Research on Short-Term Power Load Forecasting Model Based on NGO-CNN-BIGRU-AT

Qingyun Yuan, Pan Yu, Liu Tan, Yonggang Wang, Heming Zhang

***Abstract*—A CNN-BIGRU-AT model for power load forecasting is constructed by introducing Attention Mechanism (AT) into a model combining Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (BIGRU), which has high prediction accuracy. In this model, CNN is first used to extract relevant features from power load data, and then BIGRU is applied to capture the inherent complex nonlinear dynamic patterns of power load. Additionally, Attention Mechanism (AT) is incorporated to further refine the feature extraction of power load data. The North Grey Wolf Optimizer Algorithm (NGO) is used to optimize parameters such as the number of hidden layer nodes, initial learning rate, and regularization coefficient in BIGRU, thus obtaining the modeling method of NGO-CNN-BIGRU-AT. Finally, the validity and feasibility of the model are substantiated using actual power load data from a certain region in China. The results demonstrate that the model has better performance than conventional models, including BP, BIGRU, CNN-BIGRU, and CNN-BIGRU-AT.**

***Index Terms*—Power load forecasting, convolutional neural network, bidirectional gated recurrent unit, attention mechanism, north grey wolf optimization algorithm.**

I. INTRODUCTION

WITH the gradual development of the next generation smart grid, the exploration of power load forecasting is also strengthening. The increasing prevalence of consumer-side decentralized power sources has brought significant obstacles due to the unpredictability and fluctuation of load. Without accurate load forecasting, the power grid will face the risk of low supply-demand matching rate. Consequently, power providers face challenges in accurately anticipating and fulfilling future energy requirements, leading to insufficient supply, power outages, or unstable power grids. Thus, it is necessary to find a power load forecasting model to reduce operating expenses and avoid inconvenience or losses to power enterprises.

The methods of power load forecasting include conventional technologies and Artificial Intelligence (AI) methodologies. Conventional technologies, such as time series analysis [1], regression analysis [2], and grey prediction [3], are adept at managing relatively stable time series and less volatile load data. However, the accuracy of nonlinear power load time series prediction is limited. Emerging AI forecasting methodologies, including Artificial Neural Networks (ANN) [4], Support Vector Machines (SVM) [5], and deep learning [6], have been widely used in load forecasting and have demonstrated commendable performance in addressing nonlinear problems. However, these methodologies encounter persistent challenges, such as the complexity of data sequences, the daunting process of identifying optimal local parameters, and insufficient feature extraction capabilities. Faced with these challenges, domestic and foreign scholars have been committed to neural network models for power load forecasting. For instance, one study enhanced the Gated Recurrent Unit (GRU) neural network model for short-term load forecasting by using the grey wolf algorithm known for its high consistency and fast convergence [7]. Another study presented an enhanced neural network load forecasting model based on multi-layer clustering, which addressed the issue of traditional BP neural networks easily falling into local optima [8]. However, the use of neural network requires a large number of training data, which inevitably brings challenges to data feature mining and leads to a decrease in prediction accuracy [9]. Another study [10] integrated attention mechanism with Bidirectional GRU (BIGRU) to extract multiple features from power load data, resulting in commendable prediction performance. To enhance the extraction of feature information, the convolutional neural network (CNN) is incorporated into relevant models. Wang Huan et al. proposed a CNN-BILSTM model for power load forecasting, which had good fitting performance [11]. On this basis, an ultra-short-term power load forecasting model with high accuracy was developed by combining CNN-BILSTM with attention mechanism [12].

In summary, the hybrid model for power load forecasting with data feature extraction demonstrates high prediction accuracy. Therefore, based on CNN-BIGRU hybrid model and attention mechanism, a CNN-BIGRU-AT model for power load forecasting is constructed in this paper. Furthermore, to enhance the prediction accuracy of the CNN-BIGRU-AT model, an intelligent optimization

Manuscript received February 8, 2024; revised July 3, 2024. This work was supported was supported by National Natural Science Foundation of China under Grant 32001415, General Project of Liaoning Provincial Department of Science and Technology under Grant 2023-MS-212, General Project of Liaoning Province Education Department under Grant LJKZ0683 and LJKMZ20221035, Doctoral Start-up Foundation of Liaoning Province under Grant 2019-BS-207 and Natural Science Foundation Funding Scheme of Liaoning Province under Grant 2019-KF-03-01.

Qingyun Yuan is an associate professor of School of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China (e-mail: yqy8 29@126.com).

Pan Yu is a postgraduate student of School of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China (corresponding author to provide phone:15710598863; e-mail: (15710598863@163.com)

Liu Tan is an associate professor of School of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China (e-mail: liutan 0822@126.com).

Yonggang Wang is an associate professor of School of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China (e-mail: wygvern@163.com).

Heming Zhang is a postgraduate student of School of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110866, China (e-mail: 2995107351@qq.com)

algorithm called Northern Goshawk Optimizer (NGO) is adopted to find the optimal values for the number of hidden layer nodes, initial learning rate, and regularization coefficient in BIGRU. Subsequently, the NGO-CNN-BIGRU-AT model is constructed for power load forecasting. Finally, the model is validated using actual power load data from a certain region in China. The results demonstrate that the model is superior to traditional BP, BIGRU, CNN-BIGRU, and CNN-BIGRU-AT models.

II. Analysis of Factors Affecting Power Load

There are numerous factors that influence power consumption, such as meteorological conditions, temperature fluctuations, holidays, and weekdays. However, utilizing all these variables as modeling inputs may lead to convergence challenges or prolong iteration time during training. Therefore, Pearson correlation analysis [13] is used to examine the factors that affect power consumption. This study utilized actual data from August 2018 as a reference for correlation examination. The factors that affect power consumption include the maximum temperature, minimum temperature, average temperature, relative humidity, rainfall volume, and historical load.

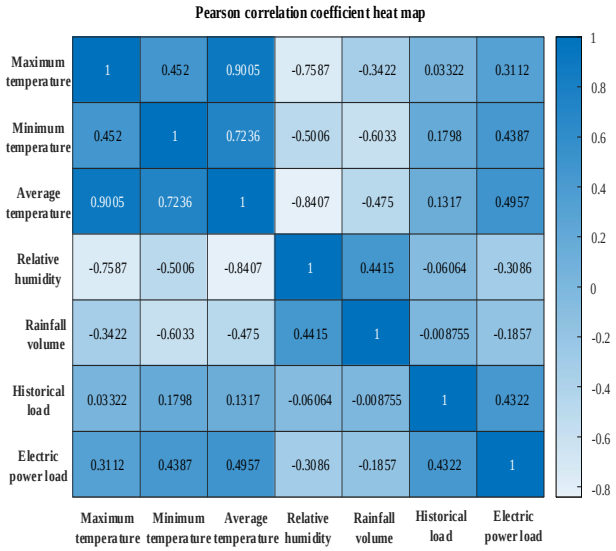
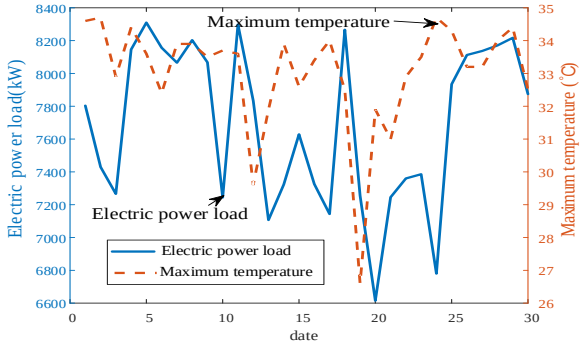
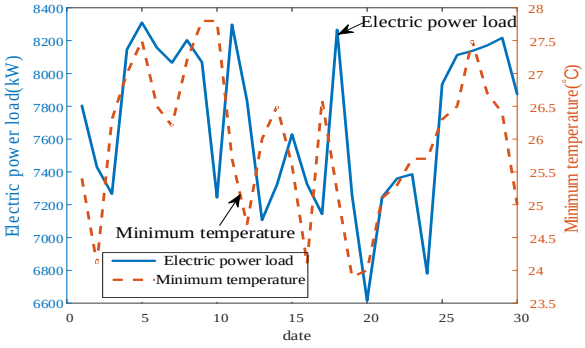


Fig.1 Heat map of factors related to power load

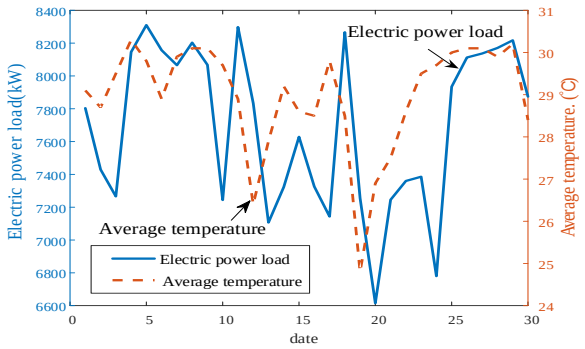
Pearson correlation coefficient between them and power consumption is shown in Fig. 1.



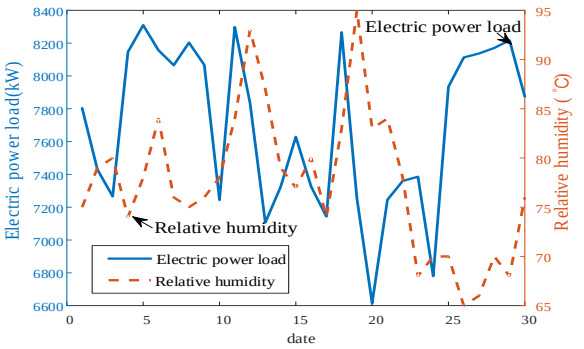
(a) Graph of power load and maximum temperature



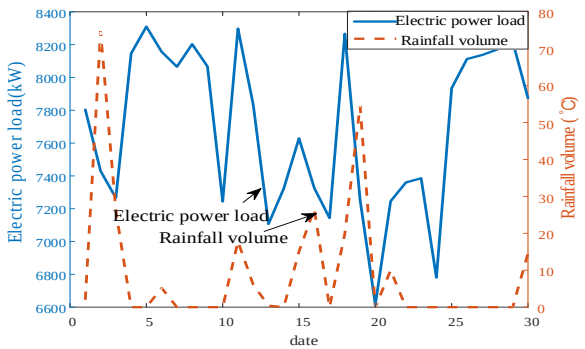
(b) Graph of power load and minimum temperature.



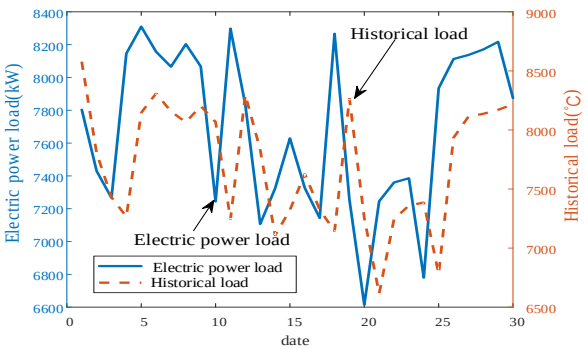
(c) Graph of power load and and average temperature.



(d) Graph of power load and relative humidity



(e) Graph of power load and rainfall volume



(f)Graph of power load and historical load .

Fig.2 Hyperbolic chart of influencing factors

It can be seen from Fig1 that the statistical analysis unveils a notable correlation between them. The correlation between maximum temperature and power load is 0.3112. while the correlation between minimum temperature and power load is 0.4387. The correlation coefficient between average temperature and power load is 0.49575.

The correlation coefficient between relative humidity and power load is -0.3086. Additionally, the correlation coefficient between historical load and power load is 0.4322, exceeding the threshold of 0.2, indicating that it has a significant impact on power load. Analysis shows that these variables are indispensable inputs for predictive models. On the other hand, the correlation between rainfall volume and power load is -0.1857, which is weak and therefore not used as an input variable.

The corresponding changes of power load with each factor are shown in Fig.2, elucidating their dynamic impact on power load.

From Fig. 2, it can be seen that there is a moderate positive correlation between historical load, average temperature, minimum temperature, and maximum temperature. Conversely, relative humidity demonstrates a strong negative correlation. The correlation with rainfall seems relatively weak. Consequently, the final input variables of the model include maximum temperature, minimum temperature, average temperature, relative humidity, and historical load.

III. Establishment of NGO-optimized CNN-BIGRU-AT Load Forecasting Model

In the field of power load forecasting, the utilization of Convolutional Neural Network (CNN) serves to distill salient features that influence the process, while the integration of attention mechanism focuses more on pivotal data points, thereby adapting to a broader spectrum of information and yielding superior predictive outcomes [14]. Nonetheless, a single prediction model invariably grapples with constrained accuracy and inherent deficiencies. To address these limitations, the present study proposes combining multiple models to create a predictive framework. The model detailed in this article is the Bidirectional Gated Recurrent Unit (BIGRU), which amalgamates the distinguishing features of bidirectional networks and gated recurrent units [15]. However, although the BIGRU model faces challenges in parameter selection and demonstrates significant randomness. Consequently, an integration of the Northern Goshawk Optimization (NGO) algorithm is proposed to fine-tune model hyperparameters, including the optimal number of hidden layer nodes, initial learning rate parameter, and regularization coefficients, thereby enhancing the convergence speed of algorithm. Furthermore, an NGO optimized CNN-BIGRU-AT model is developed for load forecasting, improving prediction accuracy.

A. Constructing CNN-BIGRU-AT Model

The modeling flowchart of CNN-BIGRU-AT is shown in Fig 3.

In Fig. 3, the specific steps for establishing the model are as follows:

Step1: Incorporate variables that affect power load fluctuations as model inputs, expressed as follows:

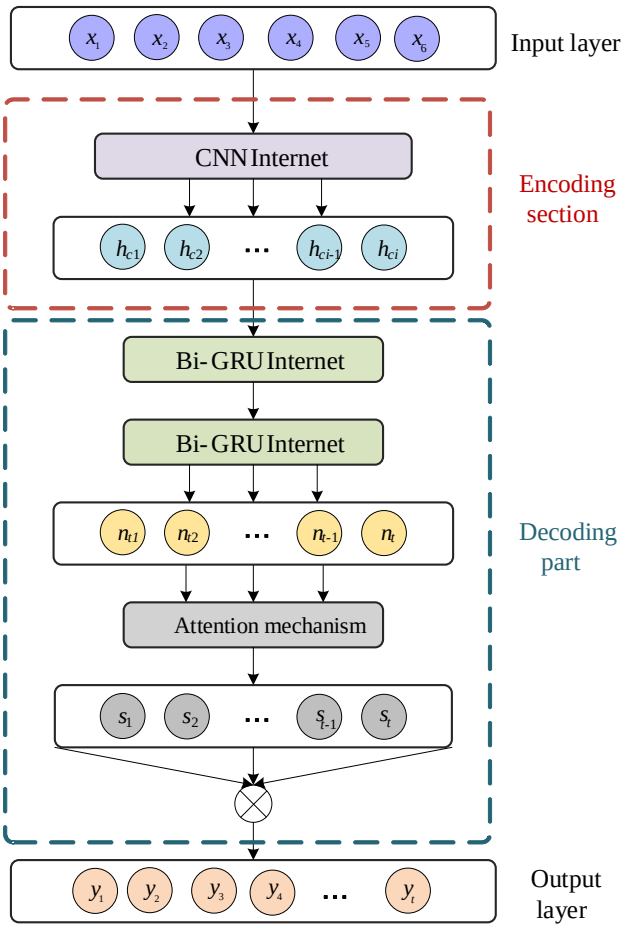


Fig. 3 Structure diagram of CNN-BIGRU-AT

$$X = \left[x_1, x_2, x_3, x_4, x_5, x_6\right]^T \tag{1}$$

where x_1 is the maximum temperature, x_2 is the minimum temperature, x_3 is the average temperature, x_4 is the relative humidity, x_5 represents the historical load, x_6 is the actual load value, and X is the input vector.

Step 2: Utilize Convolutional Neural Network (CNN) layer [16] to extract features from the power load sequence. This layer primarily consists of one-dimensional convolutional layer, pooling layer, and fully connected layer.

The one-dimensional convolutional layer (Conv1D) is responsible for extracting features from input data related to power load. Additionally, maximum pooling is employed to remove redundant information and mitigate overfitting. Finally, the output values of the CNN layer are obtained by applying the Sigmoid activation function, represented as follows:

$$C_i = f\left(X \otimes W_1 + b\right) \tag{2}$$

$$f(z) = Relu(z) = \begin{cases} 0, if.z \le 0 \\ z, if.z > 0 \end{cases} \tag{3}$$

where C_i represents the output of the convolutional layer, $f(z)$ symbolizes the ReLU activation function, W_1 symbolizes the weight value, and $\otimes$ represents the convolution operation.

From Equation (3), it can be seen that when z is less than 0, the convolutional layer function produces 0, and when z is greater than 0, it produces z.

During the maximum pooling operation, the pooling kernel traverses the input feature map to identify the maximum value in each region, which becomes the output value of the pooling layer and can be represented as follows:

$$P = \max\left(C_j\right), \left\|C_j - C_i\right\| \le \sqrt{poolh^2 + poolw^2} \tag{4}$$

$$H_c = \text{Sigmoid}\left(P \times W_2 + b_1\right) \tag{5}$$

where P denotes the output of the pooling layer, *poolh* and *poolw* represent the strides of the pooling window's height and width, respectively. H_c denotes the final output of that layer, the dimensionality of the output is *i*, and $H_c = \left[h_{c1}, \cdots, h_{ci-1}, h_{ci}\right]^T$. W_2 symbolizes the weight value, and b_1 signifies the bias.

Step 3: The BIGRU Layer [17] comprises a forward GRU layer and a backward GRU layer, facilitating bi-directional information flow. The output of the BIGRU layer is determined by both the forward and backward GRU. Utilizing newly extracted features from CNN for training, the BIGRU layer captures internal sequence patterns. Subsequently, the output of the BIGRU layer at time *t* can be represented as follows:

$$n_t = BIGRU\left(n_{t-1}, H_c\right) \tag{6}$$

where n_{t-1} indicates output at the previous time.

The input of the attention mechanism includes the amalgamated subsequence load data from the BIGRU layer [18]. By iteratively updating the weights of input features during training, the significance of crucial input information has been heightened. The weight of the attention mechanism is calculated as follows:

$$e_t = V^T \tanh\left(W_3 + U_1\right) n_t \tag{7}$$

The normalization of weights follows this procedure:

$$a_t = \text{soft}\max\left(e_t\right) = \frac{\exp\left(e_t\right)}{\sum_{i=1}^{t} e_i} \tag{8}$$

The sum of weights is expressed as follows:

$$s_t = \sum_{j=1}^{N} a_t h_t \tag{9}$$

where V, W_3 and U_1 represent weight parameters, e_t represents the probability distribution value of attention mechanism at time *t*, a_t represents the weight of attention mechanism corresponding to the output value of the BIGRU hidden layer at time *t*, h_t represents the output of the BIGRU layer, and s_t represents the output value of the attention mechanism at time *t*.

The connected layer [19] uses the Rectified Linear Unit (ReLU) activation function to compute prediction values, thereby enhancing the model's capability to extract power load related features. Assuming that the prediction value at time *t* is denoted by y_t, it can be expressed as follows:

$$y_t = Relu\left(w_0 s_t + b_0\right) \tag{10}$$

where y_t represents the predicted value at *t*, w_0 denotes the weight matrix, and b_0 represents the bias term.

Step 4: The output layer [20] is the data calculated by the model mentioned in the paper, as well as the prediction value, it is represented as follows:

$$Y_{\text{predict}} = \left[y_1, y_2, \cdots, y_m\right]^T \tag{11}$$

where *m* denotes the dimensionality of forecasting, Y_{predict} represents the predicted power load value.

Utilizing Convolutional Neural Networks (CNN) to extract features from the raw power load data. These features serve as inputs to the BIGRU model, which comprises forward and backward GRU layers to capture the nonlinear dynamic patterns of power load. Furthermore, an Attention Mechanism (AT) has been incorporated to enhance the efficiency of feature extraction from power load data by filtering out irrelevant information and emphasizing significant features.

B. Building an NGO-optimized CNN-BIGRU-AT Model

The paper proposes a NGO-optimized CNN-BIGRU-AT power load prediction model, and the establishment process of the model is shown in Fig4.

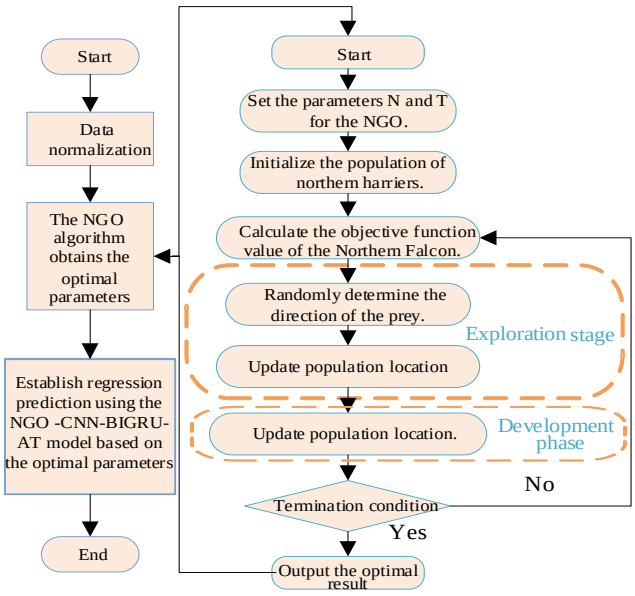


Fig.4 Structure diagram of NGO-CNN-BIGRU-AT

In Fig.4, NGO [21] is employed to optimize the optimal number of hidden layer nodes, initial learning rate, and regularization coefficient of BIGRU to establish a more optimal power load forecasting model. The main steps for NGO to optimize BIGRU model parameters are outlined as follows:

Step 1: Set the fundamental parameters of the NGO algorithm, including the population size and maximum number of iterations.

Step 2: Initialize the population of the algorithm.

Step 3: Define the fitness function F = F(X). Select Root Mean Square Error (RMSE) as the fitness function, and define the optimal number of hidden layer nodes, initial learning rate, and regularization coefficient in BIGRU as the position of the NGO. The algorithm calculates the root mean square error between the actual value and the predicted load.

Step 4: In the exploration stage [22], determine the prey location of the *i*-th Northern Harrier by Equation (12),

calculate the fitness function value F_{P_i} at the prey location, and the fitness function value F_i corresponding to the i-th Northern Harrier's position. Compare these two values to form a new position $X_i^{new,Pi}$ for the i-th Northern Harrier, as shown in Equation (13). Then calculate the fitness function value at the new position $X_i^{new,Pi}$ and compare it with F_i^{new,P_i} to determine the final position of the i-th Northern Harrier, as shown in Equation (14). This process is repeated to obtain the optimal position for the entire Northern Harrier population and determine the optimal region for parameter estimation.

$$P_i = X_k, \begin{cases} i = 1,2,\cdots,N \\ k = 1,\cdots,i-1,i+1,\cdots,N \end{cases} \tag{12}$$

$$X_i^{new,Pi} = \begin{cases} X_i + r\left(P_i - I \cdot X_i\right), F_{P_i} < F_i \\ X_i + r\left(X_i - P_i\right), F_{P_i} \geq F_i \end{cases} \tag{13}$$

$$X_i = \begin{cases} X_i^{new,P_i}, F_i^{new,P_i} < F_i \\ X_i, \quad F_i^{new,P_i} \geq F_i \end{cases} \tag{14}$$

In this scenario, X_i denotes the position of the i-th northern eagle, where N signifies the population size of the northern eagle. P_i represents the position of the prey of the i-th Northern Harrier, and F_{P_i} denotes the corresponding fitness function value. Additionally, F_i represents the fitness function value of the i-th Northern Harrier. The variable k represents a random integer within the range of [1, N].

During the exploration stage, the new position of the i-th Northern Falcon in the j-th dimensional space is denoted as $X_{i,j}^{new,P_i}$ .The fitness function value for the i-th prey position of the Northern Harrier is F_i^{new,P_i} .

The new position of the i-th Northern Harrier during the exploration stage is X_i^{new,P_i} and its corresponding fitness function value is F_i^{new,P_i} , and r is a random number within the range of [0,1].

Step 5: During the development stage [23], once the Northern Harrier attacks its prey, the prey will attempt to escape. T according to Equation (15), the new position of the Northern Harrier X_i^{new,P_2} is determined, it can be calculated as follows:

$$X_i^{new,P_2} = X_i + R \cdot \left(2r - 1\right) \cdot X_i \tag{15}$$

where $R = 0.02(1\text{-}t/T)$, t and T represent the current number of iterations and the maximum number of iterations, respectively [24].

By using Equation (16), the final position of the i-th Northern Eagle is determined by comparing the fitness function values of the new position $F_i^{new,P2}$ and the original position F_i . Following this process, the optimal position for the entire northern eagle population can be obtained [25].

$$X_i = \begin{cases} X_i^{new,P2}, F_i^{new,P2} < F_i \\ X_i, F_i^{new,P2} \geq F_i \end{cases} \tag{16}$$

Step 6: Determine whether the termination conditions of the iteration are met. If they are met, the obtained position represents the optimal parameters. Otherwise, return to Step 3 and continue the iteration process.

Step 7: Use the optimal parameters to train the BIGRU prediction model.

IV. Example Validation and Result Analysis

The data is the historical load data of a specific region in southern China from July 10, 2018, to October 10, 2018. Data is collected every 15 minutes and 96 samples is collected per day. After data processing, the obtained dataset is partitioned into a training set and a testing set in an 8:2 ratio. The fluctuation of power load is influenced by various factors, including historical load, temperature, humidity, etc., which exert a significant impact on power load. This study utilizes the previously sampled load, as well as the maximum temperature, minimum temperature, average temperature, and average humidity as input variables. To highlight the distinctions between the model and other models, models including BP, BIGRU, CNN-BIGRU and CNN-BIGRU-AT are used for load forecasting under the same conditions. Fig. 5 shows a comparison of load forecasting curves for various models over three consecutive days. In order to further compare the differences between various models, error evaluation metrics are used to compare the prediction performance of each model, and the corresponding results are shown in Table 1.

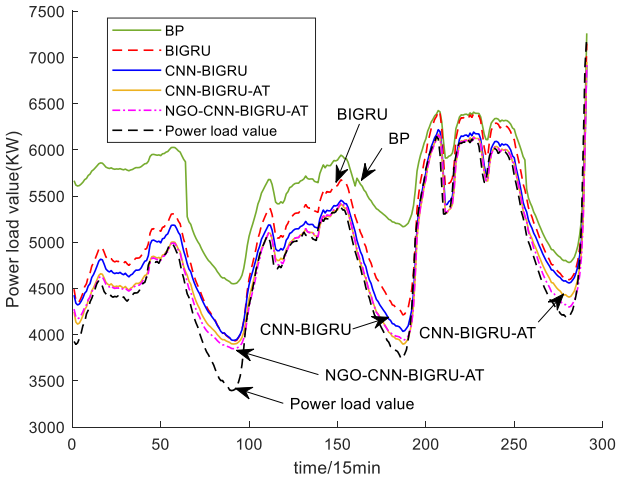


Fig.5 Comparison of model predictions

Table I
Performance Evaluation Indexes of Each Forecasting Model

	MAE	RMSE	R^2	MAPE/%
BP	825.15	911.44	0.52	18.55
BIGRU	352.20	366.86	0.75	7.75
CNN-BIGRU	222.96	262.93	0.87	5.09
CNN-BIGRU-AT	127.46	180.50	0.94	2.94
NGO-CNN-BIGRU-AT	110.58	160.19	0.95	2.53

In Table 1, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2) can be calculated as follows:

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left(y_{\mathrm{i}}-\hat{y}\right)^{2}} \tag{17}$$

$$\mathrm{MSE}=\frac{1}{\mathrm{n}}\sum_{i=1}^{\mathrm{n}}\left|\hat{y}_{\mathrm{i}}-y_{\mathrm{i}}\right| \tag{18}$$

$$\mathrm{MAPE}=\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\frac{\left|\hat{y}_{i}-y_{\mathrm{i}}\right|}{y_{\mathrm{i}}}\% \tag{19}$$

$$\mathrm{R}^{2}=\frac{\mathrm{SSR}}{\mathrm{SST}}=\frac{\sum\left(\hat{y}-\bar{\mathrm{Y}}\right)^{2}}{\sum\left(y_{\mathrm{i}}-\bar{\mathrm{Y}}\right)^{2}} \tag{20}$$

From Fig. 5 and Table 1, it is evident that the load prediction curve of the proposed hybrid model, NGO-CNN-BIGRU-AT, is very close to the actual load curve, exhibiting the highest degree of fitting. Compared with the BIGRU model, CNN-BIGRU-AT and CNN-BIGRU have significant improvements in MAE, RMSE, MAPE, and R^2. This indicates that compared with the direct prediction using a single model, the hybrid model can effectively improve the prediction performance of the model. Furthermore, compared with CNN-BIGRU, the MAE, RMSE, and MAPE of CNN-BIGRU-AT decrease by 42%, 31%, and 42.2%, respectively, while R^2 increases by 0.07. This is because using CNN networks to extract features of influencing factors and applying attention mechanisms to prioritize key information can comprehensively consider information, and further enhance the prediction performance of the model.

Compared with BIGRU, CNN-BIGRU demonstrates a reduction in MAE, RMSE, and MAPE by 36%, 28% and 34%, respectively, while an increase in R^2 by 0.13. This indicates that integrating attention mechanisms into historical load time series can enhance prediction accuracy. Furthermore, compared with CNN-BIGRU, CNN-BIGRU-AT and CNN-BIGRU exhibit excellent performance on MAE, RMSE, MAPE, and R^2. In comparison to other models, the proposed model NGO-CNN-BIGRU-AT demonstrates superior performance, with MAE of 110.58, RMSE of 160.19, MAPE of 2.53%, and R^2 of 0.95. Compared to CNN-BIGRU-AT, NGO-CNN-BIGRU-AT exhibits decrease in MAE, RMSE, and MAPE by 13%, 11.2%, and 13.9%, respectively, while R^2 increases by 0.01. Thus, it is verified that utilizing optimization algorithms to determine the optimal number of hidden layer nodes, initial learning rate parameters, and regularization coefficients of CNN-BIGRU-AT model effectively enhances prediction accuracy. Consequently, the proposed hybrid model accurately forecasts the changes in actual load data and surpasses other models, rendering it more suitable for short-term load forecasting.

The proposed NGO-CNN-BIGRU-AT model, as well as the BP, BIGRU, CNN-BIGRU, and CNN-BIGRU-AT, are simulated and applied for power load forecasting on weekends (September 29, 2018, Saturday), weekdays (September 28, 2018, Friday), and holidays (October 1, 2018, National Day). The comparison curves between predicted values and actual values are depicted in Figs. 6-8.

According to Fig. 6, the power load begins to rise at 7 o'clock, which is attributed to people waking up and the activation of electrical devices.

Around 12 o'clock, the power load begins to decline as people usually leave home or office for lunch, leading to a decrease in electricity demand. Subsequently, at 1 o'clock, when people finish their lunch break and commence returning to their homes or offices, the power load resumes to rise, leading to a surge in electricity demand.

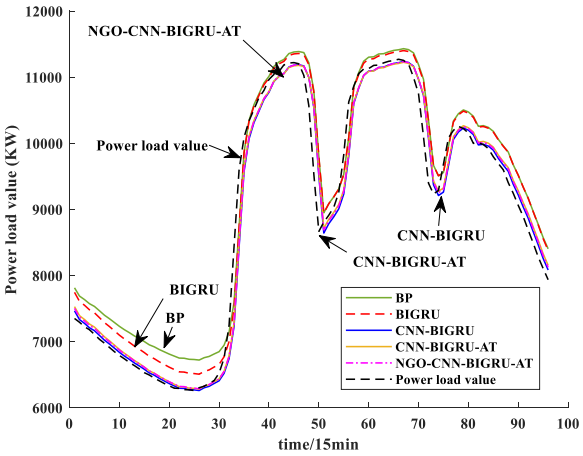


Fig.6 Comparison of load forecast results on rest days

By 7 o'clock, the power load diminishes again, coinciding with dinner time, people may leave home or workplace to eat dinner, thus reducing electricity demand. There are two peaks in the morning and afternoon, corresponding to people waking up and returning home, while there are two valleys in the afternoon and evening, consistent with lunch and dinner times. Compared with weekdays, this pattern shows a more noticeable trend on weekends, possibly attributed to the increased electricity demand for equipment in office buildings and factories during weekdays. Moreover, external factors such as seasons and weather conditions may also influence changes in power load.

TABLE II

PERFORMANCE EVALUATION INDEXES OF EACH MODEL ON REST DAY

	MAE	RMSE	R^2	MAPE/%
BP	353.95	420.85	0.951	4.34
BIGRU	295.46	371.84	0.960	3.50
CNN-BIGRU	185.54	299.24	0.978	2.05
CNN-BIGRU-AT	181.83	297.03	0.982	2.00
NGO-CNN-BIGRU-AT	175.26	295.69	0.990	1.90

According to Table 2, compared with CNN-BIGRU-AT and CNN-BIGRU, the MAPE of NGO-CNN-BIGRU-AT has decreased by 0.1% and 0.15%, respectively. RMSE has decreased by 1.34 and 3.55, and MAE has been reduced by 6.25 and 10.54. Whereas R^2 has improved by 0.008 and 0.012, respectively. These results indicate an enhancement in the accuracy of the predictive model, thereby confirming its effectiveness.

On weekdays, the trend of power load is slightly different from that on weekends. Due to work or school obligations, the load rises earlier and steeper in the morning on weekdays. At noon, the load decreases significantly as many people dine out and leave workplaces. In the afternoon, as people continue to use electricity at home or in the office, the load rises again. At 7 pm, as most people have returned home, the

load decreases again, which may reduce the demand for electricity.

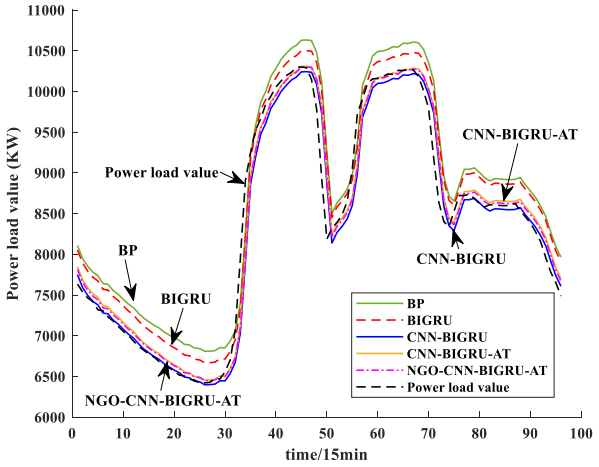


Fig.7 Comparison of load forecast results on working days

According to the observation results in Fig. 7, compared with other curves, the fitting curve of BP significantly deviates from the actual load, whereas the predictive values of NGO-CNN-BIGRU-AT are close to the actual load values and demonstrate better alignment.

The results in Table 3 indicate that compared with BP model, the MAE, RMSE and MAPE of NGO-CNN-BIGRU-AT model decreased by 62.1%, 42.3%, and 2.97%, respectively, while R^2 increased by 0.08.

In conclusion, compared with the single model, the hybrid model has higher accuracy.

Fig. 8 illustrates a notable disparity in power load fluctuations between National Day holidays and working days/rest days. Throughout the entire holiday period, the load gradually decreases at night and begins to rise around 7 am.

TABLE III

PERFORMANCE EVALUATION INDEXES OF EACH MODEL IN WORKING DAY

	MAE	RMSE	R^2	MAPE/%
BP	373.16	415.19	0.902	4.59
BIGRU	291.75	343.29	0.931	3.59
CNN-BIGRU	151.19	242.19	0.971	1.82
CNN-BIGRU-AT	144.35	242.03	0.978	1.71
NGO-CNN-BIGRU-AT	141.19	239.45	0.982	1.62

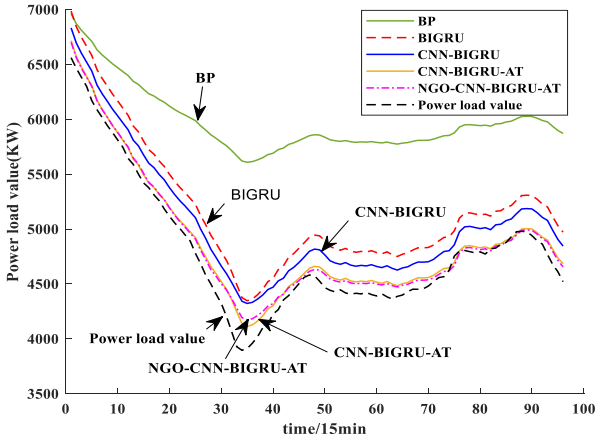


Fig.8 Comparison of load forecast results of holidays

In contrast to working days and rest days, the load variation during lunch breaks and evenings is relatively small, which may be caused by people engaging in leisure activities and socializing with friends. Consequently, the power load remains relatively stable.

TABLE IV

PERFORMANCE EVALUATION INDEX OF EACH MODEL DURING HOLIDAYS

	MAE	RMSE	R^2	MAPE/%
BP	1181.13	1220.4	0.365	25.68
BIGRU	376.75	380.38	0.604	8.25
CNN-BIGRU	256.33	263.83	0.812	5.49
CNN-BIGRU-AT	96.95	112.8	0.963	2.113
NGO-CNN-BIGRU-AT	90.52	110.86	0.978	1.982

The results in Fig. 8 and Table 4 reveal that there is a significant deviation between the predicted value of BP model and the actual load values during holidays. Traditional BP model faces challenges in accurately forecasting the load of special holidays. In contrast, the BIGRU model shows relatively stable performance in predicting load during holidays. Compared with BP model, MAE, RMSE and MAPE reduced by 68%, 68.8% and 17.43%, respectively, and R^2 increased by 0.239. The accuracy of the hybrid model has been improved in load forecasting during holidays, such as the MAE of CNN-BIGRU model is 256.33, RMSE is 263.83, MAPE is 5.490 and R^2 is 0.812. Compared with CNN-BIGRU, MAE, RMSE and MAPE of NGO-CNN-BIGRU-AT model decreased by 64.6%, 57.9% and 3.508%, respectively, and R^2 decreased by 0.166. The performance of CNN-BIGRU-AT model on MAE, RMSE, MAPE and R^2 are improved by 6.6%, 1.7%, 0.131% and 0.015, respectively. In summary, the NGO-CNN-BIGRU-AT hybrid model outperforms other models in power load forecasting.

V. CONCLUSION

To enhance the accuracy of short-term power load forecasting, this paper proposes a short-term power load forecasting model based on NGO-CNN-BIGRU-AT. And using R^2, RMSE, MAE and MAPE as performance evaluation indicators, the model is verified using actual power load data. Compared with the prediction results of BP, GRU, BIGRU, CNN-BIGRU and CNN-BIGRU-AT model, the proposed NGO-CNN-BIGRU-AT model provides a feasible method for accurately predicting power load.

REFERENCES

[1] Y. X. Huang, "Short-term load forecasting research in intelligent distribution network". *Guizhou University*. 2021.

[2] F. S. Zhang, Hong. W, Ti. H, X. Sun, Z. Zhang, J. Cao, "Short-term load forecasting based on partial least squares regression analysis". *Power System Technology*, vol. 3, no.5, pp. 36-40, 2003.

[3] W. Tang, S.Y. Zhong, J. Shu, "Research on spatial load forecasting method for distribution network based on GRA-LSSVM". *Power System Protection and Control*, vol.24, no.46, pp. 76-82, 2018.

[4] D.C. Park, M.A.E1-Sharkawi, R. J. Marks II, "Electric load forecasting using an artificial neural network". *IEEE Trans on Power Systems*, vol.2, no.6, pp.442-449, 1991.

[5] F.X. Meng, H. Qu, "A Study on power load prediction method based on GA and SVM". *Journal of Computer Science*, vol. S1, no.41, pp. 91-93+117, 2014.

[6] G.T. Chen, T. Huan. "Short-term load forecasting based on hybrid neural network deep learning". *Hydroelectric Energy Science*, vol. 4, no.38, pp. 193-196, 2020.

[7] T. Gong, H. Li, H. Liu. "Short-term power load forecasting method based on improved GRU". *Communication and Power Supply Technology*, vol.4, no.38, pp. 1-4+7, 2021.

[8] Y. Zhao, Z.P. Gao, Y. Xiao, R.M. Chang, "Short-term load forecasting based on multi-layer clustering and improved BP neural network". *Journal of Wuhan University (Engineering Edition)*, vol.07, no.52, pp. 622-629, 2019.
[9] B.X. Li, S.W. Zhang, F.H. Huang. "Power load forecasting based on LSTM and self-attention mechanism". *China Measurement and Test*, vol. S2, no.48, pp. 38-43, 2022.
[10] T.Y. Cai, X.H. Zeng. "Short-term load forecasting method based on Attention-BiGRU with multiple feature extraction". *Hebei Electric Power Technology*, vol.1, no.42, pp. 1-7, 2023.
[11] H. Wang, P. Li, M. Cao, "A method for long-term and short-term electricity load forecasting based on CNN-BiLSTM". *Journal of Computer Simulation*, vol.03, no.39, pp. 96-103, 2022.
[12] J.J. Ren, Z.L. Zou, "Super short-term electricity load forecasting based on CNN-BiLSTM-Attention". *Journal of Power System Protection and Control*, vol.8, no.50, pp. 108-116, 2022.
[13] S.J. Zhang,"Research and implementation of electricity analysis and load prediction system based on LSTM". *Shihezi University*, 2023
[14] Y. Yan, H. Sun, Z. Bai, "Analysis method of low-voltage distribution area users' electricity consumption behavior based on MEFE and GN-CNN". *Journal of Power Supply and Utilization*, vol.12, no.40, pp. 79-86, 2023.
[15] P. Zhang, X. Yin, "Short-term load forecasting of agricultural greenhouse park electricity consumption based on VMD-CNN-LSTM". *Information and Control*, pp. 1-11, 2023.
[16] S. Chen, H. Ma, "Encryption traffic classification based on attention mechanism of CNN and BiGRU". *Journal of Computer Science*, pp. 1-10, 2023.
[17] L.Zhang, X.Li, "Fault diagnosis model of wind turbine generator based on improved CNN and BiGRU with dual-channel feature fusion". *Journal of Tianjin University of Science and Technology*, vol.1, no.38, pp. 55-60+69, 2023.
[18] Q. Lian, W. Sun, "Short-term ship traffic flow prediction model based on attention mechanism of stacked LSTM". *Journal of Dalian Maritime University*, pp. 1-10, 2023.
[19] Y. Huang, Z.Li, "Short-term load forecasting using CNN-BiGRU with time pattern attention mechanism". *Journal of North China Electric Power University (Natural Science Edition)*, vol.6, no.50, pp. 11-20, 2023.
[20] Y. Xie, H. Wang, Y. Yue, "Short-term load forecasting considering VMD residual and optimized BiLSTM". *Complexity Systems and Complexity Science*, pp. 1-9, 2023.
[21] W. Fu, X. Zhang, H. Zhang, "Research on ultra-short-term wind speed forecast based on INGO-SWGMN hybrid model". *Journal of Solar Energy*, pp. 1-11, 2023.
[22] Y. Han, H. Zhu, K. Li. "Short-term photovoltaic power forecasting using clustering and stochastic configuration network". *Journal of Electronic Measurement and Instrumentation*, pp. 1-12, 2023.
[23] J. Song, S. Cui, H. Liu, "Ultra-short-term wind power forecasting based on quadratic decomposition NGO-VMD residual and long short-term memory neural network". *Science Technology and Engineering*, vol.6, no.23, pp. 2428-2437, 2023.
[24] Qingrong Wang, Kai Zhang, Changfeng Zhu, and Yutong Zhou, "A multi-regional spatio-temporal network for traffic accident risk prediction". *Engineering Letter*, vol.31, no.3, pp. 906-918, 2023.
[25] Chi Ma, Jie Wu, Hui Hu, Yue Nai Chen, and Jing Yan Li, "Predicting stock prices using hybrid LSTM and ARIMA model". *IAENG International Journal of Applied Mathematics*, vol. 54, no.3, pp. 424-432, 2024.