

# Short-Term Load Forecasting Using Hybrid GMDH-LSTM Model Optimized by ICPA

Gonggui Chen, Jie Bai, Teweï Chen, Wei Wang, Zongfu Wang, Hongyu Long\*  
and Mi Zou

**Abstract**—Power system is the primary basis for national development, and electricity load prediction is an essential part of the power system. Accurate short-term load forecasting can ensure the stable operation of the power system. To improve the accuracy of short-term load forecasting, a hybrid prediction model with an improved Carnivorous Plant Algorithm (ICPA) to Optimize the Group Method of Data Handling (GMDH) and Long Short-term Memory (LSTM) is proposed. This prediction model includes Improved Complete Ensemble Empirical Mode Decomposition Adaptive Noise (ICEEMDAN), Information Entropy (IE), LSTM, GMDH, and ICPA. Firstly, ICEEMDAN is applied to the original load data to obtain the Intrinsic Mode Functions (IMF) with different characteristic information. The obtained IMFs are reconstructed according to the IE. Then, the reconstructed IMFs are predicted using LSTM and GMDH forecasting models, respectively. Finally, the final forecasting results are weighted by the optimal weights obtained using ICAP. To verify the performance of the model, two power load datasets are selected as test datasets. And Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and inequality coefficient (TIC) are selected as indicators. The prediction model proposed in this paper outperforms other prediction models in all indicators.

**Index Terms**—short-term load forecasting, long short-term memory (LSTM), group method of data handling (GMDH), improved carnivorous plant algorithm (ICAP).

Manuscript received May 29, 2022; revised October 14, 2022. This work was supported by the National Natural Science Foundation of China (52007022), the Project funded by the China Postdoctoral Science Foundation(2021M693930).

Gonggui Chen is a professor of Key Laboratory of Industrial Internet of Things and Networked Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: chenggpower@126.com).

Jie Bai is a graduate student of Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: baijie9912@163.com).

Teweï Chen is a general manager of State Grid Chongqing Qianjiang Power Supply Company, Chongqing 409000, China (e-mail: chenteweï\_qj@163.com).

Wei Wang is a director of the operation and maintenance Department with the State Grid Chongqing Qianjiang Power Supply Company, Chongqing 409000, China(e-mail: wangwei\_qj@163.com).

Zongfu Wang is a deputy director of the operation and maintenance Department with the State Grid Chongqing Qianjiang Power Supply Company, Chongqing 409000, China(e-mail: wangzongfu\_qj@163.com).

Hongyu Long is a professor level senior engineer of Chongqing Key Laboratory of Complex Systems and Bionic Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (corresponding author to provide phone: +8613996108500; e-mail: longhongyu20@163.com).

Mi Zou is an assistant professor of Chongqing Key Laboratory of Complex Systems and Bionic Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China(e-mail: zoumi@cqupt.edu.cn).

## I. INTRODUCTION

Due to the rapid development of the social economy and technology of science, power resources have become increasingly important in the daily life of people. This also leads to increasing demand for electricity [1, 2]. The high accuracy of electric power prediction plays an irreplaceable role in ensuring a stable electric power supply. In particular, STLF not only supplies a reliable basis for the secure and stable running of the whole power system, and whole power system scheduling planning but also saves the consumption of power generation energy for society [3]. However, in the case of the rapid growth of power demand, the precision of the current load forecasting cannot guarantee the reasonable dispatch of the power system, resulting in the phenomenon that the supply of power resources usually exceeds power demand. Therefore, how to improve power load forecasting is the current trend of research.

Scholars in the field of electricity load have classified load prediction into the following three types based on the load prediction period [4]: Short-Term Load Forecasting (daily load prediction and weekly load prediction) for economical operation planning, real-time control, and safety analysis [5, 6]. Medium-Term Load Forecast (monthly and annual load forecast) for fuel supply, unit maintenance planning, etc. And Long-Term Load Forecasting (3-5 years or even longer load prediction for whole power system, equipment procurement, etc. STLF runs through the power system to ensure the stable running and reasonable dispatch of the whole power system. Besides, STLF is the basis of Medium-Term and Long-Term Load Forecasting. Therefore, Scholars have proposed many STLF methods, including heritage forecasting methods and new artificial intelligence forecasting methods [7].

In 1971, Christiaanse used an adaptive system based on general exponential smoothing to forecast hourly load data [8]. Subsequently, Charytoniuk *et al.* proposed a predictive model combining the advantages of nonparametric regression application. And the validity of the model was verified in hourly load data [9]. Considering the complex random characteristics of load data, J. Li *et al.* proposed the STLF model based on multiple linear regression (MRL) [10]. To achieve optimal distribution system operation planning, a simple load prediction method incorporating Autoregressive Integrated Moving Average (ARIMA) is proposed by J.C. Lopez [11]. These traditional prediction models are simple to use and fast to predict. However, because the nonlinearity of power load data is very complex, the prediction accuracy and stability of traditional load prediction methods are low [12]. Researchers and academics in related fields have proposed many new artificial intelligence neural network prediction

methods to solve these problems. W. Liu *et al.* proposed the STLF model combining the advantages of the GMDH neural network and proved through experiments that the forecasting precision of the proposed STLF model is higher as compared to the ARIMA forecasting model [13]. H. Bian *et al.* proved that the BP neural networks have good prediction ability and prediction stability with high prediction accuracy in the field of STLF [14]. H. Shi *et al.* proved that the load prediction model based on the Recurrent Neural Network (RNN) has a good prediction effect and prediction stability in the field of STLF [15]. Shahzad and Afshin experimentally proved that long-short-term memory (LSTM) neural network has high prediction accuracy and high adaptability in the field of short-term load prediction [16].

Although the statistical forecasting model and the artificial intelligence forecasting model can accomplish high-precision short-term load forecasting, some problems also exist, such as long convergence time, high complexity, and complexity linear modeling. To solve the above problems, various hybrid forecasting models have been widely introduced into the research of power load forecasting [17,18]. To increase the precision of power load forecasting, P. Singh used a novel evolutionary-based algorithm to solve the optimal network weights for a neural network prediction model [19]. LSTM and convolutional neural networks (CNN) are combined by Rafi *et al.* to form a new hybrid prediction mode. This load prediction model is used in the Bangladesh power system for validation, and they concluded that this hybrid power load prediction model has greater stability and higher forecasting precision compared to the single prediction model [20]. In addition, Zheng *et al.* proposed a hybrid load forecasting model that combines an adaptive network fuzzy inference system (ANFIS), multilayer perceptron (MPL), and seasonal autoregressive integrated moving average (SARIMA), in addition to demonstrating the validity and stability of the load forecasting model experimentally [21].

Due to the randomness and uncertainty of the electricity consumption of customers, the time series of the electric load has complex characteristics. In addition to this, load data with different complexity also have different feature messages, which results in the prediction accuracy of the load prediction methods will vary depending on the data information features. Therefore, the Intrinsic Mode Function (IMF) load prediction methods combined with the signal analysis method has also received a lot of enthusiasm from academics and researchers. There are various methods for signal analysis, including EMD, which decomposes complex data into multiple IMFs, and each IMF will contain different local feature information [22]. EMD has been improved to obtain CEEMDAN, which effectively solves the transmission of white noise [23]. Li *et al.* suggested a hybrid power load forecasting model with a signal decomposition process combining CEEMDAN, Gated Recurrent Unit (GRU), and Improved Grey Wolf Optimizer (IGWO) and verified that the hybrid load forecasting model with good forecasting precision and strong nonlinear fitting ability [24].

In summary, a hybrid STLF model that incorporates the advantages of LSTM and GMDH neural networks to increase the forecasting precision is proposed in this paper. The main contributions of this paper are as follows: (1) In the process of power load data handling, to more effectively capture the

features information of the power load data, this paper uses ICEEMDAN to decompose the raw electric load data into multiple IMF components containing different characteristics. (2) In the paper, Information Entropy (IE) is introduced to avoid the prolonged prediction time and increased prediction error caused by too many IMF components. The specific step is to reconstruct the IMF with the same IE. (3) This paper proposed to use LSTM and GMDH to predict the restructured components separately. Then the results of two predictions are weighted and summed to obtain better prediction results. (4) In the paper, ICPA is proposed to find the optimal weights for the prediction results of two neural networks to increase the forecasting precision. (5) This paper tests six different prediction models on two daily average load datasets. The average absolute percentage error (MAPE), the root means square error (RMSE), the mean average error (MAE), and the inequality coefficient (TIC) are used to analyze the validity of the prediction mod.

## II. DATA PROCESSING METHODS

The two load data sets in the experiment are derived from the 5-year average daily loads of two regions in Chongqing, China, and each data set contains 2119 load data. The load data in the first four years of each dataset is used as the training set, and the electric load data in the fifth year is used as the test set.

### A. Normalization

The data after the normalization process as the input value of the neural network will improve the forecasting precision and stability of the neural network [25,26]. Therefore, it is essential to normalize the two sets of raw data before making predictions. Normalization is the process of scaling a set to [0, 1] according to certain rules. The processing is shown in the equation(1).

$$P' = \frac{P - P_{min}}{P_{max} - P_{min}} \quad (1)$$

Where,  $P$  denotes the load value before normalization,  $P_{min}$  is the smallest value in the original load data, as well as  $P_{max}$  is the biggest value in the original load data.

### B. Improved Complete Ensemble Empirical Mode Decomposition Adaptive Noise (ICEEMDAN) Algorithm

Because they are influenced by many factors, electric load data have a high degree of complexity. If the power load data is forecasted straight, the forecast accuracy will result in low prediction accuracy due to its high complexity. This study decomposes the raw load data into multiple IMF components and then performs separate forecasting to solve this problem.

EMD, which decomposes the original data into data series with different characters, is the most commonly used method in data processing [27]. EEMD is an EMD-based resolution method that solves the modal mixing problem in the EMD decomposition process [28]. However, the results obtained in EEMD decomposition will produce reconstruction errors. In order to eliminate errors, CEEMDAN which takes the total mean value of IMF components of each order of EEMD is commonly used in related fields [29]. The IMF components decomposed by CEMMDAN still have defects and contains

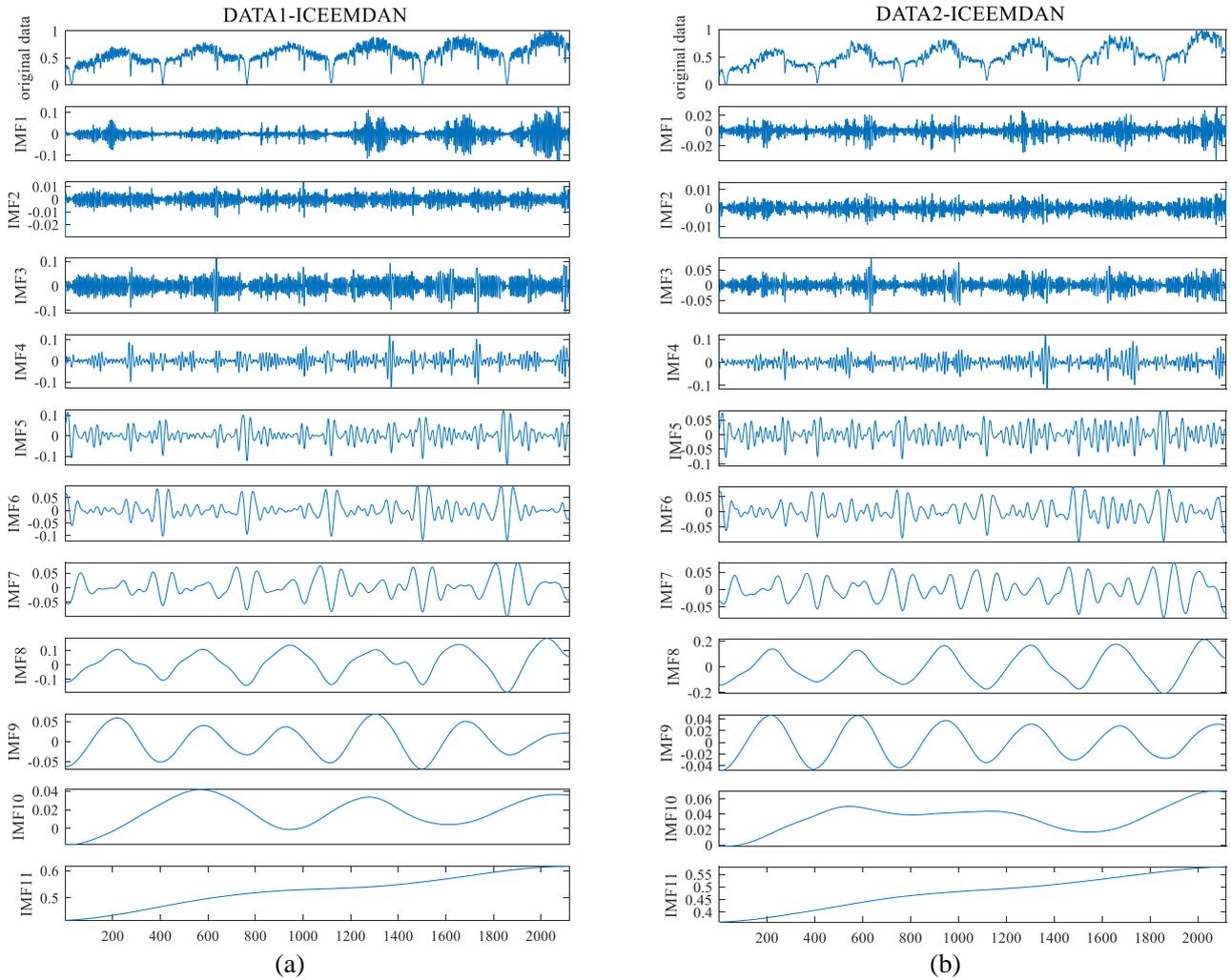


Fig.1.ICEEMDAN decomposition of dataset 1(a) and dataset 2(b)

residual noise [30]. For this problem, Colominas *at el.* proposed a new method called ICEEMDAN, which was obtained by improving on the basis of CEEMDAN. The difference between the new method and CEEMDAN is that CEEMDAN directly appends Gaussian white noise in the decomposition process, while ICEEMDAN appends the  $k$ th IMF component of white noise after EMD decomposition [31]. The original load data is defined as  $P$ .  $E_k(\cdot)$  denotes the  $k$ th order IMF component generated by the decomposition of the original data by EMD.  $N(\cdot)$  indicates the local average of the solved signal, and  $W_i$  denotes the white noise sequence with a mean of 0.  $\beta$  is the standard deviation of white noise. The decomposition steps are as follows.

Step1: Add the white noise sequence  $W_i$  to the original load time series to construct the new series.

$$P^{(i)} = P + \beta_0 E(W^{(i)}) \quad (2)$$

Step2: The EMD decomposition method is used to process the load time series after adding noise to obtain the first set of residuals ( $R_1$ ).

$$R_1 = (N(P^{(i)})) \quad (3)$$

Step3: The first IMF component is obtained by subtracting the residuals from the original data series.

$$d_1 = P - R_1 \quad (4)$$

Step4: The IMF of white noise is selected according to the order of the IMF components of the load data to be added to the residue obtained in the previous step. The EMD method is applied with the decomposition of the second set of residuals  $R_1 + \beta E(W^{(i)})$ .  $d_2$  denotes the second IMF component.

$$R_2 = (N(R_1 + \beta E(W^{(i)}))) \quad (5)$$

$$d_2 = R_1 - R_2 \quad (6)$$

Step5: Repeat the above steps until the remaining residuals are no longer decomposed. All components and residuals are exported. The original data is decomposed as the equation(7).

$$P(t) = \sum_{i=1}^K IMF_i(t) + R(t) \quad (7)$$

The IMF components of the two load data sets are shown in Fig.1.

### C. Information Entropy (IE)

Due to the high complexity of the raw load data, too many IMF components containing different feature information are

TABLE I.  
INFORMATION ENTROPY OF THE IMFS OF DATASET 1 AND DATASET 2

IMF	1	2	3	4	5	6	7	8	9	10	11
DATA 1	2.9051	3.6603	3.6486	3.5408	3.3697	3.1306	3.2407	3.5819	3.5598	3.2301	3.6679
DATA 2	3.4354	3.5883	3.5601	3.3854	3.5611	3.3883	3.4242	3.5918	3.6111	3.1035	3.6121

obtained. Since the forecasting model in this study forecasts each IMF component separately, the more IMF component there is, the longer the forecasting time will be. In addition, too many IMF components can also lead to an increase in the error of prediction results. To solve the above problem, IE is introduced, which measures the information and uncertainty of the data. The lower the IE value of a set of data means that the set of data is more orderly, on the contrary, the higher the IE value of a set of data means that the set of data is more disorderly [32]. After decomposing the raw data into multiple IMF components using ICEEMDAN during data processing, the information entropy of each IMF component is calculated according to equation (8). Then, the IMF component with a similar IE was restructured. The calculation formula of IE is as equation (8).

$$IE(X) = - \sum_{k=1}^N P_k \log_2(P_k) \quad (8)$$

IE denotes the information entropy value. X is the set of load data. N is the number of data in the set of load data. P represents the probability. P<sub>k</sub> is the percentage of the kth load data in the data set. The information entropy of each IMF component for the two load datasets is shown in TABLE I.

Based on the data in TABLE I, after reconstructing the 11 IMF quantities in Data 1, four IMF components are obtained. The first component consists of IMF5, IMF6, IMF7, IMF10, the second component consists of IMF4, IMF8, IMF9, the third component consists of IMF2, IMF3, IMF11, and the last component is IMF1. In addition, after reconstructing the 11 IMF quantities in Data 2, four IMF components are obtained. The first component consists of IMF1, IMF4, IMF6, IMF7, the second component consists of IMF3, IMF5, IMF2, IMF8, the third component consists of IMF9, IMF11, and the fourth component is IMF10.

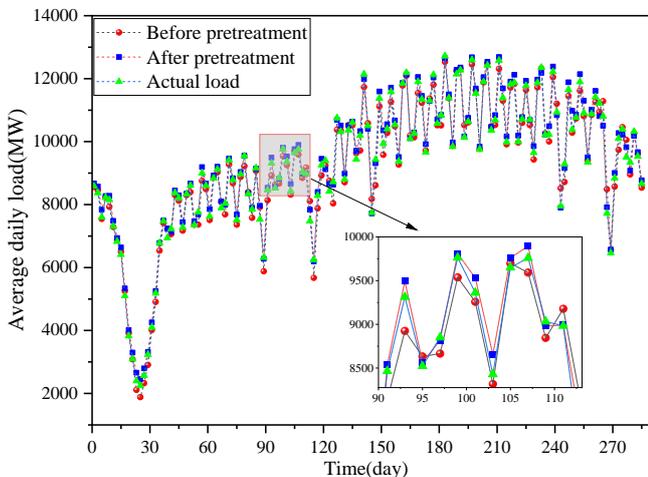


Fig.2.Forecast results before and after data processing for dataset 1

#### D. Experimental verification

To prove the applicability of the load data preprocessing

method, comparative experiments on the two load datasets are designed separately, and the experimental model LSTM was trained and predicted before and after preprocessing of the two datasets. The experimental results are shown in Fig.2 and Fig.3, and the RMSE, MAE, and MAPE of the prediction results are shown in TABLE II, and TABLE III.

According to the experimental results, both electricity load data sets have better prediction results after being processed by the processing method (ICEEMDAN) proposed in the paper. These experimental results also prove that the data processing method proposed in the paper is very effective. The following conclusions can be drawn from the three indicators (MAPE, RMSE, and MAE). Compared with the prediction results without data preprocessing using the data preprocessing method proposed in this paper, the RMSE, MAE, and MAPE of the load data set 1 preprocessed by the preprocessing method in this paper are reduced by 56.9847, 40.0003, and 0.5382%, respectively. And compared with the results of the prediction without preprocessing using the preprocessing method proposed in this paper, the RMSE, MAE, and MAPE of the load data set 2 preprocessed by the preprocessing method in this paper are reduced by 75.9663, 67.4526, and 0.8575%, respectively. In conclusion, after the preprocessing method proposed in this paper, the prediction precision and efficiency have been significantly improved.

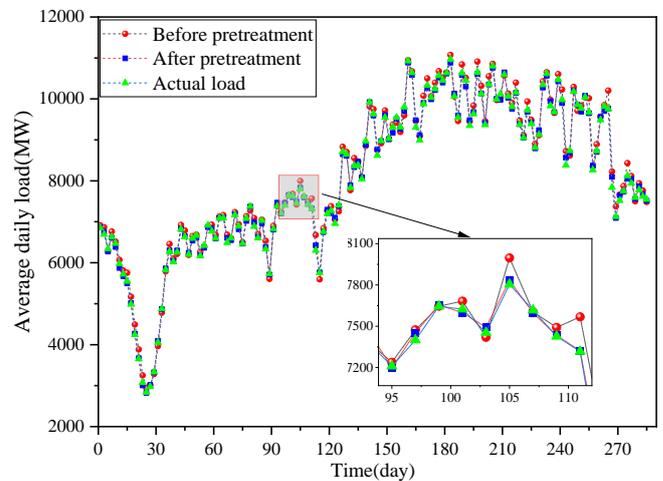


Fig.3.Prediction results before and after data processing for dataset 2

### III. PREDICTIVE MODELS

#### A. The Group Method of Data Handling (GMDH)

To predict the fish population in marine rivers, Ivakhenko proposed the GMDH prediction model [33]. GMDH is a kind of neural network that uses polynomial iteration to obtain a nonlinear relationship between input and output and therefore is also called a polynomial network. The network structure of the GMDH neural network is shown in Fig.4, and its structure is layered. The neurons in the input layer of the network are only responsible for passing the input signal to the neurons in the middle layer; each neuron in the hidden layer and each neuron in the output layer are only connected to two neurons

in the previous layer. So, the neurons in each layer of the GDMD neural network are independent [34]. The input data are defined as  $P = (P_1, P_2, P_3, \dots, P_n)$ . The predicted output is  $(\bar{O}_i)$ , and the actual output is  $(O_i)$ . The relationship of each neuron of the GDMH is shown in the equation (9).

$$O_{k,l} = a_{k,l}(O_{k-1,i})^2 + b_{k,l}O_{k-1,i}O_{k-1,j} + c_{k,l}(O_{k-1,j})^2 + d_{k,l}O_{k-1,i} + e_{k,l}O_{k-1,j} + f_{k,l}e_{k,l}O_{k-1,j} + f_{k,l} \quad (9)$$

The equation (9) is known as the Ivakhnenko polynomial. Where,  $O_{k,l}$  represents the output of the  $l$ th neurons of the  $k$ th layer of the network,  $O_{k-1,i}$  represents the output of the  $i$ th neurons of the  $k$ th layer of the network,  $O_{k-1,j}$  represents the output of the  $j$ th neurons of the  $k$ th layer of the network. And  $O_{0,i}$  represents input  $(P_i)$ .  $a_{k,l}, b_{k,l}, c_{k,l}, d_{k,l}, e_{k,l}$  are polynomial coefficients.

In order to make the error between the predicted and actual outputs as small as possible, the GMDH neural network uses regression analysis to obtain the coefficients of the quadratic polynomial in the equation (9). The least-squares method is used to optimize the coefficients of the quadratic equation.

$$R = \frac{\sum_{i=1}^N (\bar{O}_i - O_i)^2}{N} \quad (10)$$

$R$ , the target amount to be minimized, denotes the error.

The full binomial of the GMDH can be expressed as shown in the equation (11) and (12).

$$O = P \beta \quad (11)$$

$$P = \begin{bmatrix} 1 & P_{1i} & P_{1j} & P_{1i}P_{1j} & P_{1i}^2 & P_{1j}^2 \\ 1 & P_{2i} & P_{2j} & P_{2i}P_{2j} & P_{2i}^2 & P_{2j}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & P_{Ni} & P_{Nj} & P_{Ni}P_{Nj} & P_{Ni}^2 & P_{Nj}^2 \end{bmatrix} \quad (12)$$

The polynomial coefficients ( $\beta$ ) of binomial can be found using the least squares formula.

$$\beta\pi = (P^T P)^{-1} P^T O \quad (13)$$

TABLE II.  
COMPARISON OF PREDICTIVE INDICATORS BEFORE AND AFTER PRE-PROCESSING FOR DATASET 1

	Before pre-processing	After pre-processing
RMSE	196.4812	139.4965
MAE	153.7032	113.7029
MAPE (%)	2.0943	1.5561

TABLE III.  
COMPARISON OF PREDICTIVE INDICATORS BEFORE AND AFTER PRE-PROCESSING FOR DATASET 2

	Before pre-processing	After pre-processing
RMSE	146.2597	70.2934
MAE	117.9565	50.5039
MAPE (%)	1.5485	0.6910

In GMDH neural networks, any two neurons of one neural layer are used as input to produce neurons of the new neural layer. After the accumulation of multilayer neural networks, many neurons will be generated, and too many neurons will lead to a complex structure of DMGH neural networks and a long convergence time. In order to solve the above problems without degrading the low prediction accuracy of the GMDH neural network, the elite pool rule is introduced in this study. If the number of neurons  $m$  is less than or equal to the size of the elite pool  $M$ , no processing is done to the neurons of this layer of the network. On the contrary, the mean squared error  $D_i$  is calculated for the neurons of the next layer, and  $D_i$  is calculated as shown in the equation (14).

$$D_i = \sqrt{\frac{1}{2} \sum_{j=1}^2 (\bar{O}_i - O_i)^2}, i = 1, 2, \dots, C_N^2 \quad (14)$$

Then, each neuron is ranked in ascending order according to the magnitude of the mean squared error  $D_i$  of their outputs. According to the set elite pool size  $M$ , the first  $M$  neurons among the generated neurons by this layer of the network are selected as the input neurons of the next layer, and the other neurons in the layer are discarded. The simple structure of the GMDH after using the elite pool rule is shown in Fig.4. The square neurons in the figure represent the discarded neurons.

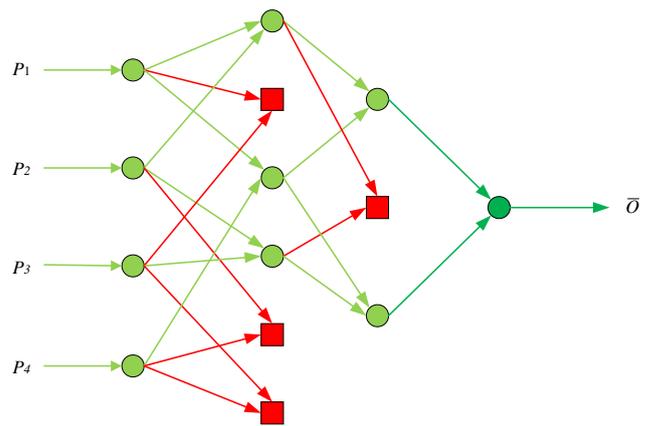


Fig.4.The simple structure of the GMDH

### B. The Long Short-Term Memory (LSTM)

LSTM neural network is a particular RNN, which is widely used in image processing, text recognition, and power load prediction due to its good ability to store long sequence data and learn long-term dependencies [35]. Compared with the general RNN, each LSTM unit has a forgetting gate, input gate, output gate, and storage unit, and the structure of LSTM is shown in Fig.5. The input gate determines the information that is stored in the unit with the newly entered information. The forget gate is used to forget some information that has an impact on the load prediction result from the output of the previous unit. The output gate is used to output the processed information to the next LSTM cell. The memory cell is used to store the useful information transmitted from the previous LSTM unit to this unit [36]. LSTM constantly optimizes the weights in the unit during training, and the weights are shared among all units. Therefore, with the model parameters fixed, the integration scale changes continuously at different times, this approach solves the problem of gradient disappearance

and reduces the frequency of gradient explosion [37]. The units of LSTM are calculated as follows.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (15)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (16)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (17)$$

$$s_t = f_t s_{t-1} + i_t \tanh(W_{xs}x_t + W_{hs}h_{t-1} + b_s) \quad (18)$$

$$h_t = o_t \tanh(s_t) \quad (19)$$

Where,  $i_t$  denotes the input gate;  $x_t$  is the input at the current moment;  $\sigma$  is the tanh activation function that converts values to  $[-1, 1]$ ;  $W_{xs}$ ,  $W_{xi}$ ,  $W_{xf}$ , and  $W_{xo}$  represent the weight matrix of the input  $X_t$ .  $H_t$  refers to the output at the current moment;  $H_{t-1}$  is the output at the previous moment;  $H_{xs}$ ,  $H_{xi}$ ,  $H_{xf}$ , and  $H_{xo}$  represent the weight matrix of  $H_{t-1}$ ;  $b_i$ ,  $b_f$ ,  $b_o$  and  $b_s$  denote the deviation vectors;  $f_t$  is the forgetting gate.  $s_t$  denotes the input cell vector at the current moment.

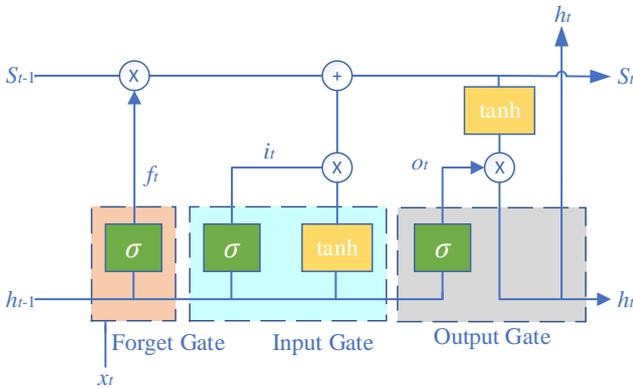


Fig.5.The structure of the LSTM

### C. Carnivorous Plant Algorithm (CPA)

Based on the ability of carnivorous plants to adapt to harsh environments, K. Meng proposed a new global optimization algorithm called the carnivorous plant algorithm (CPA) [38]. In nature, most plants are producers and are food for animals. But there are exceptions, such as carnivorous plants, which can attract, capture and eat small animals such as butterflies and mice by secreting special enzymes to obtain nutrients that growth and reproduction. CPA simulates how carnivorous plants survive in very harsh environments, such as preying on prey and growing and reproducing. CPA has been proven to be powerful in solving multi-latitude latitude variables and multiple locally optimal solutions. CPA mainly consists of four processes: carnivorous plants attract, capture, and digest prey, and carnivorous plant growth and reproduction. The process of the algorithm is as follows:

Step1: CPA is a new type of population-based intelligence optimization algorithm, so it needs to initialize the population of individuals. First, the population individuals are randomly generated and initialized according to the set total number of population individuals  $N$ . The population includes prey and carnivorous plants. The number of carnivorous plants is  $N_c$ , and the number of prey is  $N_p$ .

$$\frac{N_p}{N_c} = K \quad (20)$$

$$Pop = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & \dots & P_{1,m} \\ P_{2,1} & P_{2,2} & \dots & \dots & P_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{N,1} & P_{N,2} & \dots & \dots & P_{N,m} \end{bmatrix} \quad (21)$$

Where,  $m$  is the dimension of each individual, and  $P$  is the value of each dimension.

The objective function is found according to the problem faced. Based on the values of each individual and the function, the fitness value of each individual is calculated.

$$Fit = \begin{bmatrix} f(P_{1,1}) & P_{1,2} & \dots & \dots & P_{1,m} \\ f(P_{2,1}) & P_{2,2} & \dots & \dots & P_{2,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f(P_{N,1}) & P_{N,2} & \dots & \dots & P_{N,m} \end{bmatrix} \quad (22)$$

$Fit$  represents the fitness matrix of the population, and  $f$  is the objective function. In the follow-up process, the degree of adaptation is used as an important criterion for selecting the optimal solution.

Step2: The individuals in the population are arranged in ascending order according to the fitness values. The first  $N_c$  individuals in line were defined as carnivorous plants, and the remaining  $N_p$  individuals were considered prey. The aligned fitness matrix is shown below:

$$SortedFit = \begin{bmatrix} f_c(1) \\ f_c(2) \\ \vdots \\ f_c(N_c) \\ f_p(N_c + 1) \\ f_p(N_c + 2) \\ \vdots \\ f_p(N_c + N_p) \end{bmatrix} \quad (23)$$

Where,  $f_p$  and  $f_c$  represent the adaptation values of prey and carnivorous plants, respectively.  $SortedFit$  is the ranking of fitness.

Step 3: Carnivorous plants and prey in the population are grouped. Carnivorous plants and prey are sorted in ascending order according to fitness values. The carnivorous plants and prey in the same ranking were assigned together, and then the remaining prey was sequentially assigned to the carnivorous plants in the order of ranking.

Carnivorous plants need to hunt for prey in order to survive, so they emit sweet smells to attract prey close to them. When a predator captures prey, the prey has two possibilities, to be captured or to escape, so the attraction rate is introduced. If the random number is less than the attraction rate, the prey is captured by the carnivorous plant. On the contrary, the prey escapes. The equation for the growth of carnivorous plants after prey capture is as follows:

$$NPC_{i,j} = Gr \times PC_{i,j} + (1 - Gr) \times PP_{v,j} \quad (24)$$

$$G = Gr \times rand_{i,j} \quad (25)$$

Where,  $PC_{i,j}$  denotes the value of the  $i$ th carnivorous plant at the  $j$ th position.  $PP_{v,j}$  denotes the randomly selected prey.  $NPC$  is a newly grown carnivorous plant.  $Gr$ , a predefined value, is a growth rate. If the prey escapes predation by the carnivorous plant, the prey will continue to grow.  $NPP$  is the new growth of the prey whose growth equation is as follows:

$$NPP_{i,j} = G \times PP_{u,j} + (1 - G) \times PP_{v,j}, u \neq v \quad (26)$$

$$G = \begin{cases} G \times rand_{i,j} & f(PP_v) > f(PP_u) \\ 1 - G \times rand_{i,j} & f(PP_v) < f(PP_u) \end{cases} \quad (27)$$

Step 5: Carnivorous plants reproduce after preying on prey. However, not all carnivorous plants will reproduce, only the first-ranked carnivorous plants do. The reproduction process is as follows:

$$NPC_{i,j} = PC_{1,j} + Rr \times rand_{i,j} \times mate_{i,j} \quad (28)$$

$$mate_{i,j} = \begin{cases} PC_{v,j} - PC_{i,j} & f(PC_i) > f(PC_v) \\ PC_{i,j} - PC_{v,j} & f(PC_i) < f(PC_v) \end{cases} \quad (29)$$

Where,  $PC_{1,j}$  represents the value of the  $j$ th position of the carnivorous plant ranked first.  $PC_{v,j}$  denotes the value of the  $j$ th position of the randomly selected carnivorous plant.  $Rr$  is the set reproduction rate.

Step6: Newly produced carnivorous plants and prey are added to the population as a new population. The number of individuals in the new population is  $[N + N_G + N_R]$ , where  $N_G$  is the number of newly grown individuals;  $N_R$  is the number of new individuals for the reproduction parameter. The fitness value of each individual in the new population is calculated and then ranked again according to the fitness value. The first  $N$  individuals are selected as the next round of populations.

Step7: Step2~Step6 are repeated until the error is less than the set value or the number of iterations is greater than the set value. Finally, the optimal solution is output.

#### D. Improved Carnivorous Plant Algorithm (ICPA)

In the population of CPA, each individual represents the solution to the problem, and each fitness value represents the error value. If the initialization population is not reasonable, the solution obtained is a locally optimal solution. Meanwhile, during the growth and reproduction of carnivorous plants, the diversity of the population may be reduced if prey individuals are not reasonably selected. The above situation can lead difficult to obtain a globally optimal solution.

(1) In order to change the pseudo-randomness of the initial population and increase the diversity of the population, the traditional method of randomly generating individuals with the rand function is abandoned in the population initialization stage and the Sobol sequence is introduced to initialize the population. The population generated by the Sobol sequence has better diversity and is more evenly distributed in space. Each dimension of the population of individuals generated by the Sobol sequence is composed of radical inversions with base two, except that each dimension has its different matrix

of radical inversions.

(2) In the process of continuous update of the population of CPA, the parameter random number will be used to judge the direction of the update, so to prevent the final solution from being a locally optimal solution introduce Gaussian Mutation. Gaussian Mutation is defined as shown in the equation(30).

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (30)$$

Where,  $\mu$  denotes the expectation of the gaussian mutation, and  $\sigma^2$  denotes the variance of the Gaussian Mutation. After Gaussian Mutation is introduced into CPA, the equation (25), (27) and (28) will be changed. The changed formula is shown below.

$$G = Gr \times Guass(0,1) \quad (31)$$

$$G = \begin{cases} G \times Guass(0,1) & f(PP_v) > f(PP_u) \\ 1 - G \times Guass(0,1) & f(PP_v) < f(PP_u) \end{cases} \quad (32)$$

$$NPC_{i,j} = PC_{1,j} + Rr \times Guass(0,1) \times mate_{i,j} \quad (33)$$

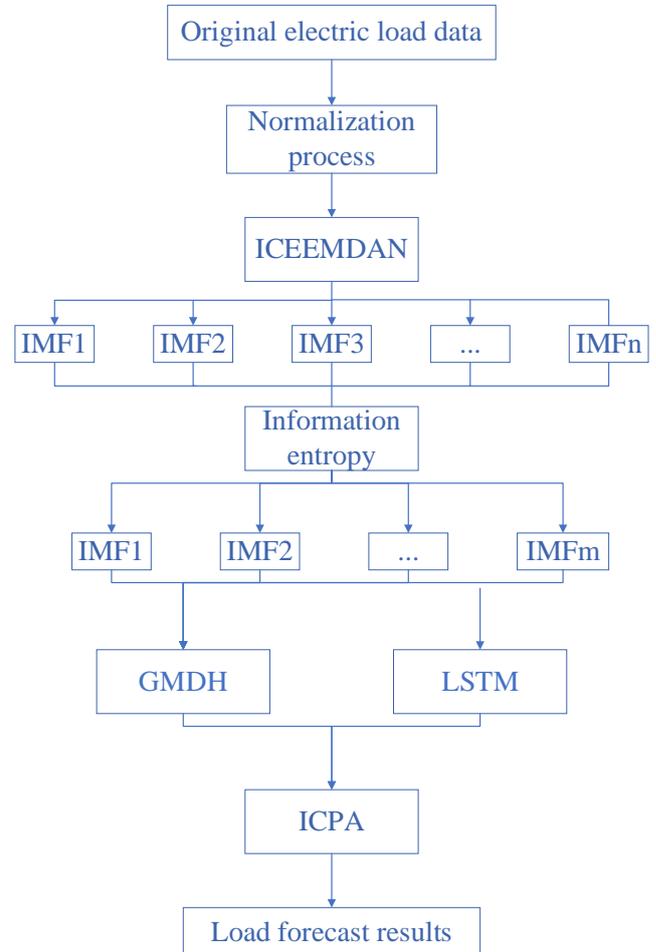


Fig.6.The workflow of the hybrid prediction model

To verify that the improved CPA has better results, a set of experiments comparing the search effect before and after the improvement in the test function. The test functions are shown in TABLE IV. The test functions in the experiment have different meanings.  $F1$  is a separable function of single-peaked variables, and  $F2$  is an inseparable function of

TABLE IV.  
TEST FUNCTIONS

Function	Formula	Boundary	Number of parameters(D)
F1	$F_1(X) = \sum_{i=1}^D X_i^2$	$[-100,100]^d$	20
F2	$F_2(x) = \left[ \sum_{i=1}^{D-1} 100(x_{i+1} - x_i)^2 + (x_i - 1)^2 \right]$	$[-50,50]^d$	20
F3	$F_3(X) = \sum_{i=1}^D [X_i^2 - 10\cos(2\pi X_i) + 10]$	$[-30,30]^d$	20
F4	$F_4(X) = \sum_{i=1}^D  X_i \sin(X_i) + 0.1X_i $	$[-24,24]^d$	20

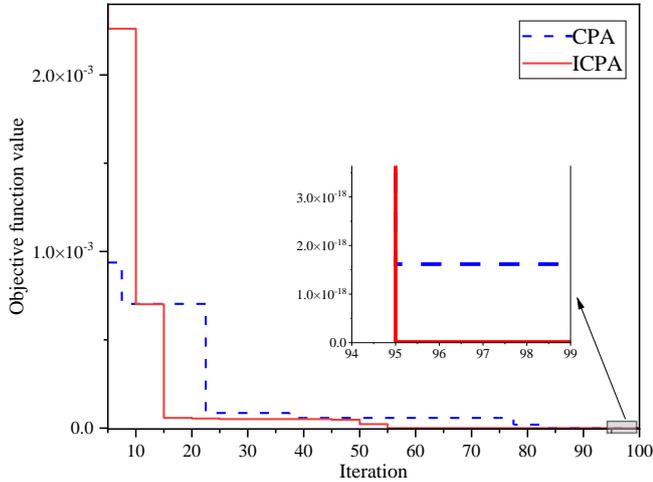


Fig.7.Convergence curve of F1

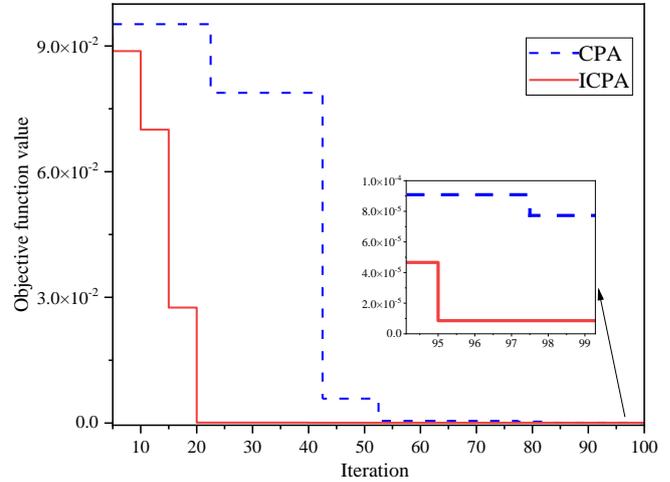


Fig.8.Convergence curve of F2

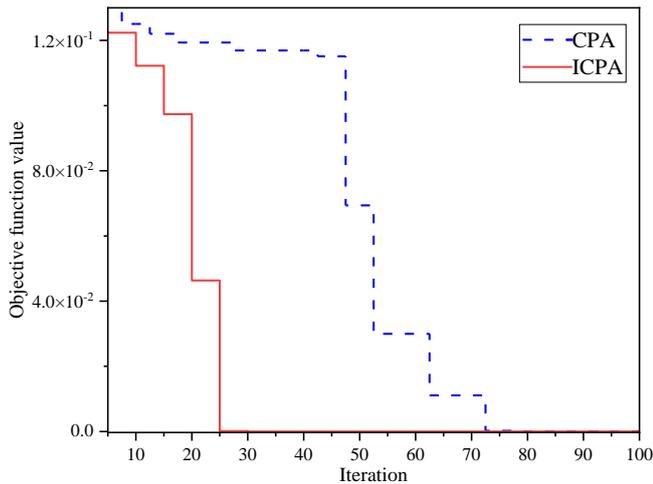


Fig.9.Convergence curve of F3

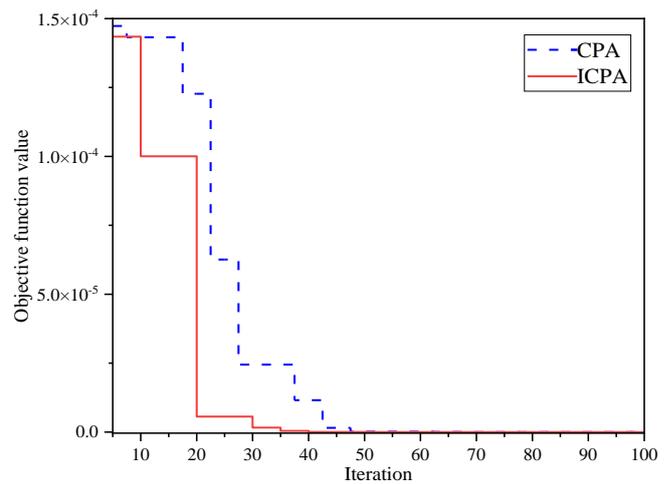


Fig.10.Convergence curve of F4

TABLE V.  
PERFORMANCE INDEX OF THE ALGORITHM

Function	Algorithm	MEAN	BEST	WORST
F1	CPA	1.8911E-4	2.9369E-19	9.3745E-4
	ICPA	<b>4.3442E-5</b>	<b>5.3028E-21</b>	<b>3.1801E-4</b>
F2	CPA	3.5618E-2	9.7638E-5	9.52E-2
	ICPA	<b>1.1795E-2</b>	<b>8.5650E-5</b>	<b>9.25E-2</b>
F3	CPA	6.1756E-2	3.4872E-7	0.1311
	ICPA	<b>2.2015E-2</b>	<b>1.2434E-14</b>	<b>0.1226</b>
F4	CPA	3.4154E-5	3.0839E-13	1.4732E-4
	ICPA	<b>2.5093E-5</b>	<b>9.2011E-16</b>	<b>1.4343E-4</b>

single-peaked variables. The main effect of these two test functions is to verify the optimization accuracy of the algorithm. F3 is a separable function of multimodal variables, and F4 is an inseparable function of multimodal variables. The main effect of these two test functions is to prove the global search ability of the optimization algorithm. Through

experiments, the iterative diagrams graph of the algorithm before and after optimization on the four test functions are obtained, and shown in Fig.7-Fig.10 respectively.

The optimal solution found by the algorithm will be different each time. To exclude the impact of the above issues on the analysis of the algorithm, The CPA before and after the

improvement is run 100 times in each of the test functions, and then the average, worst, and best value of the optimal solution of each test function is analyzed. The indicators are shown in TABLE V. From Fig.7-Fig.10, it can be clearly observed that the improved CPA has significant improvement in terms of convergence accuracy and convergence speed. In addition, the performance indices in TABLE V show that the improved CPA has a significant improvement in optimization accuracy and stability compared to the original CPA.

#### IV. FRAMEWORK OF HYBRID MODEL

The data applied in the experiments are from two regions in Chongqing, China, Firstly, the load data is normalized to ensure that the power load data are in a uniform magnitude. ICEEMDAN is used to decompose the normalized data to obtain IMFs containing different characteristic information. To reduce the errors arising from predicting multiple IMFs, the IMFs reconstructed by information entropy are predicted respectively using GMDH and LSTM, and the weights obtained from the ICAP in the training set are weighted to obtain the prediction final results. The load prediction model combines the prediction advantages of both neural networks and avoids the possibility of large errors in a single model with different feature data. The two prediction models give full advantage to their respective strengths to make the whole prediction model very reliable and make the prediction result more accurate. The workflow of the proposed hybrid load forecasting model is shown in Fig.6.

#### V. CASE AND RESULTS

##### A. Description of experimental data and experimental tools

In this paper, two sets of raw load data are applied to the load prediction model for experiments. In the experimental process, considering the existence of weekly periodicity of the historical load, the previous seven average daily load data are used as the input to predict the average daily load value on the eighth day. In this paper, the proposed load prediction model is constructed using MATLAB software in 64-bit version 2019a under Windows 10 system.

##### B. Prediction accuracy assessment metrics

Analysis of the load prediction model results is an integral part of the prediction research. And the indicators used for the analysis of forecasting results are different in different fields. However, each prediction model can be evaluated using three performance indicators: MAPE, RMSE, and MAE [39-42]. The above three metrics and TIC are selected the indicators for the analysis of load prediction results in this paper. The formulas for calculating performance indicators are equations (34) - (37):

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - \bar{O}_i| \quad (34)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (35)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\bar{O}_i - O_i}{O_i} \right| \times 100\% \quad (36)$$

$$TIC = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - \bar{O}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (O_i)^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{O}_i)^2}} \quad (37)$$

Where,  $O_i$  represents the true value;  $\bar{O}_i$  denotes the output value of the prediction model.  $N$  represents the number of output values.  $MAE$  represents the absolute error between the true and predicted load values.  $RMSE$  represents the standard deviation between the true and predicted values.  $MAPE$  can indicate the overall performance of the load prediction model, and the closer its value is to 0, the more perfect the prediction model is. The  $TIC$  value varies between 0 and 1 ( $0 < TIC < 1$ ). If the  $TIC$  value is closer to 0, the prediction model is more accurate.

##### C. Data processing

Firstly, the two sets of raw load data are normalized. Then the normalized load data are decomposed using ICEEMDAN to obtain IMF components. In the decomposition, 500 sets of Gaussian white noise with a standard deviation of 0.002 are added to the normalized load data. After the components are obtained, the components are reorganized according to the information entropy of each IMF component in TABLE I. The reorganized sequence is shown in Fig.11 and Fig.12.

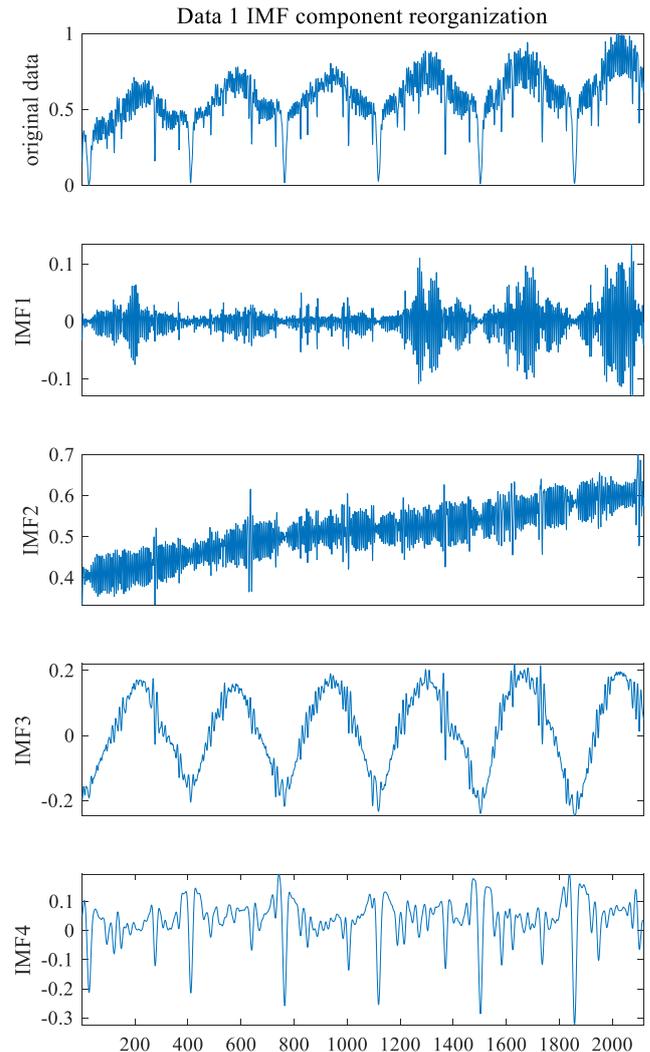


Fig.11.Sequence diagram of IMFs recombination for dataset 1.

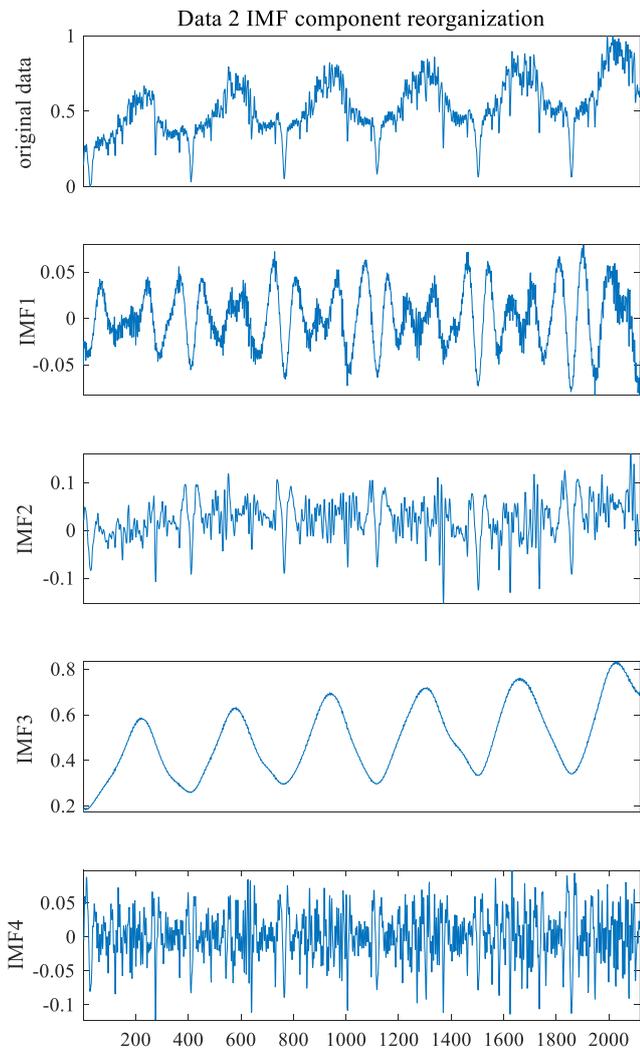


Fig.12. Restructured IMFs for dataset 2

**D. Prediction results and comparative analysis**

The four IMF components obtained after reorganization in each dataset are trained and predicted by LSTM and GMDH neural networks respectively, and the optimal weights of the two sets of load prediction results are found using ICAP. The prediction results are shown in Fig.13-Fig.16. The optimal weight is shown in TABLE VI. The MAE, RMSE and MAPE are calculated for each component of the prediction results as shown in TABLE VII and TABLE VIII. According to the data from TABLE VII and TABLE VIII, It is concluded that in Data 1, the LSTM prediction model has better predictions on the IMF1 and IMF2 components with MAPEs of 3.2114% and 0.6776%, respectively. Conversely, the IFM3 and IMF4 components are better predicted in the GMDH with MAPE of 0.9882% and 0.0929%, respectively. In Data 2, the LSTM predicted effect better for IMF1 and IMF4 components with MAPE of 1.6715% and 2.3070%, respectively. However, in GMDH, IMF2 and IMF3 components have better prediction effects, with MAPE of 0.5596% and 0.1891%, respectively.

From Fig.13-Fig.16, it can be clearly observed that in both dataset 1 and dataset 2, the main error of the prediction model derived from the prediction results of high complexity IMF components. To obtain a more accurate result, the results of the two prediction models are weighted according to the weights in TABLE VI. The prediction results for the two data sets are shown in Fig.17 and Fig.18.

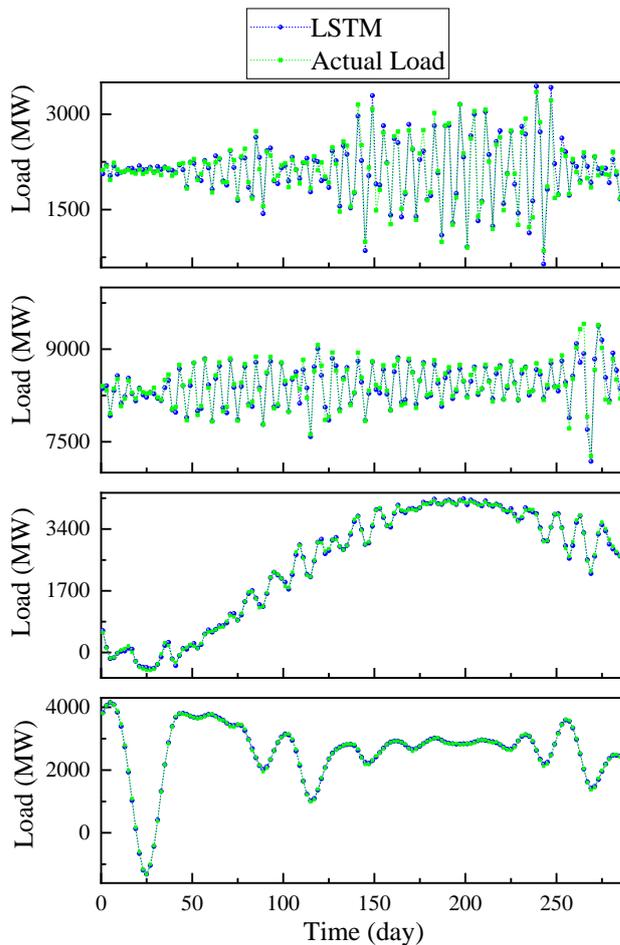


Fig.13.LSTM prediction results for dataset 1

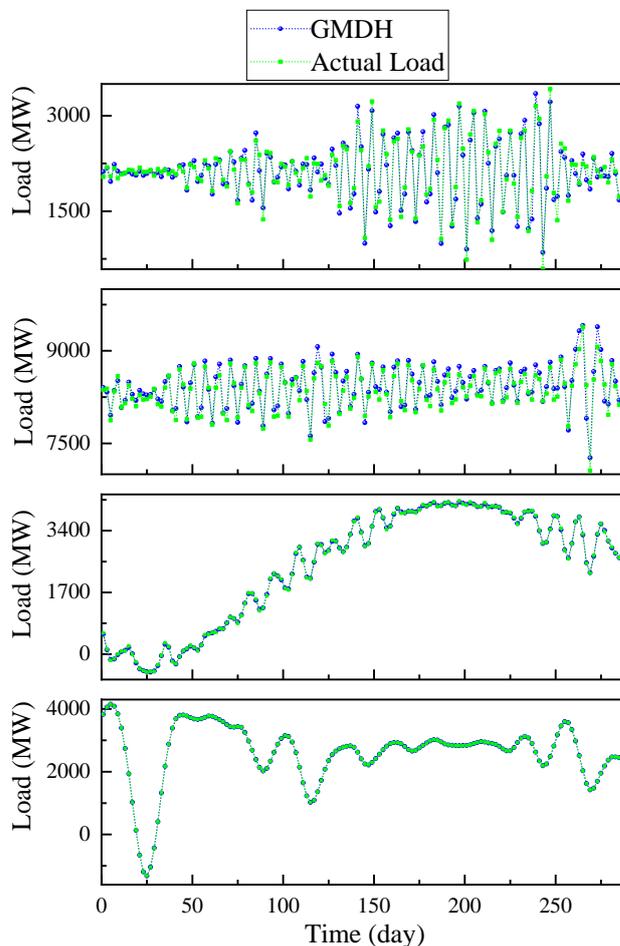


Fig.14.GMDH prediction results for dataset 1

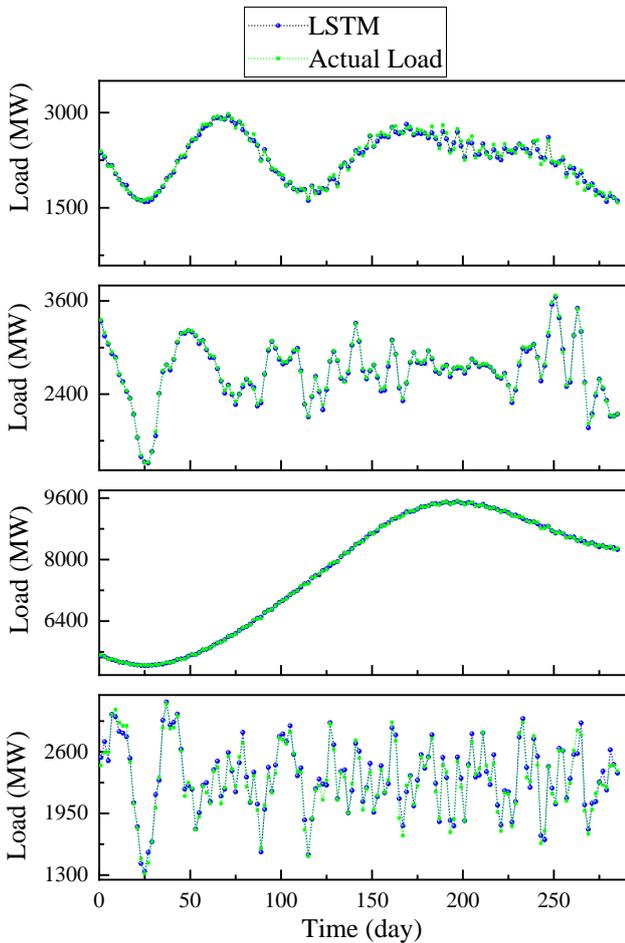


Fig.15.LSTM prediction results for dataset 2

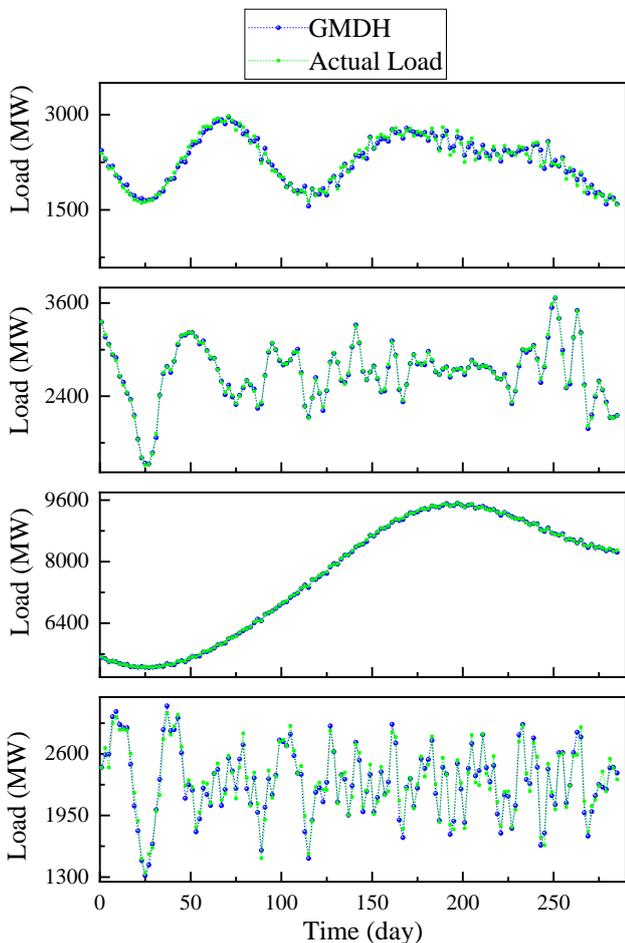


Fig.16.GMDH prediction results for dataset 2

In order to better prove the superiority and stability of the proposed hybrid short-term load prediction model, a set of comparison experiments is designed to compare the hybrid prediction model proposed in this paper with other models: Autoregressive Integrated Moving Average model (ARIMA); Back Propagation Neural Network (BPNN); ICAP optimized Back Propagation Neural Network (ICAP-BPNN); Elman neural networks (Elman); ICEEMDAN-IE-LSTM prediction model; ICEEMDAN-IE-GMDH prediction model. In taking into account that the weights obtained by the neural network prediction model during the training process are somewhat random, which can lead to differences in the results of each prediction, the load data were predicted 10 times with each prediction model and the average of the 10 results was taken as the final prediction result of that model.

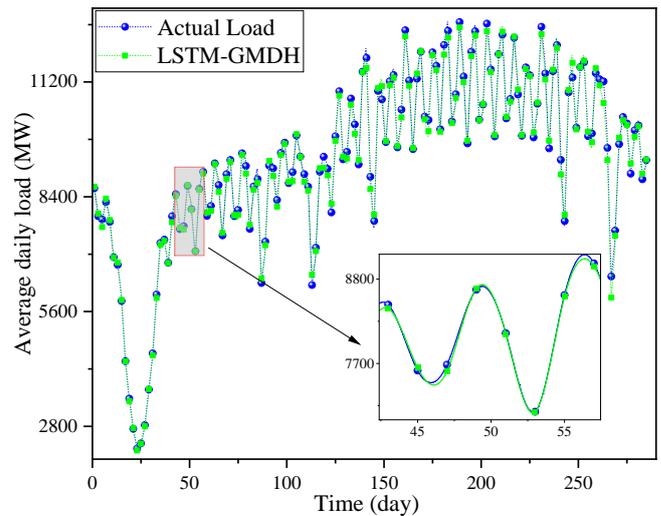


Fig.17.Predicted results for dataset 1

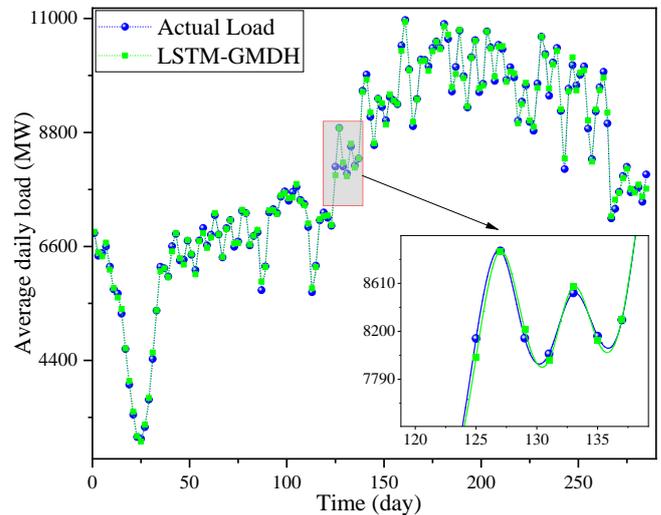


Fig.18.Predicted results for dataset 2

The values of the three prediction performance indicators (MAE, RMSE, MAPE) for the seven models in the two load data sets are shown in TABLE IX and TABLE X, separately, and The curve of the load prediction result obtained for the two load data sets in the six comparative forecasting models and the proposed load forecasting model are shown in Fig.19 and Fig.20.

From two tables, TABLE IX and TABLE X, it can be visualized that the proposed hybrid load forecasting model

TABLE VI.  
WEIGHTING OF THE RESULTS OF THE TWO MODELS

	LSTM				GMDH			
	IMF1	IMF2	IMF3	IMF4	IMF1	IMF2	IMF3	IMF4
DATA1	0.4022	0.6886	0.1250	0.0313	0.5978	0.2114	0.8750	0.9687
DATA2	0.8940	0.1875	0.3750	0.7923	0.1060	0.8125	0.6250	0.2077

TABLE VII.  
PERFORMANCE INDICATORS FOR THE COMPONENTS IN DATASET 1

	LSTM				GMDH			
	IMF1	IMF2	IMF3	IMF4	IMF1	IMF2	IMF3	IMF4
MAE	71.2300	51.0746	30.0591	19.0318	74.3556	69.2046	12.8873	1.6060
RMSE	97.3432	80.3292	37.4092	23.9529	100.1384	94.5165	15.8189	1.9960
MAPE (%)	3.2114	0.6776	2.5239	1.7035	3.4963	0.9311	0.9882	0.0929

TABLE VIII.  
PERFORMANCE INDICATORS FOR THE COMPONENTS IN DATASET 2

	LSTM				GMDH			
	IMF1	IMF2	IMF3	IMF4	IMF1	IMF2	IMF3	IMF4
MAE	36.9449	16.8356	15.8433	51.7063	53.7725	14.5963	14.2093	59.3307
RMSE	48.7821	20.7297	21.3168	64.7373	68.3134	18.0742	18.0173	73.1093
MAPE (%)	1.6715	0.6407	0.2071	2.3070	2.4258	0.5596	0.1891	2.6493

TABLE IX.  
PERFORMANCE INDICATORS OF DIFFERENT PREDICTION MODELS IN DATASET 1

	RMSE	MAE	MAPE
ARIMA	396.4881	324.5295	3.6431
BPNN	203.6553	153.5245	1.7335
ICPA-BPNN	148.4271	107.2128	1.2486
Elman	689.3004	531.0372	6.3195
ICEEMDAN-IE-LSTM	170.5295	138.9977	1.7208
ICEEMDAN-IE-GMDH	282.5200	220.6250	2.5221
Proposed Method	<b>134.0026</b>	<b>98.3204</b>	<b>1.0716</b>

TABLE X.  
PERFORMANCE INDICATORS OF DIFFERENT PREDICTION MODELS IN DATASET 2

	RMSE	MAE	MAPE
ARIMA	172.5263	137.5452	1.7699
BPNN	194.8758	141.9265	1.6554
ICPA-BPNN	142.1766	114.2722	1.4128
Elman	460.4889	374.9198	4.9040
ICEEMDAN-IE-LSTM	70.2934	53.5039	0.6910
ICEEMDAN-IE-GMDH	123.7725	92.8659	1.1955
Proposed Method	<b>68.6581</b>	<b>52.6895</b>	<b>0.6875</b>

outperforms the other kinds of comparison models in three performance indicators. The RMSE, MAE, and MAPE of the proposed load forecasting model are 134.0026, 98.3204, and 1.0716% in dataset 1, and 68.6581, 52.6895, and 0.6875% in dataset 2, respectively.

In both datasets, the prediction models that incorporate the ICEEMDAN method, such as the ICEEMDAN-IE-LSTM and the proposed forecasting model, have higher prediction accuracy than the prediction models that do not incorporate the ICEEMDAN method, such as Elman, ARIMA and BPNN. Particularly, compared with the Elman model, the RMSE, MAE, and MAPE of the proposed load prediction model improved by 80.56%, 81.49%, and 83.04%, separately, in dataset 1, and by 85.09%, 85.95%, and 85.98%, separately, in dataset 2. The above results indicate can indicate that the hybrid forecasting model combining the ICEEMDAN and the machine learning methods proposed in this paper has a remarkable advantage in power load forecasting. In addition, the result also highlights the efficacy of ICEEMDAN. And the proposed hybrid load forecasting model achieves a huge improvement in terms of forecasting accuracy compared with the single traditional prediction model.

BPNN has a relatively strong non-mapping ability and strong self-adaptive capability, and is a common prediction model in the field of prediction. However, BPNN has the disadvantages such as easy falling into local minima and slow convergence speed. To solve the above problems, academics often use intelligent optimization algorithms to optimize the internal structure of BP to reach better prediction results. In TABLE IX and TABLE X, the MAE to predict results of the BPNN and ICPA-BPNN prediction models are 153.5245 and 107.2128 in dataset 1 and 141.9265 and 114.2722 in dataset 2, respectively. The conclusion that the prediction precision of the BP optimized with ICPA has been significantly improved, and ICAP is very effective in the actual prediction process can be clearly summarized.

The proposed load hybrid forecasting model not only has data processing to enhance the forecasting precision but also combines the advantages of two single sets of models, LSTM and GMDH. Compared with the LSTM neural network, the MAPE, MAE, and MAPE of the proposed hybrid short-term load forecasting model in electricity load dataset 1 are greatly improved by 37.73% 29.26%, and 21.42%, respectively. And in electricity load dataset 2 the RMSE, MAE, and MAPE

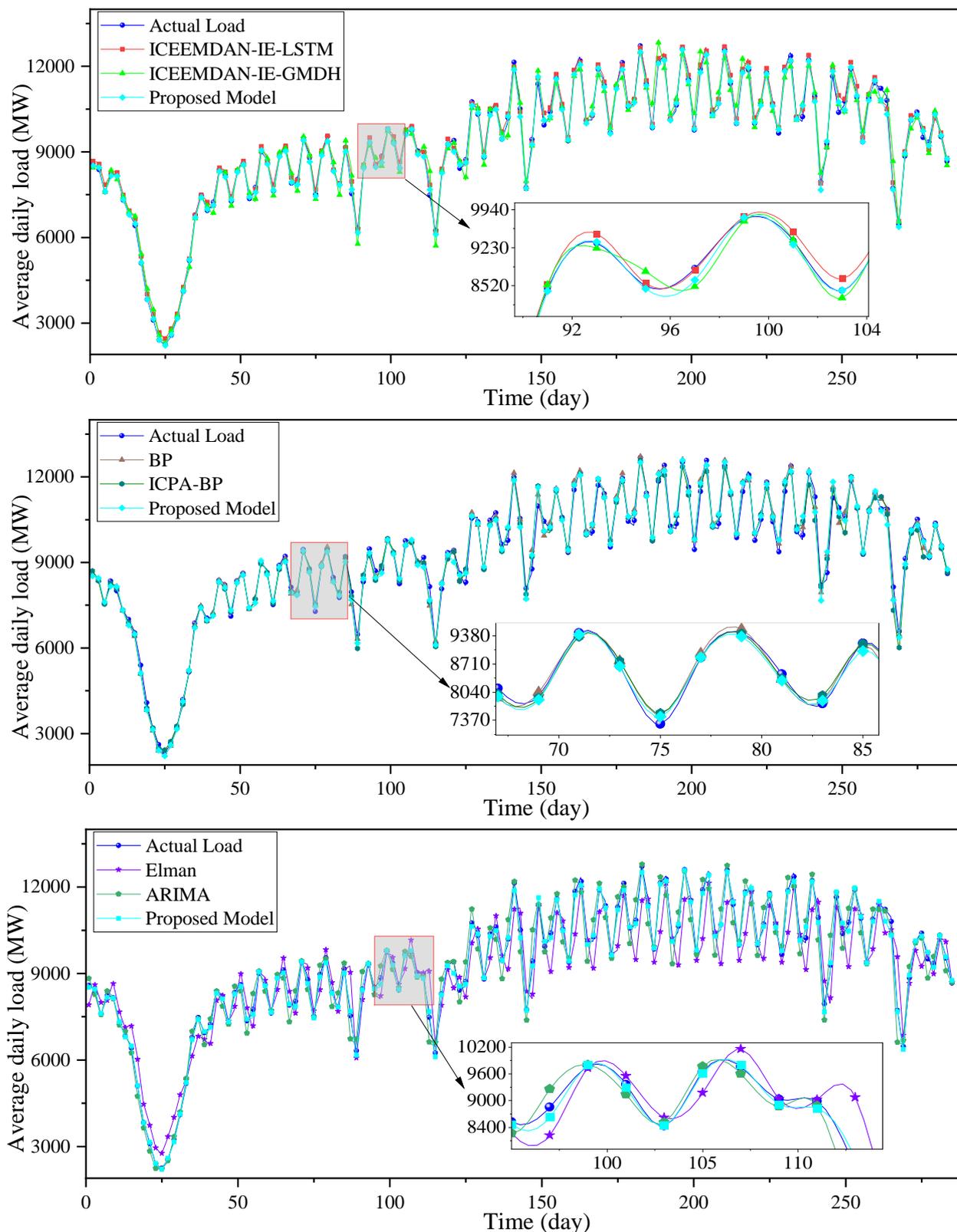


Fig.19. Predicted results for dataset 1

improved by 2.330%, 1.520% and 0.510%, respectively. In addition, compared with the GMDH neural network, the RMSE, MAPE, and MAE of the proposed load prediction model improved by 52.57%, 55.44%, and 57.51%, separately. And in dataset 2 RMSE, MAE and MAPE improved by 44.53%, 43.25%, and 42.49%, respectively. These results verify that the proposed hybrid load forecasting model is significantly effective in using ICPA to combine LSTM and

GMDH, and then compared with single method, the three results indicators of the proposed hybrid forecasting model are significantly improved.

From the raw data, it can be observed that the complexity of dataset 1 is significantly higher than that of dataset 2. The ARIMA prediction model has a significant difference in its prediction results when predicted in dataset 1 and dataset 2. In addition to, in dataset 1, ICEEMDAN-IE-GMDH model

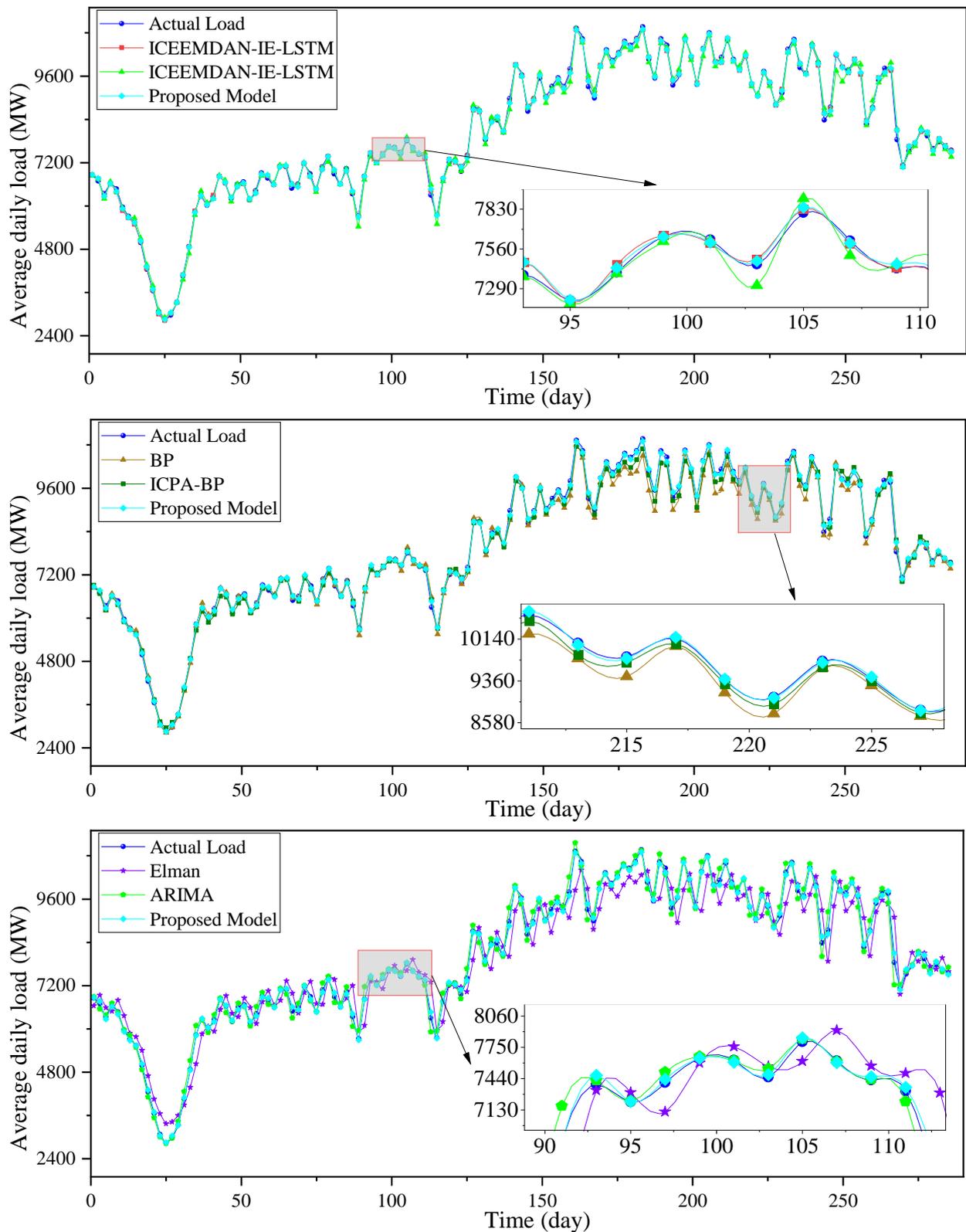


Fig.20. Predicted results for dataset 2

TABLE XI.  
MAXIMUM AND MINIMUM ABSOLUTE ERROR PERCENTAGES FOR EACH MODEL

	Data1		Data2	
	Max	Min	Max	Min
ARIMA	0.1428	2.4190e-04	0.0785	3.5680e-04
BPNN	0.0897	4.4974e-06	0.0639	2.1517e-06
ICAP-BPNN	0.1003	2.8989e-05	0.0453	<b>1.1901e-05</b>
Elman	0.2989	2.3071e-04	0.1886	4.3506e-05
ICEEMDAN-IE-LSTM	0.1079	4.5425e-05	0.0398	9.3641e-05
ICEEMDAN-IE-GMDH	0.1237	6.4915e-06	0.0651	6.4704e-05
Proposed Method	<b>0.0795</b>	<b>1.3812e-04</b>	<b>0.0345</b>	7.9465e-05

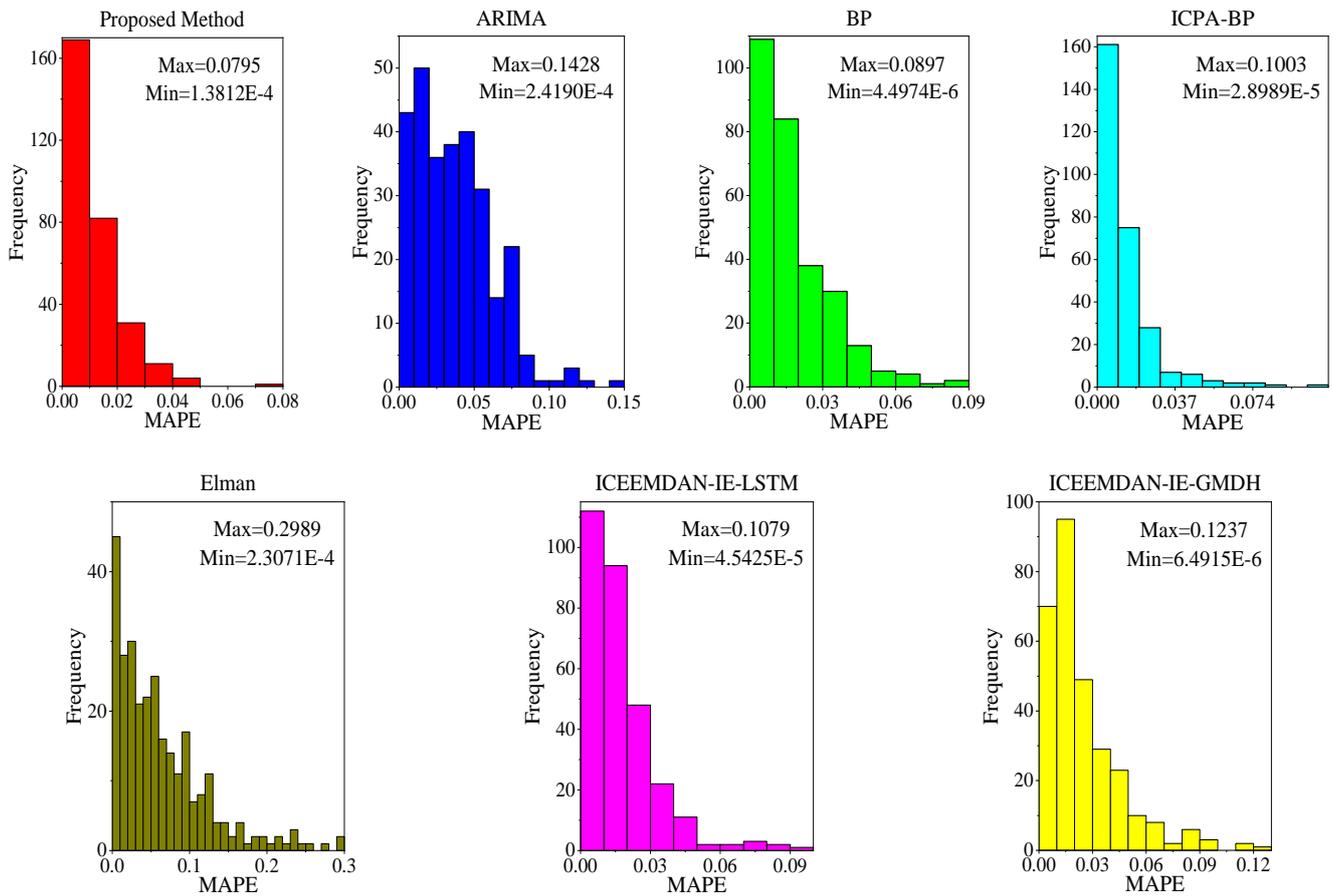


Fig.21.MAPE for each model in dataset 1

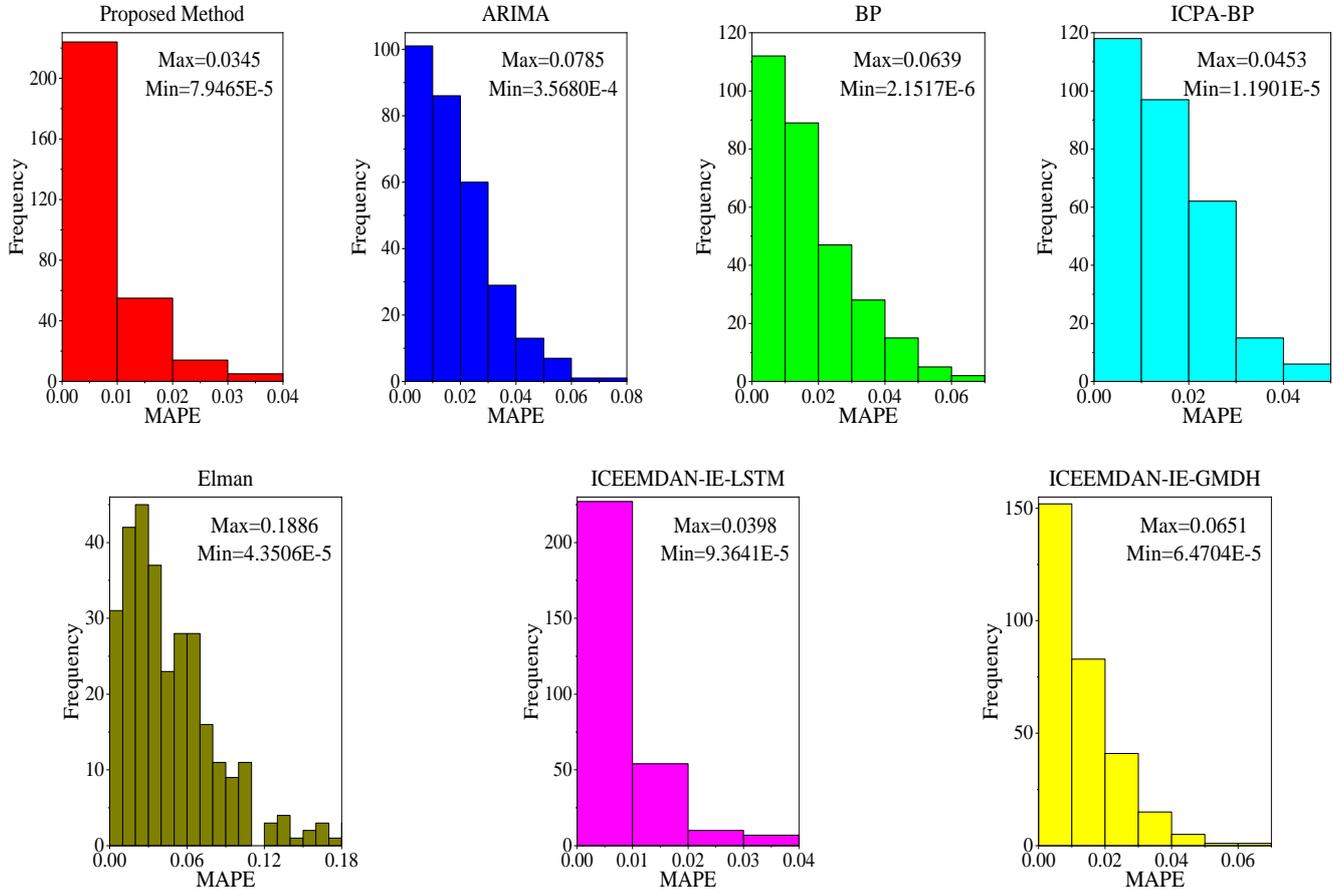


Fig.22.MAPE for each model in dataset 2

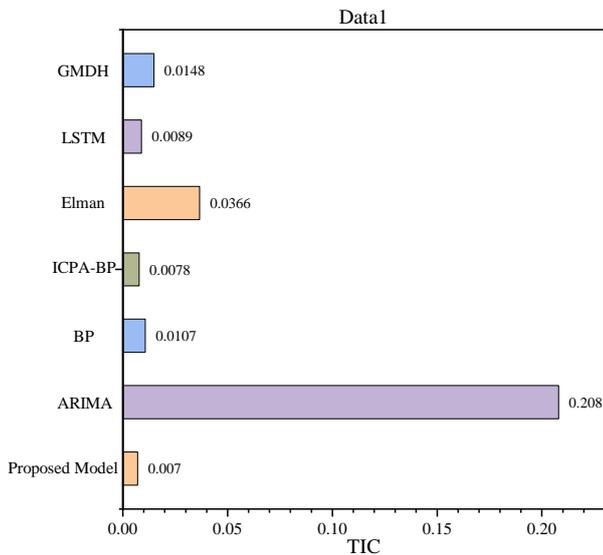


Fig.23.TIC for each model in dataset 1

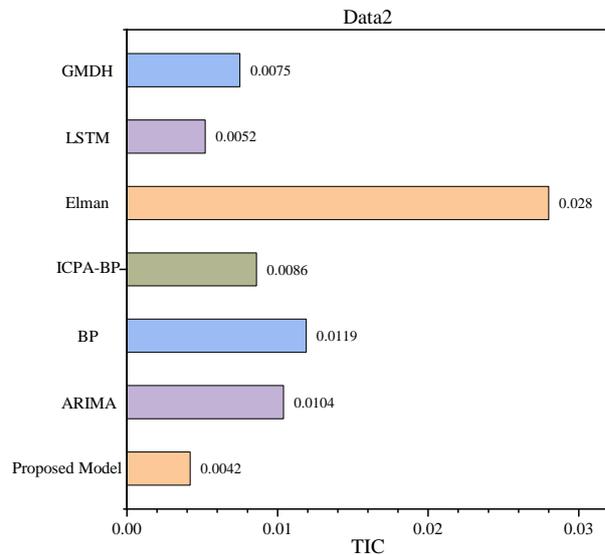


Fig.24.TIC for each model in dataset 2

TABLE XII.  
THE SPEARMAN VALUE OF THE LOAD FORECASTING MODEL

	Dataset 1	Dataset 2
BP	0.9842	0.9853
ICPA-BP	0.9901	0.9914
ICEEMDAN-IE-LSTM	0.9868	0.9959
ICEEMDAN-IE-GMDH	0.9883	0.9960
Proposed Model	<b>0.9974</b>	<b>0.9989</b>
Elman	0.9311	0.9566
ARIMA	0.9756	0.9633

has significantly poorer prediction results compared to other models, with the MAPE of 2.5221%, ranking fifth among all models. However, in dataset 2 with smaller complexity, the prediction results of the ICEEMDAN-IE-GMDH have the MAPE of only 1.1955%, ranking third among all models. For load data of different complexity, the prediction results of general prediction models will have instability, and if load data of higher complexity is encountered, the load prediction result will have larger errors. Nevertheless, the proposed load prediction model combines the advantages of two neural networks (LSTM and GMDH), and its load prediction results outperform the other six prediction models in both datasets with excellent prediction precision. The proposed hybrid load forecasting model has good stability and precision for both high-complexity and low-complexity power load data.

To better prove the superiority of the proposed short-term load forecasting model, the absolute error percentages and the inequality coefficient (TIC) of each prediction model in different data sets are calculated as shown in Fig.21-Fig.24. In addition to this, as shown in TABLE XI, the maximum and minimum values of the absolute error percentage of each load prediction model are selected for analysis.

Fig.21 illustrates the AMPE error distribution of the load prediction model proposed and the six comparison prediction models in dataset 1. And Fig.22 illustrates the AMPE error distribution of the load prediction model proposed and the six comparison load prediction models in dataset 2. The smallest values of MAPE for the proposed prediction model in this paper are 0.0795 and 0.0345 in the two data sets, which are both smaller than the values of other prediction models. The MAPE values of the prediction models proposed are mainly

in the range of 0-0.01. And the experimental load prediction results are 285 days in total, in data set 1, the MAPE value of the proposed hybrid load prediction model is less than 0.01 for 165 days, in dataset 1, the MAPE value of proposed load prediction model is less than 0.01 for 210 days. In the two datasets of different complexity, the floating range of MAPE of other load prediction models varies widely, but the floating range of MAPE of the proposed hybrid load prediction model is basically unchanged, which also demonstrates the strong stability of the proposed hybrid load prediction model. Other than that, the prediction model combining the ICEEMDAN method has less variation in the MAPE float range compared to other prediction models. The above results indicate that processing the data by the ICEEMDAN method improves the adaptability of the model and can make the model prediction more stable.

Fig.23 and Fig.24 represent the TIC values between the predicted and true values of load seven prediction models in the two experimental data sets, respectively. From the above two figures, it can be visually visualized that the proposed forecasting model has a TIC of 0.007 in dataset 1 and 0.0042 in dataset 2. The TIC values of the proposed load forecasting models are minimal compared with the TIC values of other load forecasting models. Due to the fact that the smaller TIC value of the forecasting model indicates the higher prediction precision of the forecasting model, and by analyzing the TIC values of the prediction models, it can be clearly concluded that the proposed hybrid load forecasting model has a better prediction performance than the other six prediction models.

The correlation between the predicted and actual values of each load forecasting model is calculated by the formula of

the correlation coefficient. The correlation values between the predicted and actual values of each forecasting model are shown in TABLE XII. From the data in TABLE XII, it can be visually observed that the highest correlation between the predicted and actual values of the short-term load forecasting model is 99.89% in the less complex dataset 2. In addition, the correlation between the predicted and actual values of the prediction model is lower in the more complex dataset 1 than in dataset 2, but it is still the highest correlation among all the models, reaching 99.74%. The above results fully prove the validity of the short-term load model proposed in this paper.

## VI. CONCLUSION

In this paper, a hybrid short-term load forecasting model that is based on ICEEMDAN, IE, LSTM, GMDH, and ICAP is proposed. ICEEMDAN is used in raw load data processing to decompose the original daily average load data into several IMF components with different characteristics of information. The value of IE for each IMF component is calculated. Then IMF components with similar IEs are reorganized to avoid cumulative prediction errors. In the power load prediction, the proposed load prediction model combines the prediction advantages of the LSTM and GMDH models to obtain the best power load prediction results. The reconstructed IMF components are predicted by both methods respectively, and then the final prediction results are obtained by weighting the prediction results of both by ICAP. After the validation of the experiments on two power load datasets, the hybrid load forecasting model proposed in this paper is very effective and outperforms other prediction models in four aspects: MAE, RMSE, MAPE, and TIC. In summary, the hybrid short-term load forecasting method proposed in this paper effectively reduces the prediction error, increases the prediction stability, and improves the prediction accuracy. In this paper, only the weekly periodicity of power load data is considered when studying short-term load forecasting, and the effects of other factors are not considered. Therefore, in future research, the influence of more factors on short-term load forecasting, such as weather, humidity, and holidays, can be considered.

## REFERENCES

- [1] J. Browell and M. Fasiolo, "Probabilistic Forecasting of Regional Net-Load With Conditional Extremes and Gridded NWP," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5011-5019, 2021.
- [2] M. Malekizadeh, H. Karami, M. Karimi, A. Moshari and M. J. Sanjari, "Short-term load forecast using ensemble neuro-fuzzy model," *Energy*, vol. 196, p. 117127, 2020.
- [3] M. Kazemzadeh, A. Amjadian and T. Amraee, "A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting," *Energy*, vol. 204, p. 117948, 2020.
- [4] K. Wang, J. Wang, B. Zeng and H. Lu, "An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization," *Applied Energy*, vol. 314, p. 118938, 2022.
- [5] X. Tang, Y. Dai, Q. Liu, X. Dang and J. Xu, "Application of Bidirectional Recurrent Neural Network Combined With Deep Belief Network in Short-Term Load Forecasting," *IEEE Access*, vol. 7, pp. 160660-160670, 2019.
- [6] X. Tang, H. Chen, W. Xiang, J. Yang and M. Zou, "Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network," *Electric Power Systems Research*, vol. 205, p. 107761, 2022.
- [7] Y. Nie, P. Jiang and H. Zhang, "A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting," *Applied Soft Computing*, vol. 97, p. 106809, 2020.
- [8] W. R. Christiaan, "Short-Term Load Forecasting Using General Exponential Smoothing," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-90, no. 2, pp. 900-911, 1971.
- [9] W. Charytoniuk, M. S. Chen and P. Van Olinda, "Nonparametric regression based short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 725-730, 1998.
- [10] J. Li, D. Deng, J. Zhao and D. Cai *et al.*, "A novel hybrid short-term load forecasting method of smart grid using mlr and lstm neural network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2443-2452, 2021.
- [11] J. C. Lopez, M. J. Rider and Q. Wu, "Parsimonious Short-Term Load Forecasting for Optimal Operation Planning of Electrical Distribution Systems," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1427-1437, 2019.
- [12] Y. Wang, Q. Chen, N. Zhang and Y. Wang, "Conditional Residual Modeling for Probabilistic Load Forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7327-7330, 2018.
- [13] W. Liu, Z. Dou, W. Wang and Y. Liu *et al.*, "Short-Term Load Forecasting Based on Elastic Net Improved GMDH and Difference Degree Weighting Optimization," *Applied Sciences*, vol. 8, no. 9, p. 1603, 2018.
- [14] H. Bian, Y. Zhong, J. Sun and F. Shi, "Study on power consumption load forecast based on K-means clustering and FCM-BP model," *Energy Reports*, vol. 6, pp. 693-700, 2020.
- [15] H. Shi, M. Xu and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271-5280, 2017.
- [16] S. Muzaffar and A. Afshari, "Short-Term Load Forecasts Using LSTM Networks," *Energy Procedia*, vol. 158, pp. 2922-2927, 2019.
- [17] B. Li, J. Zhang, Y. He and Y. Wang, "Short-Term Load-Forecasting Method Based on Wavelet Decomposition With Second-Order Gray Neural Network Model Combined With ADF Test," *IEEE Access*, vol. 5, pp. 16324-16331, 2017.
- [18] J. Zhang, Y. Wei, D. Li, Z. Tan and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774-781, 2018.
- [19] P. Singh and P. Dwivedi, "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," *Applied Energy*, vol. 217, pp. 537-549, 2018.
- [20] S. H. Rafi, Nahid-AI-Masood, S. R. Deeba and E. Hossain, "A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network," *IEEE Access*, vol. 9, p. 1, 2021.
- [21] J. Zhang, Y. Wei, D. Li, Z. Tan and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774-781, 2018.
- [22] A. Alsalah, D. Holloway, M. Mousavi and J. Lavroff, "Identification of wave impacts and separation of responses using EMD," *Mechanical Systems and Signal Processing*, vol. 151, p. 107385, 2021.
- [23] Y. Zhou, T. Li, J. Shi and Z. Qian, "A CEEMDAN and XGBOOST-based approach to forecast crude oil prices," *Complexity*, vol. 2019, p. 15, 2019.
- [24] T. Li, Z. Qian and T. He, "Short-Term Load Forecasting with Improved CEEMDAN and GWO-Based Multiple Kernel ELM," *Complexity*, vol. 2020, pp. 1-20, 2020.
- [25] J. Shao, K. Hu, C. Wang, X. Xue and B. Raj, "Is normalization indispensable for training deep neural network?" *Advances in Neural Information Processing Systems*, vol. 33, pp. 13434-13444, 2020.
- [26] Y. Chen, X. Tang, X. Qi, C. Li and R. Xiao, "Learning graph normalization for graph neural networks," *Neurocomputing*, vol. 493, pp. 613-625, 2022.
- [27] X. Wang, G. Song, J. Gao and X. Wei *et al.*, "High impedance fault detection method based on improved complete ensemble empirical mode decomposition for DC distribution network," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 538-556, 2019.
- [28] Y. Shrivastava and B. Singh, "A comparative study of EMD and EEMD approaches for identifying chatter frequency in CNC turning," *European Journal of Mechanics - A/Solids*, vol. 73, pp. 381-393, 2019.
- [29] Y. Gao, Y. Hang and M. Yang, "A cooling load prediction method using improved CEEMDAN and Markov Chains correction," *Journal of Building Engineering*, vol. 42, p. 103041, 2021.
- [30] B. Punithkumar, "Improved ECG Denoising using CEEMAN based on Complexity Measure and Nonlocal Mean Approach," *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp. 606-615, 2022.
- [31] H. Zheng, J. Yuan and L. Chen, "Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation," *Energies*, vol. 10, no. 8, p. 1168, 2017.
- [32] K. Wang, J. Wang, B. Zeng and H. Lu, "An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization," *Applied Energy*, vol. 314, p. 118938, 2022.

- [33] A. G. Ivakhenko, E. A. Savchenko and G. A. Ivakhenko, "Problems of future GMDH algorithms development," *Systems Analysis Modelling Simulation*, vol. 43, no. 10, pp. 1301-1309, 2003.
- [34] K. Elbaz, S. Shen, A. Zhou, Z. Yin and H. Lyu, "Prediction of Disc Cutter Life During Shield Tunneling with AI via the Incorporation of a Genetic Algorithm into a GMDH-Type Neural Network," *Engineering*, vol. 7, no. 2, pp. 238-251, 2021.
- [35] G. Memarzadeh and F. Keynia, "Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm," *Electric Power Systems Research*, vol. 192, p. 106995, 2021.
- [36] K. Zhu, Y. Li, W. Mao, F. Li and J. Yan, "LSTM enhanced by dual-attention-based encoder-decoder for daily peak load forecasting," *Electric Power Systems Research*, vol. 208, p. 107860, 2022.
- [37] G. Chen, B. Tang, X. Zeng and P. Zhou *et al.*, "Short-term wind speed forecasting based on long short-term memory and improved bp neural network," *International Journal of Electrical Power & Energy Systems*, vol. 134, p. 107365, 2022.
- [38] K. M. Ong, P. Ong and C. K. Sia, "A carnivorous plant algorithm for solving global optimization problems," *Applied Soft Computing*, vol. 98, p. 106833, 2021.
- [39] E. Ofori-Ntow Jnr, Y. Y. Ziggah, M. J. Rodrigues and S. Relvas, "A hybrid chaotic-based discrete wavelet transform and Aquila optimisation tuned-artificial neural network approach for wind speed prediction," *Results in Engineering*, vol. 14, p. 100399, 2022.
- [40] A. Agga, A. Abbou, M. Labbadi, Y. E. Houm and I. H. Ou Ali, "CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Systems Research*, vol. 208, p. 107908, 2022.
- [41] J. X. Tee, A. Selvarajoo and S. K. Arumugasamy, "Prediction of carbon sequestration of biochar produced from biomass pyrolysis by artificial neural network," *Journal of Environmental Chemical Engineering*, vol. 10, no. 3, p. 107640, 2022.
- [42] J. De-Prado-Gil, C. Palencia, N. Silva-Monteiro and R. Martínez-García, "To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models," *Case Studies in Construction Materials*, vol. 16, p. e1046, 2022.