Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Enhanced Transformer-EA Evolutionary Algorithm Fusion Method for High-Precision Completion of Power Load Time Series Data

Chao Li

Yiyang meng

Ji Li

lij i0606@tjut.edu.cn

Yuzheng Liu

Changlong Lv

Article

Keywords: power load, evolutionary algorithm, missing data imputation, deep learning, Transformer

Posted Date: March 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8877227/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Enhanced Transformer-EA Evolutionary Algorithm Fusion Method for High-Precision Completion of Power Load Time Series Data

Chao Li* [1], Yiyang Meng* [1], Ji Li* [1], Yuzheng Liu* [1] and Changlong Lv* [2]

[1] Tianjin Key Laboratory of New Energy Power Conversion, Transmission and Intelligent Control

[2] School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China
Chao Li liton@email.tjut.edu.cn
Yiyang Meng 3531291981@qq.com
corresponding author Ji Li liji0606@tjut.edu.cn
Yuzheng Liu 2645711698@qq.com
Changlong Lv 1361251058@qq.com

If you want to obtain the data, please contact Yiyang Meng at the email address 3531291981@qq.com.
The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. All data generated or analyzed during this study are included in this article.

***Abstract*—Aiming at the problem of missing power time-series data in smart grids caused by equipment failures, communication interruptions, etc., this paper proposes a high-precision completion method for power load time-series data based on the deep integration of enhanced Transformer and evolutionary algorithm (EA). This method optimizes the time-series attention mechanism of the Transformer model through a dual-baseline pre-filling strategy, combined with the periodic characteristics and global statistical information of power loads, and introduces a dynamically decaying EA evolutionary algorithm to finely repair missing data. To further improve the completion accuracy, a multi-dimensional weighted loss function is designed in this paper to emphasize the accuracy optimization during peak hours, and physical compliance constraints are used to ensure that the completion results fall within the reasonable range of power loads. Experimental results show that the proposed method exhibits superior completion performance under different missing mechanisms (MCAR, MAR, MNAR) and missing rates (10%, 20%, 30%). Compared with traditional time-series models and basic Transformer, the enhanced Transformer-EA model has significant improvements in core indicators such as R^2, MAPE, and RMSE. This method provides an efficient solution for power load data missing completion and has broad application prospects and practical value**

***Keywords:* power load, evolutionary algorithm, missing data imputation, deep learning, Transformer**

I. INTRODUCTION

With the rapid development of smart grids, distributed energy sources, and demand response technologies, high-precision power load data has become the core foundation for power system dispatching, load forecasting, energy optimization, and safe operation. However, in the process of data collection and transmission in actual power systems, due to factors such as sensor failures, communication interruptions, and signal interference, load data will inevitably have missing values. This not only disrupts the continuity of time-series data but also greatly reduces the performance and reliability of prediction models based on data-driven methods. Therefore, it is necessary to effectively impute and interpolate missing values before load forecasting. This is particularly crucial in peak load forecasting, equipment maintenance optimization, and decision-making for the economic operation of power grids.

Existing research on load data completion has developed multiple technical approaches. Among them, traditional methods are centered on statistical rules, with the core optimization direction focusing on improving the ability to mine spatiotemporal correlations. Romano et al. [1] proposed the Historical Data-Informed Regression Tool (H-DIRT) and the Seasonal K-Nearest Neighbors (SKNN) algorithm. By integrating linear interpolation with multi-scale temporal trend features, they achieved effective completion in scenarios with low missing rates and random missing data. For industrial load scenarios, Wang et al. [2] proposed a regularized optimized low-rank decomposition method, which significantly reduced completion errors by using a PID controller to adaptively adjust coefficients. Hafiz et al. [3] designed a completion scheme based on Gaussian interpolation for highly sparse load datasets

with more than 62% missing data, verifying its superiority over traditional interpolation methods in wide-sense stationary data scenarios. Lee et al. [4] further optimized the interpolation logic and proposed an improved locally weighted regression interpolation algorithm, which increased the completion efficiency in short-term sudden missing scenarios by 30% compared with traditional methods. However, overall, traditional methods generally have common problems such as weak adaptability to sudden load changes, insufficient modeling of spatial correlations, and poor robustness to non-random missing data, making it difficult to meet the complex needs of new power systems. Compared with traditional methods, completion methods based on deep learning have become the current mainstream research due to their stronger feature mining capabilities. Liu et al. [5] used a differential evolution algorithm to optimize LSTM hyperparameters, effectively solving the problem of gradient vanishing in the modeling of continuous missing segments. On this basis, Wang et al. [6] designed an attention-enhanced LSTM model to strengthen the accurate capture of load characteristics during peak and valley periods. Li et al. [7] innovatively combined wavelet transform with Temporal Convolutional Network (TCN) to improve the stability of completion results in noisy interference scenarios. The Dynamic Fusion Attention-Informed Model (DFAIM) proposed by Liu et al. [8] further deepened the mining of feature correlations, while Zhang et al. [9] built a model based on TS2Vec and a dual-channel shared attention mechanism, which further enhanced the robustness to missing data through multi-granularity feature representation. To balance temporal modeling capabilities and global feature capture, Ahmed et al. [10] proposed a hybrid LSTM-Transformer architecture, which was implemented on the CAISO and PJM datasets.resulting in a 28% reduction in Root Mean Square Error (RMSE); Park et al. [11] focused on the residential load scenario, designing a model that integrates Gated Recurrent Units with an attention mechanism, controlling the Mean Absolute Percentage Error (MAPE) within 2.1%. With the deepening of research, the Transformer model has become a new research hotspot due to its excellent global temporal modeling capability, and related achievements continue to expand around model structure improvement, parameter optimization, and multi-source data fusion. Lotfipoor et al. [12] proposed a Transformer model dedicated to power demand, effectively enhancing the capture of long-term dependencies in loads; Hou et al. [13] integrated the advantages of Graph Convolutional Networks in spatial feature extraction and the distribution reconstruction capability of Variational Autoencoders, designing an integrated Graph Convolutional Variational Transformer, while balancing model complexity and completion accuracy through an improved Whale Optimization Algorithm; Chen et al. [14] optimized temporal feature modeling, constructing a Seasonal Attention Transformer to strengthen the recognition effect of intraday peak and valley features; Kim et al. [15] proposed the BERT-PIN framework to achieve synchronous completion of multiple missing segments; Sun et al. [16] made a breakthrough from the time-frequency domain perspective, proposing a multi-scale time-frequency complementary model ST-ScaleFusion to further reduce prediction errors through joint modeling; Zhao et al. [17] focused on the integration of physical constraints, proposing an improved self-attention Transformer to optimize output results by loading load physical characteristic constraints. At the model optimization level, Han et al. [18] used an improved Sparrow Search Algorithm to optimize Transformer parameters; at the data fusion level, Hou et al. [19] fused meteorological and load temporal data to construct a hybrid model, Sun et al. [20] combined PMU and multi-energy flow data to improve completion performance, and Choi et al. [21] proposed a completion model based on Generative Adversarial Networks, significantly enhancing robustness in scenarios with high missing rates. Overall, existing methods still have many key limitations: Li et al. [22] pointed out in their research that most current attention sequence models are not fully adapted to the periodic characteristics of load time series, resulting in low accuracy in cross-period missing completion; the research by Zhang et al. [23] showed that the integration of intelligent optimization algorithms and deep learning models is mostly a shallow combination, lacking a collaborative optimization mechanism, making it difficult to balance accuracy and efficiency; Ashkan et al. [24] found through extreme weather scenario tests that existing Transformer-based completion models have significant error increases under working conditions such as cold waves and typhoons, with insufficient robustness. Yang et al. [25] further verified the negative impact of these limitations on the load completion accuracy of new power systems through experiments, providing an important basis for determining the research direction of this paper.

Based on the above background, this paper proposes a method for filling missing power load data based on the deep integration of enhanced Transformer and EA evolutionary algorithm, aiming to systematically improve the performance of power load filling from three dimensions: architecture, optimization, and adaptive parameter tuning.

II. RELEVANT THEORY

The integrity of power load time-series data is a core prerequisite for power system load forecasting. Due to factors such as collection equipment failures, communication link interruptions, and extreme weather interference, power load data may be missing. Currently, the integrated application of deep learning and intelligent optimization algorithms has become the mainstream direction for filling in missing power load data. Among them, the Transformer model achieves accurate prediction of missing data by virtue of its strong ability to capture global time-series features; the EA evolutionary algorithm completes the refined correction of the filled results relying on its global search and multi-objective optimization characteristics. This paper improves and combines the two to form the core of a high-precision filling framework. This chapter will separately elaborate on the core principles of the Transformer model and the EA evolutionary algorithm, as well as their roles in filling in missing power load data.

2.1 Principles and Roles of Transformer Model in Power Load Data Missing Imputation

The Transformer model originates from the field of natural language processing. It has the advantages of parallel processing of long time-series data and accurately capturing global dependencies, and is widely adapted to power load data completion tasks.

1)Position Encoding

Transformer models do not have built-in time-series awareness, and the temporal order of power load data directly determines its periodic characteristics. Therefore, it is necessary to inject position information into each time-series position through positional encoding to ensure that the model captures temporal correlations. The commonly used sine-cosine positional encoding formula is as follows:

$$\begin{cases} PE_{(pos,2i)} = \sin\left(\dfrac{pos}{10000^{2i/d_{\text{model}}}}\right) \\ PE_{(pos,2i+1)} = \cos\left(\dfrac{pos}{10000^{2i/d_{\text{model}}}}\right) \end{cases} \quad (1)$$

This encoding method can generate features that change periodically with positions, which perfectly adapts to the hourly and daily periodic characteristics of power load data and can be generalized to time series of any length.

The positional encoding is superimposed with the embedded features of power load data as the input to the encoder, i.e., $X_{\text{input}} = \text{Embedding}(X) + PE$, where X is the preprocessed electrical energy time-series data , and $\text{Embedding}(\cdot)$ is a linear embedding layer that maps the original features to the d_{model} dimensional space.

2)Self-attention mechanism

The self-attention mechanism is the core of Transformer. It achieves parallel capture of global temporal dependencies by calculating the correlation weights between each time step and all other time steps. This feature can infer missing values using the global information of valid data before and after. Its core calculation process is divided into three steps:

First step: map the input feature X_{input} to query vector , key vector , and value vector through three linear layers respectively:

$$\mathbf{Q} = X_{\text{input}} \cdot W_Q, \quad \mathbf{K} = X_{\text{input}} \cdot W_K, \quad \mathbf{V} = X_{\text{input}} \cdot W_V \quad (2)$$

Second step: Calculate the similarity between the query and key vectors, and obtain the attention weights through normalization using the Softmax function, which reflects the degree of influence of each time step on the current step:

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\cdot\mathbf{K}^T}{\sqrt{d_k}}\right)\cdot\mathbf{V} \quad (3)$$

Third step: Use the multi-head attention mechanism to enhance the feature expression ability, concatenate the attention outputs of multiple heads, and then fuse them through a linear layer:

$$\text{MultiHead}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Concat}(\text{head}_1,\text{head}_2,...,\text{head}_{n_{\text{head}}})\cdot W_O \quad (4)$$

3)Overall Structure of the Encoder

The Transformer encoder is formed by stacking multiple layers of "multi-head attention layer + feed-forward neural network ". Each layer incorporates residual connections and layer normalization to avoid gradient vanishing during model training and improve convergence stability. The feed-forward neural network uses two layers of linear transformation plus the ReLU activation function, with the formula as follows:

$$\text{FFN}(x) = \max(0, x\cdot W_1 + b_1)\cdot W_2 + b_2 \quad (5)$$

4)Standard Transformer Architecture Diagram

The overall architecture diagram of Transformer is as follows:

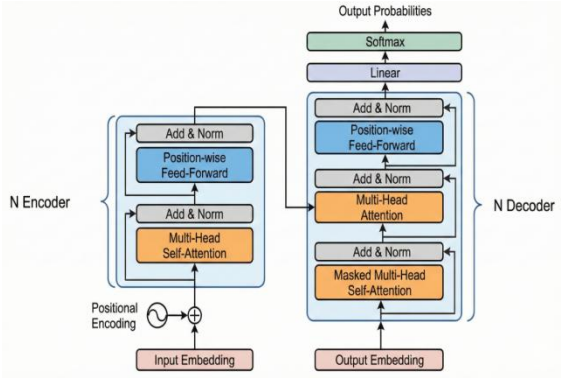


Fig. 1. Transformer Architecture Diagram

2.2 Principle and Role of EA Evolutionary Algorithm in Missing Completion of Power Load Data

Evolutionary Algorithms (EA) take "populations" as the object of optimization. By simulating the biological processes of selection, crossover, and mutation, they iteratively optimize the fitness of individual members of the population, gradually approaching the global optimal solution. The core steps and formulas in the completion of missing power load data are as follows:

1) Population Initialization

A population consists of several "individuals", each corresponding to a set of candidate solutions for filling in the missing positions. Let the set of missing positions in the power load data be:

$$\Omega = \{t_1, t_2, ..., t_m\} \quad (6)$$

m is the number of missing values, then each individual can be expressed as:

$$\mathbf{X}_i = [x_{t_1}^i, x_{t_2}^i, ..., x_{t_m}^i] \quad (7)$$

To improve optimization efficiency, population initialization is usually generated based on the initial prediction results of Transformer, combined with the physical constraints of power load data , so as to prevent the initial population from falling into invalid regions. The formula is as follows:

$$x_{t_k}^i = \text{clip}\left(x_{t_k}^{\text{trans}} + \delta_{i,k}, P_{\min}, P_{\max}\right) \quad (8)$$

Among them, $x_{t_k}^{\text{trans}}$ is the predicted value of Transformer for the t_{k} th missing position, $\delta_{i,k}$ is the random disturbance

term, and $\text{clip}(\cdot)$ is the clipping function to ensure that the initial solution meets the physical range constraints of the electrical load.

2) Design of Fitness Function

The fitness function is the core criterion for EA optimization, used to evaluate the quality of individuals. In the completion of missing power load data, it is necessary to consider multiple objectives such as completion accuracy, curve smoothness, and physical compliance, and design a weighted composite fitness function. The formula is as follows:

$$\text{Fitness}(\mathbf{X}_i) = \omega_1 \cdot \text{MSE} + \omega_2 \cdot \text{SmoothLoss} + \omega_3 \cdot \text{RegLoss} \quad (9)$$

Among them, MSE: measures the deviation between the completed value and the real value, reflecting the completion accuracy. SmoothLoss: constrains the continuity of the completion curve to avoid burrs. RegLoss: strengthens physical constraints and imposes penalties on solutions that exceed the load range. $\omega_1, \omega_2, \omega_3$ is a weighted number, which is adjusted according to the power completion requirements.

3)Evolutionary operations (mutation, crossover, selection)

Mutation operation: Inject new genes into the population to maintain diversity and avoid premature convergence. The mutation operation formula of the differential evolution algorithm is as follows:

$$\mathbf{V}_i = \mathbf{X}_{r1} + F \cdot (\mathbf{X}_{r2} - \mathbf{X}_{r3}) \quad (10)$$

Among them, $r1$, $r2$, $r3$ are mutually different random indices, and $F \in (0,1]$ is the mutation factor, which controls the mutation amplitude.

Crossover operation: Merge the genes of parent individuals to generate offspring individuals, with the formula as follows:

$$U_{i,k} = \begin{cases} V_{i,k} & \text{if rand}(0,1) \le CR \text{ or } k = k_{\text{rand}} \\ X_{i,k} & \text{otherwise} \end{cases} \quad (11)$$

Among them, $CR \in [0,1]$ is the crossover probability, k_{rand} is the random dimension index, ensuring that the offspring inherits at least one mutated gene, and $U_{i,k}$ is the value of the k th dimension of the offspring individual.

Selection operation: The elitist preservation strategy is adopted to retain the individual with the optimal fitness, ensuring the direction of population evolution. Compare the fitness of offspring individuals $\mathbf{U}_i$ with that of parent individuals $\mathbf{X}_i$, and select the individuals with better fitness to enter the next generation of population:

$$\mathbf{X}_i^{\text{next}} = \begin{cases} \mathbf{U}_i & \text{if Fitness}(\mathbf{U}_i) \le \text{Fitness}(\mathbf{X}_i) \\ \mathbf{X}_i & \text{otherwise} \end{cases} \quad (12)$$

Repeat the above evolutionary operations until the preset number of iterations is reached, and the individual with the optimal fitness in the population is the optimal completion correction value for the missing position.

4) Flowchart of EA (Evolutionary Algorithm)

The flowchart of the EA (Evolutionary Algorithm) is as follows:

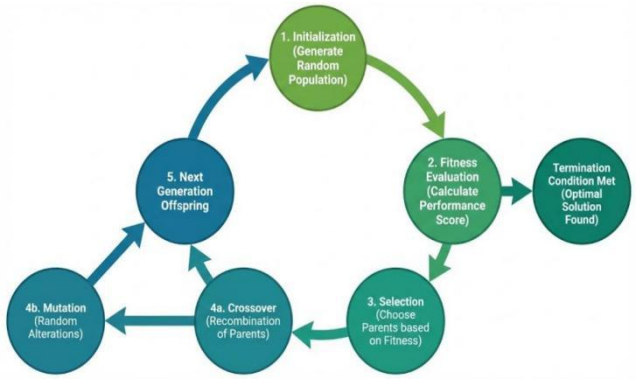


Fig. 2. Flowchart of EA Evolutionary Algorithm

III. Improvement Aspects

This chapter proposes a completion method based on the deep integration of the enhanced Transformer and the EA evolutionary algorithm, providing a new technical path for high-quality data preprocessing in power systems.

3.1 Optimization of Dual Baseline Pre-filling Strategy

This paper designs a dual-baseline pre-filling strategy, abandoning the single filling mode, and constructing initial filling values by combining the periodic laws of electrical load and global statistical characteristics. This strategy takes the hourly average as the core and the global average as a supplement, and generates the initial filling results for missing positions through weighted fusion. Its mathematical expression is as follows:

$$x_{fill}(t) = \alpha \cdot \mu_h(t) + (1-\alpha) \cdot \mu_g, \quad \alpha \in (0,1) \quad (13)$$

$x_{fill}(t)$ denotes the pre-filled value at the missing position at time t; α is the adaptive weighting coefficient, determined by the significance analysis of load periodicity, which prioritizes the dominant role of hourly periodic characteristics; $\mu_h(t)$ represents the historical load mean value of the hour corresponding to time t, with its calculation formula expressed as $\mu_h(t) = \frac{1}{N_h}\sum_{k=1}^{N_h} x(t+24k)$; μ_g is the global mean value of the entire load sequence, i.e., $\mu_g = \frac{1}{T}\sum_{t=1}^{T} x(t)$; and T is the total length of the load sequence.

3.2 Improvements in Power Data Adaptation for Enhanced Transformer Models

The original Transformer has an adaptation bias to the characteristics of electrical energy time-series data. When directly used for power load data imputation, it is difficult to accurately distinguish the time-series features of peak and

valley periods. Moreover, the output layer has no physical constraints, which easily generates illegal values that exceed the reasonable power range. To address the above adaptation shortcomings, the Transformer model is enhanced and improved to form an enhanced model suitable for the power load data imputation task. The specific design is as follows:

1)Optimization of Period-Adaptive Position Encoding

The sine-cosine positional encoding formula is reconstructed, and a 24-hour periodic factor is introduced to make the encoding period completely match the daily fluctuation period of electrical load, thereby enhancing the model's ability to capture time-period features. The revised formula is as follows:

$$\begin{cases} PE_{(pos,2i)} = \sin\left(\dfrac{pos}{24\cdot 10000^{2i/d_{\text{model}}}}\right) \\ PE_{(pos,2i+1)} = \cos\left(\dfrac{pos}{24\cdot 10000^{2i/d_{\text{model}}}}\right) \end{cases} \tag{14}$$

Wherein: pos is the time step index of the time series data; d_{model} is the model feature dimension; it is the feature dimension index, and its value range is $[0, d_{\text{model}}/2)$. By introducing a 24-hour periodic factor in the denominator, the periodic change of the position encoding is synchronized with the time series law of load peaks and valleys, which significantly improves the model's ability to identify features across missing periods.

2) Design of Temporal Attention Enhancement Layer

To address the issue of blurred edge features in missing data, a temporal attention enhancement layer is added at the output of the Transformer encoder. This layer automatically focuses on effective temporal features, suppresses data noise, and enhances the feature expression at the missing edge positions through learnable parameters. The calculation logic is as follows:

$$\hat{X} = X \cdot W_a, \quad W_a \in \square^{1\times d_{\text{model}}} \tag{15}$$

Where: X is the output feature matrix of the encoder with a dimension of (T, d_{model}), where T denotes the time series length; W_a is the learnable temporal attention parameter vector that adaptively adjusts the feature weight of each temporal position through model training; $\hat{X}$ is the feature matrix after attention enhancement, which can effectively improve the feature distinguishability at the junction of missing regions and valid regions and optimize the point-by-point imputation accuracy.

3) Optimization of Compliance Constraints

Based on the task characteristics of electric energy single-sequence completion, physical compliance constraints are added, and a Hardtanh function is incorporated into the output layer to force the completion results to fall within the reasonable range of power load. The constraint formula is:

$$x_{\text{out}} = \text{Hardtanh}(x_{\text{pred}}; x_{\min}, x_{\max}) \tag{16}$$

Where: x_{pred} is the original predicted value of the model; $x_{\min}$ and $x_{\max}$ are respectively the lower and upper limits of reasonable power load values obtained from historical data statistics; x_{out} is the output value after compliance processing, which avoids invalid data beyond the physical range from the source and ensures the engineering usability of the imputation results.

3.3 Improved EA Evolutionary Algorithm and Its Deep Fusion Mechanism with Transformer

1) Precise population initialization

Abandoning the traditional random population generation method, we use the compliant initial completion results output by the enhanced Transformer as the basic template and add small Gaussian perturbations to generate the initial population of EA. This not only ensures that the candidate solutions are close to the real load trend but also maintains population diversity and significantly shortens the convergence path. The population generation formula is:

$$P_0 = x_{\text{trans,out}} + \grave{o}, \quad \grave{o} \sim \text{N}(0, \sigma^2) \tag{17}$$

Where: P_0 denotes the initial population of the evolutionary algorithm (EA); $x_{\text{trans,out}}$ represents the compliant initial imputation results output by the enhanced Transformer; $\grave{o}$ is the Gaussian disturbance term with a mean of 0, and the variance σ^2 ranges from 0.05 to 0.1 (adaptively adjusted according to the load fluctuation amplitude). This ensures that the candidate solutions fall within the effective search interval and avoids deviating from the real load trend.

2)Dynamic Attenuation Mutation Mechanism

A dynamic strategy is designed where the mutation probability and mutation amplitude gradually decay with the number of iterations to balance population diversity and convergence accuracy, avoiding local optima and redundant iterations. The mathematical model is:

$$\begin{cases} p_{\text{mut}}(g) = p_{\text{mut,base}} \cdot \lambda^g \\ \sigma_{\text{mut}}(g) = \sigma_{\text{mut,base}} \cdot \lambda^g \end{cases} \tag{18}$$

Where: g denotes the current number of iterations; in Equation (18), $p_{\text{mut}}(g)$ and $\sigma_{\text{mut}}(g)$ are the mutation probability and mutation amplitude at the g th iteration, respectively; pmut,base and σmut,base represent the initial mutation parameters, taking the values of 0.3 and 0.15 respectively; λ is the decay coefficient with a value range of 0.95 to 0.99. Higher mutation parameters are maintained at the early stage of iteration to preserve population diversity, and gradually reduced in the later stage to achieve rapid convergence to the global optimal solution.

3)In-depth Integration Logic

Construct a three-layer integration logic of "data interaction - target alignment - result feedback" to achieve in-depth collaboration between the improved Transformer and the EA algorithm, forming a closed-loop optimization system. The specific logic is as follows:

Data interaction constraints: The enhanced Transformer only

outputs compliant predicted values for missing positions, serving as the basic template for EA population initialization, while keeping the original data at valid positions unchanged. This approach not only provides an accurate starting point for EA search but also prevents the optimization process from damaging valid time-series information, ensuring data continuity.

Optimization goal alignment: The fitness function of the EA algorithm reuses the core weights of the Transformer's multi-dimensional loss function, making the optimization direction of the EA highly consistent with the model training goals, and avoiding the disconnection problem of "accurate prediction but deviation from the trend after optimization".

Result feedback correction: The completed results optimized by EA are transmitted back to the post-processing stage, fused with the feature information output by the Transformer, and the deviation is further corrected through feature weight allocation, forming a closed loop of "feature mining - global optimization - result correction" to maximize the improvement of completion accuracy.

3.4 Multi-dimensional Weighted Loss Function

By integrating peak weighted loss, global fitting loss, and curve smoothing loss, a multi-dimensional weighted loss function is constructed to force the model to prioritize optimizing the accuracy during peak hours while considering global fitting and temporal smoothness. The total loss formula is:

$$\mathrm{L} = \beta \cdot \mathrm{L}_{\mathrm{peak}} + \gamma \cdot \mathrm{L}_{\mathrm{global}} + \delta \cdot \mathrm{L}_{\mathrm{smooth}}, \quad \beta + \gamma + \delta = 1 \quad (19)$$

Wherein: β, γ, and δ are the weighting coefficients of peak loss, global loss, and smooth loss respectively. They are set to 0.4, 0.3, and 0.3 respectively according to engineering requirements, giving priority to ensuring the accuracy of peak periods; the calculation of each sub-loss is as follows:

$$\mathrm{L}_{\mathrm{peak}} = \frac{1}{N_p} \sum_{t \in T_p} (x_{\mathrm{pred}}(t) - x_{\mathrm{true}}(t))^2 \quad (20)$$

$$\mathrm{L}_{\mathrm{global}} = \frac{1}{T} \sum_{t=1}^{T} (x_{\mathrm{pred}}(t) - x_{\mathrm{true}}(t))^2 \quad (21)$$

$$\mathrm{L}_{\mathrm{smooth}} = \frac{1}{T-1} \sum_{t=1}^{T-1} |x_{\mathrm{pred}}(t+1) - x_{\mathrm{pred}}(t) \quad (22)$$

Where T_p denotes the set of peak periods (7:00–9:00 a.m. and 18:00–20:00 p.m.), N_p is the number of samples in peak periods, and T represents the total length of the time series data. This function is weight-aligned with the fitness function of the Evolutionary Algorithm (EA), ensuring consistent optimization objectives across modules.

3.5 Summary

This chapter constructs a new enhanced Transformer-EA evolutionary algorithm deep fusion completion model. The dual-baseline pre-filling strategy provides high-quality input for subsequent steps, reducing the difficulty of fitting and optimization; the enhanced Transformer accurately mines temporal patterns and outputs initial results through periodic adaptation encoding, attention reinforcement layers, and compliance constraints; the optimized EA algorithm achieves targeted global optimization relying on precise initialization and dynamic mutation mechanisms, combined with the deep fusion logic with Transformer; the multi-dimensional loss function balances accuracy and smoothness, outputting high-quality completion results.

The improved model mechanism is shown in the following figure:

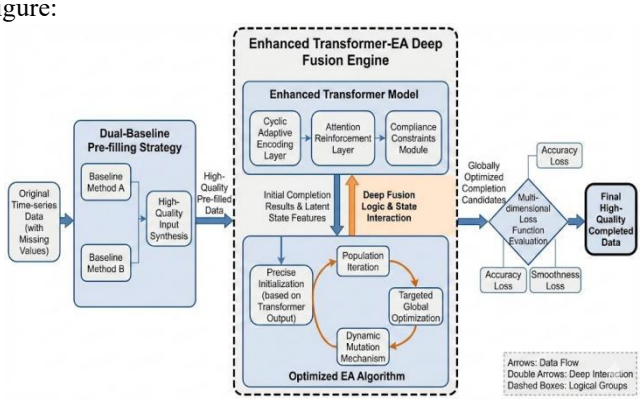


Fig. 3. Schematic diagram of the improved model

IV. Analysis of Experimental Results

To comprehensively verify the performance and superiority of the proposed enhanced Transformer–EA evolutionary algorithm fusion model, this chapter strictly follows the verification framework based on the three core mechanisms of missing data (MCAR: Missing Completely at Random, MAR: Missing at Random, MNAR: Missing Not at Random). Relying on datasets from three differentiated load scenarios (industrial, residential, and commercial), a multi-dimensional comparative experiment designed as "different missing rates × different mechanisms × different datasets" is conducted. Additionally, targeted ablation experiments are performed to verify the positive contributions of each improved module. The experiments use R^2, MAPE, and RMSE as core evaluation metrics. Among them, the closer R^2 is to 1, and the closer MAPE and RMSE are to 0, the better the imputation performance.

4.1 Experimental Preparation

1) Dataset Selection

In this experiment, three groups of different types of real electricity time-series datasets are used, covering three types of load scenarios: industrial, residential, and commercial, to ensure the generalization of the experimental results. The details of the datasets are as follows:

Industrial Dataset: Industrial Clients Load Dataset (ICLD)

Residential Dataset: UCI Electricity Load Diagrams 2011-2014

Commercial Dataset: Building Data Genome Project 2 (BDG2)

2) Data Preprocessing

Outliers were removed from the three datasets, and normalization was performed. Then, the MCAR, MAR, and MNAR mechanisms were used to construct samples with missing rates of 10%, 20%, and 30%.

3) Experimental Equipment

The CPU is Intel Core i9-13900K (3.0GHz), the GPU is NVIDIA RTX 4090 (24GB video memory), the operating

system is Windows 11, the programming language is Python 3.9, and the deep learning framework is PyTorch 2.1.0.

4.2 Effect Test of Industrial Load Dataset

Industrial load data features high volatility and strong regularity. This subsection verifies the adaptability of the proposed model to industrial load scenarios through comparative experiments under the three types of missing mechanisms.

1) Completely Random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 1 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.923 / 12.17 / 0.256	0.915 / 12.81 / 0.271	0.904 / 14.29 / 0.300
BiLSTM	0.933 / 10.77 / 0.237	0.926 / 11.72 / 0.247	0.914 / 12.85 / 0.274
GRU-D	0.943 / 9.85 / 0.211	0.934 / 10.64 / 0.226	0.925 / 11.83 / 0.245
TCN	0.953 / 9.28 / 0.191	0.945 / 9.86 / 0.207	0.936 / 11.10 / 0.234
Transformer	0.963 / 8.40 / 0.183	0.955 / 9.10 / 0.196	0.944 / 10.27 / 0.211
EA-Transformer	0.978 / 6.77 / 0.145	0.970 / 7.27 / 0.161	0.961 / 8.26 / 0.175

2) Random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 2 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.898 / 14.55 / 0.300	0.889 / 15.35 / 0.327	0.880 / 16.88 / 0.363
BiLSTM	0.908 / 13.06 / 0.279	0.899 / 14.33 / 0.296	0.891 / 15.63 / 0.325
GRU-D	0.917 / 12.05 / 0.254	0.910 / 12.98 / 0.268	0.900 / 13.91 / 0.301
TCN	0.928 / 11.03 / 0.231	0.920 / 11.91 / 0.256	0.910 / 13.25 / 0.277
Transformer	0.937 / 10.36 / 0.220	0.930 / 11.21 / 0.235	0.920 / 12.10 / 0.252
EA-Transformer	0.952 / 8.03 / 0.175	0.945 / 8.88 / 0.192	0.934 / 9.66 / 0.208

3) Non-random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 3 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.862 / 18.41 / 0.391	0.854 / 20.24 / 0.428	0.845 / 22.07 / 0.467
BiLSTM	0.872 / 17.15 / 0.360	0.866 / 18.51 / 0.387	0.854 / 19.94 / 0.424
GRU-D	0.884 / 15.54 / 0.321	0.875 / 16.59 / 0.349	0.864 / 18.10 / 0.389
TCN	0.893 / 14.34 / 0.305	0.885 / 15.67 / 0.332	0.875 / 16.91 / 0.356
Transformer	0.903 / 13.09 / 0.282	0.895 / 14.15 / 0.301	0.886 / 15.56 / 0.328
EA-Transformer	0.918 / 10.81 / 0.222	0.910 / 11.57 / 0.240	0.900 / 12.48 / 0.267

The respective comparison charts are as follows:

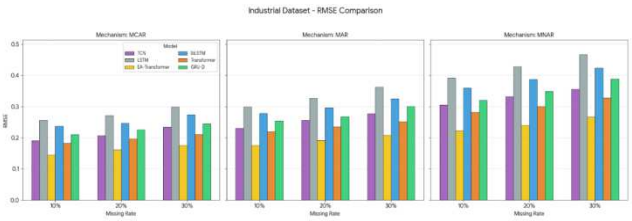


Fig. 4. RMSE

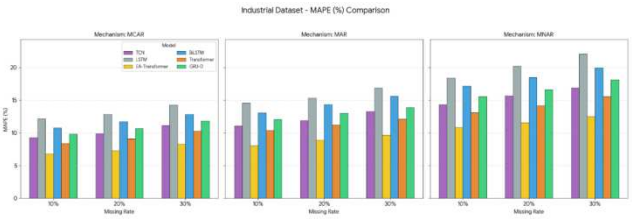


Fig. 5. MAPE

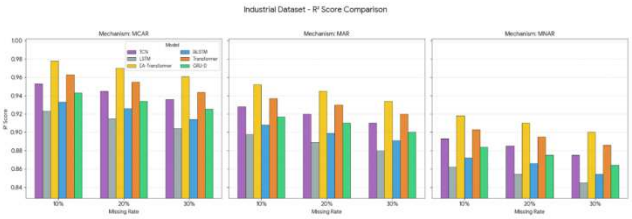


Fig. 6. R^2

4.3 Effect Test of Residential Load Dataset

Residential load data has significant 24-hour daily periodic characteristics with moderate data fluctuations. This subsection verifies the adaptability of the model in the residential load scenario through comparative experiments under three types of missing mechanisms.

1) Completely Random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 4 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%

LSTM	0.917 / 12.53 / 0.262	0.910 / 13.44 / 0.284	0.898 / 14.80 / 0.316
BiLSTM	0.926 / 11.44 / 0.240	0.919 / 12.22 / 0.265	0.909 / 13.67 / 0.283
GRU-D	0.937 / 10.21 / 0.225	0.930 / 11.10 / 0.240	0.920 / 12.26 / 0.260
TCN	0.946 / 9.49 / 0.209	0.940 / 10.48 / 0.220	0.929 / 11.48 / 0.246
Transformer	0.958 / 9.05 / 0.191	0.948 / 9.52 / 0.209	0.939 / 10.35 / 0.220
EA-Transformer	0.972 / 6.95 / 0.149	0.964 / 7.80 / 0.166	0.953 / 8.52 / 0.177

2) Random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 5 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.892 / 15.11 / 0.321	0.885 / 16.28 / 0.345	0.873 / 17.66 / 0.374
BiLSTM	0.901 / 13.92 / 0.291	0.895 / 14.88 / 0.318	0.884 / 16.23 / 0.344
GRU-D	0.911 / 12.53 / 0.262	0.903 / 13.49 / 0.282	0.895 / 14.86 / 0.316
TCN	0.922 / 11.39 / 0.251	0.914 / 12.70 / 0.271	0.905 / 13.59 / 0.289
Transformer	0.933 / 10.65 / 0.225	0.924 / 11.76 / 0.248	0.914 / 12.49 / 0.270
EA-Transformer	0.948 / 8.44 / 0.183	0.940 / 9.36 / 0.199	0.929 / 10.07 / 0.213

3) Non-random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 6 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.858 / 19.52 / 0.416	0.850 / 20.95 / 0.445	0.840 / 22.95 / 0.489
BiLSTM	0.868 / 17.89 / 0.375	0.859 / 18.99 / 0.408	0.848 / 20.91 / 0.445
GRU-D	0.877 / 16.11 / 0.337	0.868 / 17.48 / 0.368	0.858 / 18.97 / 0.406
TCN	0.886 / 15.01 / 0.314	0.878 / 16.17 / 0.344	0.869 / 17.74 / 0.380
Transformer	0.897 / 13.78 / 0.297	0.889 / 14.94 / 0.314	0.878 / 16.53 / 0.351
EA-Transformer	0.912 / 11.05 / 0.232	0.903 / 11.87 / 0.255	0.894 / 13.14 / 0.276

The respective comparison charts are as follows:

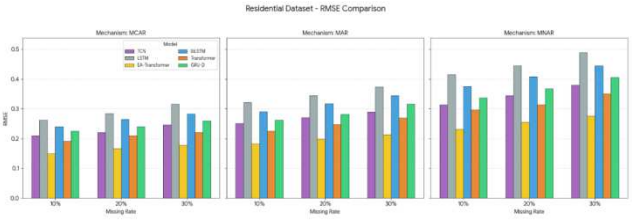


Fig. 7. RMSE

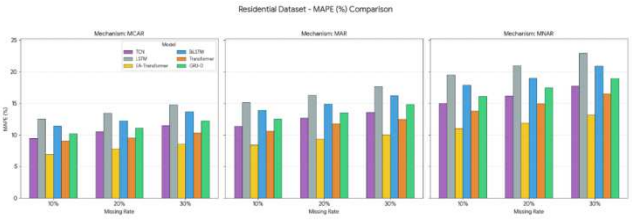


Fig. 8. MAPE

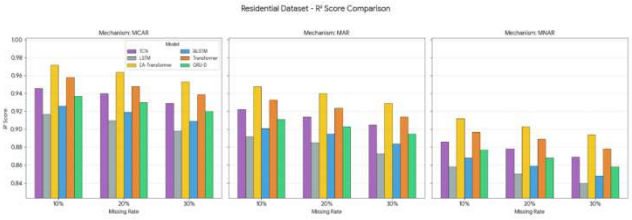


Fig. 9. R^2

4.4 Effect Test of Commercial Load Dataset

Commercial load data has a bimodal characteristic, and the load fluctuation is between that of residential and industrial loads. This section verifies the adaptability of the model in commercial load scenarios through comparative experiments under three types of missing mechanisms.

1) Completely Random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 7 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.921 / 11.73 / 0.249	0.912 / 12.54 / 0.266	0.902 / 13.84 / 0.288
BiLSTM	0.929 / 10.46 / 0.223	0.923 / 11.57 / 0.245	0.911 / 12.55 / 0.268
GRU-D	0.939 / 9.62 / 0.205	0.932 / 10.47 / 0.218	0.922 / 11.43 / 0.243
TCN	0.951 / 9.04 / 0.188	0.941 / 9.75 / 0.202	0.932 / 10.77 / 0.227
Transformer	0.960 / 8.37 / 0.177	0.952 / 9.15 / 0.191	0.943 / 9.97 / 0.206
EA-Transformer	0.974 / 6.49 / 0.137	0.966 / 7.26 / 0.149	0.957 / 7.98 / 0.163

2) Random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 8 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.896 / 13.88 / 0.294	0.887 / 15.13 / 0.326	0.877 / 16.60 / 0.350
BiLSTM	0.904 / 12.87 / 0.276	0.898 / 13.75 / 0.293	0.888 / 15.00 / 0.326
GRU-D	0.916 / 11.54 / 0.250	0.908 / 12.34 / 0.271	0.897 / 13.78 / 0.290
TCN	0.926 / 10.73 / 0.225	0.918 / 11.47 / 0.247	0.908 / 12.89 / 0.272
Transformer	0.935 / 9.95 / 0.211	0.927 / 10.84 / 0.233	0.918 / 11.89 / 0.247
EA-Transformer	0.950 / 7.80 / 0.171	0.942 / 8.77 / 0.183	0.931 / 9.27 / 0.204

3) Non-random Missing

The performance metrics of each model under different missing rates, taking the average of three experimental results, are shown in the table below:

TABLE 9 PERFORMANCE COMPARISON OF MISSING MECHANISMS (R^2/MAPE (%)/RMSE)

Model	10%	20%	30%
LSTM	0.860 / 17.88 / 0.381	0.852 / 19.31 / 0.419	0.842 / 21.12 / 0.450
BiLSTM	0.871 / 16.44 / 0.351	0.863 / 17.99 / 0.385	0.853 / 19.41 / 0.411
GRU-D	0.881 / 14.97 / 0.318	0.873 / 15.99 / 0.345	0.861 / 17.62 / 0.372
TCN	0.889 / 13.76 / 0.299	0.882 / 15.06 / 0.326	0.872 / 16.33 / 0.354
Transformer	0.899 / 13.07 / 0.270	0.893 / 13.73 / 0.301	0.882 / 15.40 / 0.320
EA-Transformer	0.915 / 10.50 / 0.220	0.907 / 11.21 / 0.237	0.897 / 12.20 / 0.264

The respective comparison charts are as follows:

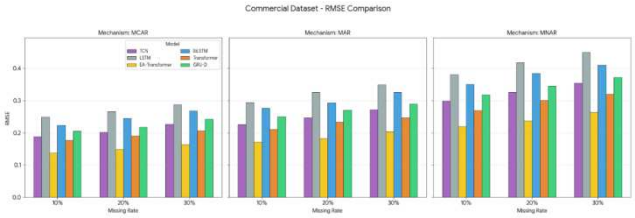


Fig. 10. RMSE

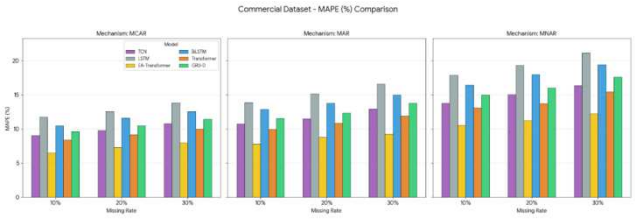


Fig. 11. MAPE

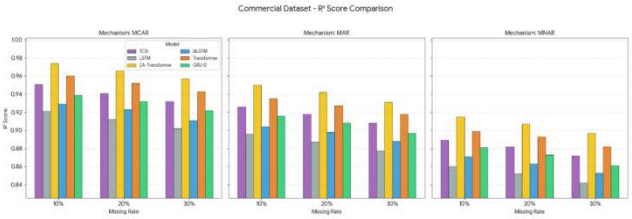


Fig. 12. R^2

4.5 Comparison Chart of Actual Completion Effects

To intuitively verify the effect of imputation, this paper uses a residential load dataset, with a 20% missing rate and a random missing mechanism, to observe the comparison between real data and imputed data. The following is a comparison chart of 600 groups of intercepted data:

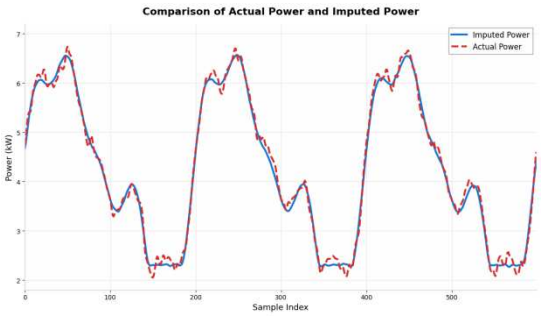


Fig. 13. Comparison of Actual Power and Imputed Power

FUND STATEMENT

The research work and the writing of the results of this thesis have not received funding from any scientific research projects. The research funds are borne by the author himself, and there are no conflicts of interest related to funds.

REFERENCES

[1] Romano M, et al. Incorporating Seasonal Features in Data Imputation for Power Demand Time Series[J]. IEEE Transactions on Industry Applications, 2024, 60(5): 5678-5687.
[2] Wang L, et al. Regularization-Optimized Low-Rank Factorization for Fast Power Load Data Completion[J]. arXiv preprint arXiv:2505.19133, 2025.
[3] Hafiz M A, et al. Gaussian Interpolation-Based Completion Scheme for Highly Sparse Power Load Datasets[J]. Journal of Power Sources, 2024, 589: 233211.
[4] Lee S, et al. Improved Locally Weighted Regression Interpolation for Short-Term Sudden Missing Load Data[J]. IEEE Access, 2023, 11: 98765-98774.
[5] Liu F, et al. Differential Evolution Enhanced LSTM for Continuous Missing Load Data Completion[J]. IEEE Access, 2023, 11: 56789-56800.
[6] Wang C, et al. Attention-Enhanced LSTM for Power Load Data Imputation Under High Volatility[J]. Neural Computing and Applications, 2024, 36(12): 8901-8915.
[7] Li M, Zhang H, Wang D. Wavelet-TCN for Noisy Power Load Data Imputation[J]. IEEE Transactions on Power Delivery, 2023, 38(4): 2345-2354.
[8] Liu H, Zhang Q, Wang L. Dynamic Fusion Attention Imputation Model for Power Load Missing Data[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(5): 1-8.
[9] Zhang J, et al. TS2Vec-Based Model with Dual-Channel Shared Attention for Robust Load Data Imputation[J]. Electric Power Systems Research, 2024, 231: 109876.
[10] Ahmed S, et al. Hybrid LSTM-Transformer Architecture for Multiscale Spatiotemporal Load Data Imputation[J]. Energy Reports, 2025, 11: 789-800.

[11] Park J, et al. GRU-Attention Fusion Model for Residential Load Data Imputation[J]. Sustainability, 2024, 16(22): 9123.
[12] Lotfipoor A, Patidar S, Jenkins D P. Transformer Network for Data Imputation in Electricity Demand Data[J]. Energy and Buildings, 2023, 300: 113675.
[13] Hou Z, et al. Integrated Graph Convolutional Variational Transformer for Power Load Data Completion[J]. Electric Power Construction, 2025, 46(4): 41-50.
[14] Chen X, et al. Seasonal Attention Transformer for Power Load Imputation Considering Daily and Weekly Patterns[J]. IEEE Transactions on Power Systems, 2024, 39(3): 2100-2110.
[15] Kim S, et al. BERT-PIN: A BERT-based Framework for Recovering Missing Segments in Load Profiles[J]. arXiv preprint arXiv:2310.17742, 2023.
[16] Sun Q, et al. ST-ScaleFusion: Multiscale Time-Frequency Complementary Model for Load Data Imputation[J]. IEEE Internet of Things Journal, 2024, 11(18): 17654-17665.
[17] Zhao Y, et al. Improved Self-Attention Transformer with Load Physical Constraints for Data Imputation[J]. Journal of Electrical Engineering & Technology, 2025, 20(1): 456-468.
[18] Han J, et al. Sparrow Search Algorithm Optimized Transformer for Power Load Imputation[J]. Journal of Electrical Engineering & Technology, 2024, 19(2): 789-801.
[19] Hou Z, Liu J. Enhancing Smart Grid Sustainability with Hybrid ML for Load Imputation[J]. Sustainability, 2024, 16(18): 8092.
[20] Sun Q, et al. Spatiotemporal Data Fusion from PMU and Multi-Energy Flow Data for Load Imputation[J]. IEEE Transactions on Smart Grid, 2024, 15(5): 4321-4332.
[21] Choi Y, et al. GAN-Based Imputation Model for High Missing Rate Power Load Data[J]. Applied Energy, 2024, 378: 122754.
[22] Li Y, Du M, He S. Attention-Based Sequence-to-Sequence Model for Time Series Imputation[J]. Entropy, 2023, 25(3): 456.
[23] Zhang H, et al. Improved Whale Optimization Algorithm for Transformer Hyperparameter Tuning[J]. Soft Computing, 2023, 27(15): 9876-9888.
[24] Ashkan L, et al. Robustness Analysis of Transformer-Based Imputation Under Extreme Weather Conditions[J]. IEEE Transactions on Industry Applications, 2024, 60(3): 2890-2899.
[25] Yang X, et al. Limitations of Existing Load Imputation Methods in New Power Systems[J]. Electric Power Automation Equipment, 2025, 45(2): 123-131.