International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-10, Issue-5; May, 2023
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.105.13

Improved Extreme Learning Machine Power Load Forecasting Based on Firefly Optimization Algorithms

Riyadzh Mahmudh[1,2], Mst Sharmin Kader[1,2], Han Xioaqing[1,2,*]

[1]College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China
[2]Shanxi Key Laboratory of Power System Operation and Control, Taiyuan University of Technology, Taiyuan 030024, China
Corresponding author: Han Xioaqing

Received: 09 Apr 2023,
Receive in revised form: 12 May 2023,
Accepted: 20 May 2023,
Available online: 31 May 2023

Keywords*— Power Load Forecasting, Firefly Algorithm, Extreme Learning Machine, Optimization.

Abstract*— Power load forecasting is a crucial safeguard for the reliable, efficient, and safe operation of the power system, which is connected to the smooth operation of society at all levels. In practical applications, extreme learning machines have advantages like quick learning rates and minimal training error, but they have poor stability and generalization skills. The learning of the sample lacks relevance because the weight matrix between the hidden layer and the output layer in the limit learning is chosen at random. Because of its simple algorithm flow and strong global optimization capabilities, the firefly algorithm helps to simplify the calculation process. To address the drawbacks of the extreme learning machine and combine its benefits, this paper integrates the firefly algorithm into limit learning and makes use of its potent optimization capabilities to determine the connection weight between the extreme learning machine's hidden layer and output layer when the training error is minimal.

I. INTRODUCTION

With the continuous progress of human society and the continuous development of computer technology, scholars are no longer satisfied with knowing the current state of things, but full of curiosity about the coming future. For the power system, in addition to the real-time monitoring of the operation status of each component, the prediction of power load is also important for the whole power system. For example, for short-term load prediction, if the prediction result is low, it indicates that the power demand in the next period is not high. For the whole power plant unit, it is necessary to adjust the corresponding unit load to save energy consumption. On the contrary, if the short-term power load forecasting result is high, it indicates that the next period may be the peak period of power consumption, and the unit needs to be in full load operation state to meet the large power demand. Therefore, power load forecasting has gradually become an important basis for power plants to formulate operation plans. Only through accurate and real-time load forecasting information and corresponding adjustment of the power generation capacity of the whole power plant can the reasonable scheduling of power generation and transmission mode be ensured, and the energy-saving and economic operation of the power plant be realized. The long-term power load forecasting is not only related to the operation scheme of the power generation system itself, but also related to the capital start-up of the power plant construction and the design planning of the whole power plant. Whether the power load forecasting is accurate or not will directly affect the operation time of each equipment in the system and the coordinated operation scheme between the equipment. For each power plant, accurate load forecasting is always important. An appropriate load forecasting is closely related to the investment and construction of the power plant. It is related to whether the investment can be returned and how to obtain more economic benefits. Therefore, power load forecasting plays a vital role in the efficient and energy-saving operation of the entire system. However, the accuracy of the actual power load forecasting results cannot fully meet people's needs, which is related to the error of the power system itself for the load measurement. A larger part of the reason is that the accuracy and adaptability of the soft sensing method need to be further explored and improved.

Power load forecasting is to combine the actual data of power system operation and consider the factors affecting power load, analyze a series of factors affecting load change through the mining and sorting of historical data, and find out the change law of power load in a certain period, to realize the scientific prediction of power load in the future [1-2]. "High precision load forecasting is an important basis for making correct decisions, and an important guarantee for relevant power departments to formulate more accurate power generation plans, carry out infrastructure construction, and achieve economic and effective power dispatching." Under the current situation, the power load is the correct reflection of the supply and demand of the power market. Therefore, accurate prediction is related to the economy, reliability, and stability of the whole power grid [3-4].

II. THEORITICAL BASIS

2.1 Basic concepts of power load forecasting

Power load forecasting is based on scientific and correct theory, with the help of specific forecasting models or forecasting methods, comprehensively considering historical power load data, economic and social environment, temperature and weather and unexpected events between forecast dates, finding out the influence degree of these factors on power load changes, and further analyzing and mining the influence degree, So as to make a more accurate inference on the future trend of power load. The key points of studying power load forecasting are as follows: first, the power load is greatly affected by random factors, which is not only related to the natural environment at that time, but also affected by the policy, market, and production management level; Secondly, the restriction of forecasting model, the uncertainty and nonlinearity of power load forecasting make many mathematical methods difficult to adapt. Power load forecasting refers to two aspects, one is hardware equipment, which refers to the equipment installed at each user, and the other is specific digital, that is, the amount of electricity consumed by the electrical equipment.

There are many methods for power load forecasting. From different perspectives, load forecasting has different classification methods. In general, from the perspective of time, the following categories are discussed in Table 1.

Table 1: Basic categories of power load forecasting

Categories	Details
Long Term	The prediction time of long-term power load forecasting is usually greater than or equal to 5 years. Due to the long prediction time, long-term power load forecasting is used for the planning and construction of power system.
Mid term	The time range of medium-term power load forecasting is wide, ranging from several weeks to several months. This type of load forecasting is aimed at the operation stage of the power system, to help dispatchers to conduct scientific dispatching of power generation capacity.

Short term	The short-term power load forecasting is shorter in time than the medium and long-term load forecasting. It forecasts the load of the next day, and the longest time is the load of each day in the next week.

The load forecasting technology needs to comprehensively consider many aspects, the most important part of which is to find the development law of the load of the measured system through the collation and analysis of the historical load data, to find the mathematical model that can describe the measured system. Driven by the continuous efforts of scholars at home and abroad, load forecasting has made great breakthroughs. At present, there are many mature load forecasting methods. Table 2 describes typical prediction methods.

Table 2: Load forecasting methods

Method	Description	Literatures
Grey prediction method	The most significant advantage of the grey prediction method is that it requires fewer sample data and does not consider the distribution and change laws of the samples. Therefore, the grey prediction method has low computational complexity and high prediction accuracy.	[5-6]
Neural network prediction method	Neural network is a new artificial intelligence algorithm proposed by scholars by simulating the neural structure and function of human brain. This algorithm has many characteristics and functions of human neural structure, including memory, autonomous learning, and knowledge reasoning.	[7-9]
Wavelet analysis method	The principle of wavelet analysis is to use a variety of “wavelet basis functions” to decompose the “original signal”, to realize the processing, storage, transmission, or reconstruction of the signal. Wavelet analysis has been widely used in signal processing, pattern recognition, fault diagnosis and language recognition.	[10-12]
Fuzzy logic method	Fuzzy logic is to use fuzzy sets and fuzzy rules to infer the system with difficult model determination or the controlled object with strong nonlinearity and serious delay by imitating the reasoning thinking mode and uncertainty concept of human brain. The core idea of fuzzy control is the theory of fuzzy mathematics.	[13-15]
Support vector machine prediction method	Support vector machine (SVM) is often used in classification, recognition, and prediction. Later, many prediction fields began to use support vector machine technology, and it has been well applied in practical problems.	[16-18]

2.2 Principal of extreme learning machine

In the past few decades, scholars have made extensive research in the field of neural networks, focusing on multilayer perception (MLP) and radial basis function (RBF) networks. Single hidden layer neural network has been widely studied because of its strong generalization ability and nonlinear approximation ability. Article [19] prove that N different training data of the same continuous system can be infinitely approximated by a single hidden layer neural network (with N neurons). Article [20] further proved that the single hidden layer neural network with n neurons can learn any N samples of continuous system by any bounded nonlinear excitation function. Then, many scholars concluded that if an excitation function satisfying certain conditions is selected, the output of the neural network can approach the objective function with arbitrary accuracy [21], wherein the excitation functions include sine, sigmoid, triangular basis, and radial basis functions. In addition, many scholars have strictly proved that when the excitation function satisfies some given conditions, the input matrix of the neural network can be infinitely close to the expected error. In the traditional method, the hidden layer neuron function and the output weight of the neural network need to be calculated and adjusted. Only after the number of hidden layer neurons and the output weight matrix are adjusted to a certain global optimal value can the neural network approach the given objective function [22].

Let the training set samples be $[x_i, y_i]$ (i =1, 2, ..., N, N is the number of training samples), the number of hidden layer units of ELM is k, and the excitation function is $g(x)$, then the output model of ELM is:

$$O_i = \sum_{j=1}^{k} \beta_j g(a_j x_i + d_j) \quad (1)$$

In formula (1), β_j is the weight connecting the jth hidden layer node and the output node, a_j is the weight matrix connecting the j^{th} hidden layer node and the input node, and d_j is the offset value of the jth hidden layer node. $g(x)$ can be sigmoid, sine or RBF function.

In the training process, find α, β, d satisfies the following equation:

$$\sum_{j=1}^{k} \beta_j g(a_j x_i + d_j) = y_i, i = 1,2, \dots, \mathrm{N} \quad (2)$$

Equation (2) can be expressed by matrix as:

$$H\beta = Y \quad (3)$$

$$H = \begin{bmatrix} g(a_1 x_1 + d_1) & L & g(a_k x_1 + d_k) \\ M & 0 & M \\ g(a_1 x_N + d_1) & L & g(a_k x_N + d_k) \end{bmatrix}_{N \times k} \quad (4)$$

Where $\beta = [\beta_1^T, \beta_2^T, \dots, \beta_k^T]^T, Y = [y_1^T, y_2^T, \dots, y_N^T]^T$

Thus, the connection weight between the hidden layer and the output layer β the minimum 2-norm least squares solution of equation (5) can be obtained:

$$'\beta = H^{+} Y \quad (5)$$

Where H^+ is the Moore Penrose generalized inverse matrix of the hidden layer output matrix H.

To sum up, the specific steps of the extreme learning machine are:

1. The excitation function $g(x)$ and the number of hidden layer neurons k are determined according to the training sample set $[x_i, y_i]$ (i =1, 2, ..., N, N is the number of training samples).
2. Randomly generate input weight matrix α and a hidden layer bias matrix d.
3. According to the known quantity, the output matrix H of the hidden layer is obtained.
4. Calculate the connection weight according to formula (5) β.

To understand the extreme learning machine more intuitively, its network model is shown as follows:

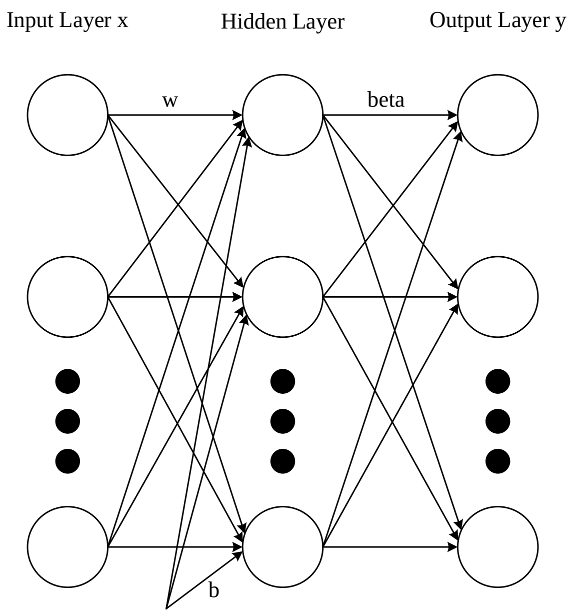

Fig.1 Network diagram of extreme learning machine

It can be seen from Figure 1 that the network structure of the extreme learning machine can be represented as input layer, hidden layer, and output layer. The input layer is used to accept external input variables, the hidden layer is used to complete calculation and identification functions, and the output layer is used to output calculation results.

Compared with the single hidden layer neural network, the extreme learning machine network has no output layer bias value, and the input weight and hidden layer bias value are randomly generated, so that the entire network only needs to determine the output weight, which simplifies the complexity of the traditional neural network and improves the training speed. Therefore, this paper adopts the prediction method based on the extreme learning machine model, which has good practicability.

The extreme learning machine is proposed on the premise of the proved general limit theorem and interpolation theorem. These two theorems show that if the mapping function of a single hidden layer satisfies the condition of infinitely differentiable, the learning ability of a single hidden layer feedforward neural network is not necessarily related to the values of input weights or thresholds, but only closely related to the current network structure. If the selected network structure is suitable, the neural network can fit any continuous function without error. At present, many extreme learning machines models obtain the input weights and thresholds in a random way, which can reduce the phenomenon of overfitting the selected training samples.

2.3 Artificial firefly algorithm

Firefly is a magical product of nature. It is said that it is magical not only because there are many firefly species, which are more than 2000 according to statistics [23], but also because firefly is a natural luminescent body. The fluorescence generated by the tail of firefly is used to attract other small partners to gather in their own area to complete a task.

The artificial firefly optimization (FAO) algorithm is a new swarm intelligence bionic algorithm [24]. Its idea is derived from the fact that firefly adults can show their behaviors of foraging or courtship through the biological characteristics of luminescence. According to the location of the firefly, the algorithm describes the brightness of the firefly and its attraction to other fireflies. The higher the brightness of the firefly, the better its location and the greater its attraction. Each firefly moves and updates according to the brightness and attractiveness of its peers in its own neighborhood structure to achieve the goal of optimizing its position. Once proposed, firefly algorithm has been widely recognized. After continuous in-depth research by many scholars, firefly algorithm has been successfully applied to combinatorial optimization, path planning, image processing, economic scheduling, and other fields [25].

III. DATA PREPARATIONS

3.1 Preprocessing of data

The establishment of the prediction model of the extreme learning machine and the learning ability with high accuracy depend on the learning samples, so the quality of the sample model directly affects the prediction accuracy of the model. If there are errors or large errors in the learning samples, the prediction model may not converge to the ideal error or not. Even if the network can converge, it is difficult to reflect the real change law in the case of defective sample data, and the output stability of the model will be poor. Therefore, sample data should be preprocessed before model prediction. For example, missing data in the sample should be filled according to certain rules, and bad data in the sample should be deleted or adjusted.

Before the experiment, the sample data were first repaired with defective load data and processed horizontally (the horizontal processing makes the sample sequence smooth), and then all historical data were normalized. In this paper, only the normalization processing of samples is described in detail.

3.2 Data normalization

The real value of the power load data will affect the learning accuracy of the extreme learning machine model, increase the learning time, and affect the learning efficiency of the model. Therefore, the real power load data needs to be further normalized. The normalization formula is as follows:

$$x\% = x\%_{min} + (\frac{(x-x_{min})}{(x_{max}-x_{min})(x\%_{max}-x\%_{min})} \quad (6)$$

In formula (6), x represents the real load data, x_{max} represents the maximum value in the real load data, x_{min} represents the minimum value in the real load data, $x\%$ represents the normalized value, $x\%_{max}$ represents the normalized maximum value, and $x\%_{min}$ represents the normalized minimum value. In this paper, $x\%_{max}$ and $x\%_{min}$ is taken as 1 and 0.1 respectively, so the normalization formula in this paper is:

$$x\% = 0.1 + 0.9 \times \frac{x-x_{min}}{x_{max}-x_{min}} \quad (7)$$

The numerical composition matrix after normalization processing is directly applied to the training model of the extreme learning machine. The function of normalization is to narrow the sample data range and reduce the training time of the model, to accelerate the convergence speed, improve the prediction accuracy, and play an optimization role.

3.3 Date and temperature data

In recent years, with the continuous development of the global economy and the improvement of people’s material living standards, people’s requirements for the comfort of living environment and office environment are constantly improving. Therefore, the influence of meteorological factors on power load is becoming increasingly important. When analyzing the area studied in this paper, it is found that the temperature and the date type have the greatest impact on the power load in this area, while the consideration of other factors will not improve the accuracy of the forecast results. Therefore, this paper takes the temperature and the date type of the load day as the main influencing factors in the daily load forecast.

Temperature has an important influence on power load, and the load changes are different under different temperature conditions. When the temperature fluctuates slightly in a certain range, the influence of the temperature on the power load will not be obvious, but when the temperature change range is large, especially in the case of seasonal transition, the temperature will have a profound influence on the power load. Therefore, to increase the accuracy of the prediction results, the influence factor of temperature is included in the analysis scope.

In this paper, according to the characteristics of the considered time date type, it is divided into working days and rest days for quantitative processing. After quantitative processing, the working days are taken as 0 and the rest days are taken as 1.

3.3 Test data

Select the power load data of 56 days in autumn (September to November) in a certain area of Bangladesh, record the data every hour, and record the temperature at that time. It can be seen from Figure 2 that the load fluctuation in autumn is stable, because the temperature change range in autumn is small, between 18℃ and 32℃.

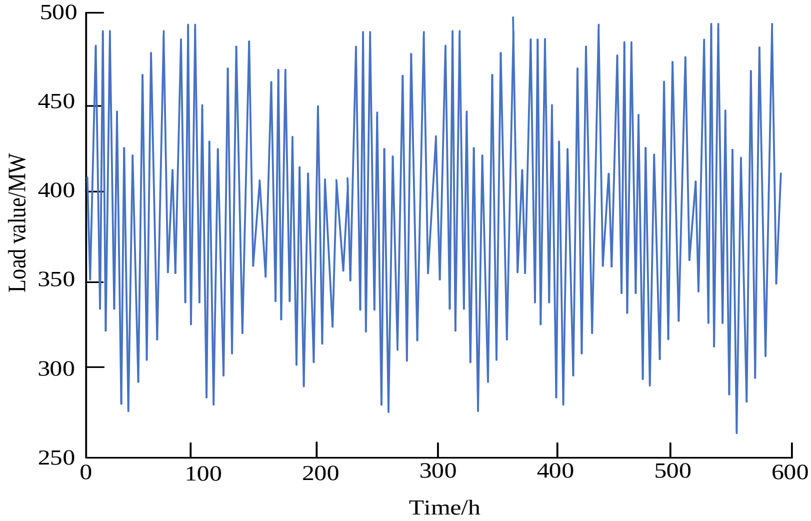

Fig.2 Load curve

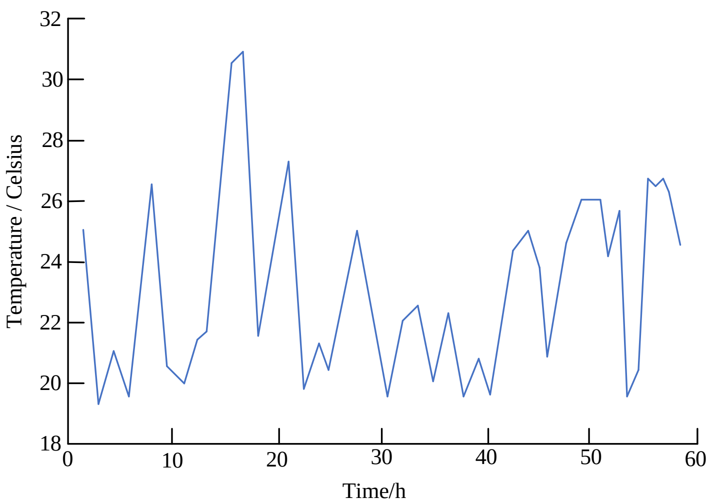

Fig.3 Temperature curve

$$x(i,d) = 0.1 + 0.9 \times (\frac{(x(t,d)-x_{min})}{(x_{max}-x_{min})}) \quad (8)$$

Where, $x(t,d)$ represents the real power load value at time *t* on day *d*, $x(i,d)$ represents the normalized power load value at time i on day d, x_{max} represents the normalized power load value at time *i* on day *d*, and x_{min} represents the minimum value of power load in all real sample data.

After comprehensive analysis, this paper considers two factors: the temperature of the forecast day, whether the forecast day is a national legal holiday and weekend (to save the calculation time of the model, the temperature in the program is taken according to the temperature quantification table, and the date is taken as 0 on the normal working day and 1 on the weekend and legal holiday). The factors affecting the forecast value at a certain time on the forecast day are the day before the forecast day, the two days before the forecast day The load value currently seven days before the forecast date, and the predicted value at the forward time on the forecast date, one day before the forecast date, two days before the forecast date, and seven days before the forecast date. Therefore, the input matrix of the extreme learning machine is $MATRIX_{in}$ and the output matrix is $MATRIZ_{out}$:

$$MATRIX_{in} = \begin{cases} [Q_d, T_i, X_{24,d-8}, X_{i,d-7}, X_{24,d-3}, X_{i,d-2}, X_{24,d-2}, X_{i,d-1}, X_{24,d-1}] \\ i = 1 \\ [Q_d, T_i, X_{i-1,d-7}, X_{i,d-7}, X_{i-1,d-2}, X_{i,d-2}, X_{i-1,d-1}, X_{i,d-1}, X_{i-1,d}] \\ i = 2, \ldots, 24 \end{cases} \quad (9)$$

$$MATRIX_{OUT} = X_{id} \quad (10)$$

Where, X_{id} represents the normalized power load value at time *i* on day *d*, Q_d represents the date type of day *d*, the value of normal working days is 0, and the value of weekends and legal holidays is 1. T_i represents the temperature quantization value corresponding to the i^{th} time predicted by this model.

IV. MODEL

4.1 Power load forecasting model based on artificial firefly algorithm

This section introduces the basic principle of the artificial firefly algorithm to improve the extreme learning machine, that is, the strong global optimization ability of the artificial firefly algorithm is used to find the connection weight matrix between the input layer and the hidden layer and the bias vector of the hidden layer that minimize the training error of the extreme learning machine.

The specific implementation steps of FA-ELM prediction model are as follows:

1. Initialize the given training sample set $[x_i, y_i]$ ($x_i \in R_n$, n is the number of input neurons, i=1, 2, …, N, N is the number of training samples), set the number of hidden layers k of the extreme learning machine and the excitation function g(x). Initializes NP parameter vectors $t_{r,g}$ (r=1, 2, ..., NP), with dimension D(D = k*(n + 1), where the value range of any one dimension is [-1,1], and g represents the number of iterations. The individual T of firefly population is represented by the input weight matrix of ELM $\alpha(\alpha_1, \alpha_2, \ldots, \alpha_k)$ And the hidden layer bias matrix d, t=[a_{11}, a_{12}, …,a_{1n}, …, a_{k1}, a_{k2}, …, a_{kn}, d_1, …, d_k] for each population individual t_{ig}, calculate the hidden layer output matrix H according to formula (4), and then obtain the output weight according to formula (5) β, Finally, the root mean square error (RMSE) of each individual is calculated according to formula (11). The root mean square error is taken as the fitness function of the firefly algorithm to find the minimum value of the root mean square error.

$$RMSE = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{k} \|\beta_j g(a_j x_i + d_j) - y_i\| / N} \quad (11)$$

2. The fitness value of each firefly is converted into the corresponding fluorescence brightness value according to formula ($l_i(\mathrm{g}) = (1-\rho) * l_i(\mathrm{g}-1) + \gamma * f(x_i(\mathrm{g}))$).
3. Determine neighbors' stage: Fireflies look for neighbors within their sensing radius and determine the neighbor set.
4. Moving probability update stage: determine the moving direction of each individual according to the roulette mode in the determined neighborhood set.
5. Move the firefly to move the firefly toward the selected object according to formula ($x_i(\mathrm{g}+1) =$

$x_i(\mathrm{g}) + \mathrm{step} * \left(x_j(\mathrm{g}) - x_i(\mathrm{g})\right)/\|x_j(\mathrm{g}) - x_i(\mathrm{g})\|)$)

6. Update the adaptive sensing radius of Fireflies: after the fireflies move, they need to modify and update the adaptive sensing radius of each firefly according to the neighbor set. The update formula is ($r_d^i(\mathrm{g}+1) = min\left\{r_0, max\{0, r_d^i(\mathrm{g}) + \beta * (n_t - |N_i(\mathrm{g})|)\}\right\}$)

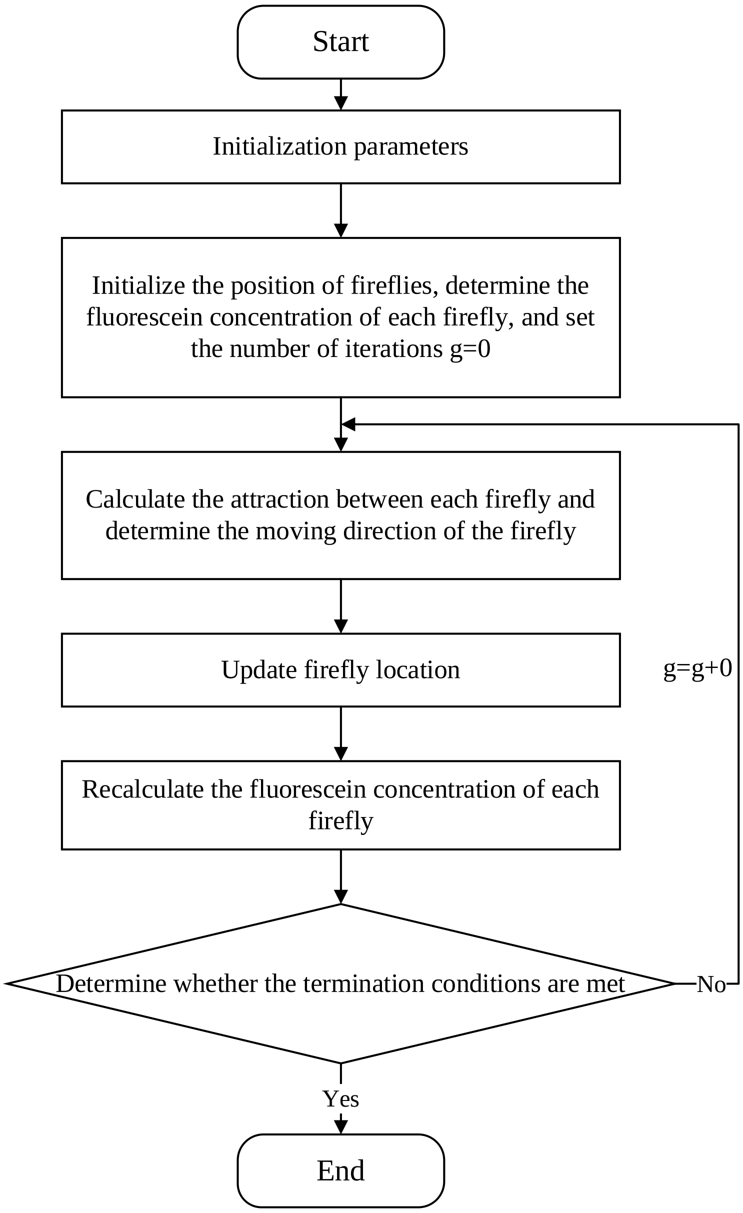

Fig.4 flow chart of FA-ELM prediction model

In this paper, the number of neurons of the extreme learning machine is n=9, so the number of hidden layer nodes is k=2*n+1=19. The transfer function of the hidden layer and the output layer is set as' sin 'function. In the experiment, the number of hidden layers of the extreme learning machine is set as 50. The parameter settings of the firefly algorithm in the text are shown in table 3.

Table 3: FA-ELM parameter setting

ρ	γ	β	set	iter_max
0.3	0.5	0.07	0.02	500

ρ Represents the fluorescein Volatilization Coefficient, γ Represents the fitness extraction ratio, β Represents the change rate of the field, step represents the step size, iter_max represents the number of iterations.

V. RESULTS AND ANALYSIS

5.1 Experimental results

The power load data of 56 days are normalized and put into the program for use. The data of the first 51 days are used as the training sample data. The ELM model and FA-ELM model are used to predict the power load from 1:00 to 12:00 in the next 5 days. To reduce the computational complexity of the extreme learning machine, the prediction for the next 5 days and 12 hours is divided into 12 groups. The output of each group is the value of a certain time unified in the next 5 days, and each group is predicted to run independently for 20 times. The training error and test error of elm and FA-ELM are recorded during each operation. Finally, the average value of 20 operation results is obtained as the final prediction result. The predicted values of 12 hours on the first day and the last day are selected as the results.

It can be seen from Figures that the measurement accuracy of FA-ELM is higher than that of elm algorithm in terms of test error and training error. The relative error of elm prediction model is mostly about 12, while the relative error of FA-ELM prediction model is mostly about 7. The learning ability and generalization ability of FA-ELM model are better than ELM.

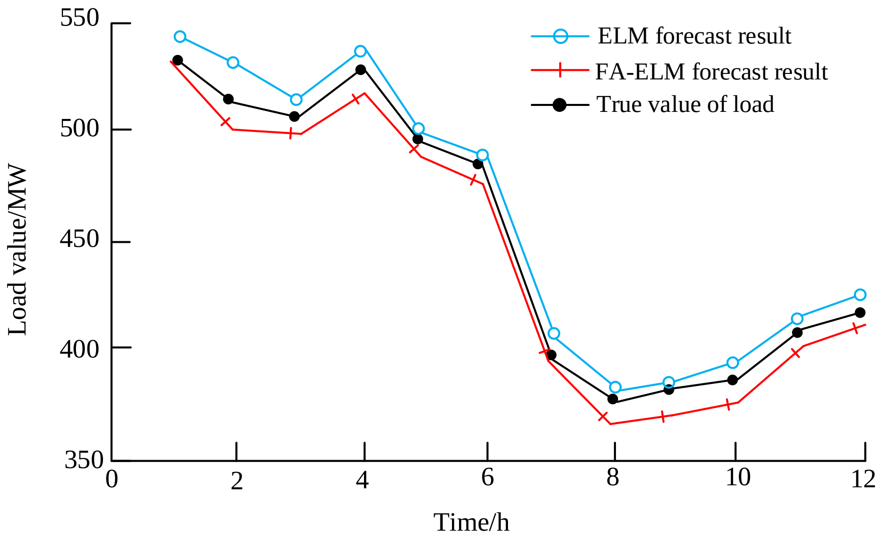

Fig.5 Comparison of prediction results on day 52

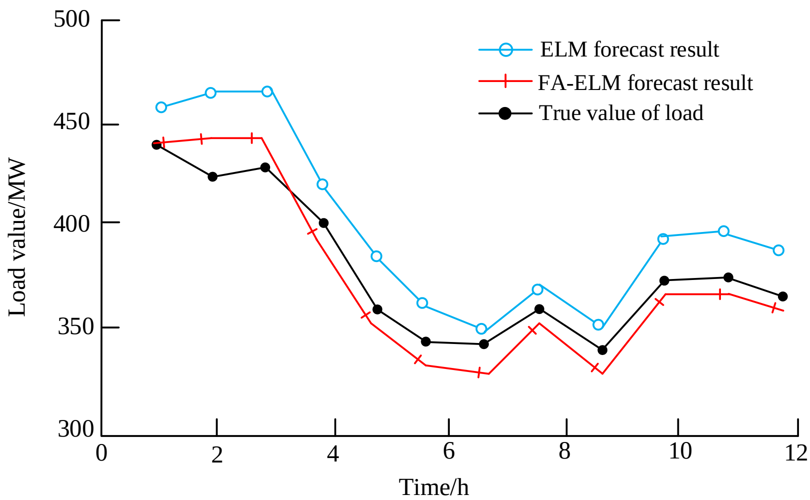

Fig.6 Comparison of prediction results on day 53

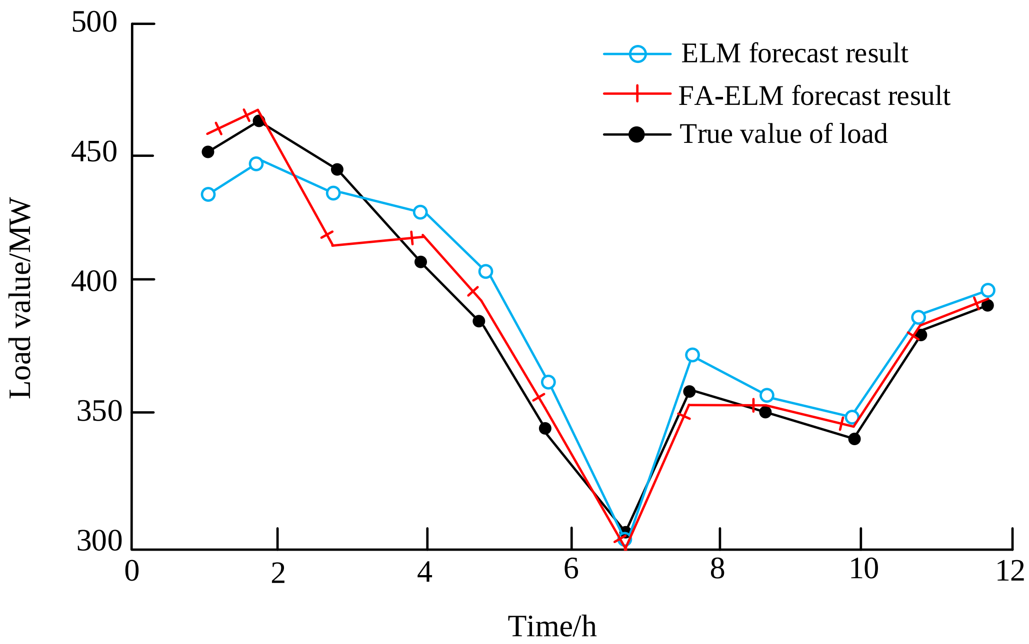

Fig.7 Comparison of prediction results on day 54

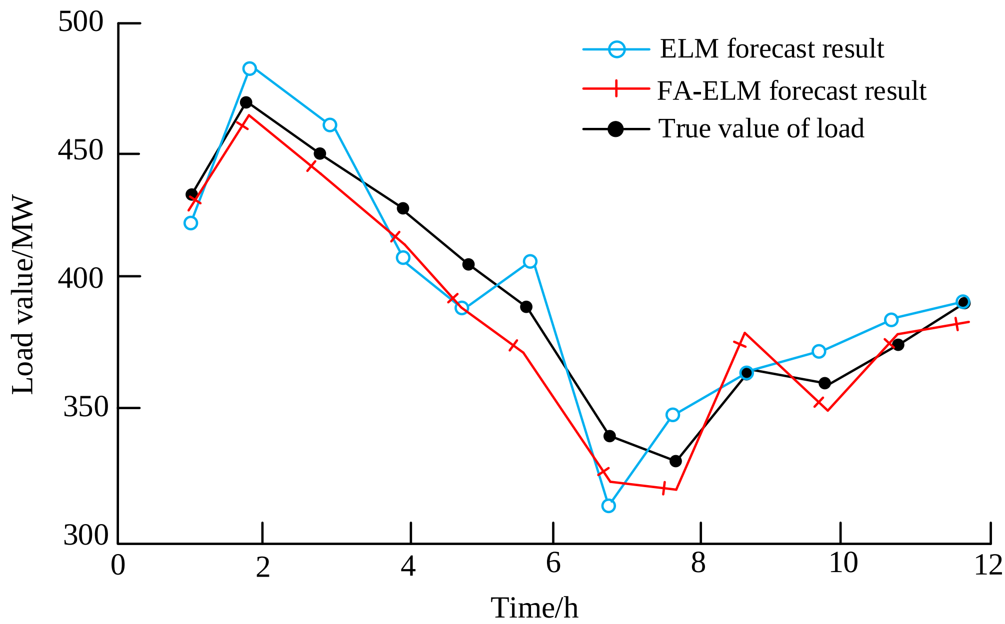

Fig.8 Comparison of prediction results on day 55

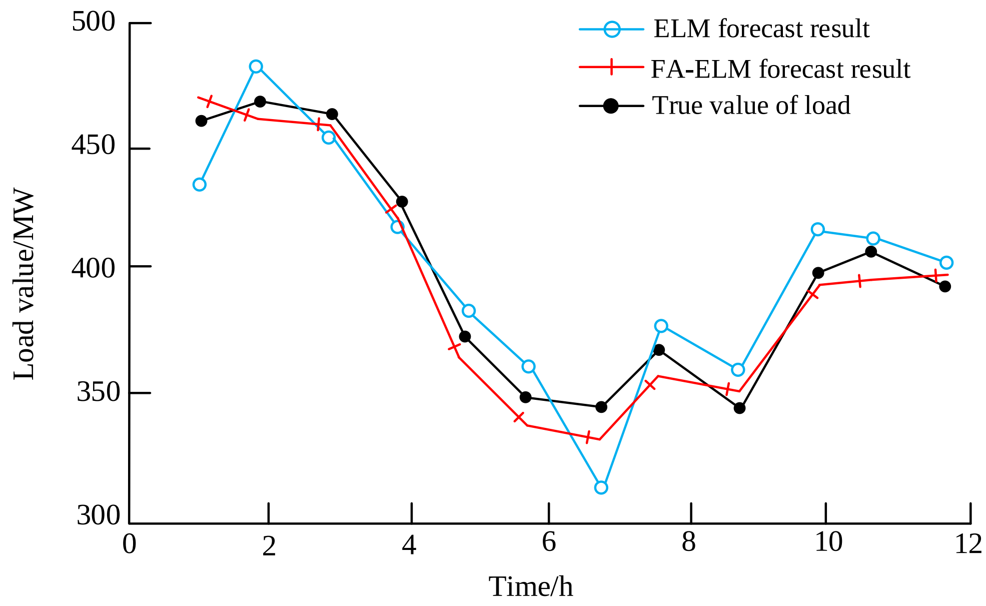

Fig.9 Comparison of prediction results on day 56

In the above five figures, the horizontal axis represents 12 times per day, and the vertical axis represents the load value at the corresponding time. It can be seen from the above five figures that the tracking effect of FA-ELM is better than that of elm in the prediction results of the next five days. The reason is that FA-ELM uses the global optimization ability of the firefly algorithm to find the connection weight matrix *a* and the hidden layer bias vector *d* that match the training samples, thus avoiding the random selection of the elm model, thus greatly reducing the training error, and thus reducing the test error.

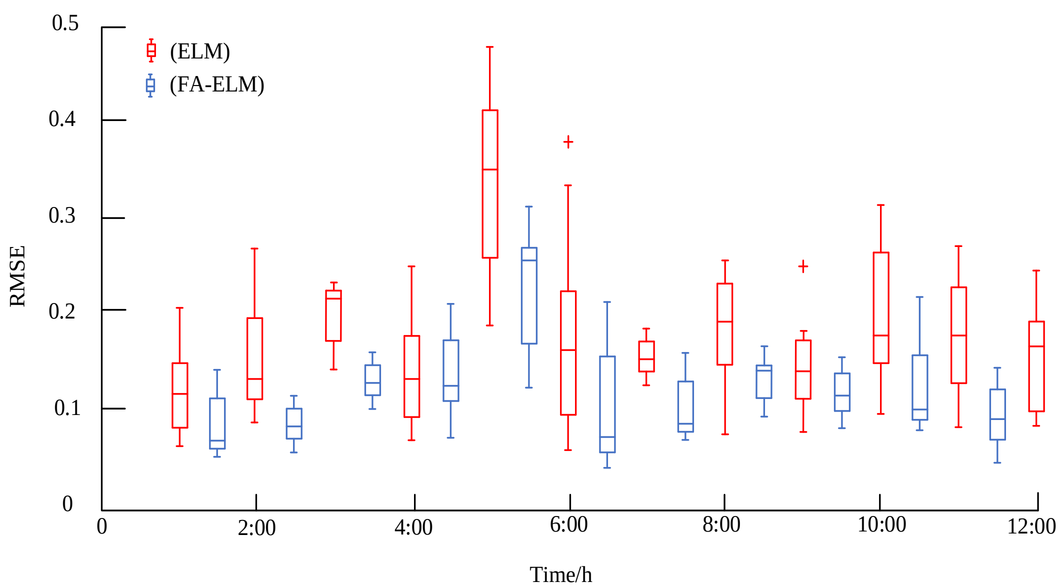

Fig.10 ELM and FA-ELM box line diagram

To analyze the stability of FA-ELM algorithm, the results of 20 measurements are shown by boxplot. Through the analysis of Figure 11, it can be concluded that FA-ELM only has higher measurement accuracy, but also its stability is better than elm.

To fully illustrate the advantages of FA-ELM algorithm, the prediction results of FA-ELM are compared with the traditional BP neural network and support vector machine (SVM). Because BP neural network and support vector machine (SVM) are mature load prediction algorithms, this paper will not give a detailed description. Only the comparison chart of prediction results on day 52 and day 56 is shown for illustration.

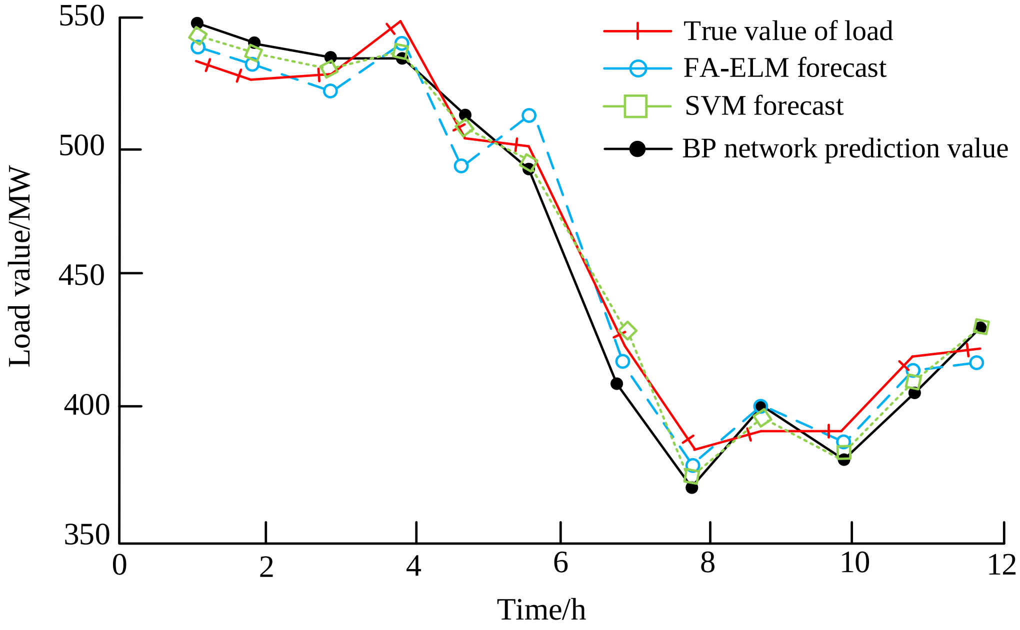

Fig.11 Comparison of prediction results on day 52

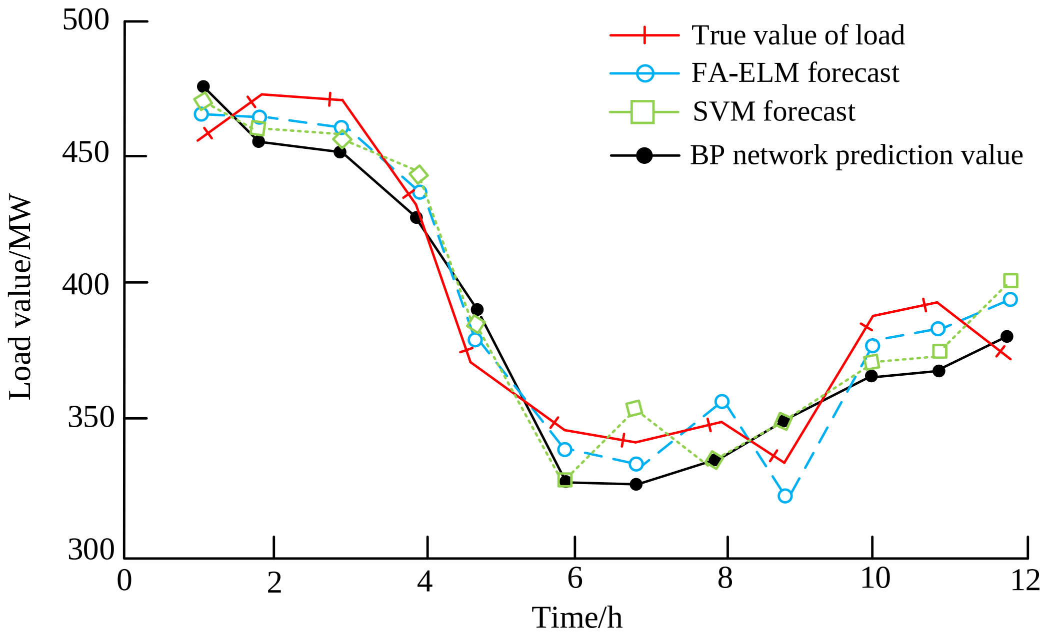

Fig.12 Comparison of prediction results on day 56

From the above graphic analysis, the traditional BP neural network has the largest prediction relative error, and the stability of the BP neural network is poor. The relative error in the test is small, which is related to the defects of the neural network itself. The prediction effect of SVM method is better than that of BP neural network, because SVM has rigorous theoretical and mathematical basis, so its generalization ability is better than that of BP neural network, and the algorithm has global optimization. It can be seen from the graph analysis that FA-ELM algorithm is superior to both in stability and test error. Thus, the effectiveness of the algorithm is proved.

VI. CONCLUSION

This paper introduces the electric power load forecasting model based on the improved extreme learning machine (FA-ELM) of artificial firefly algorithm. Before that, the data preprocessing method in this paper is first introduced, involving the normalization processing of samples and the corresponding inversion formula, as well as other specific processing methods of historical data in the experiment. Then it introduces the specific implementation steps of FA-ELM and shows them with flow chart. The last part of this paper is the display of experimental results. The prediction error of FA-ELM model and traditional elm model on a certain prediction day is compared, and the prediction results of the two models for all prediction days are shown in the form of simulation figures. The experimental results show that the prediction effect of FA-ELM model is better than that of traditional elm model. Finally, the FA-ELM model is compared with the current mature load forecasting model to illustrate the superiority of its algorithm.

REFERENCES

[1] Min Wang, Zixuan Yu, Yuan Chen, Xingang Yang, Jian Zhou, Short-term load forecasting considering improved cumulative effect of hourly temperature, Electric Power Systems Research, Volume 205,2022,107746,ISSN 0378-7796.

[2] Dong, H., Gao, Y., Fang, Y., Liu, M., & Kong, Y. (2021). The Short-Term Load Forecasting for Special Days Based on Bagged Regression Trees in Qingdao, China. Computational intelligence and neuroscience, 2021, 3693294.

[3] Singh, Arunesh & Nasiruddin, Ibraheem & Khatoon, Shahida & Muazzam, Md & Chaturvedi, D. (2012). Load forecasting techniques and methodologies: A review. ICPCES 2012 - 2012 2nd International Conference on Power, Control and Embedded Systems. 1-10. 10.1109/ICPCES.2012.6508132.

[4] S. Izudheen and A. M. Joykutty, “A methodology for Short-term Electric Power Load Forecasting,” 2019 9th International Conference on Advances in Computing and Communication (ICACC), 2019, pp. 322-325.

[5] Z. Zheng, B. Zha, H. Yuan, Y. Xuchen, Y. Gao and H. Zhang, “Adaptive Edge Detection Algorithm Based on Improved Grey Prediction Model,” in IEEE Access, vol. 8, pp. 102165-102176, 2020.

[6] W. Niu, P. Lei, W. Wang, X. Song and J. Cheng, “Multi-parameter Discrete Grey Prediction With Few Observations,” 2021 4th World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM), 2021, pp. 320-324.

[7] X. Wei and J. Dai, “Design and implementation of the data prediction model based on PSO-ELM,” 2021 7th International Symposium on Mechatronics and Industrial Informatics (ISMII), 2021, pp. 272-275.

[8] L. Zhou, R. Wang, Y. Zhu, J. Li and X. Luo, “Landslide displacement prediction based on integrated neural network,” 2020 8th International Conference on Digital Home (ICDH), 2020, pp. 57-62.

[9] J. Zhang, W. Jing, Z. Lu, Y. Wang and X. Wen, “A Hybrid Load Forecasting Method Based on Neural Network in Smart Grid,” 2021 IEEE/CIC International Conference on Communications in China (ICCC), 2021, pp. 928-933.

[10] Fan, GF., Peng, LL. & Hong, WC. Short-term load forecasting based on empirical wavelet transform and random forest. Electr Eng (2022).

[11] Peng, L.-L., Fan, G.-F., Yu, M., Chang, Y.-C. and Hong, W.-C. (2021), Electric Load Forecasting based on Wavelet Transform and Random Forest. Adv. Theory Simul., 4: 2100334.

[12] C. Z. Huan, F. J. Yu, L. Hao and W. P. Pan, “Research on short term load forecasting method of distribution network based on wavelet clustering analysis,” 2021 China International Conference on Electricity Distribution (CICED), 2021, pp. 1086-1090.

[13] Chaouki Ghenai, Omar Ahmed Abduljabbar Al-Mufti, Omar Adil Mashkoor Al-Isawi, Lutfi Hatem Lutfi Amirah, Adel Merabet, Short-term building electrical load forecasting using adaptive neuro-fuzzy inference system (ANFIS), Journal of Building Engineering, Volume 52, 2022, 104323.

[14] Olaru, LM, Gellert, A, Fiore, U, Palmieri, F. Electricity production and consumption modeling through fuzzy logic. Int J Intell Syst. 2022; 1- 17.

[15] D. V. N. Ananth, Lagudu Venkata Suresh Kumar,

Tulasichandra Sekhar Gorripotu, and Ahmad Taher Azar. 2021. Design of a Fuzzy Logic Controller for Short-Term Load Forecasting with Randomly Varying Load. Int. J. Sociotechnology Knowl. Dev. 13, 4 (Oct 2021), 32–49.

[16] Jinghua Li, Yongsheng Lei, Shuhui Yang, Mid-long term load forecasting model based on support vector machine optimized by improved sparrow search algorithm, Energy Reports, Volume 8, Supplement 5, 2022, Pages 491-497.

[17] Kader MS, Mahmudh R, Xiaoqing H, Niaz A, Shoukat MU (2022) Active power control strategy for wind farms based on power prediction errors distribution considering regional data. PLoS ONE 17(8): e0273257.

[18] Li, L. L., Cen, Z. Y., Tseng, M. L., Shen, Q., & Ali, M. H. (2021). Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic-support vector regression machine. Journal of Cleaner Production, 279, 123739.

[19] M. Y. Mikheev, Y. S. Gusynina and T. A. Shornikova, “Building Neural Network for Pattern Recognition,” 2020 International Russian Automation Conference (RusAutoCon), 2020, pp. 357-361.

[20] G. Dudek, “Data-Driven Randomized Learning of Feedforward Neural Networks,” 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8.

[21] I. Basha Kattubadi and R. Murthy Garimella, “Novel Deep Learning Architectures: Feature Extractor and Radial Basis Function Neural Network,” 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 024-027.

[22] Zhao, H. Jiang and Q. Zhang, “Large Array DOA Estimation Based on Extreme Learning Machine and Random Matrix Theory,” 2020 IEEE Radar Conference (RadarConf20), 2020, pp. 1-5.

[23] H. Yin, H. Meng and Y. Zhang, “Adaptive firefly algorithm based on reverse search strategy,” 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), 2021, pp. 80-83.

[24] P. Napalit and M. A. Ballera, “Application of Firefly Algorithm in Scheduling,” 2021 IEEE International Conference on Computing (ICOCO), 2021, pp. 336-340.

[25] . Liu, J. Shi, F. Hao, M. Dai and Z. Zhang, “A New Firefly Algorithm with Enhanced Attractiveness,” 2021 5th International Conference on Automation, Control and Robots (ICACR), 2021, pp. 74-77.