJM, 15% for AEP), while AirPassengers demonstrates an 82% MSE reduction with R^2 rising from 0.8083 to 0.9654.

3.4. Ablation studies

In this study, we conducted an ablation analysis to quantify the contribution of each architectural element of the proposed QAAR-SIREN architecture. We evaluated six configurations derived from the full model by selectively disabling the variational quantum feature extractor, removing attention, switching the attention pooling from linear to a channel-wise softmax, and replacing residual forecasting with direct prediction, while keeping the SIREN backbone and all optimization settings fixed. Experiments used a nonstationary synthetic benchmark with regime changes and the AirPassengers dataset. The network consisted of two hidden layers of

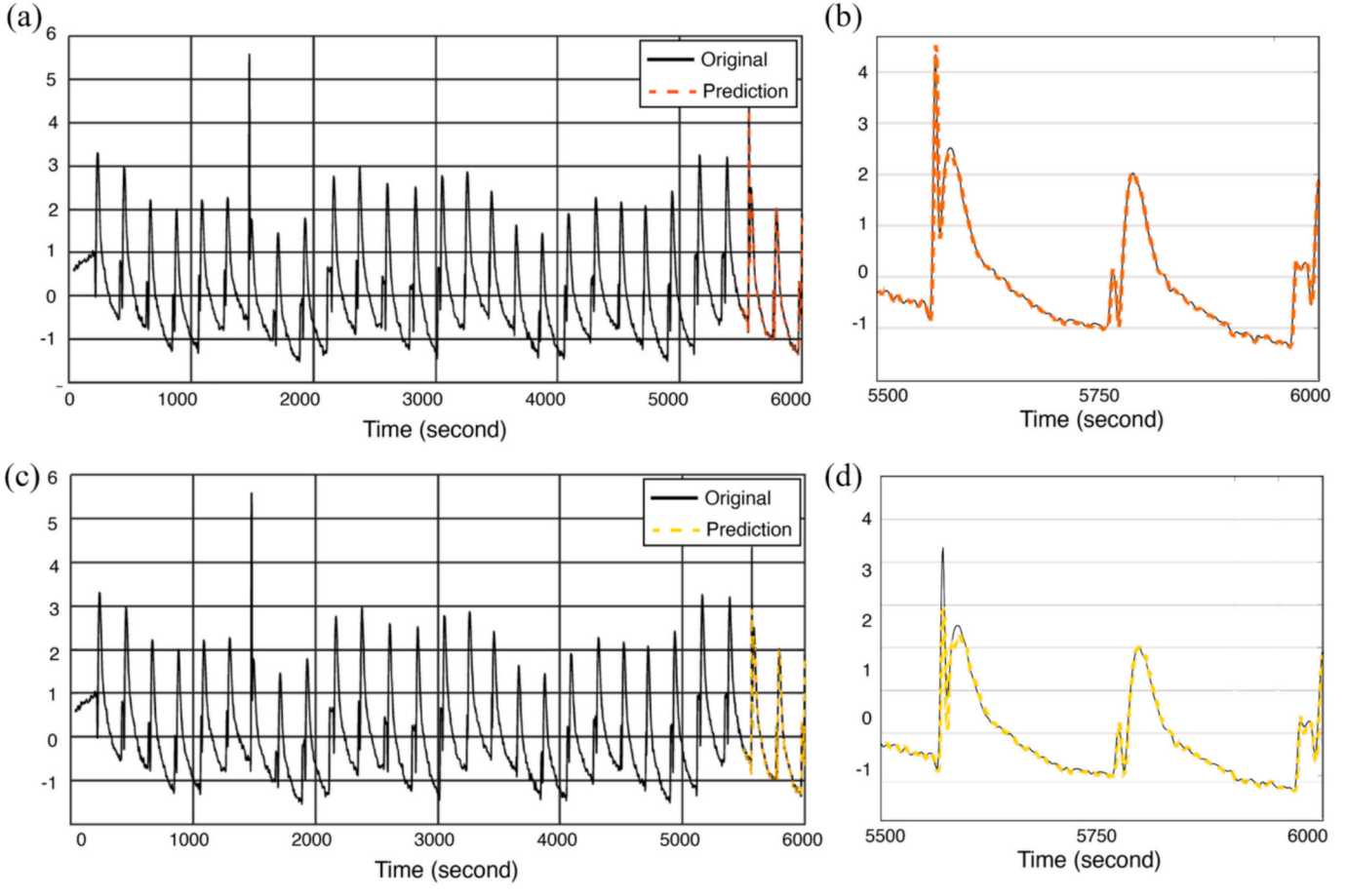

Fig. 4. Comparison of input and predicted simulated blood pressure signals using QAAR-SIREN and conventional SIREN. a) Whole signal with QAAR-SIREN prediction, b) Zoomed view of QAAR-SIREN prediction (red), c) Whole signal with SIREN prediction, d) Zoomed view of conventional SIREN prediction (yellow).

Table 4
A quantitative evaluation of the obtained predictions on the simulated blood pressure signal with both QAAR-SIREN and the original SIREN.

Method	MSE	R^2
SIREN	0.0213	0.9781
QAAR-SIREN	0.0138	0.9859

width 32 with sinusoidal activations and $w_0 = 8$, an attention branch of dimension 16, and a single quantum layer that produced a 4-dimensional feature vector through a two-qubit circuit with R_X/R_Z rotations and a CNOT entangler. The quantum parameters were trained using the parameter-shift rule.

All models were optimized with Adam $(l, r, =, 5 \times 10^{-4} \beta_1 = 0.9 \beta_2 = 0.999 \varepsilon = 10^{-8})$, using a mini-batch size of 16 and early stopping with a patience of 15 epochs on a 10% validation tail taken from the training set. Random seeds were fixed so that data splits and parameter initialization remained identical across the ablation study. For the residual variants, the final prediction was obtained by adding the estimated residual to the last observed value, avoiding any unfair disadvantage for models trained in residual form. In the softmax attention branch, back-propagation used the exact Jacobian rather than an approximation, in order to preserve gradient consistency.

Performance was evaluated on the held-out test sets using mean squared error and the coefficient of determination. This setup makes it easier to attribute any performance differences to the contribution of quantum features, attention pooling, and residual learning, rather than to changes in preprocessing or training conditions.

Table 8 shows that residual learning is the decisive component. Removing residual connections increases the error by over 20 × and drops R^2 from 96.5% to 23.4%, reflecting the difficulty of direct forecasting on a seasonal and trending series. Quantum features provide a clear and largely independent gain. When the quantum branch is removed, the error more than doubles relative to the full model and R^2 drops by about five points, which suggests that the VQC is contributing information the classical inputs alone do not capture. Attention also improves performance, though to a lesser extent: removing it increases the error by roughly half and lowers R^2 by nearly two points, indicating that this branch helps the model make better use of context within the input window. When both the quantum and attention components are removed, the non-residual model performs worst, which supports the idea that the two branches provide complementary benefits. By contrast, softmax attention pooling performs almost the same as linear pooling, suggesting that either choice is adequate once residual learning and quantum features are included.

These results become easier to interpret when looking at the role of each component in the forecasting process. Without the variational quantum feature extractor, the model loses a set of expectation-based nonlinear features tied to the most recent observation. This weakens its ability to capture abrupt changes, regime transitions, and other non-smooth temporal behavior that often appears in nonstationary series. Removing the attention mechanism affects a different part of the model. In that case, the network has less support in summarizing the input window and must depend more directly on raw lagged values, which makes it harder to highlight the most informative temporal patterns or exploit longer-range dependencies effectively.

The most severe performance degradation is observed when residual learning is removed. In this case, the model must directly predict absolute values rather than local increments, which significantly destabilizes training on seasonal and trending series and leads to a dramatic increase in error. Finally, the combined removal of both the quantum and attention components results in the weakest performance among all non-residual variants, confirming that these modules provide complementary information that jointly enhances predictive accuracy and robustness.

3.5. Comparison with state-of-the-art methods

To provide a fair evaluation, we compare QAAR-SIREN with state-of-the-art forecasting models that are commonly used or reported as competitive baselines for each dataset. Since the considered datasets originate from different application domains and exhibit distinct temporal characteristics, the selected baselines vary across datasets, following standard practice in the time series forecasting literature. This approach ensures that each comparison reflects methods that are well suited to the corresponding data regime. In all cases, performance is evaluated using the coefficient of determination (R^2) and the root mean squared error (RMSE). Table 9 reports the quantitative comparison between QAAR-SIREN and the selected baselines across six datasets.

As shown in Table 9, QAAR-SIREN demonstrates strong and consistent performance across all datasets. On the Climate dataset, it achieves an R^2 of 0.9995 and an RMSE of 0.0173, outperforming UA-LNN and all other baselines by a clear margin. On Electricity PJM, QAAR-SIREN reaches an R^2 of 0.9716 with an RMSE of 0.0728, although the lowest RMSE on that dataset is obtained by the ResNet-Stacked LSTM model. On Electricity AEP, QAAR-SIREN achieves an R^2 of 0.9837 and an RMSE of 0.1122, clearly outperforming UA-LNN and RCNN-ESN. The same pattern is largely observed in the financial datasets, where QAAR-SIREN remains competitive and often performs best. On the S&P 500 dataset, it records an R^2 of 0.9738 and an RMSE of 0.1127; while UA-LNN attains a lower absolute error, QAAR-SIREN still maintains strong predictive accuracy. On the Nikkei dataset, it reaches an R^2 of 0.9898 with an RMSE of 0.0469, matching the lowest reported error while improving overall fit. On Nasdaq 100, it delivers the best overall result, with an R^2 of 0.9964 and an RMSE of 0.0387. Taken together, these results show that QAAR-SIREN provides the most consistently strong R^2 values across datasets and achieves the best, or tied-best, RMSE on Climate, Electricity AEP, Nikkei, and Nasdaq 100, while remaining competitive on Electricity PJM and S&P 500.

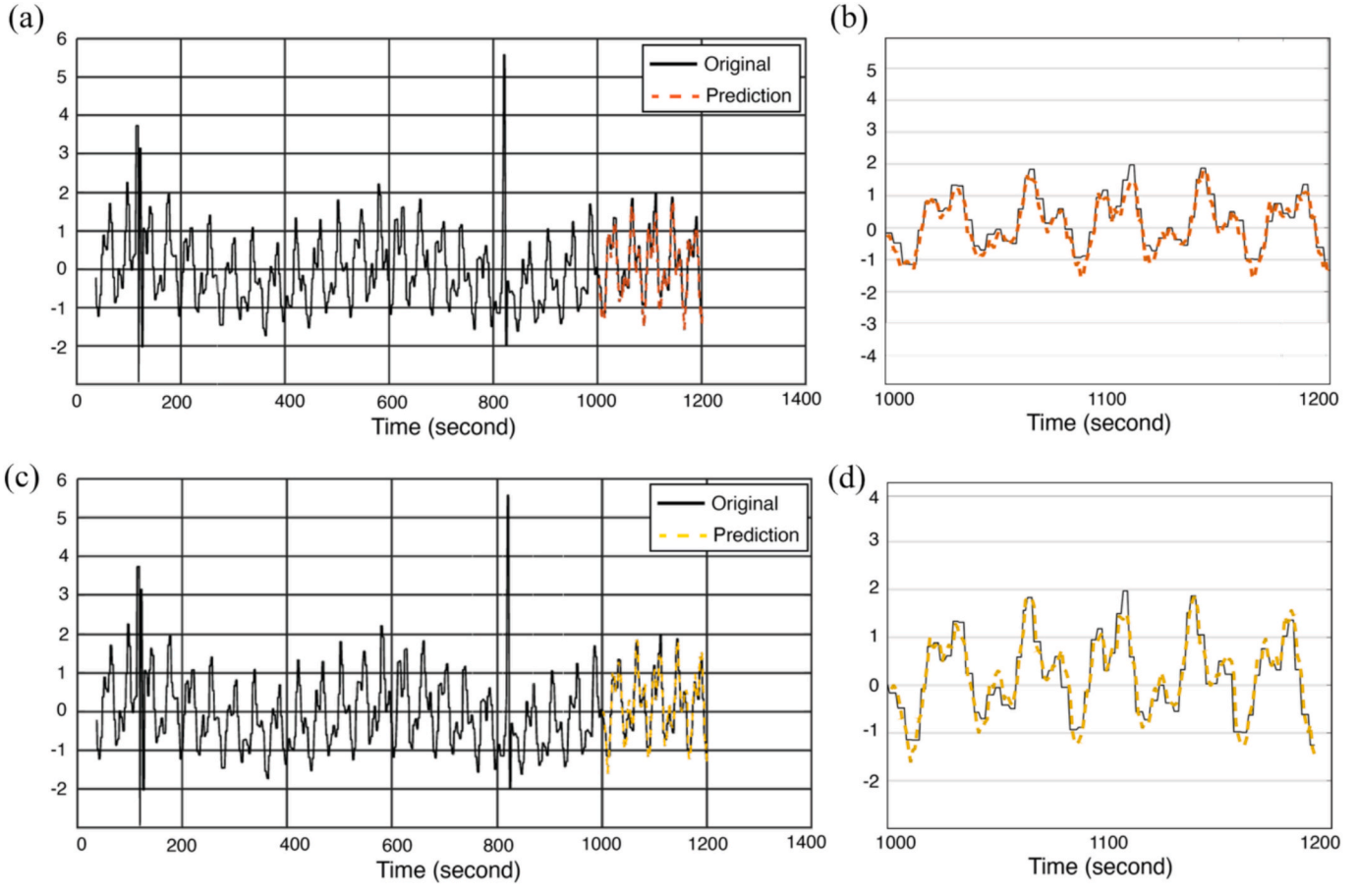

Fig. 5. Comparison of input and predicted simulated heart-rate-variability signals using QAAR-SIREN and conventional SIREN. a) Whole signal with QAAR-SIREN prediction, b) Zoomed view of QAAR-SIREN prediction (red), c) Whole signal with SIREN prediction, d) Zoomed view of conventional SIREN prediction (yellow).

Table 5
A quantitative evaluation of the obtained predictions on the simulated heart-rate-variability signal with both QAAR-SIREN and the original SIREN.

Method	MSE	R^2
SIREN	0.0856	0.8690
QAAR-SIREN	0.0866	0.8674

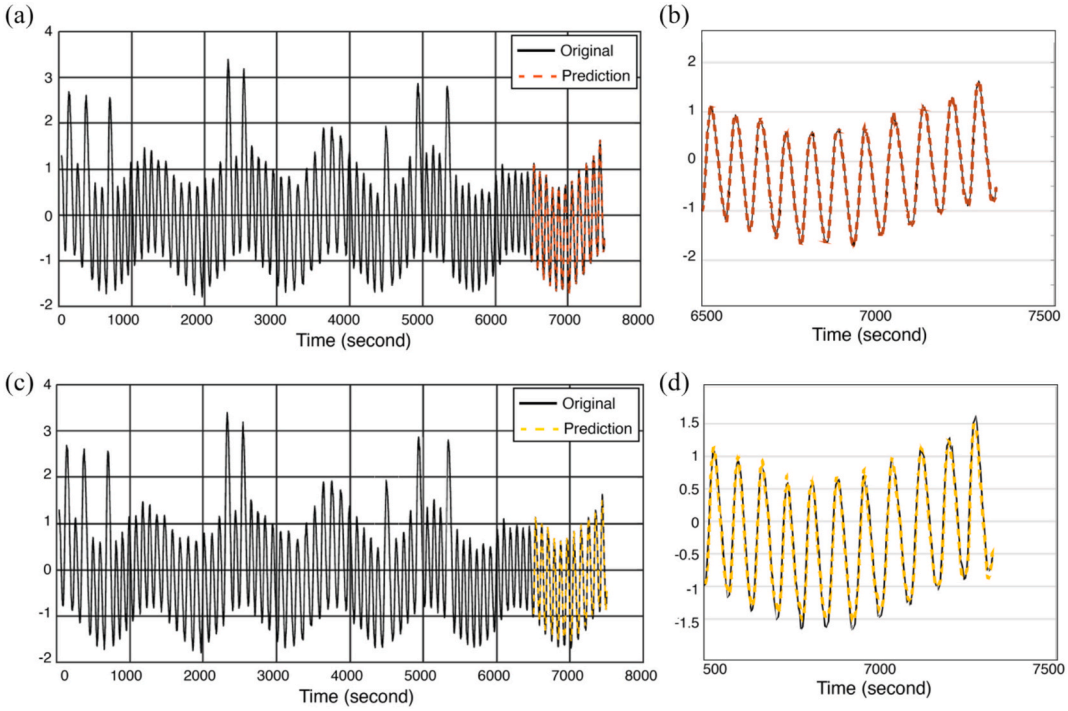

Fig. 6. Comparison of input and predicted simulated respiratory rhythm signals using QAAR-SIREN and conventional SIREN. a) Whole signal with QAAR-SIREN prediction, b) Zoomed view of QAAR-SIREN prediction (red), c) Whole signal with SIREN prediction, d) Zoomed view of conventional SIREN prediction (yellow).

Table 6
A quantitative evaluation of the obtained predictions on the simulated respiratory rhythm signal with both QAAR-SIREN and the original SIREN.

Method	MSE	R^2
SIREN	0.0073	0.9896
QAAR-SIREN	0.0009	0.9987

Alongside the classical baselines, we also included the single-qubit quantum circuit time-series prediction (QC-TSP) method introduced in [23], as a representative quantum forecasting approach that has been used widely in the literature. To make the comparison as direct as possible, QC-TSP was evaluated under the same dataset settings across all tasks. This makes it possible to compare two forecasting strategies that both involve quantum circuits but use them in very different ways within the prediction pipeline. As reported in Table 9, QC-TSP delivers competitive results on several datasets, especially when compared with some classical baselines. Even so, QAAR-SIREN consistently performs better, particularly on the Climate and Electricity datasets, where it achieves both higher R^2 and lower RMSE. Overall, this suggests that incorporating learnable quantum feature extraction directly into the forecasting architecture is more effective than relying only on quantum-generated data.

4. Discussions

In this paper, we present a compact forecasting methodology that couples a linear attention pool with a fixed-length VQC to produce four complementary features, which are then fed into a two-layer sinusoidal representation network trained on residual targets. The attention block forms a low-dimensional summary of recent dynamics; the VQC encodes the most recent sample via RX/RZ

Table 7
A quantitative evaluation of the obtained predictions on the real-world time series prediction datasets with both QAAR-SIREN and the original SIREN.

Dataset	Model	R^2	MSE
Climate (Temperature) [31]	SIREN	0.9317	0.0493
	QAAR-SIREN	0.9995	0.0003
Electricity (PJM) [32]	SIREN	0.9707	0.0055
	QAAR-SIREN	0.9716	0.0053
Electricity (AEP) [32]	SIREN	0.9807	0.0149
	QAAR-SIREN	0.9837	0.0126
S&P500 [33]	SIREN	0.9114	0.0430
	QAAR-SIREN	0.9738	0.0127
NIKKEI [24]	SIREN	0.9544	0.0099
	QAAR-SIREN	0.9898	0.0022
NASDAQ100 [34]	SIREN	0.8755	0.0534
	QAAR-SIREN	0.9964	0.0015
Air Passengers [35]	SIREN	0.8083	0.0852
	QAAR-SIREN	0.9654	0.0153

Table 8
The obtained MSE and R^2 scores for ablation studies.

Experiment	MSE	R^2
Full: QAAR-SIREN	0.01533	0.9654
No Quantum	0.03605	0.9189
No Attention	0.02341	0.9473
No Residual	0.34088	0.2339
No Quantum and No Attention	0.04321	0.9029
Softmax Attention	0.01548	0.9652

Table 9
Performance comparison of various models, including QAAR-SIREN, on time series forecasting datasets. Best performances are highlighted in bold.

Dataset	Model	R^2	RMSE
Climate (Temperature) [31]	CNN-RNN [36]	0.9870	0.1168
	LSTM [36]	0.9640	0.1168
	ARIMA [36]	0.8310	0.2092
	SVM [36]	0.8320	0.1674
	MLP [36]	0.8420	0.2348
	UA-LNN [22]	0.9955	0.0522
	QC-TSP [23]	0.9768	0.1003
	QAAR-SIREN (Proposed)	**0.9995**	**0.0173**
Electricity (PJM) [32]	ResNet-Stacked LSTM [37]	–	0.0380
	UA-LNN [22]	0.9685	0.2000
	QC-TSP [23]	0.9633	0.0770
	QAAR-SIREN (Proposed)	**0.9716**	**0.0728**
Electricity (AEP) [32]	RCNN-ESN [38]	–	0.2670
	UA-LNN [22]	0.9658	0.1937
	QC-TSP [23]	0.8305	0.2571
	QAAR-SIREN (Proposed)	**0.9837**	**0.1122**
S&P500 [33]	ICE2DE-MDL [39]	0.9900	0.1170
	UA-LNN [22]	0.9934	0.0460
	QC-TSP [23]	0.9439	0.0930
	QAAR-SIREN (Proposed)	0.9738	0.1127
NIKKEI [24]	ICE2DE-MDL [39]	0.9190	0.1640
	UA-LNN [22]	0.9719	0.0469
	QC-TSP [23]	0.9246	0.0541
	QAAR-SIREN (Proposed)	**0.9898**	**0.0469**
NASDAQ100 [34]	ICE2DE-MDL [39]	0.9660	0.2440
	UA-LNN [22]	0.9907	0.0455
	QC-TSP [23]	0.9617	0.0570
	QAAR-SIREN (Proposed)	**0.9964**	**0.0387**

layers and an entangling gate to yield expectation features; and the SIREN head models smooth yet expressive nonlinearities. This design is intentionally lightweight yet resilient to non-stationarity, regime shifts, heteroskedastic noise, and sparse spikes, delivering stable one-step forecasts without heavy preprocessing or elaborate architectures.

The experimental results suggest that the gains obtained by QAAR-SIREN are not simply a consequence of higher model complexity, but are more closely tied to the way temporal information is represented and learned. Across heterogeneous datasets, predicting temporal increments rather than absolute values leads to more stable optimization and better generalization, especially when the data contain trends, level shifts, or changes in scale. In this sense, the residual formulation makes the learning problem easier by directing the model toward local temporal changes, which helps reduce the instability often seen in direct forecasting.

Another point that emerges from the results is that the benefits of QAAR-SIREN become more evident as the temporal dynamics move further away from stationarity. On smoother or more regular signals, the proposed model tends to perform similarly to strong classical baselines. When regime changes, heteroskedastic noise, or abrupt transitions are present, however, it shows better stability and stronger generalization. This pattern reflects the complementary contribution of residual learning and attention-based contextual aggregation: the former stabilizes the prediction target, while the latter helps the model focus on the most informative parts of the input window without depending on recurrent state propagation.

These results also clarify the practical role of quantum components in hybrid forecasting architectures. Rather than functioning as standalone predictors or data generators, shallow variational quantum circuits appear most effective when used as compact, learnable feature extractors that enrich classical representations. The observed gains are achieved with minimal circuit depth and a fixed number of expectation values, suggesting that quantum augmentation can provide tangible benefits without requiring deep or hardware-intensive quantum models. At the same time, the modest improvements observed on more stable datasets indicate that quantum features are not universally advantageous and contribute most when localized nonlinear or rapidly evolving temporal patterns are present.

The advantages of the proposed methodology are:

1. Three-stream fusion enriches the input with complementary evidence from raw lags, a compact attention summary, and four features from a two-qubit quantum circuit. This improves representation without requiring a large model.
2. Residual learning with a frequency calibrated SIREN provides stable training and a strong fit on fast variations, capturing fine temporal detail with only two hidden layers and sinusoidal activations.
3. End-to-end optimization updates quantum circuit angles with the parameter shift rule in the same loop as network weights, enabling joint learning of quantum and classical parts with consistent gains across synthetic and real datasets.

And the limitations are:

1. The quantum block encodes only the most recent sample with a shallow circuit, which can limit expressivity for long-range or multivariate patterns unless circuit depth or temporal coverage is increased.
2. Performance is sensitive to design choices such as window length, attention dimension, and the SIREN frequency scale. Poor tuning can reduce stability or over-smooth high-frequency structure.
3. The evaluation focuses on one-step forecasting with mean squared error and R^2, so behavior on multi-step horizons, covariate shift, and uncertainty calibration remains less characterized.

5. Conclusions

This study presented a compact forecasting method that combines raw temporal lags, a lightweight attention summary, and quantum-derived features within a sinusoidal representation network. The model learns residuals through end-to-end training that includes exact parameter-shift updates for the quantum circuit, then adds the predicted residual to the last observed value for final forecasts. Experimental results on simulated and real data demonstrate that the proposed QAAR-SIREN model improves robustness and predictive accuracy under nonstationary and noisy conditions while maintaining a lightweight architecture. Future work will extend the model to multi-step forecasting and multivariate series with external covariates. Richer yet shallow quantum encoders, uncertainty-aware prediction, and interpretability tools will be explored to further enhance the applicability of hybrid quantum–-classical forecasting models.

CRediT authorship contribution statement

Abdulkadir Sengur: Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization. **Massimo Salvi:** Writing – original draft, Visualization, Validation. **Prabal Datta Barua:** Writing – review & editing, Validation. **Ravinesh Deo:** Writing – review & editing. **Yan Li:** Writing – review & editing. **U.R. Acharya:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] R. Zhou, Q. Feng, H. Wen, Z. Deng, W. Ai, G.G. Yen, GLPT: global-local pyramid transformer for multivariate time series forecasting, Inf. Sci. 753 (2026) 123599, https://doi.org/10.1016/j.ins.2026.123599.
[2] Z. Han, J. Zhao, H. Leung, K.F. Ma, W. Wang, A review of deep learning models for time series prediction, IEEE Sens. J. 21 (2021) 7833–7848, https://doi.org/10.1109/JSEN.2019.2923982.
[3] M.A. Morid, O.R.L. Sheng, J. Dunbar, Time series prediction using deep learning methods in healthcare, ACM Trans. Manag. Inf. Syst. 14 (2023) 1–29, https://doi.org/10.1145/3531326.
[4] R.J. Frank, N. Davey, S.P. Hunt, Time series prediction and neural networks, J. Intell. Robot. Syst. 31 (2001) 91–103, https://doi.org/10.1023/A:1012074215150.
[5] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. James Kong, S.T.S. Bukkapatnam, Time series forecasting for nonlinear and non-stationary processes: a review and comparative study, IIE Trans. 47 (2015) 1053–1071, https://doi.org/10.1080/0740817X.2014.999180.
[6] S.V. Kumar, K.C. Dogiparthi, L. Vanajakshi, S.C. Subramanian, Integration of exponential smoothing with state space formulation for bus travel time and arrival time prediction, Transport 32 (2015) 358–367, https://doi.org/10.3846/16484142.2015.1100676.
[7] F. Oueslati, A. Ben Abbes, V. Barra, Fourier-optimal loss for distortion and time in non-stationary time series forecasting, Inf. Sci. 749 (2026) 123535, https://doi.org/10.1016/j.ins.2026.123535.
[8] R. Hyndman, A. Koehler, K. Ord, R. Snyder, Forecasting with exponential smoothing, Springer Berlin Heidelberg, Berlin Heidelberg, 2008, 10.1007/978-3-540-71918-2.
[9] D. Salinas, V. Flunkert, J. Gasthaus, T. Januschowski, DeepAR: probabilistic forecasting with autoregressive recurrent networks, Int. J. Forecast. 36 (2020) 1181–1191, https://doi.org/10.1016/j.ijforecast.2019.07.001.
[10] B.N. Oreshkin, D. Carpov, N. Chapados, Y. Bengio, N-BEATS: Neural basis expansion analysis for interpretable time series forecasting, (2019). https://doi.org/10.48550/ARXIV.1905.10437.
[11] B. Lim, S.Ö. Arık, N. Loeff, T. Pfister, Temporal Fusion Transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (2021) 1748–1764, https://doi.org/10.1016/j.ijforecast.2021.03.012.
[12] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, (2020). https://doi.org/10.48550/ARXIV.2012.07436.
[13] Y. Nie, N.H. Nguyen, P. Sinthong, J. Kalagnanam, A time series is worth 64 words, Long-Term Forecasting with Transformers (2022), https://doi.org/10.48550/ARXIV.2211.14730.
[14] S. Makridakis, E. Spiliotis, V. Assimakopoulos, The M4 competition: 100,000 time series and 61 forecasting methods, Int. J. Forecast. 36 (2020) 54–74, https://doi.org/10.1016/j.ijforecast.2019.04.014.
[15] V. Sitzmann, J.N.P. Martel, A.W. Bergman, D.B. Lindell, G. Wetzstein, Implicit Neural Representations with Periodic Activation Functions, (2020). https://doi.org/10.48550/ARXIV.2006.09661.
[16] J.C.B. Gamboa, Deep Learning for Time-Series Analysis, (2017). https://doi.org/10.48550/ARXIV.1701.01887.
[17] S. Mehtab, J. Sen, A time series analysis-based stock price prediction using machine learning and deep learning models, Int. J. Bus. Forecast. Mark. Intell. 6 (2020) 272, https://doi.org/10.1504/IJBFMI.2020.115691.
[18] I.R. Parray, S.S. Khurana, M. Kumar, A.A. Altalbe, Time series data analysis of stock price movement using machine learning techniques, Soft. Comput. 24 (2020) 16509–16517, https://doi.org/10.1007/s00500-020-04957-x.
[19] A.Y. Barrera-Animas, L.O. Oyedele, M. Bilal, T.D. Akinosho, J.M.D. Delgado, L.A. Akanbi, Rainfall prediction: a comparative analysis of modern machine learning algorithms for time-series forecasting, Mach. Learn. Appl. 7 (2022) 100204, https://doi.org/10.1016/j.mlwa.2021.100204.
[20] G. Bontempi, S. Ben Taieb, Y.-A. Le Borgne, Machine Learning Strategies for Time Series Forecasting, in: M.-A. Aufaure, E. Zimányi (Eds.), Bus. Intell., Springer Berlin Heidelberg, Berlin, Heidelberg, 2013: pp. 62–77. https://doi.org/10.1007/978-3-642-36318-4_3.
[21] A. Zeroual, F. Harrou, A. Dairi, Y. Sun, Deep learning methods for forecasting COVID-19 time-series data: a comparative study, Chaos Solitons Fractals 140 (2020) 110121, https://doi.org/10.1016/j.chaos.2020.110121.
[22] M.H. Akpinar, O. Atila, A. Sengur, M. Salvi, U.R. Acharya, A novel uncertainty-aware liquid neural network for noise-resilient time series forecasting and classification, Chaos Solitons Fractals 193 (2025) 116130, https://doi.org/10.1016/j.chaos.2025.116130.
[23] M. Saghafi, L. Mili, Predictive time-series analysis of single-qubit quantum circuit outcomes for a superconducting quantum computer: forecasting error patterns, IEEE Access 13 (2025) 40115–40132, https://doi.org/10.1109/ACCESS.2025.3546259.
[24] M.A. Rivera-Ruiz, A. Mendez-Vazquez, J.M. López-Romero, Time Series Forecasting with Quantum Machine Learning Architectures, in: O. Pichardo Lagunas, J. Martínez-Miranda, B. Martínez Seis (Eds.), Adv. Comput. Intell., Springer Nature Switzerland, Cham, 2022: pp. 66–82. https://doi.org/10.1007/978-3-031-19493-1_6.
[25] D. Alaminos, M.B. Salas, M.A. Fernández-Gámez, Quantum computing and deep learning methods for GDP growth forecasting, Comput. Econ. 59 (2022) 803–829, https://doi.org/10.1007/s10614-021-10110-z.
[26] M.R. Habibi, S. Golestan, Y. Wu, J.M. Guerrero, J.C. Vasquez, Electrical load forecasting in power systems based on quantum computing using time series-based quantum artificial intelligence, Sci. Rep. 15 (2025) 7429, https://doi.org/10.1038/s41598-025-89933-x.
[27] M. Schuld, I. Sinayskiy, F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. 56 (2015) 172–185, https://doi.org/10.1080/00107514.2014.964942.
[28] Z. Zhang, M.R. Sabuncu, Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels, (2018). https://doi.org/10.48550/ARXIV.1805.07836.
[29] K. Wagner, D. Psaltis, Multilayer optical learning networks, Appl. Opt. 26 (1987) 5061, https://doi.org/10.1364/AO.26.005061.
[30] M. Benedetti, E. Lloyd, S. Sack, M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. 4 (2019) 043001, https://doi.org/10.1088/2058-9565/ab4eb5.
[31] Jena Climate Dataset, (https://www.kaggle.com/datasets/mnassrib/jena-climate).
[32] Hourly Energy Consumption, (https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption).
[33] S&P500 Stock Prices, (https://www.kaggle.com/datasets/rprkh15/sp500-stock-prices).
[34] MarketWatch News Links and Prices for NASDAQ100, (https://www.kaggle.com/datasets/hserdaraltan/marketwatch-news-sentiment-scores-for-nasdaq).
[35] Air Passengers- Time Series- ARIMA, (https://www.kaggle.com/code/sunaysawant/air-passengers-time-series-arima).
[36] A. Utku, U. Can, An efficient hybrid weather prediction model based on deep learning, Int. J. Environ. Sci. Technol. 20 (2023) 11107–11120, https://doi.org/10.1007/s13762-023-05092-4.
[37] Z.A. Khan, A. Ullah, I. Ul Haq, M. Hamdy, G. Maria Mauro, K. Muhammad, M. Hijji, S.W. Baik, Efficient short-term electricity load forecasting for effective energy management, Sustain. Energy Technol. Assess. 53 (2022) 102337, https://doi.org/10.1016/j.seta.2022.102337.
[38] M.D. Alanazi, A. Saeed, M. Islam, S. Habib, H.I. Sherazi, S. Khan, M.M. Shees, Enhancing short-term electrical load forecasting for sustainable energy management in low-carbon buildings, Sustainability 15 (2023) 16885, https://doi.org/10.3390/su152416885.
[39] Z.D. Akşehir, E. Kılıç, Multi level perspectives in stock price forecasting: ICE2DE-MDL, PeerJ Comput. Sci. 10 (2024) e2125.