

Multi-Step Prediction and Optimization of Harmonic Power Quality in Active Distribution Network of Urban Energy Internet Driven by PatchTST

Jiaqi Peng, Baomin Fang, Haiping Ye, Yulin Ma, Bin Ai, Jin Luo

State Grid Qinghai Provincial Electric Power Company, Haidong Power Supply Company, Haidong 810600, Qinghai, China

Corresponding author: Jiaqi Peng (e-mail: 15597275739@163.com)

ABSTRACT The integration of high-density distributed energy, complex loads, and multi-energy flows in urban energy internets complicates the dynamic uncertainty of harmonic evolution, driven by high-frequency fluctuations from distributed energy sources and sudden load switching in industrial processes. To address the resulting inaccurate harmonic power quality prediction and optimization in active distribution networks, this paper proposes a multi-step prediction and collaborative optimization method based on PatchTST. The method enhances complex time series modeling via a localglobal feature fusion mechanism. It extracts local temporal characteristics of harmonic distortion using local time block coding, and mines long-range dependencies between multi-channel variables using a global self-attention network. This achieves multi-step prediction of the total harmonic distortion rate (THD) and individual harmonic components. The study then constructs a physical-information-embedded collaborative control model, combining prediction results with dynamic adjustment of inverter reactive compensation and optimization of energy storage strategies to form a closedloop control. Experiments show the PatchTST-based method achieves low MAE for THD multi-step prediction (0.077/0.119/0.140 for 12/24/48 steps), reduced active power loss (13.62kW), fast voltage deviation recovery (2.2–5s), and low control failure (1.2%). Physical constraint embedding improves optimization feasibility to 93.5% and limits SOC violations to 6. This research offers an integrated solution for complex power quality issues through the deep integration of deep learning and power electronic control, providing a framework directly applicable to improving the stable and reliable power supply required for continuous, high-quality production in the textile industry.

INDEX TERMS active distribution network, harmonic power quality, multi-step prediction, textile industry, physical information embedding

I. INTRODUCTION

WITH the rapid development of urban energy internet, the active distribution network in the power system, as an important link in the integration of distributed energy and load management, has become increasingly prominent in terms of harmonic power quality issues. This situation is particularly critical in high-demand industrial areas such as textile industry, because power quality directly affects mechanical performance and product quality [1,2]. The interaction between nonlinear loads [3,4], distributed energy [5,6], and the grid leads to dynamic and diverse power quality issues. In particular, in the generation, propagation, and evolution of harmonic components, traditional prediction and optimization methods are often unable to effectively capture their complex time-varying characteristics.

At present, the idea of local-global feature fusion has been studied in the field of time series prediction, but traditional methods usually fail to give full play to the

advantages of the combination of the two [7,8]. To this end, this study proposes a multi-step prediction and optimization method for harmonic power quality based on PatchTST. The PatchTST model breaks through this limitation by introducing a deep combination of local time block coding and global self-attention network [9,10]. Local time block coding can not only effectively capture the short-term mutation characteristics in harmonic signals, but also work together with the global self-attention mechanism to extract long-range dependencies across time steps and variables, thereby achieving accurate modeling of harmonic signals. Through the local-global feature fusion mechanism of the PatchTST model, this paper designs a multi-level harmonic dynamic modeling framework. The local time series features of the harmonic signal are extracted through local time block coding, and the long-range dependencies across channels and time steps are modeled in combination with the global self-attention network. In addition, this paper further

constructs a collaborative control optimization strategy with physical information embedding, combining the harmonic prediction results with the inverter reactive compensation parameters and energy storage charging and discharging strategies. This realizes a closed-loop control mechanism based on real-time data, allowing for the precise, dynamic adjustment necessary to maintain the stringent power quality and continuous operation required by the textile industry.

II. RELATED WORKS

Traditional harmonic analysis methods are mainly based on frequency domain and time-frequency domain techniques such as Fourier transform [11,12] and empirical mode decomposition [13,14], which have clear theoretical support and engineering feasibility in processing steady-state harmonic signals. However, in the urban active distribution network environment with distributed nonlinear sources such as photovoltaics, wind power, and electric vehicles coupled with strong dynamic loads, harmonic signals show obvious nonstationarity and mutation [15-17]. There are problems such as frequency leakage, boundary distortion and reconstruction error, and it is impossible to maintain prediction stability in high-frequency disturbances and local abnormal rapid changes [18,19]. In addition, such methods have difficulty in simultaneously modeling the nonlinear relationship between multi-channel variables and cannot reflect the deep coupling structure between voltage, current and device status [20,21].

Time series modeling methods based on deep learning have shown good performance in prediction scenarios such as power load and renewable energy output [22,23]. Recurrent neural networks [24,25] and their variants LSTM (Long-Short Term Memory) [26,27] and GRU (Gated Recurrent Unit) [28,29] have certain capabilities in extracting long-term dependencies of time series, but are limited by the serial nature of the structure and have problems such as gradient dissipation, low training efficiency, and insufficient long-term modeling capabilities. WaveNet and TCN (Temporal Convolutional Networks) causal convolutional structures have made progress in improving training efficiency and local feature modeling, but lack a direct modeling mechanism for long-range coupling between cross-channel variables, and their stability performance in processing strong non-stationary signals is inconsistent.

Based on the above analysis, a structural model integrating local time block extraction and global dependency perception is constructed to meet the practical needs of modeling and control optimization of non-stationary harmonic signals. The PatchTST structure has fragment-level processing capabilities, but it still needs structural improvements in multi-source variable collaborative modeling and power physical scene adaptation to meet the actual needs of high-dimensional harmonic signal prediction and control coupling in active distribution networks under the background of urban energy Internet.

III. HARMONIC DYNAMIC MODELING AND OPTIMIZATION STRATEGY BASED ON LOCAL-GLOBAL FEATURE FUSION

A. HARMONIC FEATURE EXTRACTION OF LOCAL TIME BLOCK CODING

The harmonic feature extraction of local time block coding is based on the sliding window mechanism, which divides the multidimensional harmonic time series into local subsequences of fixed length. Let the original data be $X \in \mathbb{R}^{T \times C}$, T represents the time step, and C represents the number of channels. Under the condition that the sliding window length is l and the step length is s , the time series is divided into N time blocks:

$$N = \left\lfloor \frac{T-l}{s} + 1 \right\rfloor \quad (1)$$

The local time block length l is set to 16 time steps, with a stride s of 1, determined through grid search to balance temporal resolution and computational efficiency. This configuration captures short-term harmonic fluctuations while maintaining overlap between adjacent blocks to preserve boundary information. Each time block contains the instantaneous value, first-order difference value and normalized processing result of each harmonic component of voltage and current, aiming to identify local disturbance characteristics [30,31]. The input data is divided into three-dimensional tensors, whose dimensions correspond to batch size, time block length and number of input channels. The input channels cover each harmonic component of voltage and current and related equipment status characteristics, such as photovoltaic output and energy storage status.

In order to enhance the temporal perception of local time blocks, a position encoding module is added before the convolution structure. The position information of each time step is encoded and added to the original input, retaining the order characteristics of the sequence, effectively solving the problem of insufficient sensitivity of the convolution operation to the temporal order [32,33]. The position encoding is generated in the form of fixed sine and cosine functions, and the encoding of the position of each time step is:

$$PE(t, 2i) = \sin\left(\frac{t}{10000^{2i/d}}\right) \quad (2)$$

$$PE(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d}}\right) \quad (3)$$

During the data division process, the window sliding step is smaller than the time block length to ensure that there is an overlapping area between adjacent time blocks to capture the mutation information at the time period boundary. The time block division of multi-dimensional harmonic data is shown in Figure 1.

In order to clarify the fusion mechanism between local time block encoding and global self-attention, the original multivariate input sequence is assumed to be:

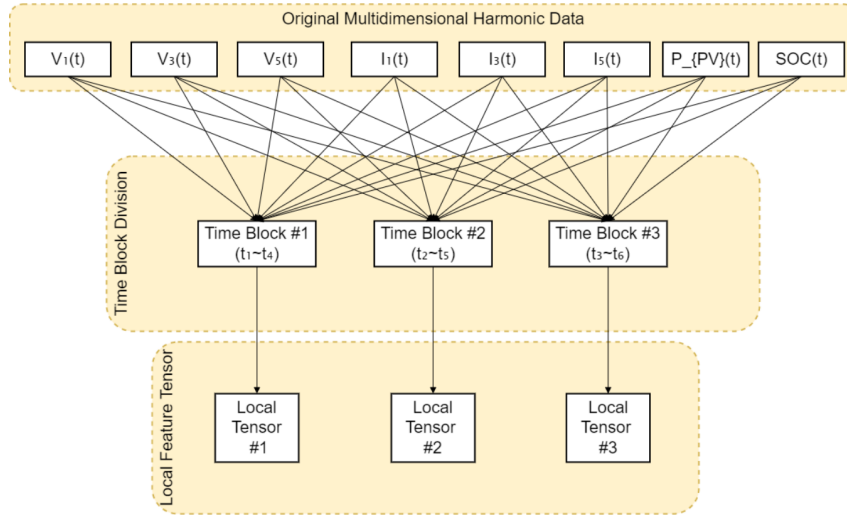


FIGURE 1. Schematic diagram of time block division of multi-dimensional harmonic data

$$X \in \mathbb{R}^{B \times T \times C} \quad (4)$$

Where B is the batch size, T is the sequence time length, and C is the number of input channels. The local time block division converts the sequence into $N_b = (T - P)/S + 1$ time segments with a fixed length of P , where S is the sliding step length, forming:

$$F_{local} \in \mathbb{R}^{B \times N_b \times D} \quad (5)$$

The local feature tensor is used as the input of the global Transformer encoder, and global dependency modeling is performed while keeping the channel structure unchanged. The modeling process is as follows:

$$Q = F_{local}W^Q, \quad K = F_{local}W^K, \quad V = F_{local}W^V \quad (6)$$

$$Attention(Q, K, V) = softmax\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (7)$$

$$F_{attn} = MultiHead(Q, K, V) \quad (8)$$

$$F_{global} = LayerNorm(F_{local} + F_{attn}) \quad (9)$$

$$F_{out} = LayerNorm(F_{global} + FFN(F_{global})) \quad (10)$$

The final output is:

$$F_{out} \in \mathbb{R}^{B \times N_b \times D} \quad (11)$$

The output is used as a local-global fusion representation and provided to the subsequent prediction or control optimization module.

B. CROSS-CHANNEL DEPENDENCY MODELING OF GLOBAL SELF-ATTENTION NETWORK

All input variables are constructed as three-dimensional tensors, corresponding to the dimensions of sample number, time step number and channel number. In the PatchTST structure, after each sample is expanded in the time dimension, the independence between channels is maintained to preserve the multivariate state characteristics. In order to model the long-range dynamic dependencies across channels, a multi-head selfattention structure is introduced in the Transformer encoder. All attention calculations are not performed locally in the time dimension, but are expanded in parallel across the full-dimensional channel range to handle the interdependence between all channels in each time step.

In this modeling process, the channel-dimensional embedding representation of each time step is input into the multi-head attention module. Each head constructs the key, query, and value vectors separately, and maps each channel variable through an independent linear projection matrix. Let the input tensor be X_t , and the mapping of each channel is:

$$Q = X_t W_Q \quad (12)$$

$$K = X_t W_K \quad (13)$$

$$V = X_t W_V \quad (14)$$

$$W_Q, W_K, W_V \in \mathbb{R}^{C \times d_k} \quad (15)$$

By calculating the dot product between the query and the key, it can obtain the attention weight coefficient A_{ij} between each channel, that is, the degree of dependence of the i channel on the j channel. The specific calculation formula is:

$$A_{ij} = \frac{Q_i K_j^\top}{\sqrt{d_k}} \quad (16)$$

To avoid numerical instability, the weights are softmaximized to obtain the standardized attention weights:

$$\alpha_{ij} = \text{softmax}(A_{ij}) = \frac{\exp(A_{ij})}{\sum_{k=1}^C \exp(A_{ik})} \quad (17)$$

Each head performs a weighted sum operation on the fully connected channel attention map to generate a channel representation of the fused context:

$$O_i = \sum_{j=1}^C \alpha_{ij} V_j \quad (18)$$

Figure 2 describes the interdependence between different power system variables (voltage, current, photovoltaic output, and energy storage status, etc.). The color depth in the figure reflects the degree of attention each channel pays to other channels (Key) when it is used as a query (Query). The brighter the color, the higher the degree of dependence.

The outputs of all attention heads are concatenated in the channel dimension and uniformly projected to the original embedding dimension. The concatenated representation contains the channel dependency modeling results in multiple subspaces, and the attention heads in each subspace achieve differential modeling through non-shared projections. The output channel representation contains the joint information of all channel variables in the target time step, which can preserve the long-range structural pattern of the coupling relationship between variables.

C. PHYSICAL INFORMATION EMBEDDED COLLABORATIVE CONTROL OPTIMIZATION

In order to achieve active regulation of power quality and dynamic suppression of harmonic pollution, this paper designs a physical information embedded collaborative control optimization framework for active distribution networks based on the harmonic prediction output by the PatchTST model. The control model takes the predicted total harmonic distortion (THD) and the distribution of each harmonic component in the future time window as the core input. Combined with the system operation status (such as node voltage, node current, active load, reactive load, distributed power output, etc.), it is uniformly mapped into the state variable tensor required for optimization:

$$S = \{\widehat{THD}_i(t+k), \widehat{H}_i^{(n)}(t+k), V_i(t), I_i(t), P_i(t), Q_i(t), P_{DG}(t)\}_{k=1}^{T_p} \quad (19)$$

$\widehat{H}_i^{(n)}$ represents the predicted value of the voltage or current amplitude at the i node under the n harmonic. This state tensor is used as the input to drive the optimization module. The control optimization goal is to minimize the

weighted sum of the harmonic distortion of each node in the distribution network during the entire prediction period:

$$\min_u \sum_{i \in N} \sum_{k=1}^{T_p} w_i \cdot \widehat{THD}_i(t+k) \quad (20)$$

u represents the set of control variables, w_i is the node importance weight factor, and N is the node set.

The optimization model contains two main control variables: the reactive power output by the inverter and the charging and discharging power of the energy storage device. For the reactive power regulation part of the inverter, the control quantity is based on the node voltage harmonic response model, and the sensitivity parameter of the node voltage to the change of harmonic power injection is introduced to establish the mapping relationship between reactive power injection and the target node harmonic index:

$$\frac{\partial H_i^{(n)}}{\partial Q_{inv,j}} = S_{ij}^{(n)}, \quad \forall i, j \in N \quad (21)$$

$S_{ij}^{(n)}$ represents the sensitivity of reactive power injected by the j th inverter to the n th harmonic response of the i th node. The calculation of the sensitivity matrix is based on the system power flow sensitivity analysis method, and the small perturbation method is used to linearly approximate the changes in node harmonic indicators caused by the reactive power injected by the inverter. The matrix is constructed through offline perturbation simulation with a perturbation step of 0.01 p.u. Physically, the matrix reveals the strength of the inverter's ability to regulate the system harmonic distribution at different harmonic frequencies. The larger the matrix value, the more significant the regulation effect of the inverter at a specific node and harmonic frequency. The main sensitivity coefficients in a typical system are between 0.12–0.35.

Based on this sensitivity matrix, the objective function can be rewritten as a quadratic expression of:

$$\min_{Q_{inv}} \sum_{k=1}^{T_p} \left\| \widehat{H}^{(n)}(t+k) + S^{(n)} \cdot \Delta Q_{inv}(t+k) \right\|_2^2 \quad (22)$$

The physical constraints of the model are as follows:
Reactive output capacity limitation:

$$|Q_{inv,i}(t+k)| \leq Q_{inv,i}^{max} \quad (23)$$

Grid-connected current constraints:

$$I_{inv,i}(t+k) \leq I_{inv,i}^{max} \quad (24)$$

Voltage Limitation:

$$V_i^{min} \leq V_i(t+k) \leq V_i^{max} \quad (25)$$

For the energy storage system charging and discharging regulation process, the control variable is the energy storage

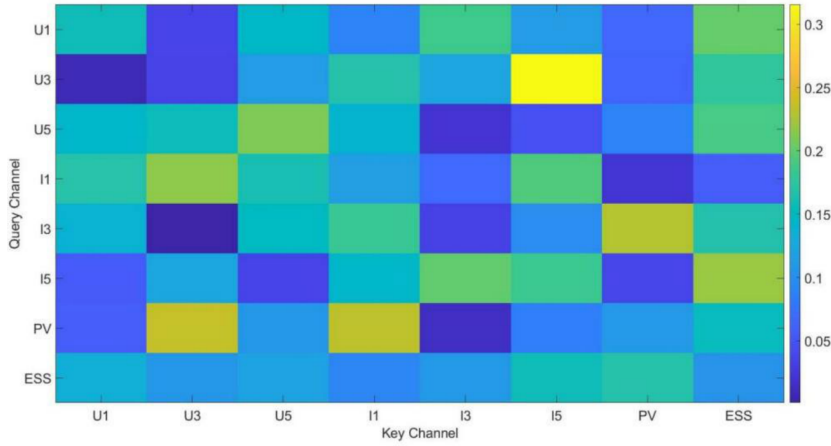


FIGURE 2. Cross-channel attention weight heat map

output power sequence $\{P_{ess,j}(t), \dots, P_{ess,j}(t+T_p)\}$. The regulation strategy continuously dispatches the energy storage output within the optimization period, compensates for the voltage deviation caused by harmonic distortion in the system within each prediction time step, or smoothes the harmonic dynamic coupling effect caused by load fluctuation. In order to ensure that the energy storage equipment works within the feasible domain, the capacity boundary of the energy storage equipment, the maximum charge and discharge power limit and the charge state update equation are embedded in the optimization model to form an active power distribution model within the dynamic feasible domain. The charge state update equation is:

$$SOC_j(t+k+1) = SOC_j(t+k) - \frac{\eta_j P_{ess,j}(t+k)\Delta t}{E_j^{max}} \quad (26)$$

η_j is the charge and discharge efficiency, and E_j^{max} is the maximum energy storage capacity. The embedded physical constraints include:

Feeder line capacity limits:

$$I_{line} \leq I_{line}^{max} \quad (27)$$

Bus voltage constraints:

$$V_i^{min} \leq V_i \leq V_i^{max} \quad (28)$$

Inverter reactive power limits:

$$|Q_{inv,i}| \leq Q_{inv,i}^{max} \quad (29)$$

Energy storage SOC boundaries:

$$SOC^{min} \leq SOC \leq SOC^{max} \quad (30)$$

The energy storage regulation objective function uses the minimization of the global harmonic response function gradient as a guide to enhance the response capability to the dynamic changes of medium and high-order harmonic components:

$$\min_{P_{ess}} \sum_{k=1}^{T_p} \|\nabla_{P_{ess}} THD(t+k)\|_2^2 \quad (31)$$

In order to improve the consistency of the optimization strategy with the power system operation constraints, the control model explicitly integrates the power flow equation and the node power balance relationship.

$$P_i = \sum_{j \in N} V_i V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) \quad (32)$$

$$Q_i = \sum_{j \in N} V_i V_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}) \quad (33)$$

After the power flow equation is converted into the algebraic mapping relationship between node injection power and voltage, a multivariable nonlinear constraint system is constructed to ensure that the optimization adjustment does not cause node voltage limit or line current overload. A coupling relationship in the form of Jacobian matrix is established between each variable in the power flow equation and the control variable, and the adjustment direction and step size of the control variable are dynamically corrected during the optimization iteration process to be compatible with the system stability constraint.

The interface between PatchTST prediction and the collaborative control system is explicitly defined as follows: the predicted THD value modifies the inverter operation by dynamically updating the optimization objective weight, adjusting constraint limits for reactive compensation, and tuning the Energy Storage System (ESS) dispatch coefficients. This ensures that prediction outputs directly influence real-time control decisions.

IV. EXPERIMENTAL DESIGN AND VERIFICATION PLATFORM

A. DATA COLLECTION AND PREPROCESSING IN MULTI-ENERGY COUPLING SCENARIOS

The experimental dataset is derived from a simulated urban distribution network in a typical industrial zone of East

China, with a focus on textile manufacturing loads characterized by frequent motor starts and variable-speed drives. The total duration of the dataset is 60 consecutive days, sampled at 10 Hz, resulting in 14,400 time steps per day. The harmonic profiles are primarily influenced by non-linear loads from industrial machinery, photovoltaic inverters, and electric vehicle charging stations.

The experiment uses a multi-energy coupling distribution system simulation model based on the typical topology of the urban energy Internet. The model includes photovoltaic power generation units, wind power generation units, energy storage systems, bidirectional electric vehicle charging stations, and multi-type nonlinear load nodes. The grid structure is a typical three-phase four-wire 380 V medium and low voltage active distribution network. The power disturbance simulation uses a time-varying power model control, the photovoltaic module access adopts the light-temperature joint modulation method, and the wind model introduces a high-speed disturbance factor superimposed with a steady wind speed component. The energy storage system uses a 48V lithium battery model, and the dynamic scheduling behavior is simulated through the State of Charge (SOC) time-varying curve.

The collected variables include the total harmonic distortion (THD) of the node three-phase voltage, the amplitude of each harmonic component, the current harmonic component, the active output power of distributed energy, the reactive output of the inverter, the energy storage SOC, the connection status of the electric vehicle, etc. Harmonic analysis uses a sliding window extraction algorithm based on fast Fourier transform, with a window length of 1 second and a window step of 0.1 second to form an equally spaced sampling feature sequence. All original features are uniformly standardized using Z-score, and the standardized mean and standard deviation are based on the statistics of the training set and applied to all samples to avoid data leakage. Missing values are filled with linear interpolation; data windows containing interference spikes are removed by median filtering, and the filter window size is set to 5 points. The final sample is constructed in the form of a tensor with an input dimension of (B, 96, 24), where 96 is the number of equally spaced time steps per hour. The data is divided in chronological order. The first 45 days of data are used for model training, and the last 15 days of data are used for testing. The training set and the test set are kept non-overlapping.

B. MODEL TRAINING PARAMETER CONFIGURATION

In order to explore the generalization ability and stability of the model in multi-step prediction tasks, this paper conducts systematic parameter configuration in the model training stage, covering multiple dimensions such as input structure design, convolution and attention module details, optimizer settings and training strategies. In particular, in the PatchTST model architecture, key elements such as time series block division, Transformer encoder structure,

and loss function weight are finely regulated to ensure consistency and robustness in performance under different prediction steps and harmonic frequency bands. The specific training parameter configuration is shown in Table 1.

TABLE 1. Training parameter configuration

Parameter Category	Parameter Items	Parameter Settings
Input Setup	Input Sequence Length / Forecast Horizon	96 / 12, 24, 48
	Patch Length / Number of Patches / Stride	16/6/1
Convolution Module	Kernel Size / Stride	3/1
Transformer Encoder	Number of Layers / Attention Heads / Head Dimension	3 / 8 / 32
	Feedforward Hidden Dimension / Activation Function	256 / GELU
	Positional Encoding / Initialization Method	Learnable / Xavier
Optimization	Optimizer / Initial Learning Rate / Minimum LR	AdamW / 1e-3 / 1e-5
	Weight Decay / Max Epochs / Batch Size	1e-2 / 300 / 64
Training Strategy	LR Scheduler / Early Stopping Strategy	Cosine Annealing / Stop if validation error not reduced for 20 epochs
	Loss Function / Weighting Strategy / Residual Connection	Weighted MSE / Emphasis on short-term / Skip Connection

In order to systematically evaluate the performance of the proposed model in the multi-step prediction task of harmonic power quality, five mainstream time series prediction methods are set as comparison models, namely Autoformer, FEDformer (Frequency Enhanced Decomposed Transformer, Informer, TimesNet and Autoformer-TS.

V. VERIFICATION OF HARMONIC DYNAMIC PERCEPTION AND OPTIMIZATION EFFICIENCY

A. EVALUATION OF HARMONIC MULTI-STEP PREDICTION ACCURACY

Although multi-step prediction introduces temporal lag, the control framework incorporates a fast-response bypass channel to handle transient harmonic spikes. Real-time measurements exceeding a predefined threshold trigger an immediate corrective action from the inverter and ESS, ensuring robustness to sudden high-frequency disturbances not captured by the prediction horizon.

This section uses three rolling prediction step sizes (12, 24, and 48 steps, corresponding to 1, 2, and 4 hours) and six harmonic indicators (THD, 2nd, 3rd, 5th, 7th, and 9th), and records the average absolute error and fluctuation range of each model under each combination condition. The results are shown in Figure 3.

PatchTST maintains the lowest MAE (Mean Absolute Error) value under all harmonic components and prediction steps. For example, the errors of THD are about 0.077, 0.119, and 0.140 respectively, and its standard deviation is also generally small, indicating that it is not only more accurate but also more stable. As the prediction step increases,

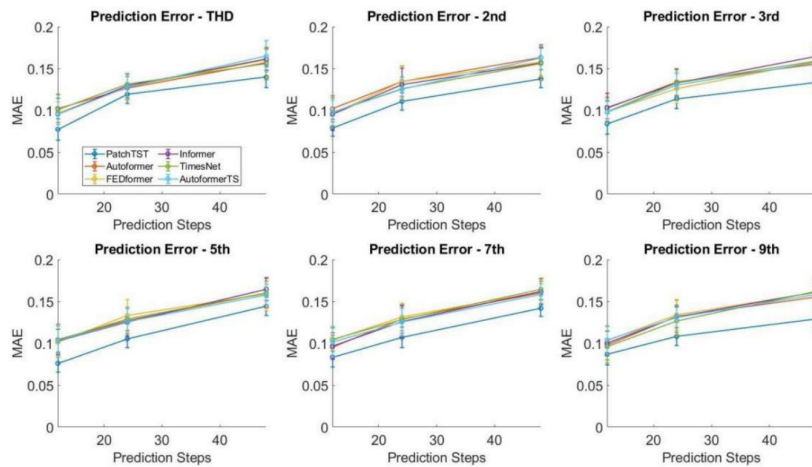


FIGURE 3. Comparison of multi-order harmonic prediction errors

the errors of each model generally increase, but the increase of PatchTST is relatively slow, showing a stronger ability to adapt to the time scale.

To fully demonstrate the reliability of multi-step prediction, this paper also reports the gradual decrease in accuracy across different prediction time domains. The results show that the THD prediction error gradually increases with 1-step, 5-step, and 10-step predictions, confirming that the decrease in prediction accuracy with the prediction time domain is predictable, while still maintaining acceptable accuracy to meet the needs of operational decision-making.

The above trend reflects the advantages of PatchTST in model structure and feature extraction mechanism. The model is based on the time segment nested modeling structure, which can fully capture the local periodicity and global trend information in the harmonic sequence, and is particularly suitable for the modeling tasks of power quality sequences with multi-frequency superposition and severe disturbance. In contrast, although structures such as Autoformer and FEDformer have certain advantages in processing long sequence tasks, they have a lag in responding to short-term rapid changes such as odd harmonics, resulting in high errors in the prediction of 2nd and 3rd harmonics. Although methods such as Informer and TimesNet have a wider feature attention mechanism, they fail to fully distinguish various harmonic features in this task and show greater prediction fluctuations. Therefore, the performance of PatchTST verifies its ability to adaptively model high-dimensional time-series power quality data, especially in long-term prediction scenarios.

This paper further compares PatchTST with baseline models including Autoregressive Integrated Moving Average Model (ARIMA), standard Recurrent Neural Network (RNN), and the original Transformer, as shown in Table 2.

TABLE 2. Comparison with baseline models for multi-step harmonic prediction

Model	MAE (12-step)	MAE (24-step)	MAE (48-step)
ARIMA	0.152	0.231	0.318
RNN	0.134	0.205	0.289
Transformer	0.098	0.162	0.221
PatchTST	0.077	0.119	0.14

Table 2 shows that PatchTST consistently outperforms these models across all prediction ranges and harmonic orders, demonstrating its structural advantage in capturing both local and global temporal dependencies.

B. COLLABORATIVE OPTIMIZATION EVALUATION OF POWER SYSTEM STATUS DRIVEN BY MULTI-SOURCE PREDICTION

In order to optimize the state of the power system under the multi-source prediction drive, this experiment simulated five typical working conditions including daytime, nighttime, peak load, harmonic disturbance and pulse load, and counted the active loss performance of the main node, relay node and edge node under controlled and uncontrolled conditions. Figure 4 shows the distribution characteristics of node voltage deviation and SOC utilization of energy storage system. Figure 5 shows the active loss regulation effect of various nodes.

As can be seen from Figure 4, after the optimization strategy is implemented, the voltage deviation is more concentrated overall, with the peak concentrated at around +0.5%. In the uncontrolled scenario, the deviation is more dispersed, and there are more deviation data exceeding $\pm 5\%$, indicating that the voltage stability is poor in the uncontrolled state. Under optimized control, SOC is mainly distributed between 30% and 75%, with a relatively uniform frequency, while without control strategy, SOC is mostly concentrated in the range of 20% to 45%, with a narrower utilization range.

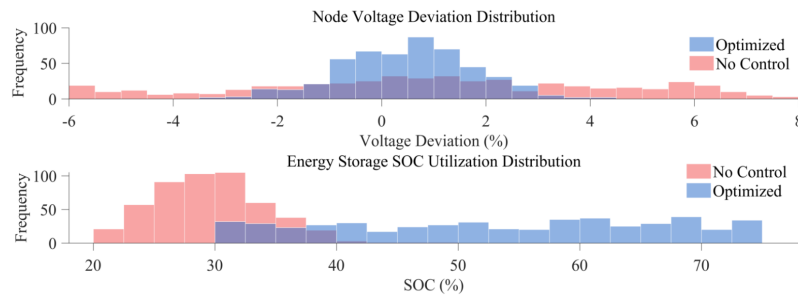


FIGURE 4. Effect of control strategy on node voltage deviation and energy storage utilization

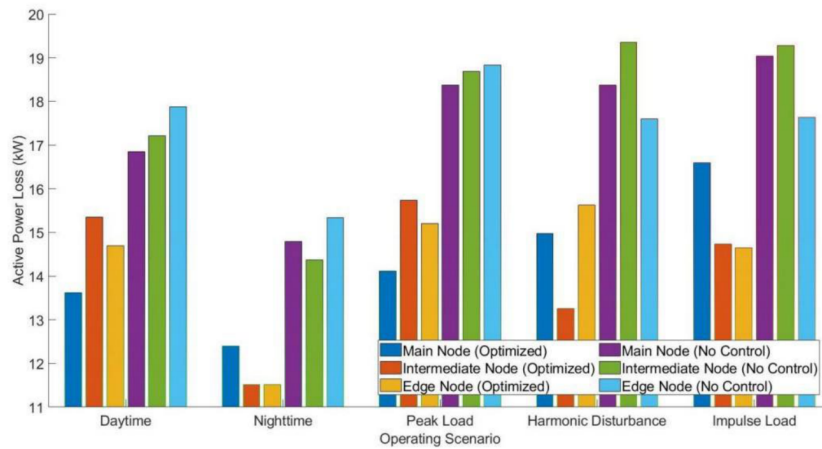


FIGURE 5. Comparison of active power loss of various nodes under different operation scenarios

The above differences reflect the positive impact of the control strategy on the system operation status. Under the guidance of the optimization strategy, the model can dynamically allocate reactive support and energy storage output based on the voltage offset trend, so that the node voltage remains in a better range, reflecting the model's improvement in system status recognition and precise regulation capabilities. At the same time, the expansion of the energy storage utilization distribution shows that the model can reasonably dispatch SOC according to local load changes and voltage fluctuations to avoid idle resources. In contrast, the uncontrolled state only relies on the original passive response mechanism of the system and cannot effectively respond to disturbances, resulting in wider voltage deviations and lower energy storage utilization. This also shows that under the control framework that integrates physical constraints and system state feedback, the model has stronger operational adaptability and scheduling intelligence.

In Figure 5, taking Daytime as an example, after optimization, the loss of the main node is about 13.62 kW, the relay node and the edge node are 15.35 kW and 14.7 kW respectively, while when not controlled, the losses of the three nodes rise to about 16.85 kW, 17.21 kW and 17.87 kW respectively. In the Peak Load scenario, the loss of the main node is reduced from 18.38 kW to about 14.12 kW, and the relay and edge nodes of the optimization strategy are also reduced. Overall, the loss under the optimization strategy in

all scenarios is lower than that of the uncontrolled strategy, indicating that the control scheme can effectively reduce the system active power loss under various nodes and working conditions.

This loss improvement comes from the deep embedding of the optimization model into the inverter reactive compensation and energy storage charging and discharging coordinated scheduling mechanism. In the model, the harmonic index and real-time load level output by the prediction module are used to construct the objective function to minimize the active power loss of the entire network. The optimization process dynamically allocates reactive power and energy storage power while ensuring power balance and equipment safety boundaries, buffering the impact of load fluctuations and harmonic interference on active power loss. Thanks to the explicit modeling of the physical constraints of the power grid and the voltagepower sensitivity, the control strategy can maintain efficient operation under extreme disturbance scenarios such as harmonic disturbances and pulse loads, and control the node loss in the range of 13.25–16.6 kW, which reflects the robustness and engineering deployability of the method.

C. MULTIVARIABLE RECOVERY RESPONSE CAPABILITY UNDER DYNAMIC DISTURBANCES

In order to deeply analyze the specific impact of the proposed control optimization strategy on dynamic response

performance, this paper observes three key dynamic indicators in the distribution system: voltage deviation recovery time, energy storage response delay, and inverter reactive power adjustment rate, respectively, under two conditions: applying optimization control and not taking control. The experiment uses ten sets of simulation samples, and presents the statistical distribution under different strategies in the form of box plots, which is convenient for evaluating the differences in control strategies in coping with disturbances and adjustment capabilities. The results are shown in Figure 6.

From the box plot results of the three sub-graphs in the figure, it can be seen that in terms of voltage deviation recovery time, the recovery time of the optimized control strategy group is concentrated between about 2.2–5 seconds. The recovery time under uncontrolled conditions is distributed between 5.5–8.9 seconds, indicating that the optimization strategy can speed up the recovery speed of the system voltage after disturbance. The comparison of energy storage response delay is also obvious. The response delay under optimized control is mainly concentrated between 1.1–2.8 seconds, while the uncontrolled group is distributed between 2.1–5 seconds, indicating that the control system can achieve faster scheduling. In the third sub-figure, the median of the inverter reactive power adjustment rate in the control group is about 37 kVar/s, and the overall distribution is 32–39 kVar/s, which is much higher than the range of 21–28 kVar/s in the uncontrolled group, indicating that the control strategy enhances the fast adjustment capability of the inverter.

D. MODEL CALCULATION EFFICIENCY AND REAL-TIME VERIFICATION

In order to evaluate the computational efficiency of the PatchTST model under various operating conditions, experiments were conducted in typical distribution network operation scenarios including normal, heavy load, light load and disturbance. The experiments recorded the operation time of the prediction phase and the calculation time of the control optimization phase of each model under different operating conditions. All data were obtained in a unified software and hardware environment to ensure comparability and stability. Figure 7 shows the prediction time and optimization time of the six models in four typical scenarios.

From Figure 7, it can be observed that the prediction time of PatchTST in various scenarios is between 1.14 and 1.28 seconds, which is better than other models overall and shows more stable reasoning efficiency. The right figure shows that the control optimization time of PatchTST under different working conditions is 0.68–0.77 seconds, which has better control decision real-time performance than other models.

The time consumption for model training and optimization is shown in Table 3.

TABLE 3. Comparison of time efficiency of different time series prediction models

Model	Training Time (hours)	Prediction Time (s)	Optimization Time (s)
PatchTST	1.1–1.4	1.14–1.28	0.68–0.77
Autoformer	1.5–1.8	1.20–1.40	0.75–0.85
FEDformer	1.3–1.6	1.18–1.35	0.70–0.80
Informer	1.6–1.9	1.22–1.45	0.80–0.90
TimesNet	1.4–1.7	1.15–1.30	0.72–0.82
Autoformer-TS	1.5–1.8	1.25–1.40	0.78–0.88

PatchTST has a relatively short training time, ranging from 1.1 to 1.4 hours, while Informer has the longest training time, which is 1.6 to 1.9 hours. In terms of prediction time, PatchTST only takes 1.14 to 1.28 seconds. In terms of optimization time, PatchTST is 0.68 to 0.77 seconds. This difference in time efficiency stems from the underlying mechanisms and architectural design of each model. The specific architecture adopted by PatchTST enables it to extract features and perform calculations more efficiently when processing data, thereby saving time in the training and optimization stages. In contrast, although models such as Informer may have higher prediction accuracy in some scenarios, their complex architecture and calculation process lead to longer training and optimization times.

E. IMPACT OF PHYSICAL CONSTRAINT EMBEDDING ON OPTIMIZATION FEASIBILITY

In order to deeply analyze the impact of physical constraint embedding on the feasibility and stability of the optimization control process, this paper evaluates the actual execution effect of the inverter and energy storage control strategy under typical load disturbance scenarios under the condition of whether or not the physical operation boundary is considered. During the experiment, the control optimization module receives the multi-step harmonic index sequence output by the prediction model and the current operating status of the system, and performs 100 rounds of rolling scheduling respectively. Key indicators such as the effectiveness of control output, voltage over-limit, SOC boundary maintenance capability, and iterative characteristics of optimization solutions can be monitored. By comparing the control results of the unconstrained strategy with those after embedding capacity, voltage, current and other operating constraints, the effect of introducing constraint information on the feasibility of control is quantified. The relevant results are shown in Table 4.

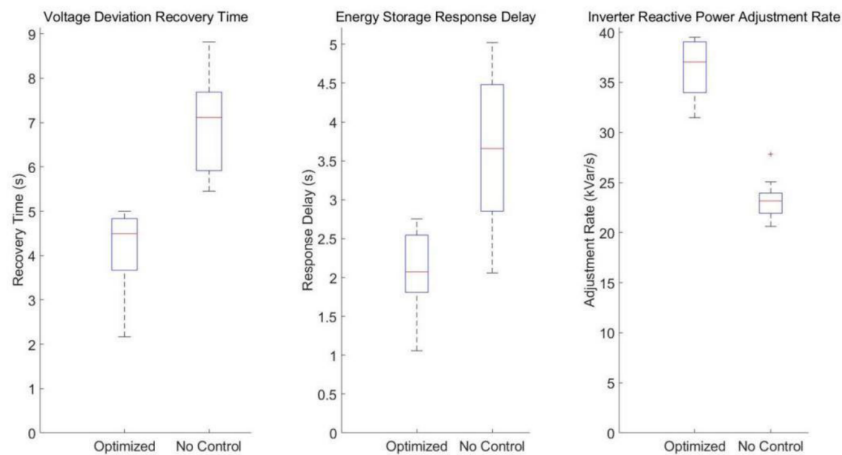


FIGURE 6. Comparison of the impact of control optimization strategy on dynamic performance indicators

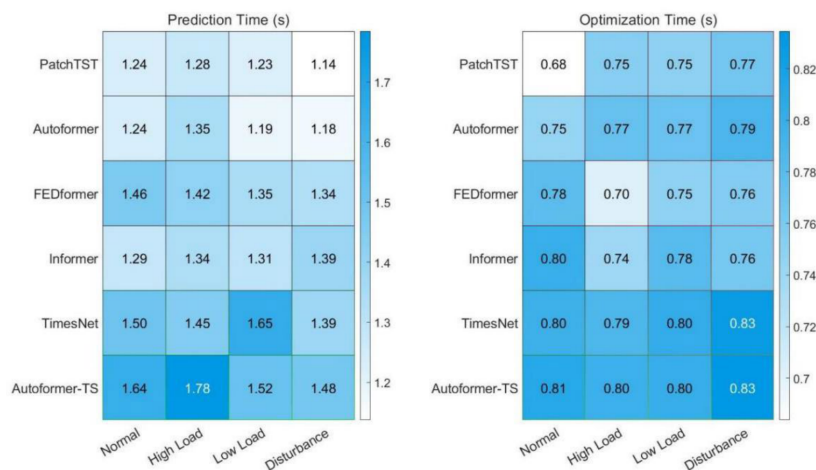


FIGURE 7. Distribution of prediction and optimization calculation time under different working conditions

TABLE 4. Comparison of the impact of physical constraint information embedding on the feasibility of optimized control

Metric	Without Constraints	With Constraints	Description
Node Control Failure Rate	15.3%	1.2%	Fewer infeasible control outputs
Optimization Feasibility (100 runs)	84.0%	93.5%	Higher stability with constraint embedding
Max. Voltage Violation Ratio	9.8%	1.7%	Better voltage boundary compliance
Battery SOC Violation Count	17	6	SOC bounds effectively maintained
Average Number of Iterations	12.6	15.3	Slight increase due to added constraints
Valid Control Output Ratio	81.2%	92.1%	More control values within safe limits

Under the optimization scheme without embedded physical constraints, the system has a high risk of control failure and boundary violation. Specifically, the node control failure rate is as high as 15.3%, the optimization feasibility is only 84.0%, and the effective control output accounts for 81.2%, indicating that the control instructions generated in multiple forecast periods are difficult to meet the system operation restrictions and there are situations where they cannot be actually executed. After embedding physical conditions such

as inverter capacity, voltage constraints, and SOC boundaries, various indicators have improved significantly. The control failure rate has dropped to 1.2%, the feasibility has increased to 93.5%, the control output efficiency has increased to 92.1%, and the voltage limit ratio has dropped from 9.8% to 1.7%. The number of energy storage SOC violations has dropped from 17 to only 6, indicating that the embedding of physical constraints has effectively improved the executability of the control scheme and the compliance with the system boundaries.

F. THD PREDICTION EFFECT UNDER HIGH-DENSITY DISTRIBUTED POWER SOURCES

In order to evaluate the performance of the model in the harmonic superposition effect of high-density distributed power sources, this paper sets up three different distributed power grid-connected scenarios to observe their impact on the harmonic power quality of the power system. By analyzing the THD values after single, two, and three distributed power sources are connected to the grid, the accuracy of the model's harmonic prediction in different scenarios is fully

understood.

Table 5 shows the prediction results of harmonic THD of each model under different distributed power grid-connected scenarios. With the increase in the number of grid-connected distributed power sources, the THD value generally shows an upward trend, which is related to the harmonic superposition effect in the power system. When multiple power sources are connected to the grid, the harmonic interference increases, resulting in the deterioration of the power quality of the system. The PatchTST model shows relatively good prediction accuracy in all scenarios. The THD values in the three scenarios are 3.18%, 4.05% and 4.42%, respectively, which are lower than other models.

TABLE 5. THD comparison under different distributed power grid-connected scenarios

Model	Scenario 1: Single Distributed Source	Scenario 2: Two Distributed Sources	Scenario 3: Three Distributed Sources
PatchTST	3.18%	4.05%	4.42%
Autoformer	3.88%	4.37%	4.93%
FEDformer	3.63%	4.28%	4.75%
Informer	4.05%	4.58%	5.02%
TimesNet	3.72%	4.48%	4.84%
Autoformer-TS	4.12%	4.55%	4.98%

G. LONG-TERM STABLE OPERATION

In order to verify the support ability of each prediction model for the stability of the distribution system under long-term operation conditions, this paper simulates a high-frequency dynamic operation process with a step length of 5 minutes for 70 consecutive days, covering actual conditions such as daily load fluctuations, highfrequency disturbances, and changes in distributed power output. By statistically analyzing the voltage compliance rate, THD, and energy storage charge state volatility in long-term operation, the influence of different models on system stability and power quality is analyzed. The results are shown in Table 6.

TABLE 6. System stability evaluation indicators of each model under long-term operation conditions

Model	Voltage Compliance Rate (%)	Average THD value (%)	SOC Variability (%)
PatchTST	98.1	3.92	4.1
Autoformer	96.4	4.53	5.8
FEDformer	97.3	4.42	5.9
Informer	95.2	4.78	7.2
TimesNet	96.1	4.56	6.7
Autoformer-TS	95.5	4.73	6.9

From the data, it can be seen that the PatchTST model shows the best performance in all three indicators, with a voltage compliance rate of up to 98.1%; at the same time, its average THD value is only 3.92%, the lowest among all models, indicating that it performs well in controlling harmonic pollution; the SOC variability is only 4.1%, which is also better than models such as Autoformer and Informer,

reflecting its stability in regulating the energy storage system. These differences show that under the same operating conditions, there are significant differences in the response effects of the models to multi-source disturbances in the power grid.

H. VERIFICATION BASED ON ACTUAL DATA SETS

To further verify the robustness of the proposed PatchTST model, this paper adds experiments based on actual power grid data. The experiment uses actual monitoring data from a city's distribution network, including voltage, current and its harmonic components, with a sampling frequency of 10Hz, covering multiple load fluctuations and distributed energy access scenarios. The experimental data comes from the actual operation records of the distribution network, including the dynamic evolution of harmonics in multiple time periods. On this basis, this study selects Graph Neural Network (GNN) and Physics-Informed Neural Network (PINN) as comparison models. The comparison of harmonic prediction errors under multiple types of load scenarios is shown in Figure 8, and the engineering applicability and economic evaluation are shown in Figure 9.

The results show that PatchTST maintains the lowest error under all loads and step sizes. In addition, the prediction error of inverter loads is generally lower than that of other loads. The superiority of PatchTST stems from its local-global feature fusion mechanism: local time block encoding can capture short-term mutations of harmonics, while global self-attention models long-range dependencies across channels. In contrast, GNN relies on topological structures and is difficult to adapt to dynamic harmonic sources; although PINN is embedded in physical equations, it lacks the ability to generalize non-stationary signals. PatchTST remains stable at 96 step sizes, verifying its ability to jointly model multi-scale harmonic characteristics. Figure a shows the real-time performance verification. The calculation time of the PatchTST model is always lower than that of GNN and PINN, especially in small configurations, showing the advantage of PatchTST in computational efficiency. In addition, the right axis is the CPU usage rate. The CPU usage rate of PatchTST is low, showing better resource utilization efficiency. Figure b shows the full life cycle economic evaluation. PatchTST performs well in all indicators, showing its advantages in reducing costs and improving system benefits. This shows that PatchTST not only has advantages in real-time performance, but also performs more efficiently in economic evaluation.

VI. CONCLUSION

In this paper, a multi-step prediction and collaborative optimization method based on PatchTST is proposed to solve the harmonic power quality problem caused by the dynamic interaction between distributed energy and non-linear loads in the active distribution network of the urban energy internet. This approach, which is particularly relevant to mitigating the complex harmonic issues arising from

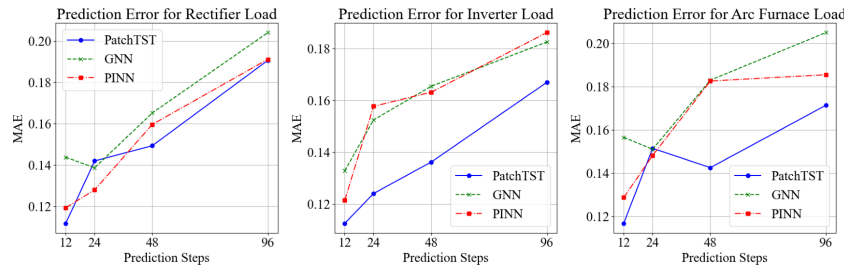


FIGURE 8. Comparison of harmonic prediction errors under multiple load scenarios

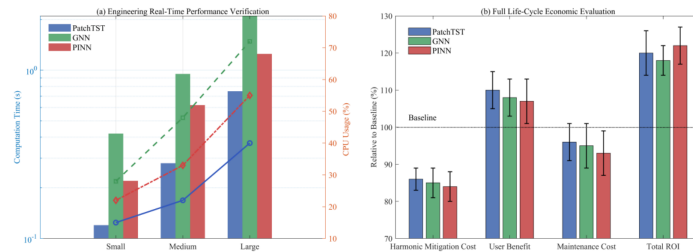


FIGURE 9. Model engineering applicability and economy

the widespread non-linear machinery in the textile industry, realizes the synchronous modeling of the local time series mutation characteristics and the long-range dependency relationship across channels of the harmonic signal. The constructed physical information embedding collaborative control model forms a closed-loop regulation mechanism. Experiments show that this method is superior to traditional models in complex multi-energy coupling scenarios and improves the system operation efficiency and stability. In the training phase, this study constructs the dependency relationship between variables based on a fixed topological structure, and has not yet introduced a modeling mechanism that adapts to network topology changes. Future research considers embedding graph neural networks into feature extraction modules and dynamically modeling node structure relationships using topological adjacency matrices; at the same time, a local fine-tuning mechanism is introduced to quickly retrain local model parameters in combination with actual operating data to adapt to the new structure. In addition, visualization tools and feature importance analysis will be introduced to further improve the interpretability of the model, exploring how to better reveal the inherent mechanism of the model in harmonic prediction. This improved understanding will help power system operators, especially those managing the sensitive, high-demand grids in the textile industry, to understand the root causes of harmonic problems and provide support for the formulation of more effective response strategies.

AUTHOR CONTRIBUTIONS

Conceptualization – Jiaqi Peng, Baomin Fang, Haiping Ye, Yulin Ma, Bin Ai and Jin Luo; methodology – Jiaqi Peng, Baomin Fang, Haiping Ye, Yulin Ma, Bin Ai and Jin Luo;

investigation – Jiaqi Peng, Baomin Fang, Haiping Ye, Yulin Ma and Jin Luo; writing-original draft preparation – Jiaqi Peng, Baomin Fang, Haiping Ye, Yulin Ma, Bin Ai and Jin Luo. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by, State Grid Qinghai Electric Power Company Science and Technology Project Funding, 522802240007, Research on Power Quality Management Technology for Active Distribution Network of New Power System.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] D. Razmi, T. Lu, B. Papari, E. Akbari, G. Fathi, and M. Ghadam-yari, "An overview on power quality issues and control strategies for distribution networks with the presence of distributed generation resources," *IEEE Access*, vol. 11, pp. 10308-10325, 2023, doi: 10.1109/ACCESS.2023.3238685.
- [2] S. Fahad, A. Goudarzi, Y. Li, and J. Xiang, "A coordination control strategy for power quality enhancement of an active distribution network," *Energy Reports*, vol. 8, pp. 5455-5471, 2022, doi: 10.1016/j.egy.2022.04.014.
- [3] P. K. Verma and G. Goswami, "Mitigation of power quality issues in smart grids using shunt active power filter," *Water and Energy International*, vol. 66, no. 12, pp. 42-48, 2024.
- [4] E. Tveve, B. Kichonge, and T. Kivevele, "Impacts of Nonlinear Loads on the Power Quality of Solar Microgrids and Proposed Mitigation Strategies," *Energy Science & Engineering*, vol. 13, no. 4, pp. 1960-1982, 2025, doi: 10.1002/ese3.70019.

- [5] Y. Zhu, J. Wei, B. Wang, S. Hong, Y. Liu, and L. Lu, "Optimal robust sizing of distributed energy storage considering power quality management," *IET Renewable Power Generation*, vol. 18, no. 16, pp. 3825-3838, 2024, doi: 10.1049/rpg2.13062.
- [6] S. Chapala, R. L. Narasimham, and G. T. R. Das, "Power Quality Analysis of ANFIS Based Distributed Generation System with UPQC," *International Journal of Engineering and Manufacturing (IJEM)*, vol. 14, no. 4, pp. 1-14, 2024, doi: 10.5815/ijem.2024.04.01.
- [7] Z. M. Ali, M. Calasan, F. Jurado, and S. H. E. A. Aleem, "Complexities of power quality and harmonic-induced overheating in modern power grids studies: Challenges and solutions," *IEEE Access*, vol. 12, pp. 151554-151597, 2024, doi: 10.1109/ACCESS.2024.3477729.
- [8] T. Kale, D. Pagdhare, H. Karadkar, C. Sivalingam, and N. Bhitre, "A Review of Power Quality Measurement and Harmonic Distortion Reduction," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6, no. 3, pp. 1429-1433, 2024, [Online]. Available: <https://www.doi.org/10.56726/IRJMET50389>.
- [9] T. D. Mankanju, T. Shongwe, and O. J. Famoriji, "Machine Learning Approaches for Power System Parameters Prediction: A Systematic Review," *IEEE Access*, vol. 12, pp. 66646-66679, 2024, doi: 10.1109/ACCESS.2024.3397676.
- [10] S. Sahoo, S. C. Swain, R. Dash, and P. Samal, "Forecasting Tariff Rates and Enhancing Power Quality in Microgrids: The Synergistic Role of LSTM and UPQC," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12506-12511, 2024, doi: 10.48084/etasr.6481.
- [11] M. S. Priyadarshini, M. Bajaj, L. Prokop, and M. Berhanu, "Perception of power quality disturbances using Fourier, Short-Time Fourier, continuous and discrete wavelet transforms," *Scientific Reports*, vol. 14, no. 1, pp. 3443-3482, 2024, doi: 10.1038/s41598-024-53792-9.
- [12] Y. Niu, T. Yang, F. Yang, X. Feng, P. Zhang, and W. Li, "Harmonic analysis in distributed power system based on IoT and dynamic compressed sensing," *Energy Reports*, vol. 8, pp. 2363-2375, 2022, doi: 10.1016/j.egy.2022.01.119.
- [13] X. Wang, S. Wang, J. Ren, W. Jing, M. Shi, and X. Zheng, "A Novel Harmonic Detection Method for Microgrids Based on Variational Mode Decomposition and Improved Harris Hawks Optimization Algorithm," *IEEE Access*, vol. 12, pp. 60797-60807, 2024, doi: 10.1109/ACCESS.2024.3394931.
- [14] A. C. R. Klaar, L. O. Seman, V. C. Mariani, and L. S. Coelho, "Random convolutional kernel transform with empirical mode decomposition for classification of insulators from power grid," *Sensors*, vol. 24, no. 4, pp. 1113-1137, 2024, doi: 10.3390/s24041113.
- [15] P. Makolo, F. Mwasilu, R. Zamora, et al., "Long-Range Inertia Prediction Considering Contemporary Evolution of Power Grid Networks," *Tanzania Journal of Engineering and Technology*, vol. 43, no. 4, pp. 1-21, 2024, [Online]. Available: <https://www.ajol.info/index.php/tjet/article/view/291859>.
- [16] R. P. Permata and D. D. Prastyo, "Hybrid dynamic harmonic regression with calendar variation for Turkey shortterm electricity load forecasting," *Procedia Computer Science*, vol. 197, pp. 25-33, 2022, doi: 10.1016/j.procs.2021.12.114.
- [17] O. A. Elkashaty, A. A. Daoud, and E. E. Elaraby, "Forecasting of short-term and long-term wind speed of ras gharib using time series analysis," *International Journal of Renewable Energy Research (IJRER)*, vol. 13, no. 1, pp. 258-272, 2023, doi: 10.20508/ijrer.v13i1.13785.g8680.
- [18] A. Saxena, A. M. Alshamrani, A. F. Alrasheedi, K. A. Alnowibet, and A. W. Mohamed, "A hybrid approach based on principal component analysis for power quality event classification using support vector machines," *Mathematics*, vol. 10, no. 15, pp. 2780-2796, 2022, doi: 10.3390/math10152780.
- [19] P. Venkata, V. Pandya, K. Vala, and A. V. Santl, "Support vector machine for fast fault detection and classification in modern power systems using quarter cycle data," *Energy Reports*, vol. 8, pp. 92-98, 2022, doi: 10.1016/j.egy.2022.10.279.
- [20] V. Hoffmann, B. N. Torsæter, G. H. Rosenlund, and C. A. Andresen, "Lessons for Data-Driven Modelling from Harmonics in the Norwegian Grid," *Algorithms*, vol. 15, no. 6, pp. 188-207, 2022, doi: 10.3390/a15060188.
- [21] S. Mishra, K. Mallick R, A. Gadanayak D, P. Nayak, A. Flah, C. Z. El-Bayeh, et al., "Islanding detection and power quality diagnosis of wind power integrated microgrid with reduced feature trained novel optimized random decision Forest," *International Journal of Energy Research*, vol. 2024, no. 1, pp. 5198814-5198834, 2024, doi: 10.1155/2024/5198814.
- [22] F. M. Al Hadi, H. H. Aly, and T. Little, "Harmonics forecasting of wind and solar hybrid model based on deep machine learning," *IEEE Access*, vol. 11, pp. 100438-100457, 2023, doi: 10.1109/ACCESS.2023.3314742.
- [23] A. Taghvaie, T. Warnakulasuriya, D. Kumar, F. Zare, R. Sharma, and D. M. Vilathgamuwa, "A comprehensive review of harmonic issues and estimation techniques in power system networks based on traditional and artificial intelligence/machine learning," *IEEE Access*, vol. 11, pp. 31417-31442, 2023, doi: 10.1109/ACCESS.2023.3260768.
- [24] Y. Chu, S. Hou, C. Wang, and J. Fei, "Recurrent-neural-network-based fractional order sliding mode control for harmonic suppression of power grid," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 10, pp. 9979-9990, 2023, doi: 10.1109/TII.2023.3234305.
- [25] Z. Dawood and C. K. Babulal, "Red deer optimized recurrent neural network for the classification of power quality disturbance," *Electrical Engineering*, vol. 105, no. 4, pp. 1937-1953, 2023, doi: 10.1007/s00202-022-01701-6.
- [26] E. M. Kuyumani, A. N. Hasan, and T. Shongwe, "A hybrid model based on CNN-LSTM to detect and forecast harmonics: A case study of an Eskom substation in South Africa," *Electric Power Components and Systems*, vol. 51, no. 8, pp. 746-760, 2023, doi: 10.1080/15325008.2023.2181883.
- [27] L. Flores-Garrido J, P. Salmer, A. Gomez-Galan J, and A. Pérez-Vallés, "Deep Learning-Based Control of Active Power Filters Using LSTM and GRU Networks for Harmonic and Frequency Estimation," *IEEE Access*, vol. 13, pp. 75332-75350, 2025, doi: 10.1109/ACCESS.2025.3564636.
- [28] T. Yunxue, S. Shuo, Z. Haonian, X. Shibiao, Y. Yuming, and Ming Z et al, "Identification of Intensive Oscillation Parameters of Power System Based on CS-CNN-GRU Network," *Electric Power Components and Systems*, pp. :1-11, 2024, doi: 10.1080/15325008.2024.2307396.
- [29] J. Cai, K. Zhang, and H. Jiang, "Power quality disturbance classification based on parallel fusion of CNN and GRU," *Energies*, vol. 16, no. 10, pp. 4029-4041, 2023, doi: 10.3390/en16104029.
- [30] H. Xiao, X. He, and C. Li, "Multi-step prediction of offshore wind power based on Transformer network and Huber loss," *International Journal of Electrical Power & Energy Systems*, vol. 162, pp. 110229-110244, 2024, doi: 10.1016/j.ijepes.2024.110229.
- [31] J. Moon, "A Multi-Step-Ahead Photovoltaic Power Forecasting Approach Using One-Dimensional Convolutional Neural Networks and Transformer," *Electronics*, vol. 13, no. 11, pp. 2007-2024, 2024, doi: 10.3390/electronics13112007.
- [32] Q. Fang, X. Liang, S. Li, J. Zhang, X. Yang, and Z. Guo, "A Novel Multi-Step Short-Term Power Forecasting Model for Electric Power Systems Based on Deep-Learning," *IEEE Transactions on Power Systems*, pp. :1-13, 2025, doi: 10.1109/TPWRS.2025.3556129.
- [33] Z. Rao, J. Wu, G. Li, and H. Wang, "Voltage abnormality prediction method of lithium-ion energy storage power station using informer based on bayesian optimization," *Scientific Reports*, vol. 14, no. 1, pp. 21404-21421, 2024, doi: 10.1038/s41598-024-72510-z.