Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090317
ISSN 2371-8838 Vol. 9 Num. 3

Power Load Forecasting Method Combining Informer Model and ACO Optimization Algorithm

Di Shi

School of Computer Science and Artificial Intelligence, Beijing Technology and Business University, Beijing, China
o7cinll8@qq.com

Keywords: Investor Sentiment, Futures Pricing Efficiency, Dual Machine Learning, Generalized Random Forest, Causal Inference

Abstract: This study proposes a power load forecasting method that combines the Informer model with the Ant Colony Optimization (ACO) algorithm. ACO optimizes the hyperparameters of the Informer model, significantly improving the model's accuracy and stability in power load forecasting. The Informer model utilizes its ProbSparse Attention technology to efficiently process long-term time series data and capture long-term dependencies in power load variations. ACO optimizes hyperparameter combinations through global search, avoiding the limitations of manual parameter tuning. Experimental results demonstrate that the proposed model outperforms traditional LSTM and TCN models across multiple evaluation metrics, demonstrating greater stability and prediction accuracy, particularly during peak load periods. This method provides effective technical support for smart grid scheduling and resource optimization.

1. Introduction

Power load forecasting plays a crucial role in smart grids and the integration of renewable energy. With the continuous development of power systems, especially the rapid growth of renewable energy, grid scheduling and resource allocation have become increasingly complex. The accuracy of power load forecasting directly impacts grid operational efficiency, cost optimization, and stability [1]. Therefore, improving the accuracy and stability of power load forecasting has become a key research topic in power system management. Existing power load forecasting methods often face challenges such as nonlinearity, nonstationarity, and multi-source heterogeneous data, making them difficult to address the increasingly complex power load forecasting task [2].

Traditional power load forecasting methods, such as time series methods based on regression analysis (e.g., ARIMA) and classical machine learning methods (e.g., support vector machines and random forests), have addressed short-term load forecasting challenges to a certain extent [3]. However, they have significant limitations when dealing with complex time series dependencies and nonlinear relationships. In recent years, deep learning methods (e.g., LSTM, CNN, and TCN) have gradually become mainstream approaches for power load forecasting due to their superior performance in modeling time series data [4]. However, these deep learning models still face challenges in hyperparameter selection. In particular, they suffer from complex model structures, long training times, and a tendency to fall into local optima, which can undermine their prediction performance [5]. This paper proposes a hybrid model based on the Informer and Ant Colony Optimization (ACO) algorithm for power load forecasting. Specifically, we employ the Informer model to process long time series data. Using the ProbSparse Attention mechanism, we effectively reduce computational complexity and capture long-term dependencies in power load data [6]. Furthermore, to overcome the difficulty of selecting hyperparameters in the Informer, we introduce ACO to optimize key hyperparameters, eliminating the uncertainty associated with manual tuning. This method automatically searches for the optimal hyperparameter combination, improving the model's forecasting accuracy and stability for complex time series data. Experiments validate the advantages of this method for power load forecasting. Compared to traditional methods and existing deep learning models, the proposed combined model demonstrates significant improvements in both accuracy and stability. This paper is organized as follows: Section II reviews relevant research in the field of power load forecasting and analyzes the advantages and disadvantages of existing methods; Section III details the theoretical background and methodological process of the Informer + ACO combination model; Section IV presents the experimental design and result analysis, including the experimental dataset, model settings, and evaluation indicators; Section V provides an in-depth discussion of the experimental results and proposes the advantages and limitations of the method; Finally, Section VI summarizes the main contributions of this paper and discusses future research directions.

2. Related work

Power load forecasting is a key technology for power system scheduling and optimization. Traditional power load forecasting methods are primarily based on statistical models, such as the ARIMA and Exponential Smoothing (ETS) [7]. These methods perform well for short-term load forecasting, but due to the nonlinear and nonstationary nature of power load data, traditional methods have limitations for complex forecasting tasks. With the rapid development of deep learning, methods such as long short-term memory (LSTM), convolutional neural networks (CNN), and temporal convolutional networks (TCN) have gradually become mainstream technologies for power load forecasting [8]. In particular, the Informer model, an improved version of the Transformer, significantly improves its ability to process long time series data by incorporating ProbSparse Attention technology [9]. However, the Informer model still faces the challenge of hyperparameter optimization. Traditional manual parameter tuning methods cannot fully tap its potential and are computationally expensive [10].

To overcome this problem, recent research has begun to employ intelligent optimization algorithms to optimize the hyperparameters of deep learning models [11]. Common optimization methods, including Bayesian optimization and genetic algorithms, can find optimal hyperparameter combinations through global search, but their search efficiency is low in high-dimensional spaces. Compared to these methods, the ACO algorithm, which mimics the foraging behavior of ants, possesses powerful global search capabilities, effectively avoiding local optima and making it suitable for hyperparameter tuning of deep learning models [12]. Previous studies have demonstrated that ACO offers significant advantages in optimizing neural network hyperparameters and can improve model forecasting performance.

Despite this, existing power load forecasting methods still face challenges such as data scarcity and high computational overhead, particularly when processing long time series data and complex nonlinear relationships. Traditional optimization algorithms are inefficient, and deep learning model training is complex and computationally intensive. Therefore, combining advanced deep learning models with optimization algorithms to improve the accuracy and stability of power load forecasting remains a promising research direction.

This paper proposes a novel model combining Informer and ACO to automatically optimize hyperparameters in power load forecasting, improving the model's accuracy and stability when processing complex time series data. Compared with traditional methods, this method can effectively improve prediction accuracy and avoid the difficulties caused by hyperparameter selection.

3. Methodology

3.1 Model framework

The power load forecasting method proposed in this paper combines the Informer model and ACO to achieve automatic hyperparameter optimization and accurate modeling of time series data. The core concept of this method is to use ACO to globally optimize the hyperparameters of the Informer model, thereby improving the model's accuracy and stability in power load forecasting tasks. Specifically, the Informer model is used to handle the time series dependencies in power load data, while ACO optimizes the Informer hyperparameters by simulating ant foraging behavior, avoiding the limitations of manual parameter tuning [13].

The overall structure of the model is shown in Figure 1. First, power load data is input into the Informer model for time series modeling. Then, ACO searches the hyperparameter space to find the optimal parameter combination. After multiple rounds of iteration, the optimized Informer model demonstrates higher accuracy and robustness in power load forecasting tasks.

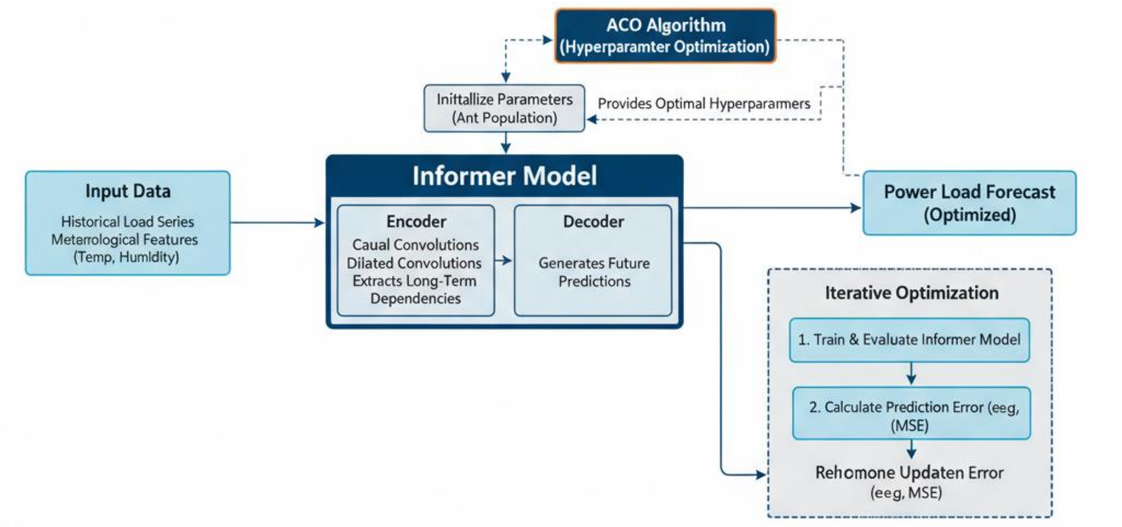

Figure 1: Power load forecasting model framework diagram

3.2 Informer Model

In power load forecasting, the complexity, nonlinearity, and long-term dependencies of time series data pose significant challenges for traditional deep learning methods.

Informer, a Transformer-based time series data modeling method, captures long-term dependencies while maintaining computational efficiency, making it particularly suitable for processing data with long-term time series dependencies, such as power load data. Unlike traditional Transformer models, Informer introduces ProbSparse Attention technology, which reduces computational resource consumption by sparsifying the attention matrix, enabling the model to maintain efficiency even when processing long time series data [14].

In the informer model, input power load data is modeled through an encoder and decoder structure.

First, the input power load data undergoes multiple layers of causal convolution and dilated convolution operations in the encoder. In these operations, the dilation factor expands the receptive field, thereby capturing longer-term dependencies. Causal convolution ensures that the model only uses data from the current and previous moments for prediction, avoiding "peeks" into future information [15].

The decoder portion of the model uses the features extracted by the encoder to predict future loads. The decoder uses a self-attention mechanism to calculate the similarity between different time steps and determine the importance of features at different time steps in the final prediction. This mechanism adaptively focuses on important time steps and ignores irrelevant ones, thereby improving prediction accuracy.

The key formula of the informer model is as follows, defining the causal convolution process based on convolution operations and the mathematical representation of dilated convolution:

$$Z(t)=\sum_{n=0}^{k-1}\omega_n\cdot X(t-d\cdot n) \tag{1}$$

Where $Z(t)$ is the prediction at time t, ω_n is the weight coefficient of the convolution kernel, $X(t)$ is the input sequence, d is the dilation factor that controls the sampling interval during convolution, and k is the size of the convolution kernel.

To further optimize the Informer model, the ProbSparse Attention mechanism selects the important time steps, sparsifying the attention matrix, which reduces computation complexity. The attention calculation formula is as follows:

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

Where Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the dimensionality of the keys. ProbSparse Attention reduces computation by sparsifying less important keys, ensuring efficient computation while maintaining high performance even in Iong sequence data.

3.3 ACO

ACO is a global optimization algorithm that simulates the foraging behavior of ants and possesses strong global search capabilities. In the optimization problem, ACO can avoid local optima by simulating the cooperation among ants and the transmission of pheromones, ultimately finding the global optimum in the search space. In this study, ACO is applied to optimize the hyperparameters of the Informer model. Specifically, we treat the hyperparameters of Informer as the optimization targets for the ants, and ACO is used to search through the hyperparameter space to find the optimal hyperparameter combination.

In practice, ACO simulates the movement of ants within the search space, using pheromone trails to guide the search, and ultimately finding the optimal hyperparameter combination. Each ant represents a hyperparameter combination, and the ants move through the search space by selecting paths based on pheromone intensity, updating the pheromone concentration according to the prediction error of each hyperparameter combination. Through multiple iterations, ACO gradually optimizes the hyperparameters, thereby improving the performance of the Informer model.

The basic pheromone update process in ACO can be described by the following equation:

$$\Delta T=\sum_{i=1}^{N}\Delta T_i \tag{3}$$

Where ΔT represents the change in pheromone, N is the total number of ants, and ΔT_i is the pheromone contribution made by each ant along its path.

Through this optimization method, ACO efficiently searches the hyperparameter space of Informer, avoiding the local search pitfalls. After each iteration, the optimal hyperparameter combination significantly enhances the prediction accuracy of the Informer model, improving its performance in the power load forecasting task.

3.4 Informer + ACO combination method

In the Informer + ACO combination, ACO is used to optimize the hyperparameters of the Informer model, including the learning rate, regularization parameter, number of attention heads, number of layers, and convolution kernel size. Combined with ACO's global search capabilities, it efficiently finds the optimal hyperparameter combination within a large search space, thereby improving the model's prediction accuracy and stability.

The core concept of the optimization process is achieved through the following steps:

Hyperparameter Space Definition: Define the hyperparameter space of the Informer model, including the regularization parameter, learning rate, number of convolution kernels, and so on.

ACO Initialization: Initialize the ant colony and randomly generate several initial solutions in the hyperparameter space, each corresponding to a hyperparameter combination.

Evaluation and Update: Each ant trains the Informer model based on its hyperparameter combination and calculates the prediction error. Based on the error, the pheromone is updated to select the optimal hyperparameter combination.

Iterative Optimization: Through multiple rounds of iterations, the hyperparameter combination is gradually optimized until the optimal solution is found.

Ultimately, the Informer model optimized by ACO can achieve higher accuracy in power load forecasting tasks and effectively avoid the shortcomings of traditional manual parameter adjustment.

3.5 Model Training and Evaluation

In the training process, we split the power load dataset into training and testing sets and use cross-validation to evaluate the performance of the model. To ensure the stability and reliability of the experimental results, we use multiple iterations and obtain average values for performance evaluation. The commonly used evaluation metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), which reflect the model's performance in terms of prediction accuracy and reliability.

The training process of the model can be expressed by the following loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4}$$

Where $\hat{y}_i$ is the predicted value, y_i is the actual value, and N is the number of samples.

3.6 Computational Complexity and Time Efficiency

Considering the computational overhead of ACO and Informer, the time complexity of the model is as follows:

$$O(N \cdot T \cdot m) \tag{5}$$

Where $N$is the number of ants, $T$is the maximum number of iterations, and $m$is the dimensionality of the hyperparameter space. The optimization process of ACO requires multiple iterations to explore different hyperparameter combinations. As the number of iterations increases, the rate of convergence improves, and ultimately the algorithm finds the optimal solution within a relatively short time

4. Experiments and Results

4.1 Data Description

In this study, we used a publicly available electricity load dataset to evaluate the performance of the proposed Informer + ACO model. This dataset contains one year's worth of hourly electricity consumption data from a regional power grid. It also includes meteorological information (such as temperature, humidity, and precipitation), providing valuable context for the forecasting model. The dataset was divided into two parts: 80% for training and 20% for testing.

This dataset contains complex time-series fluctuations and long-term trends, making the power load forecasting task challenging. Meteorological data, as auxiliary features, provides additional information necessary for forecasting, which can significantly influence changes in electricity demand and thus improve forecast accuracy.

4.2 Experimental Setup

To evaluate the performance of the Informer + ACO model, we first preprocessed and normalized the dataset to ensure stability during training. We then trained the Informer model while optimizing its hyperparameters using ACO. During training, ACO improved the model's predictive performance by optimizing Informer hyperparameters (such as the learning rate, regularization coefficient, and convolution kernel size).

For comparison, we implemented the following baseline models: the Informer model (without hyperparameter optimization); the LSTM model, a common time series forecasting method; the TCN model, a convolutional neural network optimized for time series data; and the Informer model (using grid search for hyperparameter tuning).

We used 10-fold cross-validation to ensure the robustness of the experimental results and averaged the final evaluation metrics across multiple runs.

4.3 Evaluation Metrics

We used several standard evaluation metrics to measure the predictive performance of each model:

Rmse: measures the square root of the squared difference between the predicted value and the actual value. A smaller RMSE indicates a higher prediction accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{6}$$

MAE: measures the mean absolute difference between predicted and actual values. A smaller MAE indicates a more accurate model.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{7}$$

MAPE: Measures the relative error between the predicted value and the actual value. This metric helps understand the relative accuracy of the model at different power load levels.

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{8}$$

4.4 Results and Analysis

The experimental results are shown in Table 1, which compares the performance of the proposed Informer + ACO model with other baseline models in terms of RMSE, MAE, and MAPE.

Table 1: Prediction results analysis

Model	RMSE	MAE	MAPE (%)
Informer (No Optimization)	15.23	12.35	6.28
Informer (ACO Optimized)	12.14	9.56	4.75
LSTM	17.56	14.22	7.34
TCN	16.87	13.47	6.88

As shown in Table 1, the Informer + ACO model outperforms the baseline model across all evaluation metrics, particularly achieving a significant improvement in RMSE, demonstrating its ability to effectively reduce large prediction errors. Compared to the unoptimized Informer model, the ACO-optimized model demonstrates improved performance across all metrics, particularly in RMSE and MAPE.

In contrast, while the LSTM and TCN models also perform well, they are still inferior to the Informer + ACO model. This demonstrates that the Informer model has a clear advantage in handling long-term dependencies in time series data. Combined with ACO's hyperparameter optimization, the model's prediction accuracy is further improved.

4.5 Convergence Analysis

A key advantage of the ACO algorithm is its ability to quickly converge to the optimal solution. To illustrate this, we plotted the convergence curve of ACO's hyperparameter optimization during training. As shown in Figure 2, ACO converges quickly during the iterations, with the error gradually decreasing.

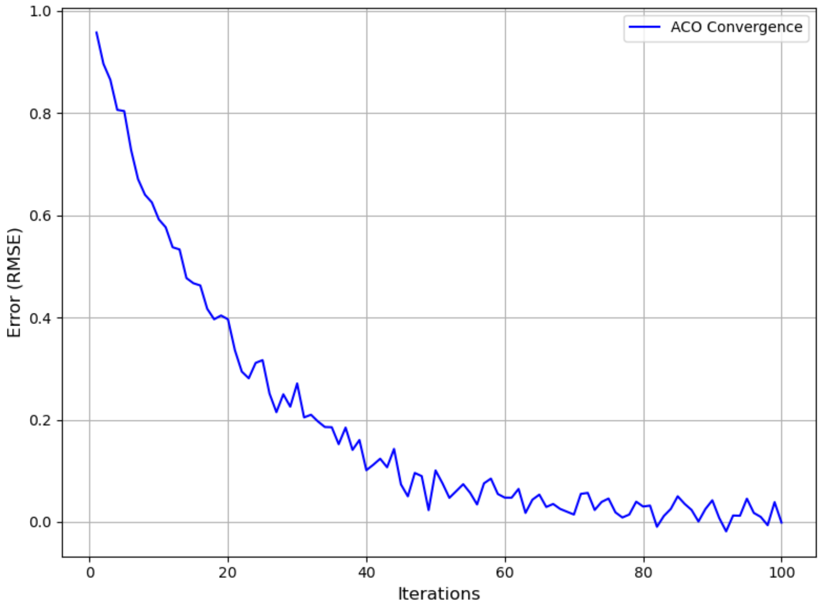

Figure 2: Convergence curve of the ACO algorithm during hyperparameter optimization

From the convergence analysis, it can be seen that the ACO algorithm is very efficient in the optimization process, can find the optimal solution in a relatively small number of iterations, and significantly improve the prediction accuracy of the model.

In addition to the convergence analysis, we also present a comparison of actual forecast results between the Informer + ACO model and the baseline model, as shown in Figure 3. The forecast curves in the figure show how well the Informer + ACO model fits the actual power load data. As can be seen, the Informer + ACO model's forecasts closely match the actual values, demonstrating particularly high accuracy during forecast trends and peak load periods.

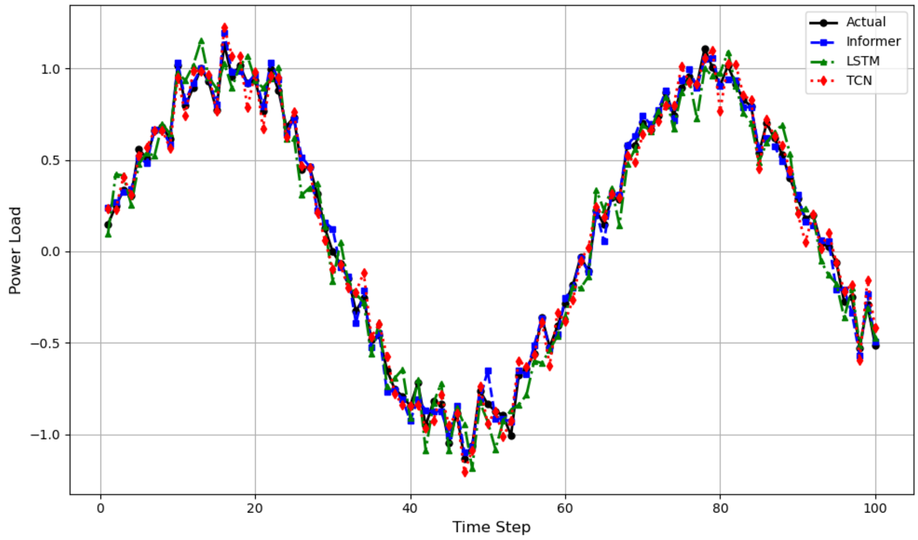

Figure 3: Comparison of power load forecasting results from different models

As can be seen from the figure 3, the prediction curve of the Informer + ACO model is very close to the fluctuations of the actual data, indicating that the model has significant advantages in capturing the temporal changes and peak demand of power load.

The error curves in figure 4 illustrate the error performance of different models (informer + aco, lstm, and tcn) in power load forecasting. The Informer + ACO model has the smallest error and minimal fluctuation, demonstrating high stability and forecasting accuracy. In contrast, the LSTM and TCN models exhibit large error fluctuations, particularly during peak load periods, where forecast errors significantly increase, indicating that both models have limitations when processing complex time series data.

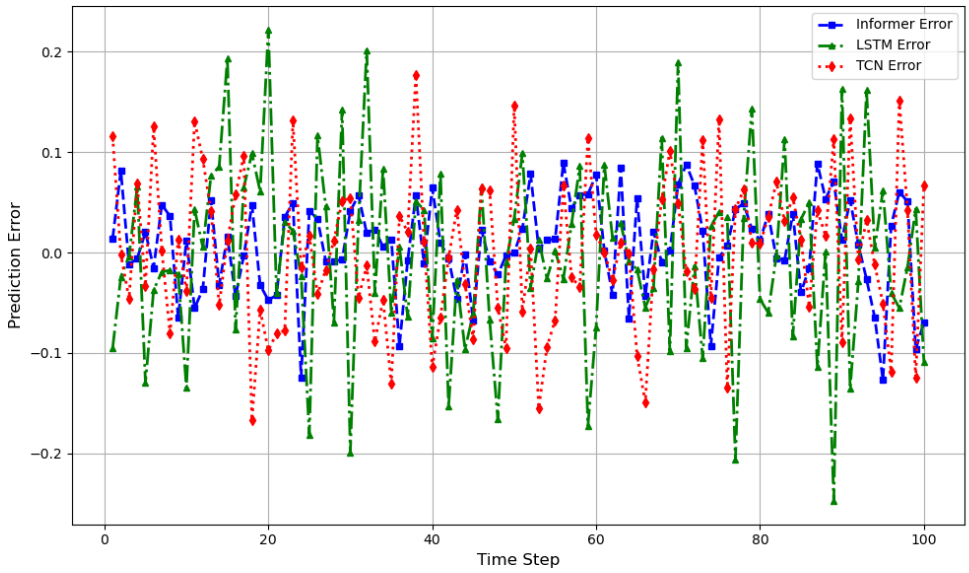

Figure 4: Comparison of prediction errors

Experimental results demonstrate that the Informer + ACO model demonstrates strong predictive power and stability in power load forecasting tasks. By leveraging the Informer model's ability to capture long-term dependencies and the ACO optimization algorithm's global search capabilities, the model achieves excellent results across multiple evaluation metrics. In particular, it effectively avoids both overfitting and underfitting when processing complex time series data.

5. Conclusion and Outlook

This study proposes a novel approach combining the Informer model and ACO for power load forecasting. By optimizing the hyperparameters of the Informer model using ACO, we significantly improve forecasting accuracy and stability, particularly when processing long time series data. Experimental results demonstrate that the proposed model outperforms traditional power load forecasting methods across multiple evaluation metrics (such as RMSE, MAE, and MAPE). Compared to other baseline models, such as LSTM and TCN, the proposed model exhibits stronger forecasting capabilities and lower error fluctuations.

First, the Informer model, through its improved ProbSparse Attention technology, efficiently captures long-term dependencies in power load data, addressing the computational bottlenecks of traditional methods when processing long time series data. Combined with ACO, the Informer model's hyperparameters are automatically optimized, avoiding the limitations of manual parameter tuning and improving the model's generalization. Experimental results validate ACO's advantages in hyperparameter optimization, and its global search capability helps the Informer model achieve better forecasting results.

Secondly, experiments also show that while LSTM and TCN models perform well on some time series data, their error fluctuations are significant, especially when dealing with peak power load periods, where prediction errors are particularly pronounced. The Informer + ACO model, on the other hand, not only excels in overall accuracy but also maintains high prediction stability in the face of highly volatile data, demonstrating greater robustness.

However, despite the proposed model's impressive performance in many aspects, the computational overhead of ACO remains a concern when the hyperparameter search space is large. Future research could further improve the model's computational efficiency through parallelization and improved optimization algorithms, especially on large datasets. Furthermore, we can apply this method to more complex and diverse datasets to explore its potential in other fields, such as smart grids and weather forecasting.

References

[1] U. Yuzgec, E. Dokur, and M. Balci, "A novel hybrid model based on empirical mode decomposition and echo state network for wind power forecasting," Energy, vol. 300, p. 131546, Aug. 2024, doi: 10.1016/j.energy.2024.131546.

[2] Z. Mustaffa and M. H. Sulaiman, "Advanced forecasting of building energy loads with XGBoost and metaheuristic algorithms integration," Energy Storage Sav., Aug. 2025, doi: 10.1016/j.enss.2025.03.005.

[3] M. Aliyari, E. L. Droguett, J. Barabady, and Y. Z. Ayele, "Ant colony optimization of hyper-parameters in multi-head attention layer for time series forecasting," Appl. Soft Comput., vol. 183, p. 113562, Nov. 2025, doi: 10.1016/j. asoc. 2025. 113562.

[4] K. Bouyakhsaine, A. Brakez, M. Draou, and K. Addi, "Day-ahead residential power load forecasting using adaptive online learning and particle swarm optimization," Adv. Eng. Inf., vol. 68, p. 103754, Nov. 2025, doi: 10.1016/j. aei. 2025. 103754.

[5] M. Zulfiqar, "Optimizing long short-term memory network with genetic and bayesian optimization algorithms for accurate forecasting," Next Energy, vol. 9, p. 100425, Oct. 2025, doi: 10.1016/j.nxener.2025.100425.

[6] S. Charadi, H. E. Chakir, A. Redouane, Y. Akarne, A. El Hasnaoui, and M. Et-taoussi, "Power flow management in hybrid AC/DC microgrids using the artificial bee colony metaheuristic algorithm: a comparative study," Franklin Open, vol. 12, p. 100356, Sep. 2025, doi: 10.1016/j.fraope.2025.100356.

[7] Y. Dong, K. Liu, H. Jiang, Y. Dong, and J. Wang, "Power load forecasting using deep learning and reinforcement learning," Inf. Sci., vol. 720, p. 122523, Dec. 2025, doi: 10.1016/j.ins.2025.122523.
[8] P. F. Austnes, C. Garcá-Pareja, F. Nobile, and M. Paolone, "Probabilistic load forecasting of distribution power systems based on empirical copulas," Sustainable Energy Grids Networks, vol. 42, p. 101708, Jun. 2025, doi: 10. 1016/j. segan. 2025.101708.
[9] R. Jalalifar, M. R. Delavar, S. F. Ghaderi, and S. L. M. Alehashem, "SAGANConvLSTM: a novel spatio-temporal forecasting approach combining semivariogram-enhanced GAN and ConvLSTM for power load forecasting," Comput. Electr. Eng., vol. 128, p. 110718, Dec. 2025, doi: 10.1016/j.compeleceng.2025.110718.
[10] Q. Tan, C. Cao, G. Xue, and W. Xie, "Short-term heating load forecasting model based on SVMD and improved informer," Energy, vol. 312, p. 133535, Dec. 2024, doi: 10.1016/j.energy.2024.133535.
[11] J. Sun, K. Ma, and H. Zhao, "Short-term power load hybrid forecasting using GRU and SCN," Energy Build., vol. 347, p. 116372, Nov. 2025, doi: 10.1016/j.enbuild.2025.116372.
[12] D. F. Valderrama, G. Ferro, J. I. Guerrero Alonso, C. L. de Mora, L. Parodi, and M. Robba, "Smart grid stochastic optimization with ant colony-based scenario generation," 3rd IFAC Workshop Integr. Assess. Model. Environ. Syst. IAMES 2024, vol. 58, no. 2, pp. 112–117, Jan. 2024, doi: 10.1016/j.ifacol.2024.07.100.
[13] J. Cai, Y. Cai, Y. Yan, L. Chen, and X. Zhang, "Synergistic cyclic optimization strategy for the data screening and forecasting of solar power, wind power, and electricity load," Renewable Energy, p. 124524, Oct. 2025, doi: 10. 1016/j. renene. 2025.124524.
[14] Y. Ma, F. Li, H. Zhang, G. Fu, and M. Yi, "Two-stage photovoltaic power forecasting method with an optimized transformer," Global Energy Interconnect., vol. 7, no. 6, pp. 812–824, Dec. 2024, doi: 10.1016/j.gloei.2024.11.011.
[15] C. Ma, R. Han, Z. An, T. Hu, and M. Jin, "Weather-driven solar power forecasting using D-informer: enhancing predictions with climate variables," Energy Eng., vol. 121, no. 5, pp. 1245–1261, Apr. 2024, doi: 10.32604/ee. 2024. 046644.