

## ARTICLE

# Advanced Predictive Analytics for Green Energy Systems: An IPSS System Perspective

Lei Shen<sup>1,2</sup>, Chutong Zhang<sup>2</sup>, Yuwei Ge<sup>1</sup>, Shanyun Gu<sup>1</sup>, Qiang Gao<sup>1</sup>, Wei Li<sup>1</sup> and Jie Ji<sup>2,\*</sup>

<sup>1</sup>Electric Engineering Department, Huaian Hongneng Group Co., Ltd., Huaian, 223002, China

<sup>2</sup>Faculty of Automation, Huaiyin Institute of Technology, Huaian, 223002, China

\*Corresponding Author: Jie Ji. Email: jijie@hyit.edu.cn

Received: 14 November 2024; Accepted: 20 January 2025; Published: 31 March 2025

**ABSTRACT:** The rapid development and increased installed capacity of new energy sources such as wind and solar power pose new challenges for power grid fault diagnosis. This paper presents an innovative framework, the Intelligent Power Stability and Scheduling (IPSS) System, which is designed to enhance the safety, stability, and economic efficiency of power systems, particularly those integrated with green energy sources. The IPSS System is distinguished by its integration of a CNN-Transformer predictive model, which leverages the strengths of Convolutional Neural Networks (CNN) for local feature extraction and Transformer architecture for global dependency modeling, offering significant potential in power safety diagnostics. The IPSS System optimizes the economic and stability objectives of the power grid through an improved Zebra Algorithm, which aims to minimize operational costs and grid instability. The performance of the predictive model is comprehensively evaluated using key metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination ( $R^2$ ). Experimental results demonstrate the superiority of the CNN-Transformer model, with the lowest RMSE and MAE values of 0.0063 and 0.00421, respectively, on the training set, and an  $R^2$  value approaching 1, at 0.99635, indicating minimal prediction error and strong data interpretability. On the test set, the model maintains its excellence with the lowest RMSE and MAE values of 0.009 and 0.00673, respectively, and an  $R^2$  value of 0.97233. The IPSS System outperforms other models in terms of prediction accuracy and explanatory power and validates its effectiveness in economic and stability analysis through comparative studies with other optimization algorithms. The system's efficacy is further supported by experimental results, highlighting the proposed scheme's capability to reduce operational costs and enhance system stability, making it a valuable contribution to the field of green energy systems.

**KEYWORDS:** Advanced predictive analytics; green energy systems; IPSS system; CNN-transformer predictive model; economic and stability optimization; improved zebra algorithm

## 1 Introduction

In modern society, stable electricity supply is indispensable for almost all industries and daily life, ranging from household electricity to industrial production, from commercial operations to public services. The stable supply of electricity is the foundation for maintaining the normal operation of society. Therefore, ensuring the stability of electricity is not only the core objective of power system operation but also a key guarantee for the stable development of the social economy.

With the rapid development and increasing installed capacity of new energy sources such as wind and solar energy, the operational characteristics and structure of the power grid are undergoing significant



changes. These changes bring new challenges to the fault diagnosis of the power grid. Firstly, the intermittency of wind and solar energy leads to drastic fluctuations in grid load, increasing the uncertainty of grid operation and making fault diagnosis more complex. Secondly, the integration of new energy into the grid poses higher requirements for grid stability and voltage control, necessitating more precise monitoring and diagnostic technologies to detect and handle potential faults in a timely manner. Liu et al. [1] proposed a dynamic vulnerability assessment method that addresses the uncertainty issues in the safety and stability of hybrid power systems by considering the random fluctuations of wind power. Eskandari et al. [2] introduced a comprehensive six-layer multilayer model that diagnoses, classifies, and identifies the severity of electrical faults in photovoltaic arrays, solving the problem of early detection of photovoltaic system faults to improve their service life and reliability. Ajagekar et al. [3] proposed a hybrid QC-based deep learning framework solving the issues of computational workload and diagnostic performance quality, and improving computational efficiency. Hao et al. [4] proposed a method for the operation status assessment and fault diagnosis of photovoltaic arrays based on an enhanced Adaptive Neuro-Fuzzy Inference System (ANFIS), which effectively improves the precision and stability of the evaluator by improving ANFIS, solving the problem of fault diagnosis in photovoltaic arrays.

Existing methods seldom apply predictive technology to fault diagnosis, and the application of predictive technology in fault diagnosis remains a relatively new and underutilized field. There is a wide variety of predictive technologies that cover multiple domains and scenarios, demonstrating strong potential and flexibility. The continuous development and innovation of these technologies have led to their extensive application across various industries.

Liu et al. [5] developed the ResConvLSTM-Att model, which integrates residual networks, ConvLSTM, and attention mechanisms, effectively improving the spatiotemporal prediction accuracy of forest cover changes. Bharatheedasan et al. [6] suggested an integrated method that merges the capabilities of feedforward Multilayer Perceptrons (MLP) and Long Short-Term Memory networks (LSTM) to tackle the challenges of fault identification and Remaining Useful Life (RUL) prediction in rolling bearings, which in turn improves the efficacy of proactive maintenance plans. Hu et al. [7] proposed a deep learning method based on 1D-CNN and 1D-CNN-LSTM hybrid neural networks, which leverages existing big data on rolling bearing lifespan, targeting bearing wear for lifespan prediction, and solving the problem of time-consuming and costly bearing lifespan testing. Lu et al. [8] proposed a CNN-LSTM prediction model that addresses the accuracy of forecasting chaotic time series and the construction of a more plausible phase space configuration through traversal search in phase space, parameter judgment by incremental attention layer, and spatiotemporal feature extraction.

With the advancement of deep learning technology, the Transformer model has shown tremendous potential across various industries. Zhao et al. [9] introduced a Transformer supported by Spatiotemporal Graph Attention (STGA) for multivariate, multi-step residential load forecasting. This method bridges the gap between modeling long-term dependencies in load sequences and capturing dynamic spatial correlations among multiple households, solving the issue of accurately predicting residential load across variables and time steps. Ata et al. [10] proposed a novel hybrid CNN-SKIPGRU architecture based on the multi-head attention transformer architecture for predicting traffic and its long-term periodic dependencies. This method effectively combines CNNs and GRUs to weigh feature correlations, utilizing CNNs for spatial features and GRU-SKIP for long-term dependencies, and assigns weights through multi-head attention, addressing the challenge of traffic flow prediction with long-term dependencies. Ding et al. [11] presented a hybrid encoder-decoder model (CTH-Net) based on Convolutional Neural Networks (CNN) and Transformer, leveraging the advantages of these approaches. The proposed CTH-Net provides better skin lesion segmentation performance in both quantitative and qualitative testing compared to state-of-the-art methods,

solving the problem of accurate segmentation in medical imaging. Wu et al. [12] introduced a novel hybrid deep learning model named CTransCNN. The framework exhibits strong competitiveness, and its robust generalization ability makes it applicable to other medical multilabel image classification tasks, addressing the challenge of handling multi-label classification in medical imaging. Ahn et al. [13] proposed a technology based on artificial intelligence deep learning to directly predict the quantity of harmful cyanobacteria cells. The authors conducted advanced analysis by combining CNN and Transformer algorithms, solving the problem of accurately predicting harmful algal blooms. Hendria et al. [14] proposed a fusion of transformer-based and convolutional neural network (CNN) models, integrating Swin Transformer and DetectoRS with the ResNet backbone, addressing the challenge of integrating diverse object detection models for improved performance. The advantages and disadvantages of each model are shown in Table 1.

**Table 1:** Comparison of the advantages and disadvantages of each model

Model	References	Advantages	Disadvantages
LSTM	Reference [6]	1. The model has a simple structure 2. The model is good at handling sequential data	The ability to learn from long sequence dependencies is limited
LSTM-Attention	Reference [5]	The attention mechanism enhances the ability to capture key information in long sequences	High computational complexity
CNN-LSTM	Reference [7,8]	CNN extracts local features, LSTM handles temporal information, suitable for spatiotemporal feature fusion tasks	The model has a large number of parameters, resulting in relatively long training times
Transformer	Reference [9]	1. Suitable for long sequence tasks 2. Strong parallelization capability, fast training speed	The model has a large number of parameters and requires high computational resources for training
CNN-Transformer	Reference [12–14]	CNN extracts local information, Transformer captures global information	The computational load and storage requirements are high

It can be observed that the integration of Convolutional Neural Networks (CNN) and Transformer models is rarely applied in power safety diagnostics. Combining CNN and Transformer can construct a more powerful model for power safety diagnostics. This hybrid model enhances the modeling capability of global dependency relationships while maintaining CNN's local feature extraction ability.

Grid stability and economic analysis are two important indicators for power system planning and operation. With the large-scale integration of renewable energy sources such as wind and solar, grid stability and economic analysis have become more complex. On one hand, the volatility and uncertainty of renewable energy pose challenges to grid stability, requiring advanced forecasting technologies, energy storage systems, and flexible scheduling strategies to address; on the other hand, the low-cost characteristics of renewable energy help reduce generation costs and improve grid economic performance.

Bagherzadeh et al. [15] introduced a market clearing price model based on the coordination between Transmission System Operators and Distribution System Operators. The model addressed the challenge of optimizing network performance in the context of renewable energy integration. Jia et al. [16] proposed a Finite-Time Economic Model Predictive Control (MPC) strategy aimed at managing frequency control and achieving optimal power distribution across various regions within interconnected power systems. Yoon et al. [17] demonstrated the criticality of real-time detection of Power Quality Disturbances (PQD) in Electrical Power Systems (EPS) for swift action to protect against cascading damages, addressing the need for immediate response to power quality issues. Titz et al. [18] developed an interpretable machine learning model using Gradient Boosting Trees and SHAP values for hourly rescheduling and counter-trading volumes. This model reveals driving and mitigating factors and quantifies their post-event impact, addressing the challenge of understanding and managing the factors influencing power system operations. Wu et al. [19] designed a risk-value-based optimization model for power systems, replacing the continuous variables in the PSO algorithm with discrete quantities and solving it using a Discrete Binary Particle Swarm Optimization algorithm. The model addressed the challenge of achieving sustainable and stable grid operations. Therefore, when conducting stability and economic analysis of power grids, it is essential to consider various factors and employ scientific analysis methods to achieve sustainable development of the power grid.

This paper proposes a new framework to cleaner energy: intelligent power stability and scheduling (IPSS) system for integrated green energy systems. Firstly, this paper achieves the diagnosis of microgrid faults by constructing a CNN-Transformer prediction model; secondly, this paper takes the economic and stability of the power grid as the objective function and solves it using the improved Zebra Algorithm, thereby obtaining the minimum operating cost of the power grid and the minimum power grid instability. Finally, this paper comprehensively evaluates the performance of the prediction model using three indicators: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination ( $R^2$ ), which reveal the prediction accuracy and fit degree of the model from different perspectives. By comparing different prediction models, the economic and stability of the system are analyzed to verify the rationality and accuracy of this system. The system framework of this paper is shown in Fig. 1.

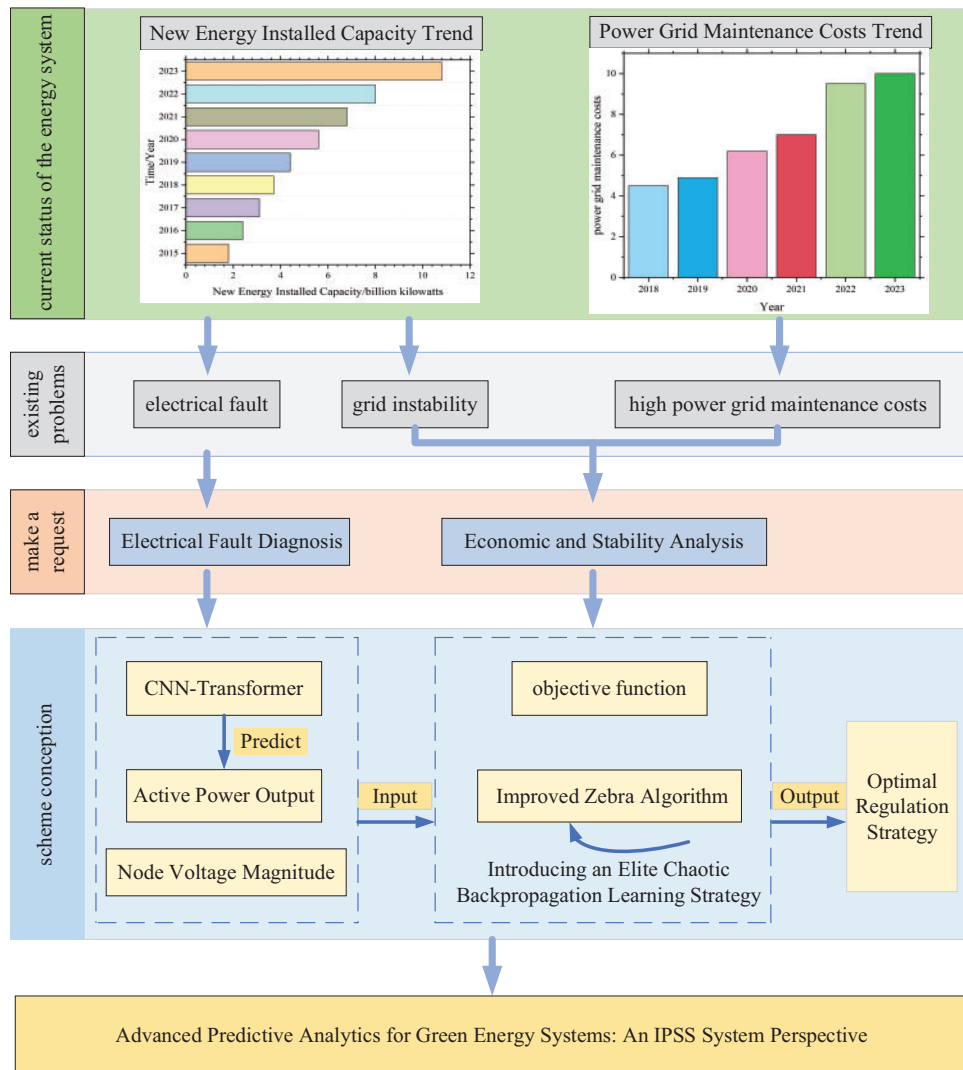


Figure 1: Overall research roadmap

## 2 CNN-Transformer Model

### 2.1 Convolutional Neural Networks

Convolutional neural network (CNN) uses the convolution layer to extract the temporal and spatial features of the voltage data, and its output  $O_I$  can be expressed as:

$$O_I = f(W_I^* I + b_I) \quad (1)$$

where  $W_I$  and  $b_I$  represent the weight and bias of the first convolutional layer, respectively;  $*$  represents the convolution operation;  $f$  is an activation function, usually using the corrected linear unit.

### 2.2 Transformer Model

The Transformer model, through its attention mechanism, can effectively handle long sequences of data, avoiding the vanishing gradient problem present in traditional RNNs, such as LSTMs.

The multi-head attention mechanism of the Transformer allows for simultaneous attention to different parts of the input sequence, thereby extracting more layered feature information.

The main body of the Transformer consists of an encoder and a decoder. The encoder and decoder primarily contain three components: input, multi-head attention, and fully connected feedforward neural networks.

Considering the temporal order in the input sequence, positional encoding is added after word encoding. Positional encoding mainly encodes the input sequence through sine and cosine functions to generate a fixed position matrix representing the temporal order in the sequence. The expression is as follows:

$$P(p, 2i) = \sin\left(\frac{p}{10000^{\frac{2i}{d_m}}}\right) \quad (2)$$

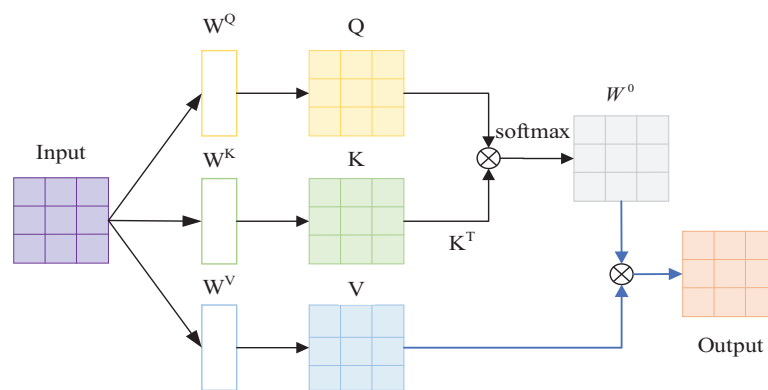
$$P(p, 2i + 1) = \cos\left(\frac{p}{10000^{\frac{2i}{d_m}}}\right) \quad (3)$$

where  $P$  represents the position coding matrix;  $p$  represents the fixed position in the input sequence;  $i$  represents the number of the input sequence;  $d_m$  represents the model dimension.

The multi-head attention section is constructed by combining the output results of multiple self-attentions. Each set of attention maps the input into distinct sub-representational spaces, allowing the model to concentrate on various positions within these spaces. The computation of the Transformer's multi-head attention is shown in Fig. 2. The formula for the self-attention mechanism is as follows:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where  $Attention$  represents the attention function;  $Q, K, V$  represent the vector converted into each element of the input sequence;  $\text{softmax}$  represents the activation function;  $d_k$  represents the dimension of the input sequence.



**Figure 2:** Transformer Multiple-head attention calculation process diagram

The result from the multi-head self-attention is acquired by merging the results from each individual attention head and then multiplying the concatenated result with a weight matrix, which can be expressed

as follows:

$$MultiHead(Q, K, V) = Concat(H_1, \dots, H_h) W^0 \quad (5)$$

$$H_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (6)$$

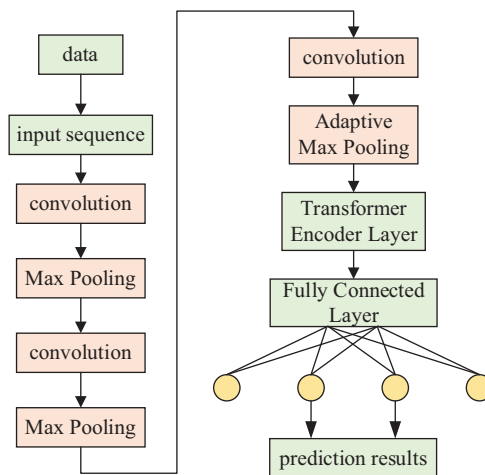
where  $W^0$  represents the weight matrix;  $W_i^Q$ ,  $W_i^K$ ,  $W_i^V$  represent the projection matrices for the  $i$ -th head;  $i \in [1, h]$ ,  $h$  represents the number of heads.

The fully connected feedforward neural network is used to transform the output of the multi-head attention in the spatial domain. The expression is as follows:

$$FFN(x) = \max(0, xW_1 + b_1) W_2 + b_2 \quad (7)$$

### 2.3 CNN-Transformer Model

In traditional CNNs, the convolution operation can effectively capture local features and spatial relationships, but it lacks the ability to model global context. The Transformer model, on the other hand, achieves global context modeling through its self-attention mechanism. By combining CNNs with the Transformer model, multi-dimensional extraction of information from voltage data can be realized. The structure of the CNN-Transformer model is shown in Fig. 3.



**Figure 3:** Structural diagram of the CNN-Transformer model

The workflow of this combined model is roughly divided into the following three steps:

- (1) The convolutional network is used to receive the input feature sequence and perform convolution operations to process the data and extract feature information;
- (2) Encoding and decoding are performed through the Transformer, ultimately mapping the encoded output to the predicted results;
- (3) The predicted results are output.

### 3 Comprehensive Evaluation System

#### 3.1 L-Index and Simplified L-Index

##### 3.1.1 L-Index

The L-index is a voltage stability indicator for power systems, with a value range of 0–1, which can provide normalized numerical values for different power system models. The L-index has the advantages of fast calculation speed, clear physical concept, fixed upper and lower limits, and can provide normalized values for different systems. It has been applied in practical power grids. The expression for the L-index is as follows:

$$L_j = \left| \frac{\sum_{k \in \alpha_G} H_{LG_{jk}} \dot{V}_{Gk}}{\dot{V}_{L_j}} \right| = \frac{\left| \sum_{i \in \alpha_L} Z_{LL_{ji}} \left( \frac{S_i}{\dot{V}_i} \right)^* \right|}{V_{L_j}} \quad (8)$$

where  $L_j$  represents the local voltage stability index of the load bus  $j$ ;  $\alpha_G$  represents the set of PV buses;  $\alpha_L$  represents the set of load buses,  $j \in \alpha_L$ ;  $H_{LG_{jk}}$  represents the element in the  $j$ -th row and  $k$ -th column of the load participation factor matrix  $H_{LG}$ ;  $\dot{V}_{Gk}$  represents the voltage phasor of the  $k$ -th PV bus;  $\dot{V}_{L_j}$  and  $\dot{V}_i$  represent the voltage phasors of load buses  $i$  and  $j$ , respectively;  $V_{L_j}$  represents the voltage magnitude of bus;  $Z_{LL_{ji}}$  represents the mutual impedance between load buses  $i$  and  $j$ ;  $S_i$  represents the system load power of the node; The upper corner landmark  $*$  represents the conjugation.

The stability index of all load nodes jointly forms the stability index vector ( $L' = [L_1, L_2, \dots, L_n]$ ,  $n \in \alpha_L$ ) of the power grid, and the L-index of the system is defined as follows:

$$L = \max_{j \in \alpha_L} |L_j| = \| L' \|_{\infty} \quad (9)$$

where  $L$  represents the L-index of the system.

The relationship between the L-index and the voltage stability of the system is:

$$\begin{cases} L < 1, \text{ voltage stability} \\ L = 1, \text{ The system is at the critical state of voltage stability} \\ L > 1, \text{ voltage instability} \end{cases} \quad (10)$$

##### 3.1.2 Simplified L-Index

Due to the complex number operations involved in the expression of the L-index, the calculation is relatively complex, and the computational load increases rapidly with the expansion of the power grid scale. However, in actual low-voltage power grids, the resistance of the lines is often not negligible compared to the reactance, while the change in voltage phase is relatively small. The simplified L-index formula, which neglects the change in voltage phase, is as follows:



$$\left\{ \begin{array}{l} L_{sj} = \frac{1}{V_j} \sqrt{(f_s)^2 + (g_s)^2} \\ f_s = \sum_{i \in \alpha_L} \frac{f_{si}}{V_i} \\ f_{si} = P_i R_{ij} + Q_i X_{ij} \\ g_s = \sum_{i \in \alpha_L} \frac{g_{si}}{V_i} \\ g_{si} = Q_i R_{ij} - P_i X_{ij} \end{array} \right. \quad (11)$$

where  $L_{sj}$  represents the simplified  $L$  index obtained after ignoring the voltage phase;  $f_s$  and  $g_s$  represent the real part and the imaginary part after ignoring the voltage phase change amount change  $\sum_{i \in \alpha_L} Z_{LLji} \left( \frac{S_i}{V_i} \right)^*$ ;  $f_{si}$  and  $g_{si}$  represent the real part and the imaginary part after ignoring the voltage phase change amount change  $\sum_{i \in \alpha_L} Z_{LLji} \left( \frac{S_i}{V_i} \right)^* \times V_i$ ,  $V_i$  represents the mode of  $\dot{V}_i$ ;  $P_i$  and  $Q_i$  represent the active power and reactive power injected into the node, respectively;  $R_{ij}$  and  $X_{ij}$  represent the resistance and reactance between nodes  $i$  and  $j$ , respectively.

### 3.2 Objective Function

#### 3.2.1 Economicality

Since the generation costs of wind and solar energy are very low, this paper mainly considers the generation costs of traditional motors using non-renewable energy sources. To meet the economic requirements of the power grid, the minimum operational cost of the system is taken as the objective function:

$$f_1 = \min \sum_i^m (C_{0i} P_{Gi}^2 + C_{1i} P_{Gi} + C_{2i}) \quad (12)$$

where  $m$  represents the number of traditional generators in the microgrid;  $P_{Gi}$  the active output of the  $i$ -th traditional generator in the microgrid;  $C_{0i}$ ,  $C_{1i}$ ,  $C_{2i}$  represent its fuel consumption curve parameters.

#### 3.2.2 Stability

The intermittency and instability of wind and solar power generation significantly affect the stability and voltage quality of the power grid. To increase the static stability margin of the grid, this paper uses the sum of the squares of the simplified  $L$ -indices of all weak nodes in the system as the objective function:

$$f_2 = \min \sum_{j=1}^l L_j^2 \quad (13)$$

where  $l$  represents the number of weak nodes in the power grid;  $L_j$  represents the simplified  $L$  index of the load node  $j$ .

### 3.3 Constraints

#### 3.3.1 Control Variable Constraints

The constraints on the active power output of each generator in the microgrid and the voltage magnitude at each node are as follows:

$$u_{\min} \leq u \leq u_{\max} \quad (14)$$

where  $u$  represents the control variable;  $u_{\min}$  and  $u_{\max}$  represent the lower and upper limits of the control variable, respectively.

#### 3.3.2 Dependent Variable Constraints

The constraints on bus voltage magnitudes, reactive power output at the buses of each generator, and the maximum transmission power between nodes are as follows:

$$h_{\min} \leq h(x, u, D, p, A) \leq h_{\max} \quad (15)$$

where  $u$  represents the control variable;  $u_{\min}$  and  $u_{\max}$  represent the lower and upper limits of the control variable, respectively.  $x$  represents a compliance variable;  $u$  represents a control variable;  $D$  represents an interference variable or uncontrollable variable;  $p$  represents a network component parameter;  $A$  is a structural variable of the network, represented by the association matrix;  $h$  represents a vector composed of compliance variables;  $h_{\min}$  and  $h_{\max}$  represent the vector consisting of the lower and upper limits of the compliance variables, respectively.

### 3.4 Solving with Improved Algorithm

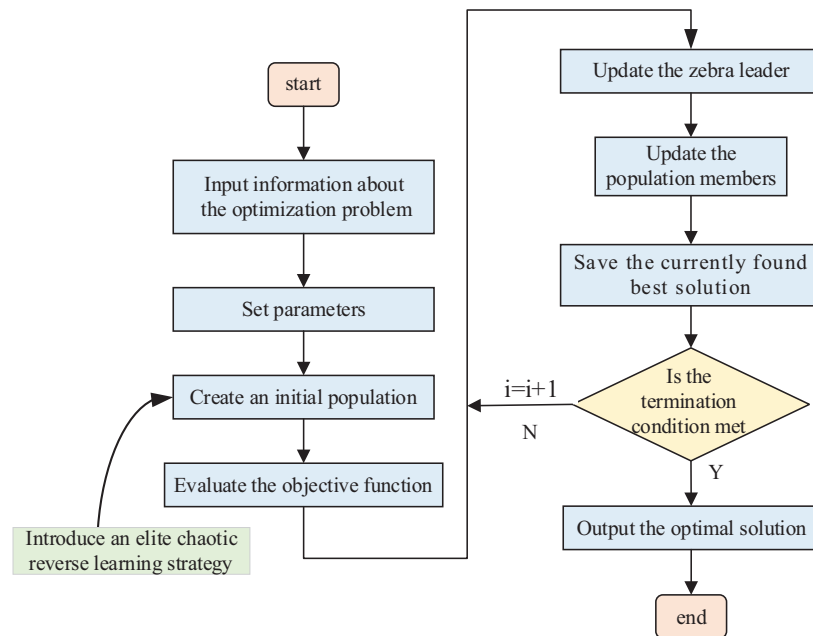
The Zebra Algorithm simulates the foraging behavior of zebras and their defensive strategies against predators' attacks. Mathematical models were based from the behaviors, as shown in Fig. 4. The specific process is as follows:

#### (1) Initialize parameters and population

Zebras are members of the population. The population matrix of the Zebra Algorithm is as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}_{N \times m} = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nj} & \cdots & x_{Nm} \end{bmatrix}_{N \times m} \quad (16)$$

where  $X$  represents the zebra population;  $X_i$  represents the  $i$ -th zebra;  $x_{ij}$  represents the value of the  $j$ -th problem variable proposed by the  $i$ -th zebra;  $N$  represents the full number of members of the zebra population;  $m$  represents the number of decision variables.



**Figure 4:** Flow chart of the improved Zebra Optimization Algorithm

(2) Use the randomly generated positions of the zebras to evaluate the objective function

Each zebra represents a candidate solution to the optimization problem, that is, the lowest economic cost and the weak points of the power grid. The formula for the objective function is as follows:

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \tag{17}$$

where  $F$  represents the vector of the objective function value;  $F_i$  represents the objective function value obtained for the  $i$ -th zebra.

(3) Update the population members

In the Zebra Algorithm, the best member of the population is considered the lead zebra and guides the other population members towards its position in the search space. That is, to find the most suitable active power and voltage magnitude. The formula is as follows:

$$x_{ij}^{new,P_1} = x_{ij} + r \cdot (PZ_j - I \cdot x_{ij}) \tag{18}$$

$$X_i = \begin{cases} X_i^{new,P_1}, & F_i^{new,P_1} < F \\ X_i, & else \end{cases} \tag{19}$$

where  $X_i^{new,P_1}$  represents the new position of the  $i$ -th zebra in the first phase, with  $x_{ij}^{new,P_1}$  being its  $j$ -th dimensional value;  $F_i^{new,P_1}$  represents the objective function value;  $PZ$  represents the pioneer zebra, represents the  $j$ -th dimensional value of the pioneer zebra;  $r$  represents a random number uniformly

distributed between 0 and 1;  $I = \text{round}(1 + \text{rand})$ ,  $\text{rand}$  represents a random number uniformly distributed between 0 and 1.

(4) Updates the location of the zebra population members in the search space

When updating the zebra position, the zebra accepts the new position if the objective function of a new position has better values. The formula is as follows:

$$X_{ij}^{\text{new}, P_2} = \begin{cases} S_1: x_{ij} + R \cdot (2r - 1) \cdot \left(1 - \frac{t}{T}\right) \cdot x_{ij}, P_s \leq 0.5 \\ S_2: x_{ij} + r \cdot (AZ_j - I \cdot x_{ij}), \text{else} \end{cases} \quad (20)$$

$$X_i = \begin{cases} X_i^{\text{new}, P_2}, F_i^{\text{new}, P_2} < F_i \\ X_i, \text{else} \end{cases} \quad (21)$$

where  $X_i^{\text{new}, P_2}$  represents the new position of the  $i$ -th zebra in the second stage,  $X_{ij}^{\text{new}, P_2}$  represents the  $j$ -th dimensional value;  $F_i^{\text{new}, P_2}$  represents the value of the objective function;  $t$  represents the current number of iterations;  $T$  represents the maximum number of iterations of the algorithm;  $R$  is a constant equal to 0.01;  $P_s$  represents the random number evenly distributed between 0 and 1;  $AZ$  represents the state of the attacked zebra,  $AZ_j$  represents the  $j$ -th dimensional value.

(5) Determine whether the termination conditions are met.

If the termination conditions are not satisfied, return to step (3); if the constraints are satisfied, proceed to step (6).

(6) Output the optimal solution

(7) Improved Zebra Optimization Algorithm

The improvement of IZOA lies in the modification of step (1), by introducing an elite chaotic reverse learning strategy to increase the probability of selecting better solutions, thereby enhancing the quality of the initial population individuals. The improved formula is as follows:

$$X_i^* = \frac{a_j + b_j}{2} + \frac{a_j + b_j}{2k} - \frac{X_i}{k} \quad (22)$$

where  $X_i^*$  represents the zebra individual reverses the lens;  $a_j$  and  $b_j$  represent the maximum and minimum values of the first dimension of the current population;  $k$  represents the scaling coefficient of the lens.

## 4 Results Analysis

### 4.1 Forecasting Results Analysis

The forecasting results are shown in Fig. 5. In this paper, we used the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination ( $R^2$ ) as three metrics to comprehensively evaluate the performance of the forecasting model. These metrics reveal the model's predictive accuracy and fit from different perspectives, and the results are shown in Fig. 6.

A smaller Root Mean Square Error (RMSE) value means the model's predictive accuracy is higher. The formula for the Root Mean Square Error is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (23)$$

where  $n$  represents the number of samples;  $y_i$  represents the true value;  $\hat{y}_i$  represents the predicted value.

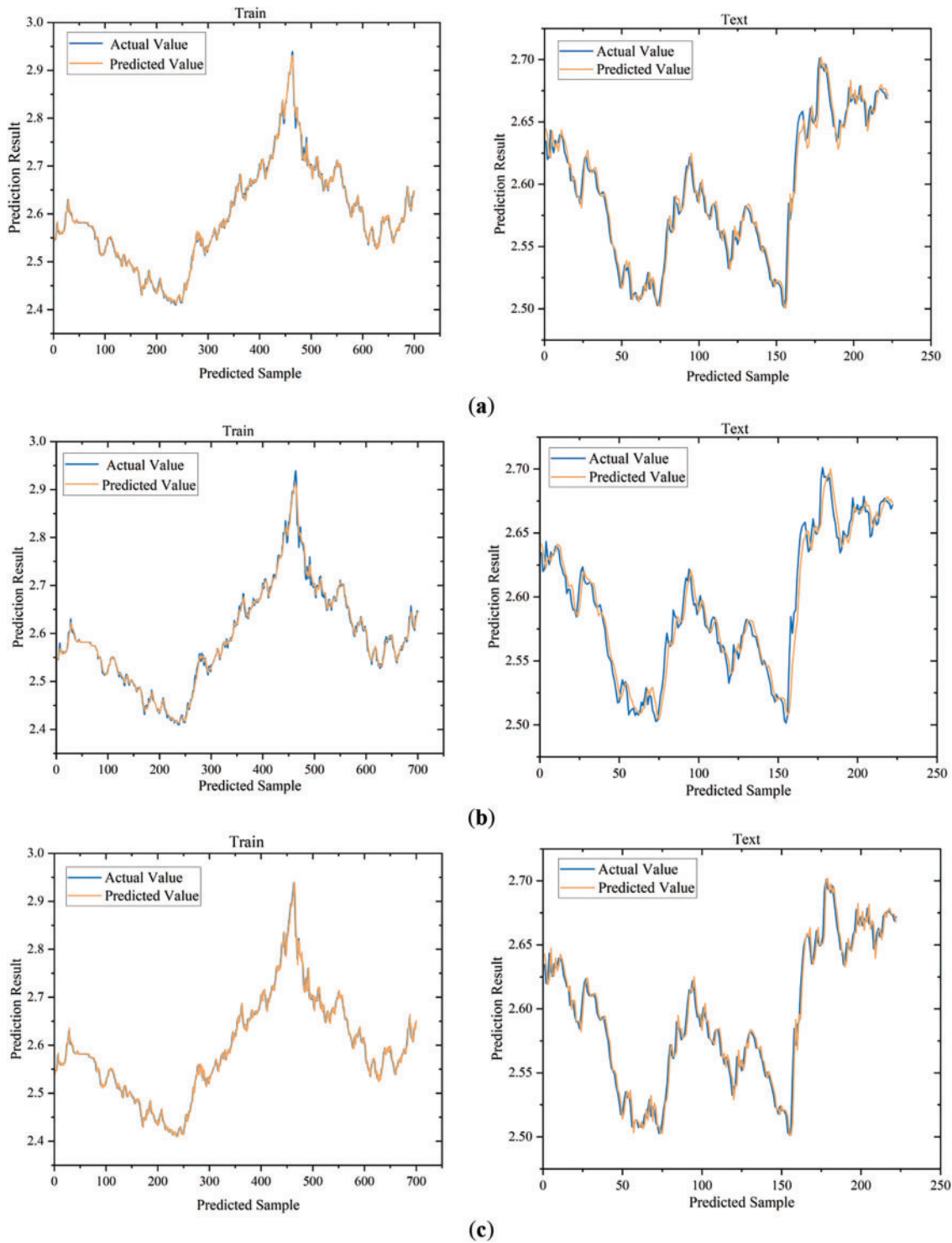
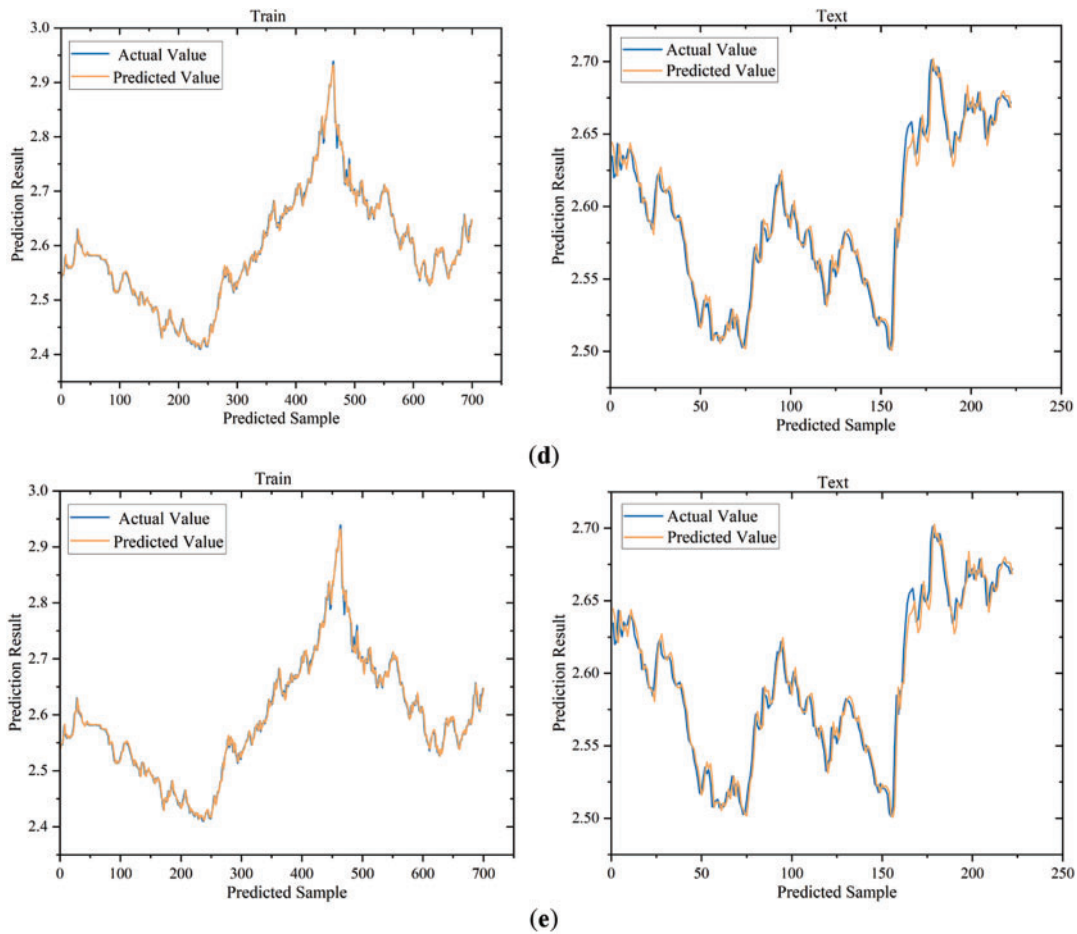


Figure 5: (Continued)



**Figure 5:** Comparison of the true and predicted values of each model. (a) Comparison of the true and predicted values of Transformer; (b) Comparison of the true and predicted values of LSTM-Attention; (c) Comparison of the true and predicted values of CNN-LSTM; (d) Comparison of the true and predicted values of CNN-BiLSTM-Attention; (e) Comparison of the true and predicted values of CNN-Transformer

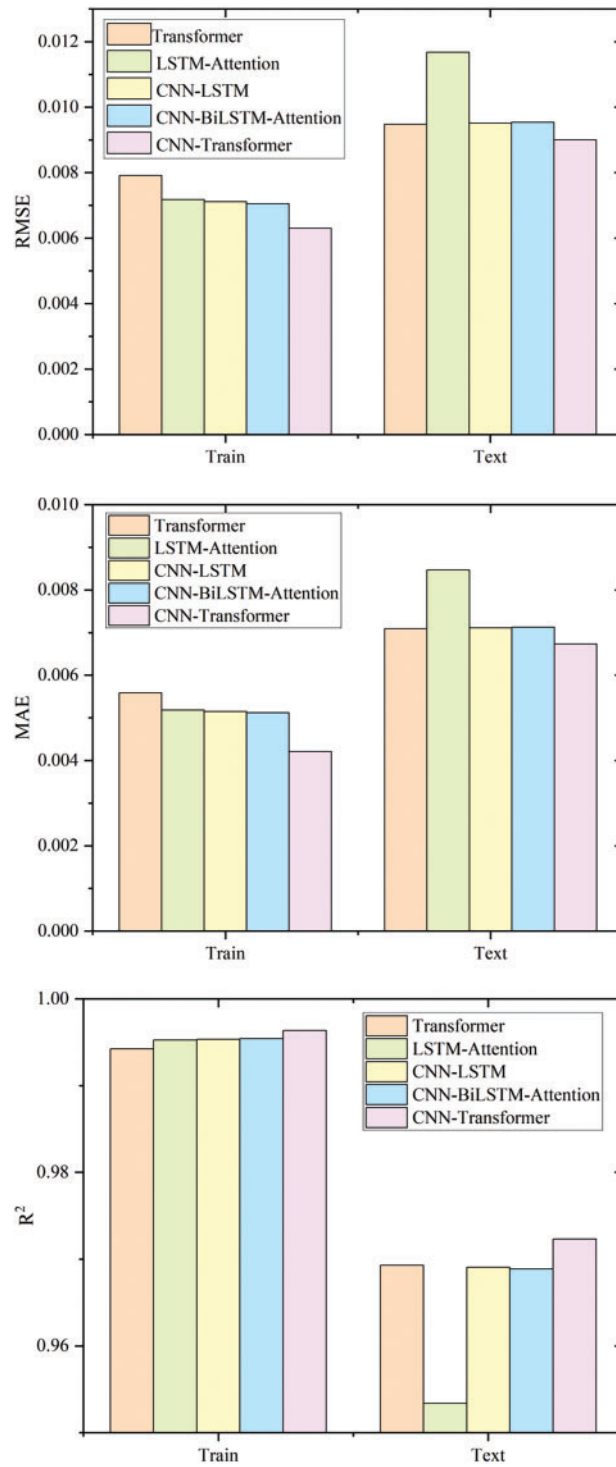
The smaller the Mean Absolute Error (MAE) value, the closer the forecasting model's predictions are to the true values, indicating higher forecasting accuracy. The formula for the Mean Absolute Error is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (24)$$

The closer the value of the Coefficient of Determination ( $R^2$ ) is to 1, the better the model fits the data. The formula for the Coefficient of Determination is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2} \quad (25)$$

where  $\bar{y}$  represents the sample mean value.



**Figure 6:** Comparison diagram of each model index

This paper compares the Transformer, LSTM-attention, CNN-LSTM, CNN-BiLSTM-Attention and CNN-Transformer models, as shown in Table 2. The experimental results indicate that the CNN-Transformer model performs best in terms of predictive accuracy, with the lowest RMSE and MAE values of 0.0063 and 0.00421, respectively, indicating the smallest prediction error. At the same time, the CNN-Transformer model's  $R^2$  value is closest to 1, reaching 0.99635, demonstrating the strongest data explanatory power. Following that is the CNN-BiLSTM-Attention model, with RMSE and MAE values of 0.00718 and 0.00512, respectively, and an  $R^2$  value of 0.99543. The Transformer model performs relatively weaker in these three metrics, with RMSE and MAE values of 0.00705 and 0.00559, respectively, and an  $R^2$  value of 0.99425. Considering these evaluation metrics, the CNN-Transformer model has the best overall performance on the test set, indicating that it is superior to the other two models in both predictive accuracy and model interpretability.

**Table 2:** Comparison of the evaluation indexes of each model test set

Evaluating indicator	RMSE	MAE	$R^2$
Transformer	0.00791	0.00559	0.99425
LSTM-Attention	0.00718	0.00518	0.99526
CNN-LSTM	0.00711	0.00515	0.99535
CNN-BiLSTM-Attention	0.00705	0.00512	0.99543
CNN-Transformer	0.00630	0.00421	0.99635

As shown in Table 3, the CNN-Transformer model performs best in terms of prediction accuracy, with the lowest RMSE and MAE values of 0.009 and 0.00673, respectively, indicating the smallest prediction error. At the same time, the CNN-Transformer model has the highest  $R^2$  value, reaching 0.97233, demonstrating the strongest data explanatory power. Following closely is the CNN-LSTM model, with RMSE and MAE values of 0.00952 and 0.00711, respectively, and an  $R^2$  value of 0.96906. The LSTM-attention model performs relatively weaker in these three metrics, with RMSE and MAE values of 0.01168 and 0.00847, respectively, and an  $R^2$  value of 0.95336. Considering these evaluation metrics, the CNN-Transformer model has the best overall performance on the test set, indicating that it is superior to the other two models in both predictive accuracy and model interpretability.

**Table 3:** Comparison of the evaluation indexes of each model test set

Evaluating indicator	RMSE	MAE	$R^2$
Transformer	0.00948	0.00709	0.9693
LSTM-Attention	0.01168	0.00847	0.95336
CNN-LSTM	0.00952	0.00711	0.96906
CNN-BiLSTM-Attention	0.00954	0.00713	0.96888
CNN-Transformer	0.00900	0.00673	0.97233

#### 4.2 Analysis of Algorithm Fitness before and after Improvement

To verify the performance of the improved Zebra Optimization Algorithm (IZOA), IZOA is compared with the ZOA and PSO (Particle Swarm Optimization) intelligent optimization algorithms, using the worst



value, best value, average value, and standard deviation as comparative data [20]. The worst and best values reflect the stability of the algorithm, the average value demonstrates the precision of the algorithm, and the standard deviation shows the robustness of the algorithm.

As can be seen from Table 4, the IZOA algorithm demonstrated the best optimal values and stability in most cases, indicating its significant advantage in the optimization problems of these specific functions. Specifically, IZOA found the optimal solutions for functions  $f_1$  and  $f_6$ , and its standard deviation was zero, which means its performance on these functions was very stable. The PSO algorithm performed best on function  $f_3$ , while the ZOA algorithm excelled on function  $f_2$ . Furthermore, the average values and standard deviations of the IZOA algorithm in most cases were much smaller than those of PSO and ZOA, indicating its potential superiority in global search capability and stability.

**Table 4:** The comparison results of different algorithms

Function	Indicator	PSO	ZOA	IZOA
$f_1$	Optimum value	4.05E+00	3.43E−86	0
	Mean standard	1.11E+01	9.70E−72	0
	Standard deviation	5.73E+00	3.96E−71	0
$f_2$	Optimum value	5.98E+00	1.18E−56	9.34E−319
	Mean standard	1.22E+01	1.61E−51	5.71E−276
	Standard deviation	3.32E+00	5.58E−51	0
$f_3$	Optimum value	2.46E+02	1.43E+04	0
	Mean standard	7.70E+02	4.81E+04	0
	Standard deviation	4.40E+02	1.53E+04	0
$f_4$	Optimum value	4.17E+00	0.03E+00	0
	Mean standard	6.58E+00	3.99E+01	4.71E−289
	Standard deviation	1.66E+00	3.28E+01	0
$f_5$	Optimum value	2.48E+02	2.74E+01	2.60E+01
	Mean standard	1.32E+03	2.81E+01	2.68E+01
	Standard deviation	1.01E+03	4.61E−01	1.91E−01
$f_6$	Optimum value	1.95E−01	7.39E−05	4.06E−06
	Mean standard	6.30E−01	4.30E−03	5.94E−05
	Standard deviation	2.28E−01	5.52E−03	4.92E−05

To better showcase the performance of the improved algorithm, as well as its convergence speed and precision, the performance of the three algorithms on six test functions is presented in the form of convergence curves, as shown in Fig. 7. It is evident that the improved algorithm outperforms other optimization algorithms in optimization capability.

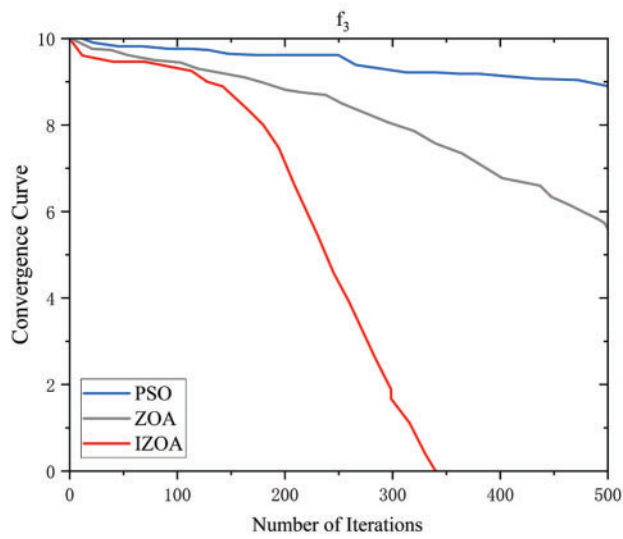
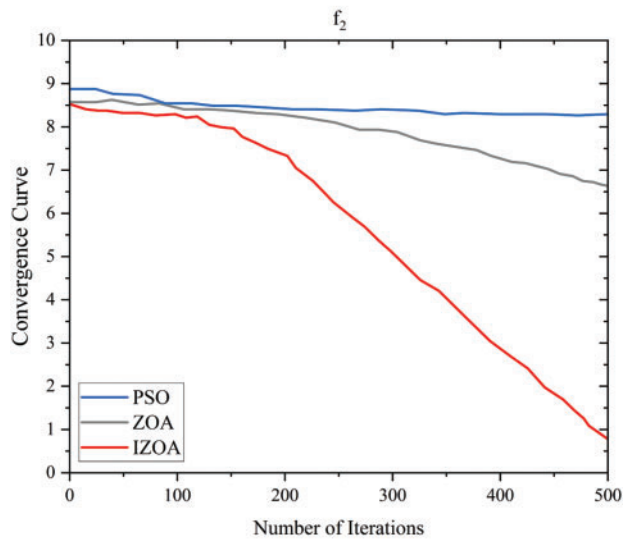
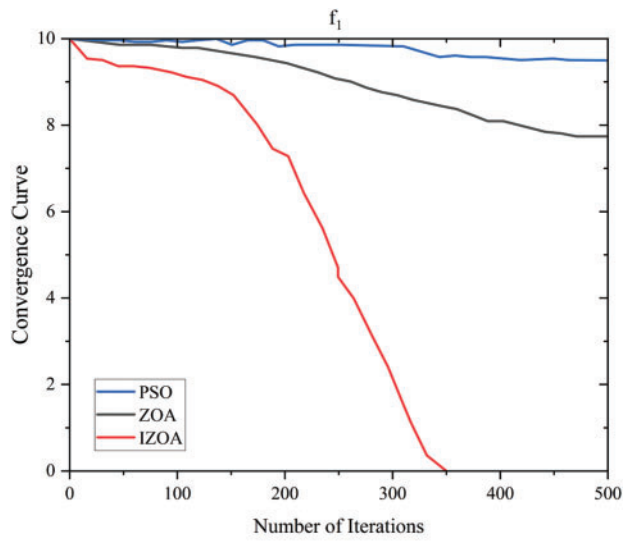
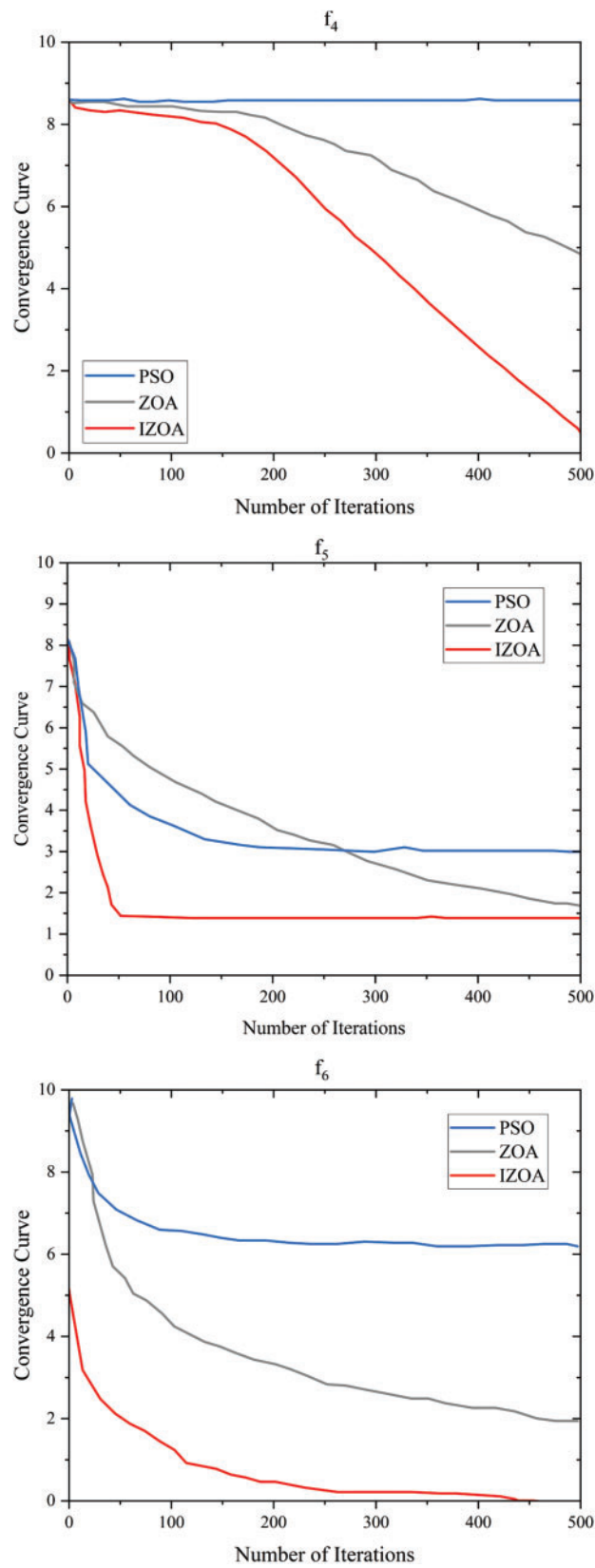


Figure 7: (Continued)



**Figure 7:** Convergence plot of each algorithm

### 4.3 Economic and Stability Analysis

As shown in Fig. 8, the system has achieved remarkable results in total cost control. Comparing Scenario 1, Scenario 2, and the method proposed in this paper, their costs are 4160 yuan, 3200 yuan, and 2890 yuan, respectively. It can be seen that the scheme proposed in this paper can effectively reduce operational costs. As shown in Fig. 9, the system stability is superior. Comparing Scenario 1, Scenario 2, and the method proposed in this paper, their simplified L-indices are 0.242, 0.189, and 0.156, respectively. It can be seen that the scheme proposed in this paper can effectively improve the stability of the system.

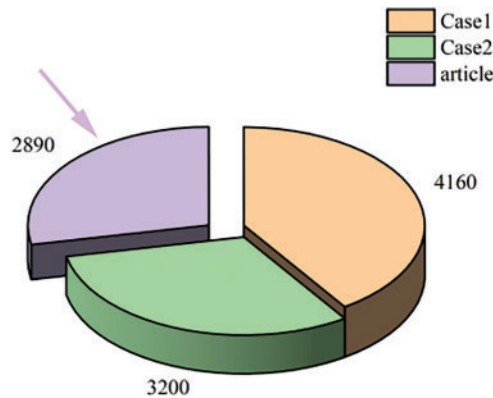


Figure 8: Economic comparison diagram

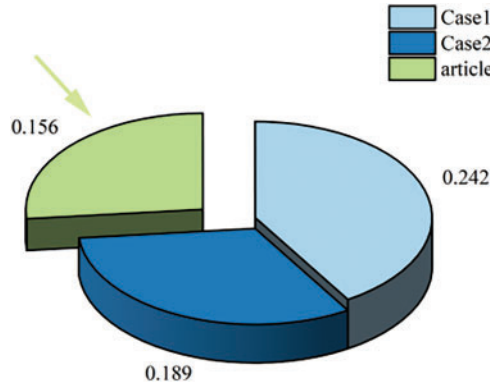


Figure 9: Stability comparison diagram

## 5 Conclusion

With the high proportion of new energy integration, the stability of power systems faces new challenges. To meet this challenge, this paper proposes a power stability system based on large language models, aiming to improve the safety and stability of power systems through advanced deep learning technology. Based on the research in this paper, we draw the following conclusions:

This paper constructs a CNN-Transformer prediction model for fault diagnosis in microgrids. The model processes the data by combining the local feature extraction capability of Convolutional Neural Networks (CNN) and the global dependency modeling capability of the Transformer model, showing great potential in power safety diagnosis. The experimental results indicate that the CNN-Transformer model demonstrated outstanding predictive performance in microgrid fault diagnosis. On the training set, the

model achieved a root mean square error (RMSE) of 0.0063, a mean absolute error (MAE) of 0.00421, and a coefficient of determination ( $R^2$ ) of 0.99635, indicating the smallest prediction error and the strongest data interpretability. On the test set, the RMSE was 0.009, the MAE was 0.00673, and the  $R^2$  was 0.97233, also showing the advantage of predictive accuracy.

This paper takes the system's economy and stability as the objective functions and uses the improved Zebra Algorithm for solving. By comparing the operating costs under different scenarios, the approach proposed in this paper can effectively reduce costs. Specifically, the cost for Scenario 1 is 4160 yuan, for Scenario 2 it is 3200 yuan, and the cost for the method presented in this paper is 2890 yuan, demonstrating the effect of cost reduction. In terms of system stability, the solution proposed in this paper also performs well. The simplified L-index for Scenario 1 is 0.242, for Scenario 2 it is 0.189, and for the method in this paper, the simplified L-index is 0.156, indicating a significant improvement in system stability.

This paper uses the improved Zebra Algorithm for solving, and the improved algorithm significantly enhances the adaptability and robustness of the algorithm by introducing an elite chaotic reverse learning strategy to increase the probability of selecting better solutions. This paper compares the improved Zebra Algorithm with other algorithms, using the worst value, best value, average value, and standard deviation as comparative data, to verify its advantages in optimization problems.

**Acknowledgement:** The authors would like to express their sincere gratitude for the support and sponsorship provided by Huai'an Hongneng Group Co., Ltd. in the research and development of the "Research on Power Safety Auxiliary Decision System Based on Large Language Models" project. This project under the Science and Technology Project Contract has significantly benefited from the insights and resources contributed by the company.

**Funding Statement:** The research project, "Research on Power Safety Assisted Decision System Based on Large Language Models" (Project Number: JSJDL24051414020001), acknowledges with gratitude the financial and logistical support it has received. This study would not have been possible without the generous backing provided by the project sponsors.

**Author Contributions:** Study conception and design: Lei Shen, Chutong Zhang, Jie Ji; data collection: Yuwei Ge, Shanyun Gu; analysis and interpretation of results: Qiang Gao, Wei Li; draft manuscript preparation: Chutong Zhang. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data that support the findings of this study are available from the corresponding author, Jie Ji, upon reasonable request.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

## References

1. Liu Y, Peng M, Gao X. Dynamic vulnerability assessment of hybrid system considering wind power uncertainty. *Energy Rep.* 2024;11:3721–30. doi:10.1016/j.egy.2024.03.033.
2. Eskandari A, Nedaei A, Milimonfared J, Aghaei M. A multilayer integrative approach for diagnosis, classification and severity detection of electrical faults in photovoltaic arrays. *Expert Syst Appl.* 2024;252(14):124111. doi:10.1016/j.eswa.2024.124111.
3. Ajagekar A, You F. Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems. *Appl Energy.* 2021;303(6):117628. doi:10.1016/j.apenergy.2021.117628.
4. Hao J, Zhan H, Xiao C, Pei H, Wang L. A method of operation status evaluation and fault diagnosis for PV arrays in the scenario of digitalization. *Energy Rep.* 2024;12(7):3020–33. doi:10.1016/j.egy.2024.08.071.

5. Liu B, Chen S, Gao L. Combining residual convolutional LSTM with attention mechanisms for spatiotemporal forest cover prediction. *Environ Model Softw.* 2025;183(4):106260. doi:10.1016/j.envsoft.2024.106260.
6. Bharatheedasan K, Maity T, Kumaraswamidhas LA, Durairaj M. Enhanced fault diagnosis and remaining useful life prediction of rolling bearings using a hybrid multilayer perceptron and LSTM network model. *Alex Eng J.* 2025;115(8):355–69. doi:10.1016/j.aej.2024.12.007.
7. Hu L, Wang J, Lee HP, Wang Z, Wang Y. Wear prediction of high performance rolling bearing based on 1D-CNN-LSTM hybrid neural network under deep learning. *Heliyon.* 2024;10(17):e35781. doi:10.1016/j.heliyon.2024.e35781.
8. Lu XQ, Tian J, Liao Q, Xu ZW, Gan L. CNN-LSTM based incremental attention mechanism enabled phase-space reconstruction for chaotic time series prediction. *J Electron Sci Technol.* 2024;22(2):100256. doi:10.1016/j.jnlest.2024.100256.
9. Zhao P, Hu W, Cao D, Zhang Z, Liao W, Chen Z, et al. Enhancing multivariate, multi-step residential load forecasting with spatiotemporal graph attention-enabled transformer. *Int J Electr Power Energy Syst.* 2024;160(6):110074. doi:10.1016/j.ijepes.2024.110074.
10. Ibrahim Mohammad Ata K, Khair Hassan M, Ghany Ismaeel A, Abdul Rahman Al-Haddad S, Alquthami T, Alani S. A multi-Layer CNN-GRUSKIP model based on transformer for spatial-TEMPORAL traffic flow prediction. *Ain Shams Eng J.* 2024;15(12):103045. doi:10.1016/j.asej.2024.103045.
11. Ding Y, Yi Z, Xiao J, Hu M, Guo Y, Liao Z, et al. CTH-Net: a CNN and Transformer hybrid network for skin lesion segmentation. *iScience.* 2024;27(4):109442. doi:10.1016/j.isci.2024.109442.
12. Wu X, Feng Y, Xu H, Lin Z, Chen T, Li S, et al. CTransCNN: combining transformer and CNN in multilabel medical image classification. *Knowl Based Syst.* 2023;281(2):111030. doi:10.1016/j.knosys.2023.111030.
13. Ahn JM, Kim J, Kim H, Kim K. Harmful cyanobacterial blooms forecasting based on improved CNN-transformer and temporal fusion transformer. *Environ Technol Innov.* 2023;32(1548):103314. doi:10.1016/j.eti.2023.103314.
14. Hendria WF, Phan QT, Adzaka F, Jeong C. Combining transformer and CNN for object detection in UAV imagery. *ICT Express.* 2023;9(2):258–63. doi:10.1016/j.icte.2021.12.006.
15. Bagherzadeh L, Kamwa I, Alharthi YZ. Hybrid strategy of flexibility regulation and economic energy management in the power system including renewable energy hubs based on coordination of transmission and distribution system operators. *Energy Rep.* 2024;12(23):1025–43. doi:10.1016/j.egyr.2024.06.061.
16. Jia Y, Zuo T, Li Y, Bi W, Xue L, Li C. Finite-time economic model predictive control for optimal load dispatch and frequency regulation in interconnected power systems. *Glob Energy Interconnect.* 2023;6(3):355–62. doi:10.1016/j.gloi.2023.06.009.
17. Yoon DH, Yoon J. Development of a real-time fault detection method for electric power system via transformer-based deep learning model. *Int J Electr Power Energy Syst.* 2024;159(4):110069. doi:10.1016/j.ijepes.2024.110069.
18. Titz M, Pütz S, Witthaut D. Identifying drivers and mitigators for congestion and redispatch in the German electric power system with explainable AI. *Appl Energy.* 2024;356:122351. doi:10.1016/j.apenergy.2023.122351.
19. Wu X, Ji Q, Liu F, Du Q, Xu W. Optimal operation strategy of power system based on stochastic risk avoidance. *Results Eng.* 2024;21(4):101832. doi:10.1016/j.rineng.2024.101832.
20. Ji J, Zhou M, Guo R, Tang J, Su J, Huang H, et al. A electric power optimal scheduling study of hybrid energy storage system integrated load prediction technology considering ageing mechanism. *Renew Energy.* 2023;215(1):118985. doi:10.1016/j.renene.2023.118985.