energies

MDPI

Article

Multiscale Quantum-Enhanced Forecasting Framework for Photovoltaic Power Output and Electric Load

Cheng Zhang [1,†], Jingren Zhou [1,†], Zhenhao Weng [1], Ruichong Li [1], Longze Wang [1,*], Zhehan Li [1] and Meicheng Li [1,2,*]

[1] State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China; 120232111006@ncepu.edu.cn (C.Z.); jrzhou0813@163.com (J.Z.)
[2] Tongji University, Shanghai 200092, China
* Correspondence: lzwang@ncepu.edu.cn (L.W.); mcli@ncepu.edu.cn (M.L.)
† These authors contributed equally to this work.

Abstract

Accurate day-ahead forecasting of photovoltaic (PV) power and electric load is essential for reliable power system operation under growing renewable penetration and demand complexity. Both series, however, are strongly nonstationary, with multiscale dynamics driven by external factors. This study develops a multiscale quantum-enhanced forecasting framework integrating rolling variational mode decomposition (VMD), feature attention (FA), and bidirectional quantum temporal modeling. Rolling VMD separates slower trends from rapid fluctuations, FA weights weather and calendar inputs, and the bidirectional quantum module encodes nonlinear dependencies within the observed history to represent irradiance-driven PV variability and temperature- and calendar-related load dynamics. Historical context is combined with the corresponding previous-day target for direct day-ahead forecasting. Using the Desert Knowledge Australia Solar Centre (DKASC) photovoltaic dataset and the GEFCom2014 load dataset, the seasonal reference comparisons cover four seven-day intervals per task. Across the 672 hourly evaluation points per task, the five-run mean RMSE and R^2 are 2.0299 kW and 0.9618 for PV and 7.5385 MW and 0.9633 for load. PV evaluation includes nighttime zeroing based on observed irradiance. Ablation results confirm the complementary contributions of the three modules, supporting a unified approach to supply- and demand-side day-ahead forecasting.

Keywords: photovoltaic power forecasting; electric load forecasting; quantum machine learning; multiscale representation; feature attention

Check for updates

Academic Editor: Francesco Calise

Received: 29 September 2026
Revised: 29 October 2026
Accepted: 2 November 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Power forecasting is fundamental to the secure and reliable operation of power systems, optimized dispatch, and the integration of renewable energy. Under the combined pressures of global warming and fossil-energy scarcity, the transition toward clean, low-carbon energy has become a broad global priority. Renewable generation technologies, particularly photovoltaic (PV) and wind power, have expanded rapidly in recent years [1]. However, increasing penetration of renewable generation introduces greater uncertainty into supply–demand balancing and system operation. On the supply side, PV output depends strongly on solar irradiance, ambient temperature, cloud cover, and abrupt changes in meteorological conditions, which makes it highly volatile, stochastic, and intermittent [2]. On the demand side, the growing use of electric vehicles, distributed storage, and integrated

energy systems makes load composition increasingly complex and strengthens its nonlinear, stochastic, and abrupt characteristics [3]. Accurate renewable-output forecasting facilitates reliable power system dispatch, reduces the risk of triggering curtailment schemes, and lowers operating costs through more effective reserve capacity allocation [4]. Accurate load forecasts are likewise essential for generation planning, electricity-market transactions, and demand-response optimization [5]. Therefore, improving forecasting accuracy and robustness under stochastic fluctuations on supply and demand sides remains an important research problem [6].

Traditional generation-forecasting methods are largely based on linear statistical models, including autoregressive integrated moving-average (ARIMA) [7] and autoregressive moving-average (ARMA) models [8]. These approaches can identify periodic components in stationary series, but their forecasting performance often deteriorates when applied to highly nonlinear and rapidly varying power-output trajectories. With advances in artificial intelligence, deep learning has become increasingly adopted for power forecasting, and recurrent neural networks (RNNs) [9], long short-term memory (LSTM) networks [10], and gated recurrent units (GRUs) [11] are widely used for temporal regression. Their gating structures capture both short- and long-term dependencies in historical output and provide strong fitting ability under conventional weather conditions.

Existing load-forecasting methods mainly emphasize multidimensional feature extraction, temporal dependency modeling, and hybrid forecasting architectures. One common approach combines convolutional neural networks (CNNs) [12] with temporal convolutional networks (TCNs) [13]. One-dimensional convolution extracts local associations from multivariate sequences, while dilated convolution enlarges the effective temporal receptive field and enhances the modeling of long-range dependencies. Another mainstream approach introduces conventional attention mechanisms [14] or variants such as feature attention (FA) [15], which learn feature-specific relevance weights with respect to the forecasting target and adaptively adjust the contribution of individual input variables. Hybrid forecasting has also received increasing attention for peak loads and complex operating conditions. For example, multistage frameworks combining similar-day selection, feature screening, and artificial neural networks can more effectively capture nonlinear variations in electric load [16].

Despite these advances, existing methods remain limited in the ability to represent multiscale fluctuations, abrupt local changes, and nonlinear relationships in highly nonstationary and noisy power and load series [17]. A central challenge is to suppress nonstationarity without discarding informative variations across temporal scales. To address this issue, studies [18–20] have introduced empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and variational mode decomposition (VMD) to separate trends, periodic patterns, and fluctuations across different frequency bands before forecasting. For feature utilization, an FA mechanism has been incorporated into LSTM to adaptively weight inputs and identify influential factors [21]. For temporal modeling, bidirectional recurrent networks have been applied to wind and PV forecasting to combine forward and backward contextual information within the historical window [22,23]. Although progress has been made in multiscale extraction, external-variable screening, and temporal representation, these techniques are often developed independently. Their combined use and unified applicability to both supply- and demand-side forecasting therefore require further investigation.

Quantum computing offers a new perspective on the feature-mapping limitations of classical deep learning. Through superposition and entanglement, quantum models can implement complex nonlinear mappings in high-dimensional state spaces formed by relatively few qubits, potentially strengthening the representation of multivariate interactions

and temporal dependence. Variational quantum algorithms use classical optimizers to train parameterized quantum circuits and have become a practical route toward useful quantum computation [24]. Quantum feature maps have also been introduced into classical time-series frameworks, where angle encoding transforms one-dimensional data into superposed and entangled states [25]. A recurrent quantum neural network has been developed for uncertain load forecasting [26], while a parameterized quantum recurrent model employing a Hadamard-based hybrid function has been used to superpose normalized inputs [27]. Applications have reported improved accuracy compared with conventional gated recurrent models on real PV data [28], demonstrated the value of strongly entangling circuit layers for capturing coupling among weather variables [29], and explored the use of quantum mechanisms in optimized microgrid-load forecasting [30].

In summary, prior studies have improved power forecasting through several technical approaches, but the coordinated use of multiscale information, external variables, and bidirectional temporal dependencies within a unified model remains insufficiently explored, particularly across both supply- and demand-side tasks. This study therefore develops a multiscale quantum-enhanced forecasting framework by integrating VMD, adaptive feature weighting, and bidirectional quantum temporal modeling for PV output and electric-load forecasting. The main contributions of this study are as follows:

(1) A rolling VMD-based representation strategy is developed to characterize the target series across frequency scales; the decomposed components capture trends, periodic variations, and short-term fluctuations as complementary inputs to the forecasting module.
(2) A bidirectional quantum temporal representation module is introduced to characterize nonlinear dependencies within the observed historical window through forward and reverse-order sequence encoding.
(3) A feature-attention mechanism adaptively weights meteorological and calendar-related variables at each time step, emphasizing informative external factors while suppressing redundant ones.
(4) The above modules are integrated into a unified direct multi-step forecasting framework that is validated on both PV generation and electric load, providing a common modeling pipeline for supply- and demand-side day-ahead forecasting.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical foundations of multiscale decomposition, adaptive feature weighting, and quantum-enhanced temporal modeling. Section 3 presents the proposed method, including preprocessing, rolling VMD, feature weighting, bidirectional temporal modeling, direct multistep forecasting, training, and evaluation. Section 4 presents PV and load case studies with correlation analyses, comparisons, and ablation experiments. Section 5 concludes the paper.

2. Methodology

This section introduces the theoretical foundations of the proposed framework, including multiscale decomposition, adaptive feature weighting, classical recurrent modeling, and quantum-enhanced temporal representation.

2.1. Variational Mode Decomposition

Variational mode decomposition is an adaptive, nonrecursive signal decomposition method proposed by Dragomiretskiy and Zosso in 2014 [31]. Unlike recursive methods such as empirical mode decomposition, VMD formulates the decomposition process as a constrained variational optimization problem and iteratively separates the constituent modes in the frequency domain [32].

VMD decomposes an original signal into K intrinsic mode functions with finite bandwidths, each concentrated around a center frequency. The constrained variational model is

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\Delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{1}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} u_k = f \tag{2}$$

Here, the Dirac distribution, convolution operator, and time derivative are used to estimate the analytic bandwidth of each mode. The objective minimizes the sum of mode bandwidths subject to exact signal reconstruction.

A quadratic penalty and a Lagrange multiplier transform the constrained problem into the following unconstrained augmented Lagrangian:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\Delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \rangle \end{aligned} \tag{3}$$

where u_k and ω_k denote the k-th mode and its center frequency, α is the balancing parameter of the data-fidelity constraint, and λ is the Lagrange multiplier.

The unconstrained problem is solved iteratively using the alternating direction method of multipliers. The mode components and center frequencies are updated in the frequency domain as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha \left(\omega - \omega_k^n \right)^2} \tag{4}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \tag{5}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_{k=1}^{K} \hat{u}_k^{n+1}(\omega) \right) \tag{6}$$

The hatted quantities denote the Fourier transforms of the corresponding time-domain variables, and the tolerance term specifies the allowable level of reconstruction noise.

Compared with conventional decomposition methods such as EMD, VMD has a well-defined variational foundation. The iterative optimization of mode center frequencies and bandwidths enables adaptive separation of different frequency bands. Because each mode is constrained to be band-limited, VMD mitigates mode mixing and improves frequency-domain resolution [33]. It also exhibits stable performance under noisy observations. The original signal can be expressed as the sum of K modes and a residual:

$$f(t) = \sum_{k=1}^{K} u_k(t) + r(t) \tag{7}$$

The residual is normally negligible in practical applications, so the original signal can be approximated by the sum of the modes. Low-frequency components represent long-term trends and periodic patterns, whereas high-frequency components reflect short-term fluctuations and random noise. The relatively stationary characteristics of the decomposed modes reduce the modeling difficulty for the subsequent forecasting model.

2.2. Feature Attention Mechanism

Attention was first widely used in neural machine translation to let a model focus dynamically on information most relevant to the current task [34]. Depending on the weighting dimension, attention may operate across time or features. This study uses feature

attention to weight meteorological and calendar-related variables at every time step. The procedure has three stages.

First, feature scores are computed. For the external-feature vector at a given time step, a two-layer feedforward network produces a score for each feature:

$$e_t = W_2 \cdot \tanh(W_1 z_t + b_1) + b_2 \tag{8}$$

The matrices and vectors in the scoring function are learnable parameters, and $\tanh(\cdot)$ is the activation function.

Second, Softmax normalization is applied across the feature dimension to obtain attention weights that sum to one:

$$a_{t,j} = \frac{\exp(e_{t,j})}{\sum_{l=1}^{d} \exp(e_{t,l})} \tag{9}$$

where e_t is the attention score vector at time t, $a_{t,j}$ is the attention weight of the j-th feature, and d is the number of external features.

A larger weight indicates greater relative attention to that external variable at the current time step.

Third, the normalized weights are applied elementwise to the original features:

$$\tilde{z}_t = a_t \odot z_t \tag{10}$$

The weighted external features are combined with the original target series and the VMD features to form the historical input.

2.3. Long Short-Term Memory Network

The LSTM network is a specialized recurrent neural network proposed by Hochreiter and Schmidhuber in 1997 [35]. Its gating mechanism and cell state alleviate vanishing gradients and improve the modeling of long-term dependencies. An LSTM cell contains forget, input, and output gates together with a cell state.

Let the current input, previous hidden state, and previous cell state be the inputs to the cell. The forward pass is defined as follows.

The forget gate controls how much information is discarded from the previous cell state:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{11}$$

The input gate determines how much new information is stored:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{12}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{13}$$

The cell state is then updated:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{14}$$

where $\tilde{C}_t$ is the candidate memory and C_t denotes the cell state at time t.

The output gate determines the hidden-state output:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{15}$$

$$h_t = o_t \odot \tanh(C_t) \tag{16}$$

The weight matrices and bias vectors are trainable; sigmoid and hyperbolic tangent are the activation functions, and the Hadamard product denotes elementwise multiplication.

LSTM has been widely applied to time-series forecasting, natural-language processing, and speech recognition. Nevertheless, its representational capacity can remain limited for highly nonlinear and nonstationary sequences.

2.4. Variational Quantum Circuit

A variational quantum circuit (VQC) is a core quantum-machine-learning component whose adjustable gate parameters are optimized for a specified objective [36]. In this study, the VQC is adopted as the quantum feature-mapping module of the quantum long short-term memory (QLSTM) gates; its implementation is informed by the quantum-enhanced forecasting framework reported in [37], while the gate-level integration follows the QLSTM formulation introduced in [38].

Quantum-enhanced approaches have also been extended to hybrid wind–PV project evaluation under carbon trading [39] and the collaborative operation of shared hydrogen storage and energy prosumers [40], demonstrating the broader applicability of quantum algorithms to energy-system assessment and coordination tasks.

2.5. Quantum Long Short-Term Memory Network

QLSTM replaces the gate mappings of a conventional LSTM cell with VQCs. In this study, the overall QLSTM cell configuration follows the architecture proposed in [38], with four VQC-based gates corresponding to the forget, input, candidate-memory, and output operations. The specific quantum-classical implementation is further adapted from the forecasting framework presented in [37].

Compared with classical LSTM, QLSTM offers two potential advantages. First, quantum superposition and entanglement allow a VQC with relatively few qubits to access a high-dimensional Hilbert space and capture complex nonlinear relationships [41]. Second, VQC achieves quantum state evolution and feature extraction through a limited number of trainable circuit parameters, thereby forming a compact model representation and possessing potential advantages in controlling model complexity [42].

2.6. Bidirectional Quantum Temporal Modeling

Bidirectional quantum temporal modeling extends quantum-enhanced recurrent representation to both chronological and reverse-order processing of the observed historical window [43]. In implementation, the proposed bidirectional quantum temporal module is realized using a bidirectional quantum long short-term memory (BiQLSTM) architecture consisting of two independent QLSTM branches.

The forward layer processes the historical window in chronological order, while the backward layer processes the same observed window in reverse. The backward branch processes only the observed historical sequence and does not use future target observations.

For an observed input sequence, the terminal hidden states of the forward and backward QLSTM branches are computed as follows:

$$h_T^{\rightarrow} = QLSTM_f(x_1, x_2, \ldots, x_T) \tag{17}$$

$$h_T^{\leftarrow} = QLSTM_b(x_T, x_{T-1}, \ldots, x_1) \tag{18}$$

The two terminal hidden states are concatenated to form a comprehensive representation of the historical window:

$$e = [h_T^{\rightarrow}; h_T^{\leftarrow}] \tag{19}$$

The forward branch preserves the chronological accumulation of historical information, whereas the backward branch supplies reverse-order context from the same observed window. Their fusion provides a more comprehensive representation of temporal dependence.

3. Proposed Method

Building on Section 2, this section presents the proposed multiscale quantum-enhanced forecasting framework for day-ahead PV output and electric-load forecasting. It describes the overall framework, preprocessing, rolling VMD, external-feature weighting,

model inputs and outputs, quantum-circuit configuration, multistep prediction, training, and evaluation.

3.1. Overall Framework

The proposed framework takes historical targets and external features as inputs. Rolling VMD separates variation at different frequency scales within trailing observation windows, and its features are retained alongside the original target series rather than replacing it. FA then assigns normalized, time-dependent weights to external variables before temporal encoding, allowing their relative influence to vary with the observed conditions. BiQLSTM combines nonlinear variational-circuit gate mappings with chronological and reverse-order processing of the same observed history, providing complementary temporal context without accessing future target observations. For PV forecasting, these complementary roles are intended to distinguish slower generation patterns from rapid irradiance-related fluctuations and relate them to irradiance, temperature and calendar inputs. For load forecasting, they are intended to distinguish the underlying demand level and recurring cycles from short-term changes while representing nonlinear relationships with temperature and calendar variables. The resulting historical context is shared by the direct multi-step decoder, so decomposition, feature weighting and temporal encoding address different parts of the prediction task. The overall framework is shown in Figure 1.

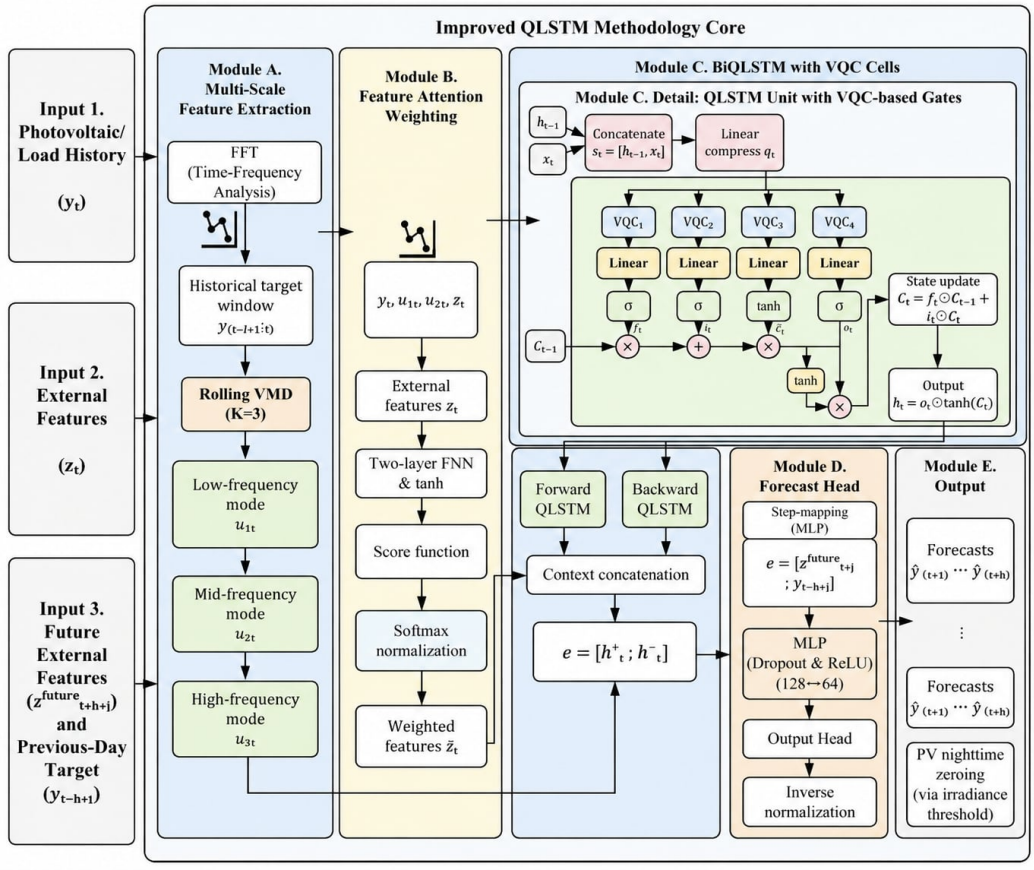


Figure 1. Overall architecture of the proposed multiscale quantum-enhanced forecasting framework.

Let T denote the forecast origin, L the historical-window length, and H the forecast horizon. The target history observed up to and including T is

$$y(t) = \{y_{T-L+1}, y_{T-L+2}, \ldots, y_T\} \quad (20)$$

where the target is either PV power or electric load at time t. After preprocessing, the model proceeds in three stages.

First, VMD is applied to a fixed-length window ending at the current historical time. The modes are sorted by their final center frequencies, and their terminal values form a VMD feature vector:

$$v_t = [u_{1(t)}, u_{2(t)}, \ldots, u_{K(t)}] \tag{21}$$

where $u_{K(t)}$ denotes the terminal value of the K-th mode at time t and K is the number of VMD modes.

Second, external features are weighted elementwise by FA and concatenated with the target and VMD features. The resulting sequence is processed by the forward and backward QLSTM branches:

$$x_t = [y(t), v_{1(t)}, \ldots, v_{K(t)}, \tilde{z}_t] \tag{22}$$

where y_t is the target value, $v_{K(t)}$ is the terminal value of the K-th VMD mode, and z_t is the attention-weighted external-feature vector at time t.

The window $[x_{T-L+1}, \ldots, x_T]$ is then fed into the BiQLSTM, whose historical-context vector e is defined in Equation (19).

Third, for each future step j, the shared historical context, external covariates available at the forecast origin, and the target at the corresponding time on the previous day are concatenated and passed through a multilayer perceptron (MLP) to predict the entire next day:

$$d_j = [e; z_{T+j}^{future}; y_{T-H+j}], \hat{y}_{T+j} = MLP(d_j) \tag{23}$$

where e is the historical-context vector, z_{T+j}^{future} denotes the external covariates available at the forecast origin, and y_{T-H+j} is the target at the corresponding time on the previous day, H is the number of time steps in one day.

Finally, forecasts are inverse-transformed to their original scale. PV forecasts are additionally set to zero at night using an observed-irradiance threshold, as described in Section 3.6. Both tasks share the VMD, FA, BiQLSTM, and direct multistep-output modules, while their input features and training settings are task-specific.

3.2. Data Preprocessing and Sample Construction

Raw observations may contain outliers and missing values, which can degrade forecasting accuracy. Task-specific preprocessing is therefore applied to the PV and load datasets.

PV records outside fixed physical validity ranges are treated as missing. Observations are sorted chronologically and resampled into right-closed, right-labeled five-minute mean bins, so readings later than a bin label cannot enter that bin. Missing values are filled only from the most recent available earlier observation; neither backward filling nor two-sided interpolation is used. Leading gaps without a prior valid observation remain missing, and requested model windows containing unresolved gaps are rejected. The existing same-timestamp target-cleaning convention sets PV power to zero when observed global horizontal irradiance is below 5 W/m^2. The processed five-minute PV and meteorological records are then aggregated to hourly resolution by taking the mean of the twelve observations within each hour. Subsequent correlation analysis, rolling VMD feature construction, and forecasting use these hourly data.

For load forecasting, historical records without observed load labels are excluded before assembling the hourly series. The retained load and temperature data are checked for missing values and an ordered, duplicate-free timeline; missing values in the assembled series raise an error rather than being interpolated or backward-filled. Hourly inputs are augmented with meteorological summaries from 25 stations, calendar-related features and

a 168 h lagged-load feature. The lag is computed by shifting the observed load series, and unavailable lag values in the requested analysis window are rejected.

Min-max normalization maps each feature to a common interval and removes scale differences:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{24}$$

Within each forecasting task, all comparison models use the same chronological data split and the same declared training interval to fit the min-max scalers. Feature and target scalers, and VMD-feature scalers where applicable, are fitted using only observations from this training interval. The fitted scaling statistics are then fixed: validation and test observations do not update them. Validation and test data are transformed using the corresponding training-fitted scalers, and predictions are subsequently inverse-transformed:

$$x_{\text{orig}} = x_{\text{norm}} \times (x_{\max} - x_{\min}) + x_{\min} \tag{25}$$

3.3. Rolling VMD-Based Multiscale Feature Extraction

The VMD module extracts multiscale features from trailing historical windows, rather than decomposing the complete training and test series jointly. For a feature attached to historical time t, the implementation decomposes a fixed-length target window ending at t and stores only the terminal mode values. Modes are ordered by their final center frequencies. The default settings are K = 3, $\alpha = 2000$, $\tau = 0$, DC = 0, init = 1 and tol = 10^{-7}.

At each forecast origin T, the historical VMD inputs use only target observations timestamped at or before T, including the most recent observation at T. The prediction horizon extends from T + 1 to T + H, and no target observation from this horizon enters the historical VMD representation. Observations from earlier test days may enter a later forecast once they have become available history. Precomputing or caching the feature sequence does not alter these trailing-window boundaries. The ordered terminal modes are concatenated with the original historical target to describe slower variation, periodic structure and more rapid fluctuations.

3.4. Construction of Weighted External Features

At every time step, attention weights are calculated across the external-feature dimension and applied elementwise before temporal modeling.

For the external-feature vector at a given time step, FA is computed as

$$a_t = softmax(W_2 \cdot \tanh(W_1 z_t + b_1) + b_2),\ \tilde{z}_t = a_t \odot z_t \tag{26}$$

where W_1, W_2 and b_1, b_2 are the learnable weight matrices and biases of the two-layer scoring network, z_t is the external-feature vector at time t, at is the attention weight vector, and d is the number of external features.

3.5. Input–Output Structure and Quantum Circuit Configuration

The bidirectional quantum temporal module receives the target, three multiscale VMD features, and attention-weighted external variables. Sliding windows of length L form the training samples. The forward branch processes each observed window chronologically and the backward branch processes the same window in reverse; their terminal hidden states are concatenated:

$$e = [QLSTM_f(X); QLSTM_b(reverse(X))] \tag{27}$$

where X denotes the input window of length L; $Xrev$, its reversed copy; and e, the fused historical-context vector (see Equation (19)).

The backward branch uses no true target from the forecast period. The hourly seasonal design uses 72 historical observations and a 24-step output for each task, corresponding to three days of history and one day ahead.

Within each BiQLSTM branch, four independent VQCs participate in the forget, input, candidate-memory, and output-gate calculations. Each VQC uses StronglyEntanglingLayers composed of parameterized single-qubit rotations and entangling connections. The circuit contains four qubits and one variational layer. The current input and previous hidden state are concatenated, compressed into four dimensions, and angle-encoded by mapping the arctangent of each component to a gate rotation:

$$\theta_i = \arctan(q_i),\ i = 1,2,\ldots,N_q \tag{28}$$

where q_i is the i-th component of the compressed input and N_q is the number of qubits.

The variational-layer weight tensor assigns three trainable rotation parameters to each qubit and couples the qubits through a strongly entangling layer. Pauli-Z expectation values are measured and mapped to the hidden dimension by a classical linear layer.

3.6. Direct Multi-Step Forecasting

The bidirectional quantum temporal module converts the historical sequence into a shared context representation.

For each future step, the MLP input comprises the shared context, external features available at the forecast origin, and the target at the corresponding time on the previous day, as defined in Equation (23).

Future external inputs include meteorological and calendar variables for PV forecasting and meteorological, calendar and lagged-load variables for load forecasting. For both tasks, meteorological inputs over the forecast horizon are estimated using five-year historical averages. Specifically, each meteorological variable at a target time step is estimated as the arithmetic mean of its observations at the corresponding time slots in the five preceding years. Only historical observations available before the forecast origin are used to construct these estimates; actual meteorological measurements from the target forecast period are not supplied to the forecasting model. The observation-based nighttime postprocessing of PV predictions is described separately below. These values are historical-average estimates of future weather conditions, whereas calendar variables are known in advance. The 168 h lagged-load input is obtained from previously observed load values.

$$\hat{y}_{T+j} = W_3 \cdot ReLU(W_2 \cdot Dropout(ReLU(W_1 d_j + b_1)) + b_2) + b_3 \tag{29}$$

where W_1, W_2, W_3 and b_1, b_2, b_3 denote the weight matrices and bias vectors of the MLP.

The MLP has hidden widths of 128 and 64, with a dropout rate of 0.1 after the first hidden layer.

$$\hat{y} = [\hat{y}_{T+1},\hat{y}_{T+2},\ldots,\hat{y}_{T+H}] \tag{30}$$

The architecture directly outputs all values for the next day, avoiding error accumulation from recursive forecasting.

Forecasts are inverse-transformed to the original scale. For PV evaluation, predicted power is set to zero at nighttime timestamps identified using actual observed irradiance and the irradiance threshold. This postprocessing is separate from the meteorological inputs to the forecasting model, which use the five-year historical averages described above. Because the nighttime rule uses observations from the evaluation period, the reported PV metrics describe an offline evaluation with observation-based nighttime zeroing. Deployment would require a nighttime rule based on information available at the forecast origin.

3.7. Model Training Strategy

The model is trained with the Adam optimizer and mean squared error loss. For PV forecasting, the learning rate is 0.001 and the batch size is set to 64 for processor-based training and 128 for graphics-accelerated training. For load forecasting, the learning rate is 0.0003, the batch size is 16, and training runs for at most 35 epochs.

$$L_{MSE} = \frac{1}{BH}\sum_{b=1}^{B}\sum_{j=1}^{H}\left(y_{b,j} - \hat{y}_{b,j}\right)^2 \tag{31}$$

where B is the batch size, H is the forecast horizon, and $y_{b,j}$ and $\hat{y}_{b,j}$ are the observed and predicted values of the j-th step in the b-th sample.

The forecasting framework was implemented in Python (version 3.12.6) using NumPy, pandas, PyTorch (version 2.12.0), PennyLane (version 0.45.0), scikit-learn, vmdpy, and Matplotlib (version 3.10.8). PyTorch was used for neural-network construction and optimization, PennyLane for variational quantum circuits with the default.qubit simulator and the PyTorch interface, scikit-learn for min-max scaling and performance evaluation, vmdpy for variational mode decomposition, and Matplotlib for visualization. The source code used in this study is available from the corresponding authors upon reasonable request.

3.8. Evaluation Metrics

Performance is evaluated using mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). All metrics are computed after inverse transformation to the original scale and the stated postprocessing. For each evaluation interval, metrics are calculated separately for each of the five independent runs and then averaged arithmetically across runs. Thus, the reported values are means of run-specific metrics, rather than metrics calculated from an average prediction curve. Overall results use the combined 672 hourly evaluation timestamps in each run before averaging across runs. The overall comparison tables report mean ± sample standard deviation across five runs, with sample standard deviations calculated using a denominator of four. The MSE-derived summaries in the sensitivity analyses are calculated separately from averaged seasonal MSEs and may therefore differ from the mean run-specific RMSE.

MSE is the mean squared prediction error and is particularly sensitive to large deviations:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{32}$$

RMSE represents the overall error magnitude and has the same unit as the original data:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{33}$$

MAE is the mean absolute deviation between predictions and observations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{34}$$

R^2 measures the agreement between predicted and observed variations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{35}$$

The symbols denote the observed value, predicted value, mean observation, and sample size. Smaller MSE, RMSE, and MAE indicate better predictive performance, while an R^2 closer to one indicates greater explanatory power.

4. Case Studies

Each season is presented using one metric table and four panels sharing the same actual series and the VMD-FA-BiQLSTM sequence. Panel (a) adds VMD-BiQLSTM and FA-BiQLSTM; panel (b) adds VMD-FA-QLSTM and VMD-FA-BiLSTM; panel (c) adds BiLSTM and QLSTM; and panel (d) adds TCN and Transformer. All metrics use 168 hourly points per season, including nighttime PV observations. The seven-day display is a sequence of daily 24 h forecast blocks, not a single 168-h-ahead forecast. The seasonal experiment driver is configured with the random seeds 101, 202, 303, 404 and 505. Every comparison model and ablation is trained independently with these five seeds. Both seasonal and combined comparison tables report mean ± sample standard deviation across the five independent runs. Seasonal results use 168 hourly observations per run, while the overall results use the combined 672 hourly observations per run, as defined in Section 3.8.

4.1. Photovoltaic Power Forecasting

4.1.1. Data Description for Photovoltaic Power Forecasting

The PV case uses the public DKASC Alice Springs dataset provided by the Desert Knowledge Australia Solar Centre (DKA Solar Centre) in Australia. The original resolution is 5 min and the target is active power. Eleven historical external features are used: ambient temperature, global horizontal irradiance, module temperature, first- to third-order sine and cosine terms for the daily cycle, and sine and cosine terms for the day of year.

The seasonal PV target series are selected from the 2025 records and named according to the local Southern Hemisphere seasons: spring covers 15–21 September, summer 15–21 December, autumn 15–21 March, and winter 15–21 June. Each interval runs from 00:00 on the first day to 23:00 on the seventh day. Fixed physical-range screening and five-minute resampling are followed by past-only forward filling, and the processed five-minute PV and meteorological records are averaged within each hour before subsequent analysis and modeling, yielding hourly series with 168 points per season and 672 overall. The existing low-irradiance target convention is retained.

4.1.2. Correlation Analysis

Pearson correlation analysis is performed after outlier handling, missing-value imputation, and temporal alignment to quantify linear associations between PV inputs, the original target, and VMD components at different scales. Coefficients are calculated only from training data before the test origin to prevent leakage. The analysis supports the physical plausibility of the selected inputs and helps interpret the FA module.

The Pearson coefficient measures the linear association between variables X and Y through normalized covariance:

$$r_{XY} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{36}$$

Here, the indexed values represent paired observations, the barred terms denote the corresponding sample means, and n is the sample size. The coefficient ranges from -1 to 1; its magnitude indicates the strength of association and its sign indicates the direction of association. No variable is removed solely by a fixed correlation threshold, because nonlinear effects and feature interactions are left for the FA module to learn.

A representative 30-day training window preceding the PV forecast origin contains 720 hourly samples, obtained after averaging the original five-minute records within each hour. Correlations among the external inputs, original target, and the aligned rolling VMD features with K = 3 are shown in Figure 2. The modes are ordered as low-frequency u1,

medium-frequency u2, and high-frequency u3, following the historical feature construction in Section 3.3.

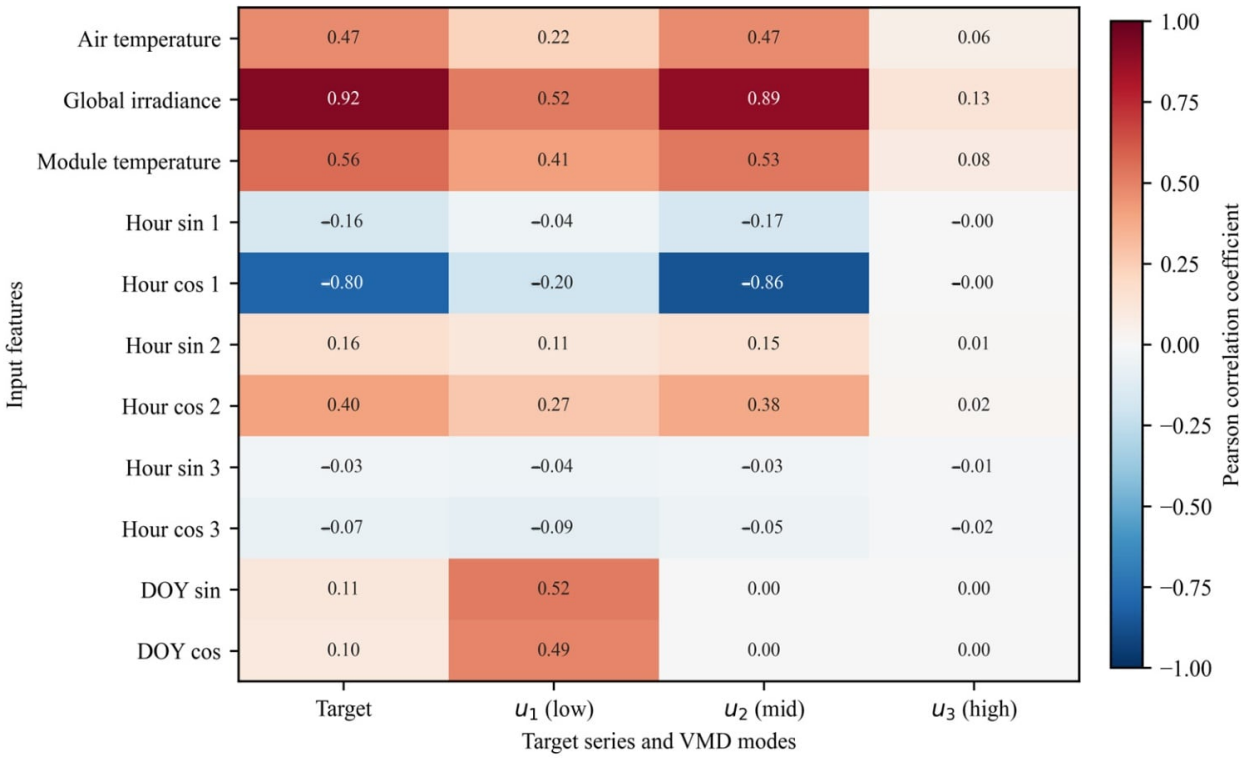


Figure 2. Pearson correlations between PV input features, the target series, and VMD modes.

Global horizontal irradiance is most strongly associated with PV power (0.92) and the medium-frequency mode u_2 (0.89). Module temperature has coefficients of 0.56 and 0.53 with these quantities, respectively. In addition to irradiance, the low-frequency mode u_1 is associated with annual-cycle terms at approximately 0.49–0.52. The high-frequency mode u_3 generally exhibits weak linear correlations with individual meteorological variables, suggesting the influence of short-term disturbances, nonlinear effects, and unobserved influences.

4.1.3. Experimental Setup

The nine sequence labels are defined in Table 1. VMD-BiQLSTM and FA-BiQLSTM are independently trained single-component ablations of the complete model; VMD-FA-QLSTM removes bidirectionality. VMD-FA-BiLSTM is the implemented classical counterpart retaining VMD, FA and the decoder. BiLSTM and QLSTM provide recurrent comparisons, while TCN and Transformer are separate modern baselines.

All PV comparison models use 72 hourly observations as input and predict the next 24 hourly values. Recurrent encoders use a hidden width of 64; bidirectional encoders use 64 units per direction. The decoder has hidden widths of 128 and 64 with a dropout of 0.1. Training settings are shared across the PV models: MSE loss, Adam with learning rate 0.001, ten epochs, gradient-norm clipping at 5, and batch size 64 on CPU or 128 on CUDA. Quantum gates use the default.qubit classical simulator. The chronological split uses 365 training days followed by 30 validation days before each seasonal test interval, matching the window lengths used for load forecasting. All PV comparison models use the same split within each seasonal experiment.

The seasonal driver specifies seeds 101, 202, 303, 404 and 505. The shared seed routine initializes Python (version 3.12.6), NumPy (version 2.3.5) and PyTorch (version 2.12.0),

including available CUDA (version 12.8) devices; training loaders use seeded generators. The driver also passes PYTHONHASHSEED to each child process. cuDNN benchmarking is disabled, deterministic cuDNN kernels are requested, and PyTorch deterministic algorithms are enabled in warning-only mode. These controls reduce variation but do not guarantee bitwise identity across devices or library versions.

Table 1. PV comparison model configurations.

Method	Configuration
VMD-FA-BiQLSTM	Implemented full architecture: rolling VMD (default K = 3), feature attention and BiQLSTM. Hidden width 64 per direction; four qubits and one entangling layer in each quantum gate. The strict classical entry inherits its data and training pipeline.
VMD-BiQLSTM	Independently trained ablation retaining VMD and BiQLSTM while removing FA. Uses the shared PV training settings and five independent runs with seeds 101, 202, 303, 404 and 505.
FA-BiQLSTM	Independently trained ablation retaining FA and BiQLSTM while removing VMD. Uses the shared PV training settings and five independent runs with seeds 101, 202, 303, 404 and 505.
VMD-FA-QLSTM	Independently trained ablation retaining VMD and FA while replacing the bidirectional encoder with a unidirectional QLSTM. Uses the shared PV training settings and five independent runs with seeds 101, 202, 303, 404 and 505.
VMD-FA-BiLSTM	Implemented matched classical counterpart: retain VMD, FA and the decoder, replacing BiQLSTM with a one-layer BiLSTM of hidden width 64 per direction. Data preparation, loss, optimizer and epoch limit are inherited from the full-model entry.
QLSTM	Unidirectional QLSTM baseline without VMD or FA. Hidden width 64; four qubits and one circuit layer per quantum gate. Uses the shared hourly PV input–output design and training settings.
BiLSTM	One-layer BiLSTM baseline without VMD or FA. Hidden width 64 per direction. Uses the shared hourly PV input–output design and training settings.
TCN	Separate TCN baseline without VMD or FA. Eight residual causal-convolution blocks, channels (32, 32, 32, 32, 32, 32, 32, 64), two convolutions per block, kernel size 3, dilations 1–128, and dropout 0.1.
Transformer	Transformer baseline without VMD or FA. Uses 72 hourly historical observations and a 24 h output horizon; embedding width 64, four heads, two encoder layers, feed-forward width 128, GELU activation, dropout 0.1, learned positional embeddings and mean pooling.

In the implemented quantum recurrent cell, a learned linear projection compresses the concatenated current input and previous hidden state to four values. Their arctangents are encoded using Y-axis AngleEmbedding, followed by one StronglyEntanglingLayers layer with weights of shape (1, 4, 3). Each qubit returns a Pauli-Z expectation value. The forget, input, candidate and output gates have separate circuit weights, and the two temporal directions have separate cells. The default.qubit QNodes use the PyTorch interface and backpropagation. Hidden and cell states start at zero. Classical weights and circuit weights retain the corresponding PyTorch and PennyLane TorchLayer default initialization, respectively.

K = 3 is a configured default, with K = 2, 3, 4 and 5 included in the sensitivity design. Run configurations record the seed, temporal split, window lengths, K, learning rate, batch size, device and parameter count.

The hourly forecasting design uses a 72 h historical input and a 24 h output horizon. The default rolling VMD representation uses K = 3 and orders modes by center frequency. Table 2 summarizes the conceptual roles of the low-, medium-, and high-frequency components.

Table 2. Interpretation of the PV VMD modes.

Mode	Relative Frequency	Interpretation	Model Input
u_1	Low	Intraday baseline and slow trend	Window-end value
u_2	Medium	Cloud variation and local fluctuation	Window-end value
u_3	High	Rapid disturbance and noise	Window-end value

4.1.4. Sensitivity to the VMD Mode Count

Table 3 compares K = 2, 3, 4 and 5 for VMD-FA-BiQLSTM over the four seasonal seven-day intervals, with 168 hourly observations per season. Bold values identify the best result for each metric within each season.

Table 3. PV mode-count comparison over four seasonal seven-day intervals.

Season/K	MSE (kW^2)	RMSE (kW)	MAE (kW)	R^2
Spring/2	5.1683	2.2734	1.2033	0.9579
Spring/3	4.8994	2.2132	1.1745	0.9601
Spring/4	5.2689	2.2954	1.2143	0.9571
Spring/5	5.3607	2.3153	1.2246	0.9564
Summer/2	6.0320	2.4560	1.3224	0.9594
Summer/3	5.7698	2.4014	1.2955	0.9612
Summer/4	5.8906	2.4271	1.3073	0.9604
Summer/5	6.1831	2.4866	1.3390	0.9584
Autumn/2	2.5190	1.5871	0.8546	0.9610
Autumn/3	2.4510	1.5642	0.8420	0.9621
Autumn/4	2.4757	1.5734	0.8469	0.9617
Autumn/5	2.5850	1.6078	0.8655	0.9600
Winter/2	3.6021	1.8979	0.9641	0.9549
Winter/3	3.3643	1.8341	0.9311	0.9578
Winter/4	3.5110	1.8738	0.9516	0.9560
Winter/5	3.5521	1.8847	0.9570	0.9555

In the tabulated comparisons, K = 3 achieves the lowest MSE, RMSE and MAE and the highest R^2 in every evaluated season. At K = 3, the seasonal mean RMSEs are 2.2132, 2.4014, 1.5642 and 1.8341 kW in spring, summer, autumn and winter, respectively. Taking the square root of the mean of the four equally sized seasonal MSEs gives RMSE-equivalent values of 2.0809, 2.0301, 2.0704 and 2.1024 kW for K = 2, 3, 4 and 5, respectively. K = 3 has the lowest combined MSE-derived error summary among the tabulated settings. The relatively small differences indicate limited sensitivity to the tested mode counts within these evaluation intervals.

4.1.5. Feature Attention Weights

The FA module uses a two-layer feedforward scoring network with a tanh activation, followed by Softmax over the external-feature dimension. The score-network input and output widths equal the number of external features: 11 for PV. At each historical time step, the resulting nonnegative weights sum to one across features and are applied elementwise before the temporal encoder. The target and VMD channels are concatenated separately; they are not part of this feature-weight normalization. These are feature weights, not attention scores normalized across historical times or seasons.

Table 4 summarizes the supplied attention results for the sequence labeled VMD-FA-BiQLSTM. Each of the 28 daily forecast origins has 72 historical-hour weight vectors, giving 2016 vectors across four seasons. Daily weights are means over those 72 vectors; seasonal weights average seven daily profiles; the overall column averages the four equally sized

seasons. All historical timestamps are at or before the forecast origin T; the forecast horizon begins at T + 1. Every column sums to one over the complete feature set before rounding. A feature row is not required to sum to one across seasons, and displayed four-decimal entries may produce small rounding differences.

Table 4. Overall and seasonal PV feature-attention weights.

Feature	Overall	Spring	Summer	Autumn	Winter
Historical global irradiance	0.2473	0.2426	0.2401	0.2568	0.2495
Historical module temperature	0.1161	0.1152	0.1181	0.1165	0.1146
Historical ambient temperature	0.1030	0.1021	0.0981	0.1061	0.1055
Daily cycle cosine	0.0923	0.0949	0.0955	0.0873	0.0914
Daily cycle sine	0.0886	0.0899	0.0929	0.0850	0.0866
Annual cycle cosine	0.0808	0.0830	0.0843	0.0763	0.0796
Annual cycle sine	0.0781	0.0787	0.0815	0.0757	0.0763
Daily second harmonic sine	0.0527	0.0538	0.0522	0.0514	0.0534
Daily second harmonic cosine	0.0522	0.0521	0.0499	0.0533	0.0535
Daily third harmonic sine	0.0452	0.0444	0.0435	0.0471	0.0459
Daily third harmonic cosine	0.0438	0.0432	0.0439	0.0445	0.0434
Sum	1.0000	1.0000	1.0000	1.0000	1.0000

In the PV profile, historical global irradiance has the largest overall weight (0.2473), followed by historical module temperature (0.1161) and ambient temperature (0.1030). Its seasonal weights are 0.2426 in spring, 0.2401 in summer, 0.2568 in autumn and 0.2495 in winter. Daily and annual cycle terms share the remaining allocation.

4.1.6. Results and Ablation Analysis

The seasonal comparisons below pair each seven-day forecast group with its own metric table. Bold values identify the lowest MSE, RMSE, and MAE and the highest R^2 within the supplied comparison sequences.

Spring: The seven-day comparison covers 15 September 2025, 00:00 to 21 September 2025, 23:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 5 and Figure 3.

Table 5. Spring PV power forecasting performance across five independent runs.

Model	MSE (kW^2)	RMSE (kW)	MAE (kW)	R^2
VMD-FA-BiQLSTM	4.8994 ± 0.1541	2.2132 ± 0.0350	1.1745 ± 0.0184	0.9601 ± 0.0013
VMD-BiQLSTM	7.3861 ± 0.2554	2.7174 ± 0.0471	1.4423 ± 0.0227	0.9399 ± 0.0021
FA-BiQLSTM	7.8355 ± 0.2414	2.7989 ± 0.0432	1.4793 ± 0.0227	0.9362 ± 0.0020
VMD-FA-QLSTM	10.2547 ± 0.4650	3.2016 ± 0.0731	1.6917 ± 0.0441	0.9165 ± 0.0038
VMD-FA-BiLSTM	12.6207 ± 0.3413	3.5523 ± 0.0478	1.8897 ± 0.0277	0.8973 ± 0.0028
QLSTM	12.8064 ± 0.3822	3.5783 ± 0.0535	1.8877 ± 0.0318	0.8958 ± 0.0031
BiLSTM	16.5998 ± 0.5329	4.0739 ± 0.0654	2.1520 ± 0.0364	0.8649 ± 0.0043
TCN	16.4808 ± 0.2131	4.0596 ± 0.0263	2.1508 ± 0.0185	0.8659 ± 0.0017
Transformer	20.6250 ± 0.4095	4.5413 ± 0.0450	2.4059 ± 0.0319	0.8321 ± 0.0033

Figure 3 presents the seven-day spring PV forecast comparison. In Table 5, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 2.2132 ± 0.0350 kW and its MAE is 1.1745 ± 0.0184 kW; the mean MSE and R^2 are 4.8994 kW^2 and 0.9601, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 2.7174 and 2.7989 kW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 3.2016 and 3.5523 kW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

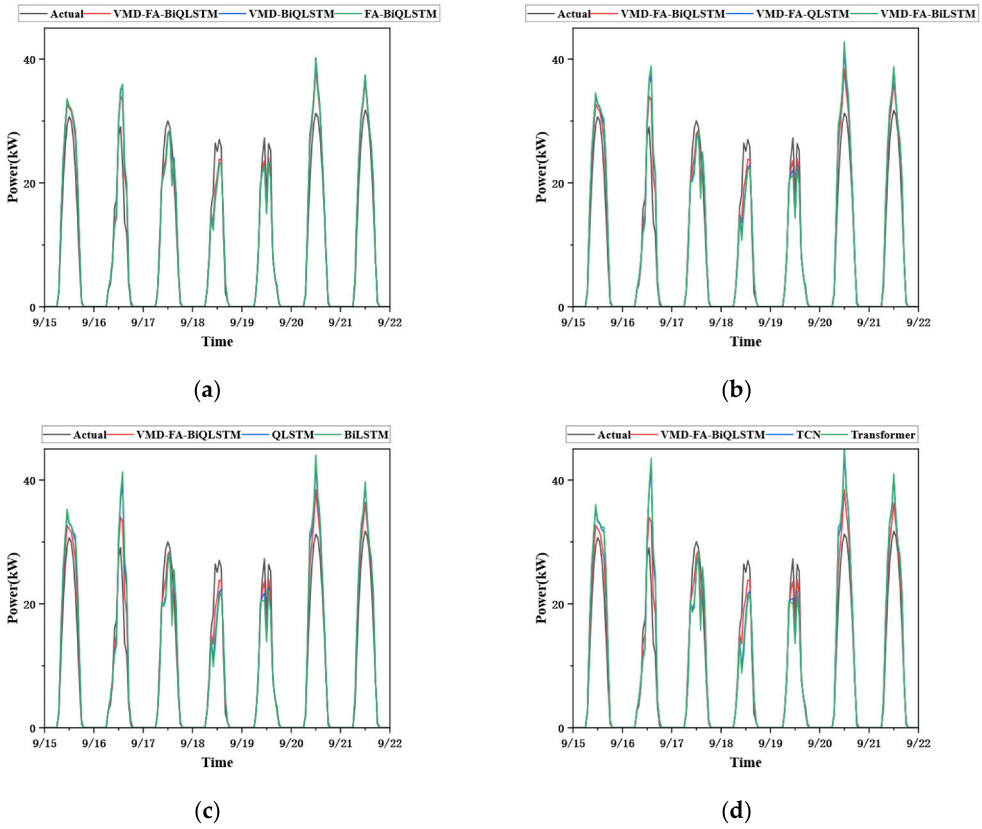


Figure 3. Spring seven-day PV power prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Summer: The seven-day comparison covers 15 December 2025, 00:00 to 21 December 2025, 23:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 6 and Figure 4.

Table 6. Summer PV power forecasting performance across five independent runs.

Model	MSE (kW^2)	RMSE (kW)	MAE (kW)	R^2
VMD-FA-BiQLSTM	5.7698 ± 0.2924	2.4014 ± 0.0619	1.2955 ± 0.0331	0.9612 ± 0.0020
VMD-BiQLSTM	8.7883 ± 0.2562	2.9642 ± 0.0435	1.5918 ± 0.0184	0.9408 ± 0.0017
FA-BiQLSTM	8.4375 ± 0.2476	2.9045 ± 0.0426	1.5706 ± 0.0246	0.9432 ± 0.0017
VMD-FA-QLSTM	11.5368 ± 0.4472	3.3961 ± 0.0660	1.8332 ± 0.0402	0.9223 ± 0.0030
VMD-FA-BiLSTM	14.4107 ± 0.4375	3.7958 ± 0.0575	2.0350 ± 0.0310	0.9030 ± 0.0029
QLSTM	14.3285 ± 0.3634	3.7851 ± 0.0480	2.0454 ± 0.0238	0.9036 ± 0.0024
BiLSTM	18.5062 ± 0.7556	4.3011 ± 0.0887	2.3176 ± 0.0438	0.8754 ± 0.0051
TCN	18.5021 ± 0.9643	4.3003 ± 0.1105	2.3078 ± 0.0579	0.8755 ± 0.0065
Transformer	23.8078 ± 0.6644	4.8789 ± 0.0681	2.6228 ± 0.0326	0.8398 ± 0.0045

Figure 4 presents the seven-day summer PV forecast comparison. In Table 6, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 2.4014 ± 0.0619 kW and its MAE is 1.2955 ± 0.0331 kW; the mean MSE and R^2 are 5.7698 kW^2 and 0.9612, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 2.9642 and 2.9045 kW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 3.3961 and 3.7958 kW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

Autumn: The seven-day comparison covers 15 March 2025, 00:00 to 21 March 2025, 23:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 7 and Figure 5.

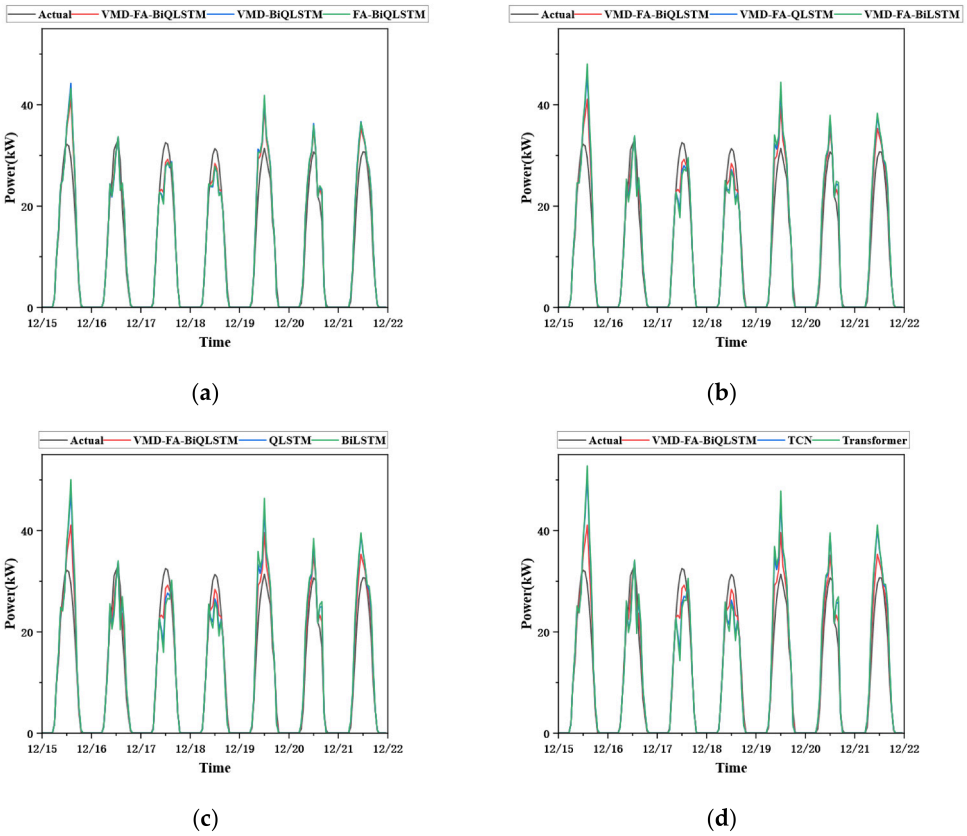


Figure 4. Summer seven-day PV power prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 7. Autumn PV power forecasting performance across five independent runs.

Model	MSE (kW^2)	RMSE (kW)	MAE (kW)	R^2
VMD-FA-BiQLSTM	2.4510 ± 0.2312	1.5642 ± 0.0730	0.8420 ± 0.0416	0.9621 ± 0.0036
VMD-BiQLSTM	3.3712 ± 0.0920	1.8359 ± 0.0252	0.9867 ± 0.0137	0.9478 ± 0.0014
FA-BiQLSTM	3.6364 ± 0.1303	1.9067 ± 0.0343	1.0239 ± 0.0173	0.9437 ± 0.0020
VMD-FA-QLSTM	4.9851 ± 0.1463	2.2326 ± 0.0328	1.1995 ± 0.0190	0.9229 ± 0.0023
VMD-FA-BiLSTM	6.1508 ± 0.2510	2.4797 ± 0.0509	1.3314 ± 0.0246	0.9048 ± 0.0039
QLSTM	6.2649 ± 0.1265	2.5029 ± 0.0253	1.3393 ± 0.0153	0.9030 ± 0.0020
BiLSTM	7.8676 ± 0.2000	2.8047 ± 0.0356	1.4927 ± 0.0176	0.8782 ± 0.0031
TCN	7.9744 ± 0.1433	2.8238 ± 0.0253	1.5174 ± 0.0173	0.8766 ± 0.0022
Transformer	9.8533 ± 0.1983	3.1389 ± 0.0316	1.6766 ± 0.0206	0.8475 ± 0.0031

Figure 5 presents the seven-day autumn PV forecast comparison. In Table 7, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 1.5642 ± 0.0730 kW and its MAE is 0.8420 ± 0.0416 kW; the mean MSE and R^2 are 2.4510 kW^2 and 0.9621, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 1.8359 and 1.9067 kW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 2.2326 and 2.4797 kW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

Winter: The seven-day comparison covers 15 June 2025, 00:00 to 21 June 2025, 23:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 8 and Figure 6.

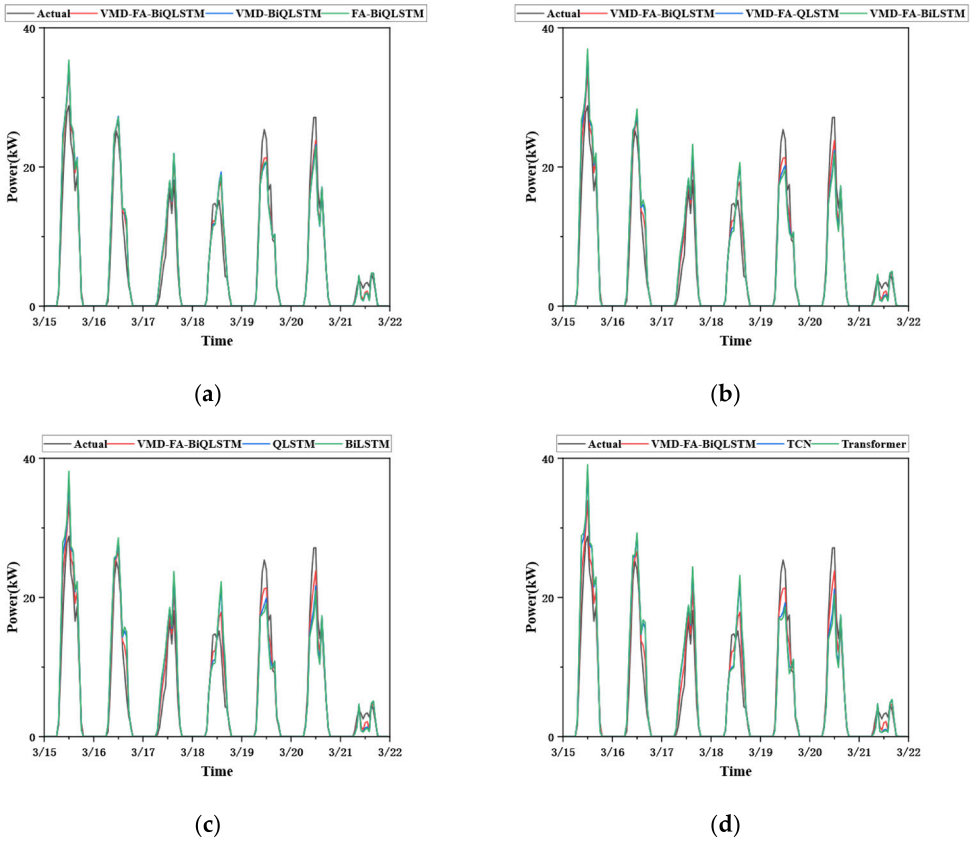


Figure 5. Autumn seven-day PV power prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 8. Winter PV power forecasting performance across five independent runs.

Model	MSE (kW^2)	RMSE (kW)	MAE (kW)	R^2
VMD-FA-BiQLSTM	3.3643 ± 0.0802	1.8341 ± 0.0218	0.9311 ± 0.0101	0.9578 ± 0.0010
VMD-BiQLSTM	5.1689 ± 0.1739	2.2733 ± 0.0384	1.1501 ± 0.0202	0.9352 ± 0.0022
FA-BiQLSTM	5.2633 ± 0.2178	2.2938 ± 0.0479	1.1666 ± 0.0221	0.9340 ± 0.0027
VMD-FA-QLSTM	7.0235 ± 0.3853	2.6494 ± 0.0731	1.3464 ± 0.0351	0.9120 ± 0.0048
VMD-FA-BiLSTM	8.4707 ± 0.4420	2.9097 ± 0.0760	1.4729 ± 0.0366	0.8938 ± 0.0055
QLSTM	8.6016 ± 0.0809	2.9328 ± 0.0138	1.4894 ± 0.0065	0.8922 ± 0.0010
BiLSTM	11.0505 ± 0.5578	3.3234 ± 0.0851	1.6796 ± 0.0444	0.8615 ± 0.0070
TCN	10.8983 ± 0.3569	3.3009 ± 0.0536	1.6758 ± 0.0260	0.8634 ± 0.0045
Transformer	13.5355 ± 0.9568	3.6772 ± 0.1300	1.8615 ± 0.0703	0.8304 ± 0.0120

Figure 6 presents the seven-day winter PV forecast comparison. In Table 8, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 1.8341 ± 0.0218 kW and its MAE is 0.9311 ± 0.0101 kW; the mean MSE and R^2 are 3.3643 kW^2 and 0.9578, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 2.2733 and 2.2938 kW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 2.6494 and 2.9097 kW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively. Across the four seasonal intervals, the complete model has mean RMSEs ranging from 1.5642 kW in autumn to 2.4014 kW in summer and mean R^2 values ranging from 0.9578

in winter to 0.9621 in autumn. The combined results and their run-to-run variability are reported in Table 9.

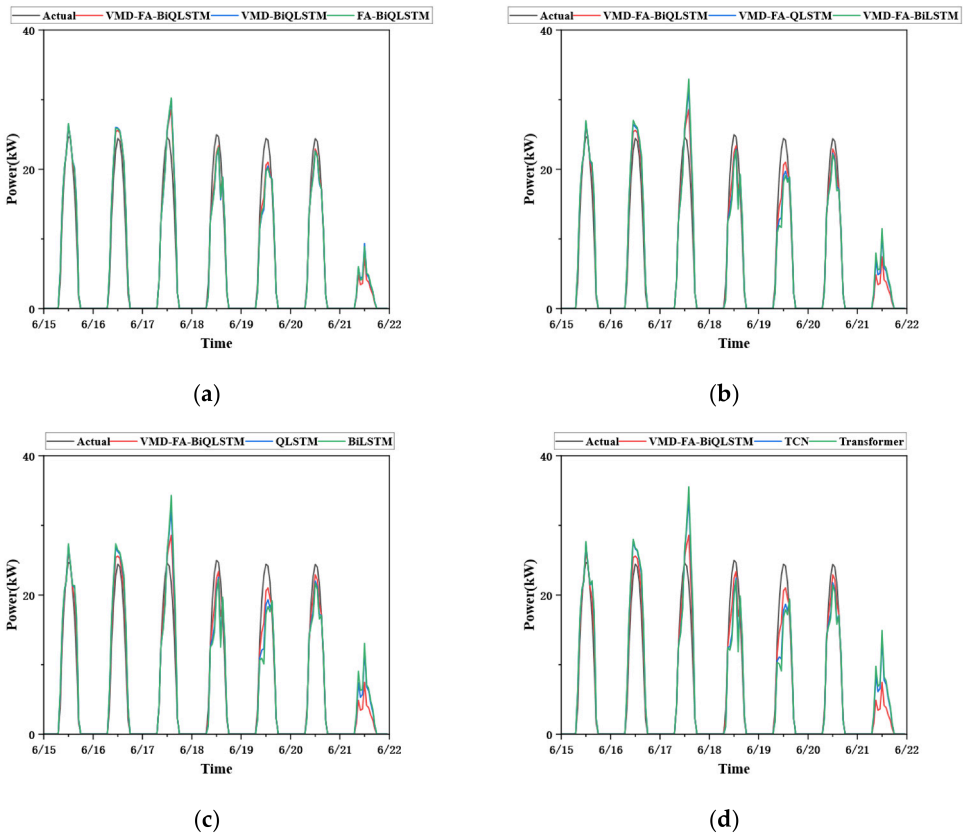


Figure 6. Winter seven-day PV power prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 9. Overall PV forecasting performance across five independent runs.

Model	MSE (kW^2)	RMSE (kW)	MAE (kW)	R^2
VMD-FA-BiQLSTM	4.1211 ± 0.1007	2.0299 ± 0.0248	1.0608 ± 0.0148	0.9618 ± 0.0009
VMD-BiQLSTM	6.1786 ± 0.0828	2.4856 ± 0.0166	1.2927 ± 0.0078	0.9428 ± 0.0008
FA-BiQLSTM	6.2932 ± 0.1367	2.5085 ± 0.0272	1.3101 ± 0.0152	0.9417 ± 0.0013
VMD-FA-QLSTM	8.4500 ± 0.2193	2.9067 ± 0.0376	1.5177 ± 0.0219	0.9217 ± 0.0020
VMD-FA-BiLSTM	10.4132 ± 0.1647	3.2269 ± 0.0255	1.6822 ± 0.0141	0.9035 ± 0.0015
QLSTM	10.5004 ± 0.1789	3.2403 ± 0.0276	1.6904 ± 0.0145	0.9027 ± 0.0017
BiLSTM	13.5060 ± 0.2302	3.6749 ± 0.0313	1.9105 ± 0.0137	0.8749 ± 0.0021
TCN	13.4639 ± 0.1990	3.6692 ± 0.0270	1.9129 ± 0.0130	0.8752 ± 0.0018
Transformer	16.9554 ± 0.2484	4.1176 ± 0.0301	2.1417 ± 0.0188	0.8429 ± 0.0023

Values are mean ± sample standard deviation across five independent runs with seeds 101, 202, 303, 404 and 505. Each run is evaluated over the combined 672 hourly observations. Bold values identify the best mean performance.

Table 9 summarizes performance and run-to-run variability over the combined seasonal intervals. VMD-FA-BiQLSTM achieves an RMSE of 2.0299 ± 0.0248 kW and an MAE of 1.0608 ± 0.0148 kW, together with the lowest mean MSE and highest mean R^2. The matched classical VMD-FA-BiLSTM has an RMSE of 3.2269 ± 0.0255 kW and an MAE of 1.6822 ± 0.0141 kW. The proposed model has the best mean performance across the four metrics, while the smallest standard deviations are distributed among different models and depend

on the metric. The results support improved mean predictive accuracy under the evaluated conditions; they do not establish uniformly lower run-to-run variability.

4.1.7. Computational Cost

Table 10 summarizes the computational cost of the PV forecasting models. Compared with the matched classical VMD-FA-BiLSTM, which retains VMD, FA and the prediction head, VMD-FA-BiQLSTM achieves lower RMSE and MAE in all four evaluated seasons. This consistent improvement supports the value of the quantum temporal module for forecasting accuracy. Table 11 additionally compares model parameter counts for both tasks. Although training takes longer, the reported training time remains on the order of minutes and prediction over each seven-day interval takes less than one second.

Table 10. Measured PV training and seven-day prediction times over four seasonal intervals.

Model	Season	Training (s)	Prediction 7 d (s)
VMD-FA-BiQLSTM	Spring	793.96	0.1398
VMD-FA-BiQLSTM	Summer	804.83	0.1908
VMD-FA-BiQLSTM	Autumn	788.58	0.1273
VMD-FA-BiQLSTM	Winter	770.45	0.1432
VMD-BiQLSTM	Spring	747.25	0.1307
VMD-BiQLSTM	Summer	755.25	0.1752
VMD-BiQLSTM	Autumn	752.74	0.1187
VMD-BiQLSTM	Winter	722.06	0.1345
FA-BiQLSTM	Spring	720.01	0.1238
FA-BiQLSTM	Summer	728.55	0.1692
FA-BiQLSTM	Autumn	732.82	0.1144
FA-BiQLSTM	Winter	703.45	0.1282
VMD-FA-QLSTM	Spring	435.90	0.0748
VMD-FA-QLSTM	Summer	442.47	0.1024
VMD-FA-QLSTM	Autumn	434.12	0.0679
VMD-FA-QLSTM	Winter	424.30	0.0776
VMD-FA-BiLSTM	Spring	49.82	0.0068
VMD-FA-BiLSTM	Summer	49.97	0.0092
VMD-FA-BiLSTM	Autumn	50.18	0.0063
VMD-FA-BiLSTM	Winter	48.39	0.0070
QLSTM	Spring	365.84	0.0599
QLSTM	Summer	373.81	0.0831
QLSTM	Autumn	366.41	0.0545
QLSTM	Winter	357.31	0.0627
BiLSTM	Spring	38.92	0.0057
BiLSTM	Summer	38.14	0.0074
BiLSTM	Autumn	39.83	0.0053
BiLSTM	Winter	37.22	0.0057
TCN	Spring	11.89	0.0059
TCN	Summer	12.26	0.0055
TCN	Autumn	11.91	0.0055
TCN	Winter	11.30	0.0060
Transformer	Spring	14.92	0.0058
Transformer	Summer	16.17	0.0103
Transformer	Autumn	15.81	0.0119
Transformer	Winter	16.02	0.0102

The reported runtime reflects the current implementation, in which quantum circuits are simulated on a classical computer. Repeated circuit evaluation within the recurrent encoder, together with gradient computation during training, introduces additional com-

putational work. For day-ahead forecasting, training can be performed offline and its cost distributed across subsequent uses of the trained model. The short reported prediction time is also small relative to the hourly sampling interval and the day-ahead forecast horizon. Taken together, the consistent accuracy gains and modest absolute prediction time support the additional computation when forecasting accuracy is the primary consideration.

The complete model contains 29,929 parameters for PV forecasting and 33,601 for load forecasting, compared with 68,105 and 80,345 for the matched classical VMD-FA-BiLSTM. This comparison indicates a more compact parameterization in the evaluated configurations. Parameter count and runtime describe different aspects of model cost: the smaller parameter count coexists with additional classical simulation work for the quantum circuits.

Table 11. Model parameter counts for PV and load forecasting.

Model	PV Parameters	Load Parameters
VMD-FA-BiQLSTM	29,929	33,601
VMD-BiQLSTM	29,665	31,977
FA-BiQLSTM	29,905	33,577
VMD-FA-QLSTM	20,089	23,693
VMD-FA-BiLSTM	68,105	80,345
QLSTM	19,813	22,057
BiLSTM	66,305	77,185
TCN	157,249	163,777
Transformer	90,689	93,953

4.2. Electric Load Forecasting

4.2.1. Data Description for Electric Load Forecasting

The load reference uses the public Global Energy Forecasting Competition 2014 (GEFCom2014) dataset at hourly resolution. The supplied actual series cover four seven-day intervals in 2011: spring from 01:00 on 1 March to 00:00 on 8 March; summer from 01:00 on 1 June to 00:00 on 8 June; autumn from 01:00 on 1 September to 00:00 on 8 September; and winter from 01:00 on 1 December to 00:00 on 8 December. Each interval contains 168 points, giving 672 points overall. The feature design uses temperature measurements from 25 stations to derive temperature summaries, heating- and cooling-degree variables, a quadratic temperature term, 20 calendar features, and a 168 h lagged load, yielding 28 external features.

4.2.2. Correlation Analysis

Figure 7 shows Pearson correlations for selected load input features, calculated using Equation (37). The 30-day training window preceding the winter forecast origin contains 720 hourly samples. The target and aligned rolling VMD features with K = 3 are constructed following Section 3.3. The figure displays a subset of the 28 external features used by the model.

Among the selected inputs, the 168 h lagged load has the largest positive correlation with the target (0.48) and is also associated with the medium-frequency mode u2 (0.43), supporting the inclusion of weekly historical information. The hourly sine term is negatively correlated with both the target (−0.44) and u2 (−0.42). The quadratic temperature term has a correlation of 0.29 with the target and 0.39 with the low-frequency mode u1, while the cooling-degree variable is associated with u2 at 0.33. Mean temperature has a target correlation of 0.00 at the displayed precision but different associations with the individual modes. These training-window associations illustrate complementary information across temporal scales and motivate feature weighting; they do not establish causal effects.

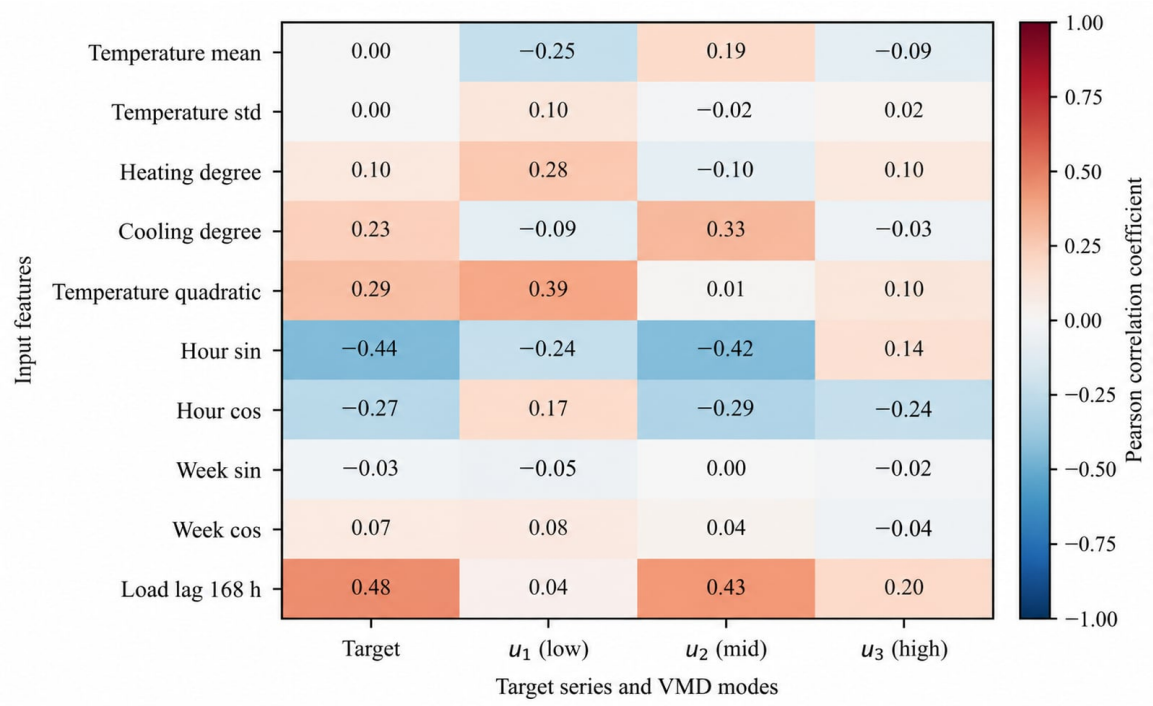


Figure 7. Pearson correlations between selected load input features, the target series, and VMD modes in the 30-day training window preceding the winter forecast origin.

4.2.3. Experimental Setup

Table 12 defines the nine load comparison models. VMD-BiQLSTM, FA-BiQLSTM and VMD-FA-QLSTM independently remove FA, VMD and bidirectionality, respectively, and each ablation is trained independently using the same five seeds and shared load training settings. VMD-FA-BiLSTM replaces only the quantum temporal encoder in the full VMD–FA pipeline. The load QLSTM and BiLSTM baselines omit both VMD and FA, while TCN and Transformer are evaluated as separate architecture types. Their roles in the four panels are the same as in the PV presentation.

All load comparison models use a 72 h history and a 24 h horizon. Recurrent hidden width is 64 per direction where bidirectional; the direct-output decoder has hidden widths 128 and 64 with dropout 0.1. All load models share MSE loss, Adam with initial learning rate 0.0003, batch size 16, up to 35 epochs and gradient-norm clipping at 5. ReduceLROnPlateau monitors validation RMSE, halves the learning rate after two non-improving epochs and uses a minimum of 0.00005. Early stopping uses patience six, and the lowest-validation-RMSE state is restored. The default rolling split uses 365 training days followed by 30 validation days before the test origin; the settings are configurable. Quantum gates use four qubits, one circuit layer and the default.qubit classical simulator.

The load experiments use the same seed controls and four-qubit gate construction described in Section 4.1.3. As defined in Equations (23) and (30), the MLP directly outputs the predicted load, with the corresponding previous-day load included as an input feature. No residual is added to the previous-day load to construct the prediction. Fixed hyperparameters are distinguished from validation-based checkpoint selection and learning-rate reduction. Run configurations additionally record the validation interval and selected epoch.

The load design applies rolling VMD with K = 3 to a 72 h observed window. Low-frequency modes describe slowly varying load levels, medium-frequency modes describe periodic variation, and high-frequency modes represent more rapid changes. Table 13 retains these conceptual roles.

Table 12. Load comparison model configurations.

Method	Configuration
VMD-FA-BiQLSTM	Implemented full architecture: rolling VMD (default K = 3), FA and BiQLSTM, with hidden width 64 per direction, four qubits and one circuit layer per quantum gate. The decoder directly predicts the next-day load, using the corresponding previous-day load as an input feature.
VMD-BiQLSTM	Independently trained ablation retaining VMD and BiQLSTM while removing FA. Uses the shared load training settings and five independent runs with seeds 101, 202, 303, 404 and 505.
FA-BiQLSTM	Independently trained ablation retaining FA and BiQLSTM while removing VMD. Uses the shared load training settings and five independent runs with seeds 101, 202, 303, 404 and 505.
VMD-FA-QLSTM	Independently trained ablation retaining VMD and FA while replacing the bidirectional encoder with a unidirectional QLSTM. Uses the shared load training settings and five independent runs with seeds 101, 202, 303, 404 and 505.
VMD-FA-BiLSTM	Implemented matched classical counterpart: retain VMD, FA and the direct-output decoder, replacing BiQLSTM with a one-layer BiLSTM of hidden width 64 per direction. Data splits, optimizer, validation and stopping rules are inherited from the full-model entry.
QLSTM	Implemented unidirectional QLSTM baseline without VMD or FA. Hidden width 64, four qubits and one circuit layer per gate; uses the shared direct-output day-ahead decoder.
BiLSTM	Implemented one-layer BiLSTM baseline without VMD or FA. Hidden width 64 per direction; uses the shared direct-output day-ahead decoder.
TCN	Separate TCN baseline without VMD or FA. Six residual causal-convolution blocks, each with 64 channels and two convolutions; kernel size 3, dilations 1, 2, 4, 8, 16 and 32, and dropout 0.1.
Transformer	Separate Transformer baseline without VMD or FA. Linear input embedding of width 64, four heads, two encoder layers, feed-forward width 128, GELU activation, dropout 0.1, learned positions for the 72 h history, and mean pooling.

Table 13. Interpretation of the load VMD modes.

Mode	Relative Frequency	Load Interpretation	Model Input
u_1	Low	Baseline load and slow trend	Window-end value
u_2	Medium	Daily cycle and working-day variation	Window-end value
u_3	High	Short-term disturbance and random fluctuation	Window-end value

4.2.4. Sensitivity to the VMD Mode Count

Table 14 compares K = 2, 3, 4 and 5 for VMD-FA-BiQLSTM over the four seasonal seven-day load intervals. Each comparison uses the same observed load values and 168 hourly timestamps as its corresponding seasonal model table. Bold values identify the best result for each metric within each season.

Among the tabulated settings, K = 3 achieves the lowest MSE, RMSE and MAE in all four evaluated seasons, together with the highest or jointly highest R^2 at the reported precision. At K = 3, the seasonal mean RMSEs are 7.8323, 8.8526, 6.9246 and 6.2932 MW in spring, summer, autumn and winter, respectively. Taking the square root of the mean of the four equally sized seasonal MSEs gives RMSE-equivalent values of 7.6033, 7.5388, 7.5819 and 7.6409 MW for K = 2, 3, 4 and 5, respectively. These results support K = 3 as the default within the evaluated intervals, although the relatively small differences do not establish its universal optimality.

Table 14. Load mode-count comparison over four seasonal seven-day intervals.

Season/K	MSE (MW^2)	RMSE (MW)	MAE (MW)	R^2
Spring/2	61.5193	7.8434	6.2191	0.9361
Spring/3	61.3815	7.8323	6.2025	0.9362
Spring/4	61.8929	7.8672	6.2284	0.9357
Spring/5	62.1287	7.8822	6.2499	0.9355
Summer/2	79.2023	8.8996	7.0536	0.9656
Summer/3	78.3930	8.8526	7.0170	0.9659
Summer/4	78.4216	8.8556	7.0187	0.9659
Summer/5	79.9868	8.9435	7.0884	0.9652
Autumn/2	50.0111	7.0719	5.6612	0.9665
Autumn/3	47.9515	6.9246	5.5448	0.9679
Autumn/4	49.5182	7.0369	5.6333	0.9668
Autumn/5	50.5065	7.1068	5.6892	0.9662
Winter/2	40.5079	6.3646	5.0570	0.9535
Winter/3	39.6087	6.2932	4.9968	0.9545
Winter/4	40.1086	6.3331	5.0320	0.9540
Winter/5	40.9092	6.3960	5.0820	0.9531

4.2.5. Feature Attention Weights

The FA module uses a two-layer feedforward scoring network with a tanh activation, followed by Softmax over the external-feature dimension. The score-network input and output widths equal the number of external features: 28 for load. At each historical time step, the resulting nonnegative weights sum to one across features and are applied elementwise before the temporal encoder. The target and VMD channels are concatenated separately; they are not part of this feature-weight normalization. These are feature weights, not attention scores normalized across historical times or seasons.

Table 15 summarizes the attention demonstrations for the sequence labeled VMD-FA-BiQLSTM. Each of the 28 daily forecast origins has 72 historical-hour weight vectors, giving 2016 vectors across four seasons. Daily weights are means over those 72 vectors; seasonal weights average seven daily profiles; the overall column averages the four equally sized seasons. All historical timestamps are at or before the forecast origin T; the forecast horizon begins at T + 1. Every column sums to one over the complete feature set before rounding. A feature row is not required to sum to one across seasons, and displayed four-decimal entries may produce small rounding differences.

Table 15. Overall and seasonal load feature-attention weights.

Feature	Overall	Spring	Summer	Autumn	Winter
Cooling-degree index	0.0657	0.0536	0.0785	0.0855	0.0453
Maximum station temperature	0.0499	0.0437	0.0488	0.0541	0.0530
Load lagged by 168 h	0.0471	0.0448	0.0520	0.0460	0.0455
Quadratic temperature term	0.0403	0.0419	0.0386	0.0387	0.0420
Annual cycle cosine	0.0399	0.0354	0.0432	0.0455	0.0354
Hourly cycle cosine	0.0382	0.0424	0.0362	0.0339	0.0403
Heating-degree index	0.0378	0.0370	0.0362	0.0350	0.0429
Weekly cycle sine	0.0375	0.0375	0.0372	0.0374	0.0379
Mean station temperature	0.0358	0.0345	0.0330	0.0380	0.0376
Weekend-transition indicator	0.0357	0.0374	0.0321	0.0349	0.0386
Monday indicator	0.0354	0.0365	0.0358	0.0353	0.0338
Weekday one-hot indicator 0	0.0348	0.0348	0.0351	0.0324	0.0367
Annual cycle sine	0.0347	0.0348	0.0333	0.0344	0.0364
Weekly cycle cosine	0.0343	0.0342	0.0366	0.0336	0.0330
Friday indicator	0.0329	0.0333	0.0331	0.0328	0.0323

Table 15. *Cont.*

Feature	Overall	Spring	Summer	Autumn	Winter
Weekday one-hot indicator 6	0.0327	0.0391	0.0306	0.0278	0.0331
Weekday one-hot indicator 2	0.0327	0.0339	0.0324	0.0308	0.0336
Hourly cycle sine	0.0324	0.0339	0.0307	0.0304	0.0347
Station temperature standard deviation	0.0323	0.0320	0.0314	0.0315	0.0342
Hourly second harmonic cosine	0.0311	0.0321	0.0301	0.0294	0.0327
Weekday one-hot indicator 5	0.0310	0.0333	0.0295	0.0297	0.0315
Weekday one-hot indicator 4	0.0310	0.0311	0.0301	0.0311	0.0315
Weekday one-hot indicator 1	0.0308	0.0317	0.0289	0.0309	0.0317
Weekday one-hot indicator 3	0.0305	0.0306	0.0313	0.0310	0.0293
Hourly second harmonic sine	0.0297	0.0309	0.0291	0.0280	0.0308
Previous-day weekend indicator	0.0297	0.0328	0.0291	0.0274	0.0294
Weekend indicator	0.0286	0.0292	0.0296	0.0267	0.0288
Minimum station temperature	0.0276	0.0276	0.0273	0.0277	0.0280
Sum	1.0000	1.0000	1.0000	1.0000	1.0000

For the load demonstration, the largest overall allocation is assigned to the cooling-degree index (0.0657), followed by maximum station temperature (0.0499) and the 168 h lagged load (0.0471). The cooling-degree weight is 0.0785 in summer, 0.0855 in autumn and 0.0453 in winter, whereas the heating-degree index receives a winter weight of 0.0429. Temperature summaries, weekly-load history and calendar indicators must be interpreted together rather than treating the largest weight as a standalone causal explanation.

4.2.6. Results and Ablation Analysis

Each load season is presented using its own 168-point metric table and four-panel seven-day forecast group, with the model grouping defined at the start of Section 4.

Spring: The seven-day comparison covers 1 March 2011, 01:00 to 8 March 2011, 00:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 16 and Figure 8.

Table 16. Spring electric load forecasting performance across five independent runs.

Model	MSE (MW^2)	RMSE (MW)	MAE (MW)	R^2
VMD-FA-BiQLSTM	61.3815 ± 3.3510	7.8323 ± 0.2121	6.2025 ± 0.1741	0.9362 ± 0.0035
VMD-BiQLSTM	82.5029 ± 1.3673	9.0829 ± 0.0756	7.7179 ± 0.0828	0.9143 ± 0.0014
FA-BiQLSTM	81.6204 ± 4.3483	9.0319 ± 0.2381	7.6799 ± 0.2102	0.9152 ± 0.0045
VMD-FA-QLSTM	105.8414 ± 4.2227	10.2863 ± 0.2040	9.1215 ± 0.1704	0.8900 ± 0.0044
VMD-FA-BiLSTM	136.2245 ± 6.2378	11.6691 ± 0.2659	10.6513 ± 0.2495	0.8585 ± 0.0065
QLSTM	137.1971 ± 5.7203	11.7111 ± 0.2423	10.6817 ± 0.2219	0.8575 ± 0.0059
BiLSTM	165.6264 ± 3.2473	12.8691 ± 0.1262	11.9845 ± 0.1139	0.8279 ± 0.0034
TCN	165.0206 ± 12.9348	12.8381 ± 0.5053	11.9710 ± 0.4731	0.8286 ± 0.0134
Transformer	207.9424 ± 3.9468	14.4197 ± 0.1370	13.5985 ± 0.1444	0.7840 ± 0.0041

Figure 8 presents the seven-day spring load forecast comparison. In Table 16, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 7.8323 ± 0.2121 MW and its MAE is 6.2025 ± 0.1741 MW; the mean MSE and R^2 are 61.3815 MW^2 and 0.9362, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 9.0829 and 9.0319 MW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 10.2863 and 11.6691 MW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

Summer: The seven-day comparison covers 01 June 2011, 01:00 to 08 June 2011, 00:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 17 and Figure 9.

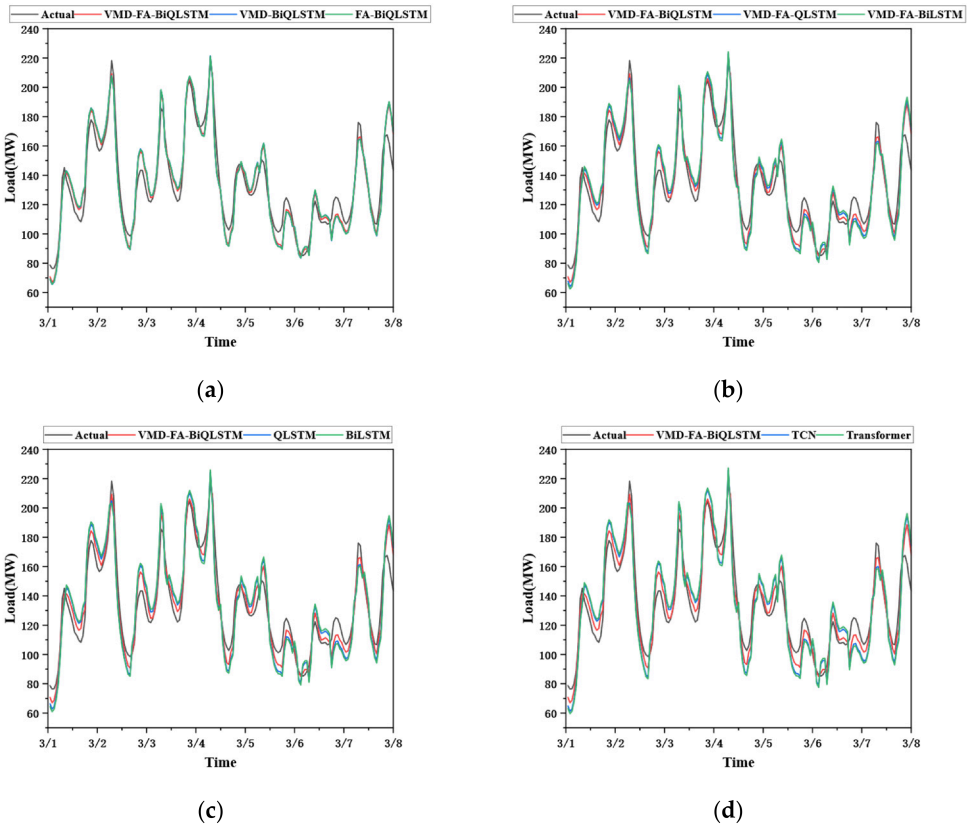


Figure 8. Spring seven-day electric load prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 17. Summer electric load forecasting performance across five independent runs.

Model	MSE (MW^2)	RMSE (MW)	MAE (MW)	R^2
VMD-FA-BiQLSTM	78.3930 ± 3.1253	8.8526 ± 0.1748	7.0170 ± 0.1469	0.9659 ± 0.0014
VMD-BiQLSTM	101.3257 ± 5.5592	10.0630 ± 0.2780	8.4939 ± 0.2375	0.9559 ± 0.0024
FA-BiQLSTM	100.1592 ± 1.9311	10.0076 ± 0.0963	8.4588 ± 0.0966	0.9565 ± 0.0008
VMD-FA-QLSTM	122.9325 ± 4.6555	11.0859 ± 0.2096	9.7524 ± 0.1830	0.9465 ± 0.0020
VMD-FA-BiLSTM	161.0659 ± 3.2686	12.6906 ± 0.1293	11.4891 ± 0.1196	0.9300 ± 0.0014
QLSTM	158.4209 ± 5.8464	12.5848 ± 0.2316	11.3965 ± 0.2115	0.9311 ± 0.0025
BiLSTM	183.1029 ± 6.2443	13.5300 ± 0.2315	12.5136 ± 0.2154	0.9204 ± 0.0027
TCN	195.0164 ± 8.7442	13.9620 ± 0.3118	12.8912 ± 0.2829	0.9152 ± 0.0038
Transformer	241.0721 ± 10.5451	15.5235 ± 0.3398	14.5406 ± 0.3074	0.8952 ± 0.0046

Figure 9 presents the seven-day summer load forecast comparison. In Table 17, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 8.8526 ± 0.1748 MW and its MAE is 7.0170 ± 0.1469 MW; the mean MSE and R^2 are 78.3930 MW^2 and 0.9659, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 10.0630 and 10.0076 MW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 11.0859 and 12.6906 MW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

Autumn: The seven-day comparison covers 01 September 2011, 01:00 to 08 September 2011, 00:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 18 and Figure 10.

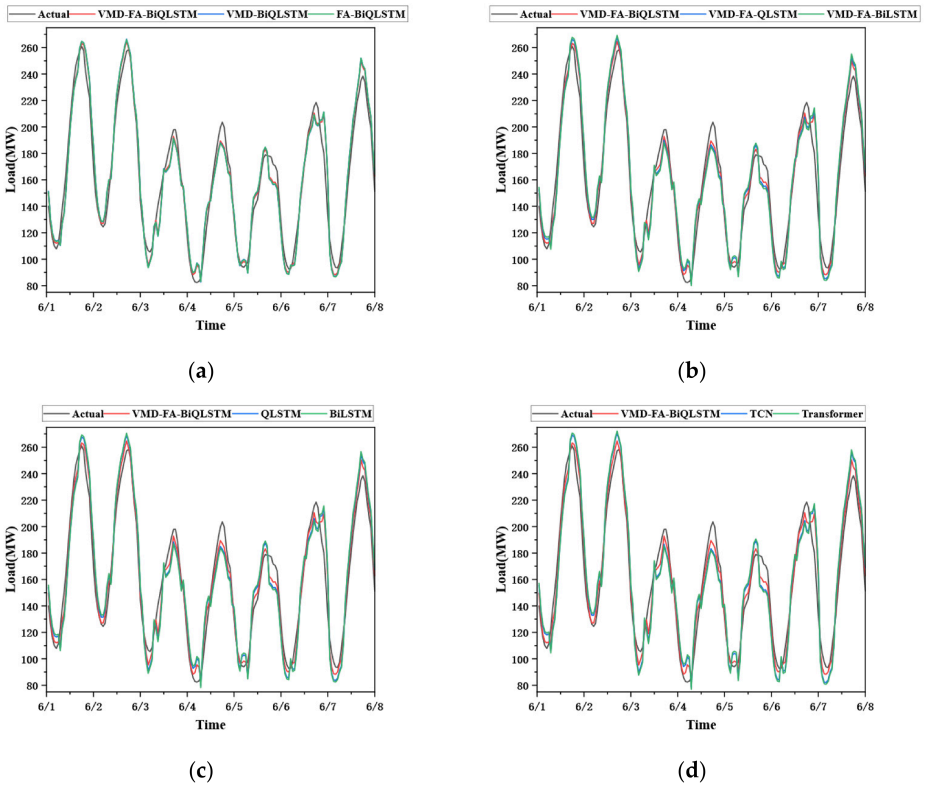


Figure 9. Summer seven-day electric load prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 18. Autumn electric load forecasting performance across five independent runs.

Model	MSE (MW^2)	RMSE (MW)	MAE (MW)	R^2
VMD-FA-BiQLSTM	47.9515 ± 0.6970	6.9246 ± 0.0504	5.5448 ± 0.0490	0.9679 ± 0.0005
VMD-BiQLSTM	68.2966 ± 3.3949	8.2621 ± 0.2051	7.1153 ± 0.1713	0.9543 ± 0.0023
FA-BiQLSTM	67.9833 ± 2.4379	8.2441 ± 0.1482	7.0991 ± 0.1360	0.9545 ± 0.0016
VMD-FA-QLSTM	89.8549 ± 2.8514	9.4782 ± 0.1493	8.5186 ± 0.1331	0.9398 ± 0.0019
VMD-FA-BiLSTM	117.6598 ± 3.6655	10.8460 ± 0.1701	9.9951 ± 0.1492	0.9212 ± 0.0025
QLSTM	122.5827 ± 5.5922	11.0694 ± 0.2540	10.2187 ± 0.2215	0.9179 ± 0.0037
BiLSTM	153.1339 ± 11.5493	12.3678 ± 0.4645	11.6371 ± 0.4428	0.8975 ± 0.0077
TCN	152.2839 ± 2.6443	12.3400 ± 0.1070	11.5875 ± 0.0994	0.8981 ± 0.0018
Transformer	195.0816 ± 5.0323	13.9662 ± 0.1801	13.2840 ± 0.1751	0.8694 ± 0.0034

Figure 10 presents the seven-day autumn load forecast comparison. In Table 18, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 6.9246 ± 0.0504 MW and its MAE is 5.5448 ± 0.0490 MW; the mean MSE and R^2 are 47.9515 MW^2 and 0.9679, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 8.2621 and 8.2441 MW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 9.4782 and 10.8460 MW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

Winter: The seven-day comparison covers 1 December 2011, 01:00 to 8 December 2011, 00:00 and contains 168 hourly observations. The corresponding metrics and forecast curves are presented in Table 19 and Figure 11.

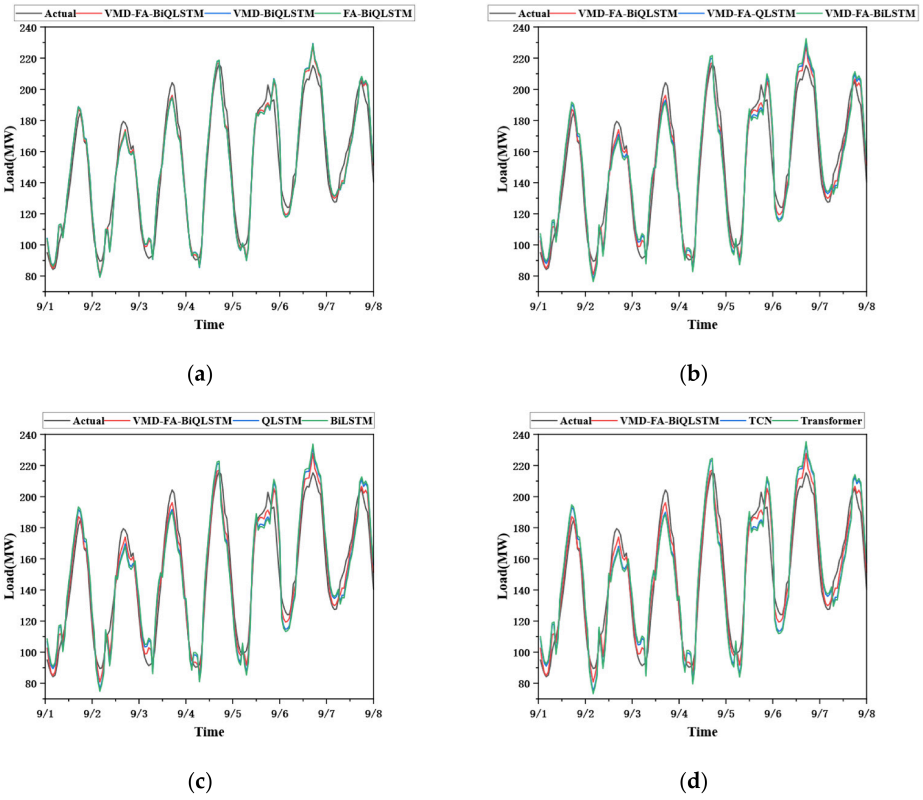


Figure 10. Autumn seven-day electric load prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 19. Winter electric load forecasting performance across five independent runs.

Model	MSE (MW^2)	RMSE (MW)	MAE (MW)	R^2
VMD-FA-BiQLSTM	39.6087 ± 0.9764	6.2932 ± 0.0775	4.9968 ± 0.0570	0.9545 ± 0.0011
VMD-BiQLSTM	57.2590 ± 3.5622	7.5640 ± 0.2350	6.5163 ± 0.2142	0.9343 ± 0.0041
FA-BiQLSTM	55.7267 ± 1.0635	7.4648 ± 0.0711	6.4332 ± 0.0634	0.9361 ± 0.0012
VMD-FA-QLSTM	79.1925 ± 2.9288	8.8978 ± 0.1644	8.0230 ± 0.1584	0.9091 ± 0.0034
VMD-FA-BiLSTM	105.4640 ± 3.8148	10.2682 ± 0.1854	9.5190 ± 0.1712	0.8790 ± 0.0044
QLSTM	107.0149 ± 3.1855	10.3439 ± 0.1537	9.5806 ± 0.1460	0.8772 ± 0.0037
BiLSTM	133.0054 ± 6.1516	11.5303 ± 0.2675	10.8921 ± 0.2550	0.8474 ± 0.0071
TCN	135.6698 ± 9.2769	11.6423 ± 0.3963	10.9635 ± 0.3657	0.8443 ± 0.0106
Transformer	176.8156 ± 4.2153	13.2965 ± 0.1577	12.6935 ± 0.1525	0.7971 ± 0.0048

Figure 11 presents the seven-day winter load forecast comparison. In Table 19, VMD-FA-BiQLSTM has the lowest mean MSE, RMSE and MAE and the highest mean R^2 among the nine models. Its RMSE is 6.2932 ± 0.0775 MW and its MAE is 4.9968 ± 0.0570 MW; the mean MSE and R^2 are 39.6087 MW^2 and 0.9545, respectively. VMD-BiQLSTM and FA-BiQLSTM have mean RMSEs of 7.5640 and 7.4648 MW. The unidirectional VMD-FA-QLSTM and matched classical VMD-FA-BiLSTM have mean RMSEs of 8.8978 and 10.2682 MW, respectively. Panels (c) and (d) provide the QLSTM/BiLSTM and TCN/Transformer comparisons, respectively.

Values are mean ± sample standard deviation across five independent runs with seeds 101, 202, 303, 404 and 505. Each run is evaluated over the combined 672 hourly observations. Bold values identify the best mean performance.

Table 20 summarizes performance and run-to-run variability over the combined seasonal intervals. VMD-FA-BiQLSTM achieves an RMSE of 7.5385 ± 0.0771 MW and an MAE of 5.9403 ± 0.0519 MW, together with the lowest mean MSE and highest mean R^2. The matched classical VMD-FA-BiLSTM has an RMSE of 11.4060 ± 0.0974 MW and an MAE of 10.4136 ± 0.0878 MW. The proposed model has the best mean performance across the four metrics, while the smallest standard deviations are distributed among different models and depend on the metric. The results support improved mean predictive accuracy under the evaluated conditions; they do not establish uniformly lower run-to-run variability.

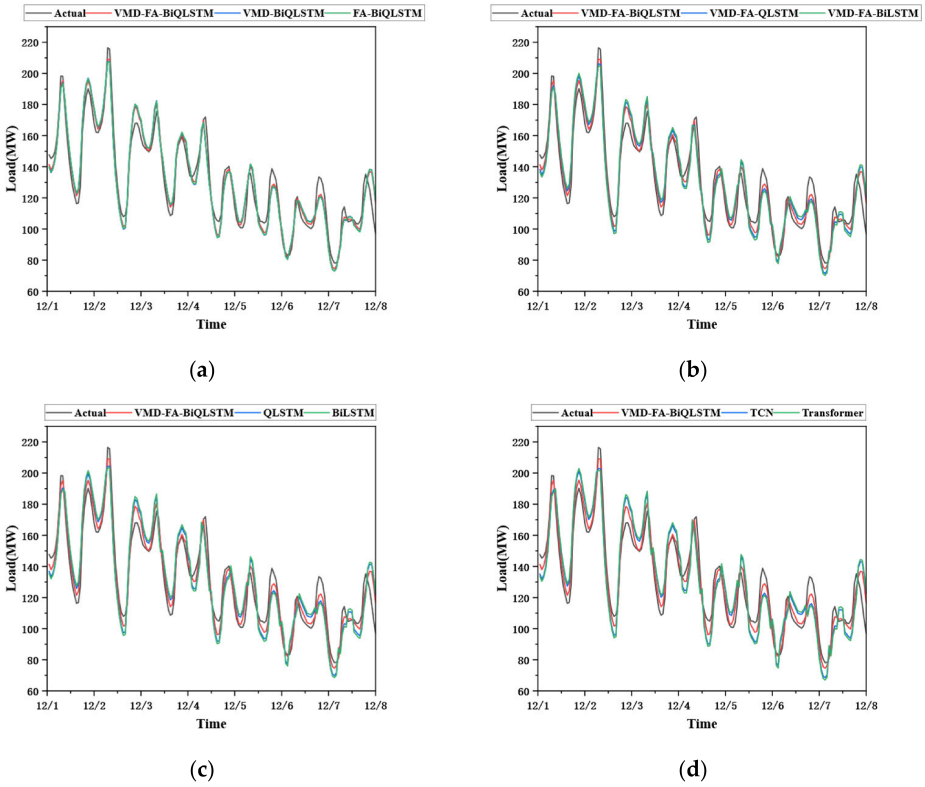


Figure 11. Winter seven-day electric load prediction comparisons. (**a**) Comparison with VMD-BiQLSTM and FA-BiQLSTM; (**b**) comparison with VMD-FA-QLSTM and VMD-FA-BiLSTM; (**c**) comparison with BiLSTM and QLSTM; (**d**) comparison with TCN and Transformer.

Table 20. Overall load forecasting performance across five independent runs.

Model	MSE (MW^2)	RMSE (MW)	MAE (MW)	R^2
VMD-FA-BiQLSTM	56.8337 ± 1.1607	7.5385 ± 0.0771	5.9403 ± 0.0519	0.9633 ± 0.0007
VMD-BiQLSTM	77.3460 ± 0.8277	8.7946 ± 0.0470	7.4609 ± 0.0550	0.9501 ± 0.0005
FA-BiQLSTM	76.3724 ± 1.0194	8.7390 ± 0.0582	7.4177 ± 0.0516	0.9507 ± 0.0007
VMD-FA-QLSTM	99.4553 ± 1.2140	9.9726 ± 0.0610	8.8539 ± 0.0538	0.9358 ± 0.0008
VMD-FA-BiLSTM	130.1036 ± 2.2223	11.4060 ± 0.0974	10.4136 ± 0.0878	0.9161 ± 0.0014
QLSTM	131.3039 ± 3.6014	11.4579 ± 0.1565	10.4694 ± 0.1460	0.9153 ± 0.0023
BiLSTM	158.7172 ± 3.4191	12.5977 ± 0.1352	11.7568 ± 0.1321	0.8976 ± 0.0022
TCN	161.9977 ± 3.4438	12.7273 ± 0.1351	11.8533 ± 0.1266	0.8955 ± 0.0022
Transformer	205.2279 ± 4.8532	14.3250 ± 0.1689	13.5291 ± 0.1538	0.8676 ± 0.0031

4.2.7. Computational Cost

Table 21 summarizes the computational cost of the load forecasting models. The complete model achieves lower RMSE and MAE than the matched classical VMD-FA-BiLSTM in every evaluated season, while retaining the same VMD, FA and prediction-head structure. These results support the contribution of the quantum temporal module to load forecasting accuracy. Table 11 also shows that the complete load model uses 33,601 parameters, compared with 80,345 for the matched classical counterpart. The reported training time is on the order of minutes, and prediction over each seven-day interval takes less than one second.

Table 21. Seasonal load timing.

Model	Season	Training (s)	Prediction 7 d (s)
VMD-FA-BiQLSTM	Spring	405.56	0.2224
VMD-FA-BiQLSTM	Summer	356.48	0.1944
VMD-FA-BiQLSTM	Autumn	283.43	0.1948
VMD-FA-BiQLSTM	Winter	419.10	0.1934
VMD-BiQLSTM	Spring	363.42	0.3154
VMD-BiQLSTM	Summer	294.67	0.2543
VMD-BiQLSTM	Autumn	210.34	0.1453
VMD-BiQLSTM	Winter	395.43	0.2341
FA-BiQLSTM	Spring	323.22	0.2361
FA-BiQLSTM	Summer	242.98	0.2046
FA-BiQLSTM	Autumn	165.41	0.1823
FA-BiQLSTM	Winter	366.77	0.2111
VMD-FA-QLSTM	Spring	212.56	0.1212
VMD-FA-QLSTM	Summer	173.67	0.0952
VMD-FA-QLSTM	Autumn	146.56	0.0965
VMD-FA-QLSTM	Winter	219.56	0.0934
VMD-FA-BiLSTM	Spring	22.45	0.0145
VMD-FA-BiLSTM	Summer	15.28	0.0098
VMD-FA-BiLSTM	Autumn	15.56	0.0091
VMD-FA-BiLSTM	Winter	24.76	0.0092
QLSTM	Spring	167.67	0.1380
QLSTM	Summer	124.78	0.1032
QLSTM	Autumn	87.78	0.0914
QLSTM	Winter	188.37	0.1034
BiLSTM	Spring	14.18	0.0145
BiLSTM	Summer	15.13	0.0156
BiLSTM	Autumn	9.67	0.0067
BiLSTM	Winter	17.23	0.0101
TCN	Spring	46.54	0.0268
TCN	Summer	35.12	0.0223
TCN	Autumn	34.64	0.0268
TCN	Winter	45.26	0.0256
Transformer	Spring	43.78	0.0123
Transformer	Summer	24.94	0.0145
Transformer	Autumn	37.90	0.0089
Transformer	Winter	43.12	0.0134

As in the PV experiments, the quantum circuits are evaluated through classical simulation, which adds circuit-evaluation and training-gradient computation to the recurrent modeling process. This implementation cost should be considered alongside the consistent improvement in forecasting accuracy. With offline training and reuse of the trained model, the training cost can be spread across repeated forecasts, while the reported prediction

time remains small relative to the hourly sampling interval and the day-ahead forecast horizon. The results therefore support a reasonable balance between forecasting accuracy and computational cost for applications that prioritize accurate day-ahead load forecasts.

5. Conclusions

This study presents a multiscale quantum-enhanced framework for day-ahead PV power and electric-load forecasting, illustrated through seasonal reference comparisons. The main conclusions are as follows:

(1) The case studies indicate that the proposed framework forms a clear multiscale representation of the forecasting targets. Low-frequency VMD modes mainly track baseline evolution and slow trends, medium-frequency modes describe periodic and local variations, and high-frequency modes capture short-term disturbances and irregular fluctuations. Feature attention emphasizes the most relevant external variables, while bidirectional quantum temporal modeling complements the decomposition by learning nonlinear dependencies within the observed historical sequence.

(2) For the PV comparisons, the five-run mean metrics over the combined 672 hourly observations are an MSE of 4.1211 kW^2, RMSE of 2.0299 kW, MAE of 1.0608 kW, and R^2 of 0.9618. The corresponding sample standard deviations are 0.1007 kW^2, 0.0248 kW, 0.0148 kW, and 0.0009, respectively. The proposed model achieves the best mean performance on all four metrics among the nine models, although it does not have the smallest standard deviation for every metric. The seasonal and ablation results support complementary contributions from multiscale representation, adaptive feature weighting, bidirectionality and quantum temporal modeling. The parameter comparison also shows a smaller model than the matched classical counterpart. In the tabulated PV mode-count comparison, $K = 3$ performs best on all four metrics in each evaluated season.

(3) For the load comparisons, the five-run mean metrics over the combined 672 hourly observations are an MSE of 56.8337 MW^2, RMSE of 7.5385 MW, MAE of 5.9403 MW, and R^2 of 0.9633. The corresponding sample standard deviations are 1.1607 MW^2, 0.0771 MW, 0.0519 MW, and 0.0007, respectively. The proposed model achieves the best mean performance on all four metrics among the nine models, although it does not have the smallest standard deviation for every metric. It also uses fewer parameters than the matched classical counterpart. These results support a common modeling approach for supply- and demand-side day-ahead forecasting within the evaluated datasets.

(4) The limited dataset coverage and seven-day seasonal intervals do not establish generalization across regions, extreme conditions or changing operating environments. Future work will extend evaluation to multiple sites and complete annual cycles, while retaining independent repeated runs and matched within-task training budgets. For PV deployment, the observation-based nighttime zeroing used in this offline evaluation must be replaced by a rule available at the forecast origin. The five-year historical-average meteorological inputs represent recurring seasonal conditions but may not capture day-specific weather changes or extreme events. Future work will quantify how uncertainty in these inputs affects PV and load forecasts and compare them with weather forecasts available at the forecast issue time. The present results characterize predictive performance and computational cost in a classical simulation environment. Further work will optimize circuit evaluation and batching and assess end-to-end latency, including rolling VMD, to improve computational efficiency within this implementation.

Author Contributions: C.Z.: conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, and writing—original draft. J.Z.: conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, and writing—original draft. Z.W.: validation, data curation, visualization, and writing—review and editing. R.L.: validation, visualization, and writing—review and editing. L.W.: methodology, validation, project administration, supervision, and writing—review and editing. Z.L.: validation, resources, and writing—review and editing. M.L.: conceptualization, methodology, resources, project administration, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported partially by Beijing Natural Science Foundation (L2609008), 2022 Strategic Research Key Project of Science and Technology Commission of the Ministry of Education, Beijing Science and Technology Project (Z211100004621010), the project of China Three Gorges Corporation named key technologies of intelligent joint regulation and operation with grid connected friendly in power station group of wind, solar photovoltaic and energy storage (WWKY-2021-0173), Huaneng Group Headquarters Science and Technology Project (HNKJ20-H88), State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (LAPS25002), the Fundamental Research Funds for the Central Universities (2025MS048) and the NCEPU "Double First-Class" Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The DKASC Alice Springs data used in this study are publicly available through the Desert Knowledge Australia Solar Centre (DKA Solar Centre) data portal at https://www.dkasolarcentre.com.au/download?location=alice-springs (accessed on 15 June 2026). The GEFCom2014 dataset is publicly available at https://ars.els-cdn.com/content/image/1-s2.0-S0169207016000133-mmc1.zip (accessed on 15 June 2026). The code used in this study is available from the corresponding authors upon reasonable request.

Conflicts of Interest: The authors declare that this study received funding from China Huaneng Group Co., Ltd. (China) (HNKJ20-H88). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Appendix A

Table A1 summarizes the abbreviations, mathematical symbols, and evaluation metrics used in this study.

Table A1. Nomenclature of abbreviations and symbols used in this study.

Abbreviation/Symbol	Definition	Unit
	Abbreviations	
PV	Photovoltaic	—
VMD	Variational mode decomposition	—
FA	Feature attention	—
LSTM	Long short-term memory	—
QLSTM	Quantum long short-term memory	—
BiLSTM	Bidirectional long short-term memory	—
BiQLSTM	Bidirectional quantum long short-term memory	—
VQC	Variational quantum circuit	—
MLP	Multilayer perceptron	—
TCN	Temporal convolutional network	—
DKASC	Desert Knowledge Australia Solar Centre	—
GEFCom2014	Global Energy Forecasting Competition 2014	—
RNN	Recurrent neural network	—
CNN	Convolutional neural network	—
GRU	Gated recurrent unit	—

Table A1. *Cont.*

Abbreviation/Symbol	Definition	Unit
ARIMA	Autoregressive integrated moving average	—
ARMA	Autoregressive moving average	—
EMD	Empirical mode decomposition	—
EEMD	Ensemble empirical mode decomposition	—
SD	Sample standard deviation	—
	Variables and units	
$f(t)$	Original signal to be decomposed	Same unit as the decomposed target
$u_k(t)$	k-th VMD mode	Same unit as f(t)
$r(t)$	VMD reconstruction residual	Same unit as f(t)
K	Number of VMD modes	Dimensionless
ω	Angular frequency in the VMD formulation	rad per time unit; normalized in discrete VMD
ω_k	Center angular frequency of the k-th VMD mode	rad per time unit; normalized in discrete VMD
α	VMD penalty parameter	Dimensionless in the discrete implementation
τ	VMD dual-ascent step size	Dimensionless in the discrete implementation
$\lambda(t)$	Numerical dual variable in the VMD optimization	—
t	Time variable in VMD; discrete time index in forecasting	Normalized time in VMD; dimensionless index in forecasting
T	Forecast-origin index	Dimensionless
L	Historical-window length (72 hourly steps)	Dimensionless (number of time steps)
H	Forecast horizon (24 hourly steps)	Dimensionless (number of time steps)
y_t	Observed target value	kW (PV) or MW (load), before normalization
$\hat{y}_t$	Predicted target value	kW (PV) or MW (load), before normalization
$\overline{y}$	Mean observed target value	kW (PV) or MW (load), before normalization
$y_{\square,j}$	Observed target at step j in batch sample b	Dimensionless during normalized training
$\hat{y}_{\square,j}$	Predicted target at step j in batch sample b	Dimensionless during normalized training
v_t	VMD feature vector	Same unit as the VMD input; dimensionless after scaling
$v_k(t)$	k-th terminal VMD feature at time t	Same unit as the VMD input; dimensionless after scaling
z_t	External-feature vector	Dimensionless after normalization
$\tilde{Z}_t$	Attention-weighted external-feature vector	Dimensionless after normalization
z_{T+j}^{future}	External covariate vector for step T+j, available at the forecast origin	Dimensionless after normalization
x_t	Temporal-encoder input vector	Dimensionless after normalization
X	Historical input window	Dimensionless after normalization
$reverse(X)$	Historical input window in reverse order	Dimensionless after normalization
e_t	Feature-attention score vector	Dimensionless
a_t	Feature-attention weight vector	Dimensionless
$a_{t,j}$	Attention weight of feature j at time t	Dimensionless
d	Number of external features (11 for PV; 28 for load)	Dimensionless
h_t	Recurrent hidden state	Dimensionless
C_t	Recurrent cell state	Dimensionless
$\tilde{C}_t$	Candidate cell state	Dimensionless
f_t	Forget-gate activation	Dimensionless
i_t	Input-gate activation	Dimensionless
o_t	Output-gate activation	Dimensionless
e	Fused bidirectional historical context	Dimensionless
d_j	Decoder input vector at forecast step j	Dimensionless
W	Learnable weight matrix; subscript identifies the layer or gate	Dimensionless in the normalized network representation
b	Learnable bias vector; subscript identifies the layer or gate	Dimensionless in the normalized network representation
x_{norm}	Min-max-normalized scalar value	Dimensionless
x_{orig}	Original-scale scalar value	Unit of the corresponding physical variable

Table A1. *Cont.*

Abbreviation/Symbol	Definition	Unit
x_{min}	Minimum fitted on the training interval	Unit of the corresponding physical variable
x_{max}	Maximum fitted on the training interval	Unit of the corresponding physical variable
q_i	i-th compressed quantum-circuit input	Dimensionless
θ_i	Encoded quantum-gate rotation angle	rad
N_q	Number of qubits (4 in this study)	Dimensionless
B	Training batch size	Dimensionless
n	Number of evaluation observations	Dimensionless
L_{MSE}	Training mean squared error loss	Dimensionless for normalized targets
MSE	Mean squared prediction error on the original scale	kW^2 (PV); MW^2 (load)
RMSE	Root mean squared prediction error on the original scale	kW (PV); MW (load)
MAE	Mean absolute prediction error on the original scale	kW (PV); MW (load)
R^2	Coefficient of determination	Dimensionless
r_{XY}	Pearson correlation coefficient	Dimensionless
X_i	i-th observation of the first variable in the Pearson correlation	Unit of the corresponding variable
Y_i	i-th observation of the second variable in the Pearson correlation	Unit of the corresponding variable
$\overline{X}$	Mean of the first variable in the Pearson correlation	Same unit as X_i
$\overline{Y}$	Mean of the second variable in the Pearson correlation	Same unit as Y_i

References

1. Saeed, M.H.; Fangzong, W.; Kalwar, B.A.; Iqbal, S. A Review on Microgrids' Challenges & Perspectives. *IEEE Access* **2021**, *9*, 166502–166517. [CrossRef]
2. Wang, J.; Zhang, Z.; Xu, W.; Li, Y.; Niu, G. Short-Term Photovoltaic Power Forecasting Using a Bi-LSTM Neural Network Optimized by Hybrid Algorithms. *Sustainability* **2025**, *17*, 5277. [CrossRef]
3. Zhao, Q.; Wang, S.; Chen, Y.; Liu, J.; Sun, Y.; Su, T.; Li, N.; Fang, J. A hybrid framework for short-term load forecasting based on optimized InMetra Boost and BiLSTM. *Energy* **2025**, *328*, 136582. [CrossRef]
4. Rabbi, M.F. Cross-framework hybrid artificial intelligence for high-penetration renewable energy integration: Multi-regional forecasting and adaptive control. *Appl. Energy* **2025**, *401*, 126834. [CrossRef]
5. Habibi, M.R.; Golestan, S.; Wu, Y.; Guerrero, J.M.; Vasquez, J.C. Electrical load forecasting in power systems based on quantum computing using time series-based quantum artificial intelligence. *Sci. Rep.* **2025**, *15*, 7429. [CrossRef]
6. Kumar, J.; Saxena, D.; Singh, A.K.; Vasilakos, A.V. A Quantum Controlled-NOT Neural Network-Based Load Forecast and Management Model for Smart Grid. *IEEE Syst. J.* **2023**, *17*, 5714–5725. [CrossRef]
7. Thota, T.; Kurumthottam, A.B.; Kumar, S.R.; Mishra, J. An Enhanced Time-Series-Based Solar Power Forecast for Microgrid System. *IEEE Access* **2025**, *13*, 144785–144797. [CrossRef]
8. Wynn, S.L.L.; Boonraksa, T.; Boonraksa, P.; Pinthurat, W.; Marungsri, B. Decentralized Energy Management System in Microgrid Considering Uncertainty and Demand Response. *Electronics* **2023**, *12*, 237. [CrossRef]
9. Abdel-Nasser, M.; Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput. Appl.* **2017**, *31*, 2727–2740. [CrossRef]
10. Joseph, L.P.; Deo, R.C.; Prasad, R.; Salcedo-Sanz, S.; Raj, N.; Soar, J. Near real-time wind speed forecast model with bidirectional LSTM networks. *Renew. Energy* **2023**, *204*, 39–58. [CrossRef]
11. Su, Y.; Wang, Z.; Dong, Z.; Hua, X.; Ye, T.; Song, Z.; Shao, Y. Frequency-aware ultra-short-term wind power forecasting using CEEMDAN–VMD–SE and Transformer–GRU networks. *Energy* **2025**, *338*, 138715. [CrossRef]
12. Rubasinghe, O.; Zhang, X.; Chau, T.K.; Chow, Y.H.; Fernando, T.; Iu, H.H.-C. A Novel Sequence to Sequence Data Modelling Based CNN-LSTM Algorithm for Three Years Ahead Monthly Peak Load Forecasting. *IEEE Trans. Power Syst.* **2024**, *39*, 1932–1947. [CrossRef]
13. Alsamraee, S.A.; Khanna, S. A hybrid transformer–TCN–GRU based model for thermal load forecasting of a large university campus. *Energy* **2026**, *344*, 140114. [CrossRef]

14. Kumar, D.; Meena, H.K.; Jain, P. Attention mechanism based solar and load joint forecasting framework with Bayesian optimization. *Comput. Electr. Eng.* **2026**, *138*, 111327. [CrossRef]
15. Su, F.; Tang, H.; Yang, S.; Zhang, T.; Tan, Q. Short-term load forecasting by a hybrid attention scheme: Multi-feature attention and context awareness. *Int. J. Electr. Power Energy Syst.* **2025**, *172*, 111065. [CrossRef]
16. Aswanuwath, L.; Pannakkong, W.; Buddhakulsomsiri, J.; Karnjana, J.; Huynh, V.-N. A Hybrid Model of VMD-EMD-FFT, Similar Days Selection Method, Stepwise Regression, and Artificial Neural Network for Daily Electricity Peak Load Forecasting. *Energies* **2023**, *16*, 1860. [CrossRef]
17. Al Mamun, A.; Sohel, M.; Mohammad, N.; Sunny, S.H.; Dipta, D.R.; Hossain, E. A Comprehensive Review of the Load Forecasting Techniques Using Single and Hybrid Predictive Models. *IEEE Access* **2020**, *8*, 134911–134939. [CrossRef]
18. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [CrossRef]
19. Zeng, W.; Li, J.; Sun, C.; Cao, L.; Tang, X.; Shu, S.; Zheng, J. Ultra Short-Term Power Load Forecasting Based on Similar Day Clustering and Ensemble Empirical Mode Decomposition. *Energies* **2023**, *16*, 1989. [CrossRef]
20. Lv, L.; Wu, Z.; Zhang, J.; Zhang, L.; Tan, Z.; Tian, Z. A VMD and LSTM Based Hybrid Model of Load Forecasting for Power Grid Security. *IEEE Trans. Ind. Inform.* **2022**, *18*, 6474–6482. [CrossRef]
21. Lin, J.; Ma, J.; Zhu, J.; Cui, Y. Short-term load forecasting based on LSTM networks considering attention mechanism. *Int. J. Electr. Power Energy Syst.* **2022**, *137*, 107818. [CrossRef]
22. Shi, J.; Teh, J.; Lai, C.-M. Wind power prediction based on improved self-attention mechanism combined with Bi-directional Temporal Convolutional Network. *Energy* **2025**, *322*, 135666. [CrossRef]
23. Huang, X.; Li, Q.; Tai, Y.; Chen, Z.; Liu, J.; Shi, J.; Liu, W. Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. *Energy* **2022**, *246*, 123403. [CrossRef]
24. Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; McClean, J.R.; Mitarai, K.; Yuan, X.; Cincio, L.; et al. Variational quantum algorithms. *Nat. Rev. Phys.* **2021**, *3*, 625–644. [CrossRef]
25. Wang, S.; Zhang, Z. Short-Term Multiple Load Forecasting Model of Regional Integrated Energy System Based on QWGRU-MTL. *Energies* **2021**, *14*, 6555. [CrossRef]
26. Zhang, Q.; Zhou, Y.; Huang, S.; Guo, C.; Xiong, L.; Li, S.; Li, Y.; Cao, Y. Short-Term Power Load Forecasting Based on the Quantum Genetic Algorithm Artificial Recurrent Memory Network. *Electronics* **2025**, *14*, 2417. [CrossRef]
27. Sahay, K.B.; Singh, M. Quantum-Inspired Deep Learning Model for Short-Term Load Forecasting in Smart Grid Energy Management. *IEEE Access* **2026**, *14*, 32702–32716. [CrossRef]
28. Zhu, W.; Li, H.; Bu, X.; Xu, L.; Jiu, A.; Dou, C. HWDQT: A hybrid quantum machine learning method for ultra-short-term distributed photovoltaic power prediction. *Comput. Electr. Eng.* **2025**, *123*, 110122. [CrossRef]
29. Hong, Y.-Y.; Medina, D.J. A quantum-assisted federated learning framework for physics-regulated wind speed forecasting under extreme weather conditions. *Expert Syst. Appl.* **2026**, *326*, 132719. [CrossRef]
30. Li, M.-W.; Geng, J.; Hong, W.-C.; Zhang, Y. Hybridizing Chaotic and Quantum Mechanisms and Fruit Fly Optimization Algorithm with Least Squares Support Vector Regression Model in Electric Load Forecasting. *Energies* **2018**, *11*, 2226. [CrossRef]
31. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
32. Carvalho, V.R.; Moraes, M.F.; Braga, A.P.; Mendes, E.M. Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification. *Biomed. Signal Process. Control* **2020**, *62*, 102073. [CrossRef]
33. Wang, Y.; Markert, R. Filter bank property of variational mode decomposition and its applications. *Signal Process.* **2016**, *120*, 509–521. [CrossRef]
34. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
35. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
36. Rodríguez-Díaz, F.; Gutiérrez-Avilés, D.; Troncoso, A.; Martínez-Álvarez, F. A Survey of Quantum Machine Learning: Foundations, Algorithms, Frameworks, Data and Applications. *ACM Comput. Surv.* **2025**, *58*, 1–35. [CrossRef]
37. Weng, Z.; Xu, Z.; Li, R.; Wang, L.; Li, Z.; Zhang, Y.; Li, M. Quantum-enhanced forecasting and optimal scheduling for energy management in photovoltaic greenhouses. *Appl. Energy* **2026**, *420*, 128142. [CrossRef]
38. Chen, S.Y.-C.; Yoo, S.; Fang, Y.-L.L. Quantum long short-term memory. In *Proceedings of the ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 8622–8626. [CrossRef]
39. Xu, Z.; Li, Z.; Wang, L.; Wang, S.; Mao, Y.; Weng, Z.; Zhang, Y.; Xie, Z.; Li, M. Quantum algorithm-enhanced evaluation method for hybrid wind-PV system projects considering carbon trading. *J. Renew. Sustain. Energy* **2026**, *18*, 016314. [CrossRef]
40. Wang, L.; Weng, Z.; Li, R.; Zhang, Y.; Wang, S.; Li, M. Collaborative operational model for shared hydrogen storage and energy prosumers: A quantum game approach. *Int. J. Hydrogen Energy* **2026**, *204*, 153266. [CrossRef]
41. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212. [CrossRef]

42. Schuld, M.; Bocharov, A.; Svore, K.M.; Wiebe, N. Circuit-centric quantum classifiers. *Phys. Rev. A* **2020**, *101*, 032308. [CrossRef]
43. Liu, Q.; Darteh, O.F.; Bilal, M.; Huang, X.; Attique, M.; Liu, X.; Acakpovi, A. A cloud-based Bi-directional LSTM approach to grid-connected solar PV energy forecasting for multi-energy systems. *Sustain. Comput. Inform. Syst.* **2023**, *40*, 100892. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.