J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 8097--8117

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

# A short-term load forecasting method for distribution networks based on multivariate information and exploratory factor analysis

Zhou Wang[1,✉], Junqiang Mu[2], Tianting Li[1], Li Zhang[3]

[1] *State Grid Gansu Electric Power Company Economic and Technological Research Institute, Lanzhou, Gansu, 730050, China*
[2] *State Grid Gansu Electric Power Company Lanzhou Power Supply Company, Lanzhou, Gansu, 730000, China*
[3] *State Grid Gansu Electric Power Company Tianshui Power Supply Company, Tianshui, Gansu, 741000, China*

ABSTRACT

Accurate short-term load forecasting of distribution networks can ensure the normal life and production of the society, effectively reduce the cost of power generation, and improve the economic and social benefits. Aiming at the multivariate information that affects the power load, this paper utilizes factor analysis to reduce the dimensionality of the original influencing factors, and obtains the main influencing factors with the highest contribution rate, so as to guarantee the accuracy of the neural network prediction. On this basis, the neural network structure is improved by combining AlexNet and GRU, and the short-term load prediction model of distribution network is finally constructed. The relevant charge data of N village in 2023-2024 is used as a research sample to analyze the main influencing factors of its short-term load change, and three main influencing factors affecting the load change in the short term are identified as temperature, air pressure, and humidity factor. Based on the real data of N-village distribution network to carry out prediction simulation experiments, the load short-term prediction curve of this paper's model has a better fitting degree and good stability, and the values of the prediction result evaluation indexes MRE, RMSE and MAE are smaller than those of the other comparative models, which are basically able to maintain a prediction accuracy of more than 90%.

*Keywords:* factor analysis, AlexNet, GRU, short-term load forecasting, distribution grid

## 1. Introduction

With the rapid development of social economy and the continuous progress of technology, the demand for energy is also increasing. Load forecasting of distribution network is particularly important. Load forecasting refers to analyzing historical data and predicting future development trends to forecast the

✉ Corresponding author.
*E-mail address:* ZhouWang2024@126.com (Z. Wang).
Received 03 March 2024; Revised 12 April 2024; Accepted 15 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-442

size and trend of load changes in a certain period of time. It plays an important role in the planning, operation and scheduling of distribution networks, and can help power system managers to better arrange the power plan, rationalize the configuration of equipment, improve the efficiency of power consumption, reduce the waste of energy, and lower the cost of power supply [1-4].

Short-term load forecasting refers to the accurate prediction of the load demand of the distribution network in a future period of time (usually a few hours or a day). Accurate short-term load forecasting can help power system dispatchers to develop reasonable generation plans and rationalize the operation and management of the grid. In contrast, inaccurate load forecasting may lead to grid overload or energy waste [5]. Currently, short-term load forecasting for distribution grids is mainly based on traditional time series analysis methods or statistically based methods. These methods often have some limitations in accuracy and robustness, especially in the face of large load fluctuations or noise interference. The study of short-term load forecasting methods based on multivariate information and exploratory factor analysis to improve the accuracy and robustness of load forecasting has become one of the research hotspots in the field of distribution networks [6-8].

Literature [9] proposes a methodology to solve the problem of probabilistic short-term load forecasting for distribution networks, which employs a Dirichlet process mixture model to deal with the uncertainty in the load pattern and extrapolates it through an MCMC-based approach. The results of the study point out that the method performs better for a given level of aggregation as opposed to the benchmark method. Literature [10] introduces a global modeling framework for load forecasting in distribution networks and proposes an unsupervised localization mechanism and an optimal integration construction strategy to improve the performance of the framework. The effectiveness of the proposed framework is demonstrated experimentally and exhibits highly desirable features. Literature [11] proposes a short-term, high-resolution distribution system load forecasting method based on support vector regression and parallel parameter optimization. It also automatically reconfigures the distribution system in a dynamic and ex-ante manner. Simulation results confirm the effectiveness of the method. Literature [12] describes the robust wavelet transform neural network load forecasting model, which utilizes both time and frequency domain information to improve the forecasting accuracy of the model. Simulation-based experiments reveal that the proposed robust time-frequency load forecasting model is superior compared to the traditional time-domain forecasting model. Literature [13] describes the application of several software for distribution network load forecasting in smart grid environment, including short-term feeder load forecasting, next year's load pocket forecasting, etc., and describes the characteristics of these eh software, indicating that their application largely improves the efficiency and reliability of the distribution network. Literature [14] proposed a short-term load forecasting model for distribution networks combining maximum information coefficient (MIC) and factor analysis. The nonlinear correlation between load and input features is quantified using MIC, and short-term load forecasting based on the GWO-GRNN model is realized by dimensionality reduction of the screened input features with the help of FA. The accuracy and applicability of the forecasting model are verified using a regional distribution network as a research object.

Literature [15] proposed a hybrid model based on STL, LSTMs and XGBoost. The load sequence is decomposed by STL method, LSTM neural network is taken as the learner and XGBoost is used as the final step of integrated prediction. The feasibility of the model is verified by applying STL-LSTMs-XGBoost to the load data of 10kV lines in distribution networks. Literature [16] launched a comparative study on the application of hybrid AI techniques, ANFIS and OP-ELM in distribution load forecasting. It was emphasized that hybrid AI techniques can be used to forecast loads in distribution networks and save costs for power utilities. Literature [17] proposes a distribution network load forecasting method that considers the impact of PV generation and conventional loads. The load data of distribution network in a region is used as an example to compare it with the traditional technical solutions, and the results show that the proposed method shows obvious advantages in terms of load forecasting accuracy and error rate. Literature [18] proposes a deep cyclic mixer model as a unified framework for accurate load forecasting. The method is based on the multilayer integration of LSTM and GRU networks, thus enabling better utilization of both techniques. Simulation results show that the deep cyclic mixer model possesses high performance. Literature [19] introduces a genetic algorithm to optimize the classical neural network algorithm to cope with the deficiencies of the traditional neural network, so as to improve the accuracy of short-term power load prediction of the power station and establish the prediction model. Based on simulation training and comparison, the prediction results are derived. And the effectiveness of the proposed algorithm is verified by taking a city as an example. Literature [20] introduced the temperature factor into load forecasting, and examined the law of load change with temperature by analyzing the characteristics of power load change. And a mathematical model of load forecasting considering the temperature effect was constructed. The results show that the time-sharing load forecasting method has good forecasting performance and can be used as an effective strategy to improve the accuracy of load forecasting.

In this paper, the factor analysis method is utilized to reduce the dimensionality of the input variables of the neural network without excessive loss of information on the influencing factors, to obtain the main influencing factors that have the highest contribution to the short-term load of the distribution network, and to guarantee the prediction accuracy of the neural network. The factors that have a greater impact on load changes are used as part of the inputs to the improved AlexNet-GRU prediction model.A short-term load prediction method based on the improved AlexNet-GRU deep learning network is proposed.Using AlexNet's characteristics of reducing the complexity of multidimensional data sets and GRU's good at dealing with time-series data, the two are combined, and the structure of the network, The network structure and the number of neurons are improved to finalize the load forecasting model. The short-term load influencing factors of the distribution network in N village, Panyu District, Guangzhou City, China, are analyzed to obtain the main influencing factors. Combining the main influencing factors of the short-term load variation of the distribution network in N village and the model of this paper, simulation experiments on short-term load forecasting of the distribution network are carried out to test the load forecasting performance of the model of this paper.

## 2. Multivariate information analysis model for short-term loads in distribution networks

In the process of short-term power load forecasting in the distribution network, there are a variety of direct or indirect influencing factors, and with the continuous development of society, the influence of power load factors are gradually increasing. But in the past the actual load forecast, and can not consider too many influencing factors. On the one hand, it is difficult to collect information, and on the other hand, too many factors will cause modeling difficulties and a large number of calculations. Therefore, when considering the input variables of the neural network, it is necessary to consider the dimensionality of the input variables. In the actual load forecasting process, it is necessary to comprehensively consider various influencing factors, extract the main components, and reduce the influencing factors from high-dimensional to low-dimensional, so as to maintain the adequacy of the influencing factors and guarantee the fit of the neural network.

As a multivariate statistical analysis method, the basic idea of factor analysis is to make the original multiple variables into a few representative public factors by combining them through dimensionality reduction. The basic principle of the factor method is to deal with the original variables in a standard way, and analyze the relationship between the variables through the matrix of correlation coefficients, so as to reduce the negative impact of correlation between the indicators on the empirical analysis. The original proposed variables are grouped through the analysis of correlation, and those with high correlation between variables are classified as a group, representing a common factor. By this means, the problem composed of multiple indicators is represented by a linear combination of a few special factors and a common factor, which achieves the effect of dimensionality reduction of multiple indicators.

Let the observation matrix of sample composition values be as shown in equation (1), where $m$ is the number of sample observations, $n$ is the number of indicators, and $x_{mn}$ is the score of the $m$ th evaluator on the $n$ th indicator:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \tag{1}$$

The core of the factor analysis method is to factor analyze a number of composite indicators and extract the common factors, and then construct a score function using the variance contribution of each factor as the sum of the weights and the score multiplier for that factor. There is a $m$-dimensional observable random vector $x = (x_1, x_2, \ldots, x_m)T$ , where $f_1, f_2, \cdots, f_n$ is the public factor and $u_1, u_2, \ldots, u_m$ is the special factor, which are unobservable random variables. The mathematical representation of the factor analysis method is the matrix: $X = AF + u$ , i.e:

$$\begin{cases} x_1 = \mu_1 + a_{11} f_1 + a_{12} f_2 + \cdots + a_{1n} f_n \\ x_2 = \mu_2 + a_{21} f_1 + a_{22} f_2 + \cdots + a_{2n} f_n \\ \cdots\cdots\cdots \\ x_m = \mu_m + a_{m1} f_1 + a_{m2} f_2 + \cdots + a_{mn} f_n \end{cases} \tag{2}$$

and the following conditions are met.

1) $n \le m$ .

2) $COV(f, u) = 0$ (i.e. $F$ and $U$ are uncorrelated).

3) $E(f) = 0COV(f) = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$ i.e. $f_1, f_2, \ldots, f_n$ are uncorrelated and both variances are 1 and means are 0.

$A(a_{ij})$ is the coefficient of the common factor $F(f_1, f_2, f_3, \cdots, f_n)$, called the factor loading matrix, and $a_{ij}\,(i = 1, 2, \ldots.., m; j = 1, 2, \ldots, n)$ is called the factor loadings, which are the loadings of the $i$ th original variable on the $j$ th factor, or $a_{ij}$ can be thought of as the weights of the $i$ th variable on the $j$ th common factor. $a_{ij}$ is the covariance between $x_i$ and $f_j$ , and the correlation coefficient between $x_i$ and $f_j$, indicating the degree of $x_i$ dependence or correlation on $f_j$. The greater the absolute value of $a_{ij}$, the greater the loading of the common factor $f_j$ on $x_i$.

The basic steps of the factorization method are mainly four steps as follows.

1) First test whether the original variables are amenable to factor analysis. The correlation test is used to analyze the original variables, the greater the correlation of the original variables, the more appropriate to use the factor analysis method, if the correlation of the original variables is lower than 0.3, it is considered that the original variables are not suitable for factor analysis:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y^2\right)}} \tag{3}$$

Let, $r$ be the correlation coefficient.

Partial correlation coefficient the partial correlation coefficient at level $m-2$ is calculated as follows.

Let $m$ correlation variables $x_1, x_2 \ldots\ldots x_m$ have a total of $n$ arrays.

First, calculate the simple correlation coefficient $r_{ij}$:

$$r_{ij} = \frac{SP_{ij}}{\sqrt{SS_i SS_j}}, (i, j = 1, 2 \cdots m) \tag{4}$$

Where, $SP_{ij} = \sum(x_i - \bar{x}_i)(x_j - \bar{x}_j)$, $SS_i = \sum(x_i - \bar{x}_i)^2$, $SS_j = \sum(x_j - \bar{x}_j)^2$ when the matrix of correlation coefficients $R$ consisting of simple correlation coefficients $r_{ij}$ is:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix} \tag{5}$$

Accordingly, find the inverse matrix of $R$ and set it to $C$:

$$C = R^{-1} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & r_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix} \tag{6}$$

The formula for the level $m-2$ partial correlation coefficient $r_{ij}$ of the correlation variables $x_i$ and $x_j$ is then:

$$r_{ij} = \frac{-c_{ij}}{\sqrt{c_{ii} c_{ij}}} (i, j = 1, 2 \cdots, m, i \neq j) \tag{7}$$

If the value of the test statistic is large and Sig is less than 0.01, the original hypothesis is rejected, proving that the original variable to be analyzed can be factor analyzed.

2) Constructing factor variables. Before constructing the factor variables, it is necessary to extract the factors, and there are more methods to extract the factors, among which, the application of principal component analysis is more common. This method firstly simplifies all the current combination of variables into the first principal component, which can explain the largest one categorized variable, and secondly, separates the remaining variables and performs the linear combination in order to calculate the second principal component. The above steps were repeated for calculation until no more common principal components could be extracted. There are two main methods for determining the number of factors, the first sets 1 as the criterion and extracts the factors with an eigenroot greater than 1 and removes the factors with an eigenroot less than 1. The second is based on the steep slope test method, where all factors are sorted according to the size of the characteristic root and the factors are plotted on a two-dimensional coordinate paper, and when at a certain point, the characteristic root changes more sharply, this point is set as the critical point, and the number of factors is determined accordingly.

3) Apply rotations to make the factor variables more interpretable.

4) Calculate the factor variable scores.

The calculation process of the DRF method has four steps as follows:

1) Standardize the original data and calculate the correlation matrix of the standard data and its corresponding eigenvalues and eigenvectors.

2) Calculate the contribution rate of variance and the contribution rate of cumulative variance.

3) Extract the raw indicators and determine the factors for analysis.

4) Determine the score of each factor.

## 3. Short-term load forecasting model for distribution networks

Load forecasting is a key part of power grid planning, which plays a very important role in ensuring the normal operation of the power system, realizing the balance between power supply and demand, and optimizing the operation of the distribution network by implementing a reasonable operation mode. Due to the issue of power is difficult to large reserves, power production, consumption is a simultaneous process, so the size of the forecast load on the grid operation is of great significance. Along with the development of power marketization, the transactions between power companies become more and more frequent, and the accuracy of load forecasting is required more and more.

In the previous chapter, this paper proposes the short-term load factor analysis method for distribution networks, through which the factor analysis makes the dimensions of the influencing factors input into the neural network greatly reduced, so that the accuracy of the neural network prediction is guaranteed under the premise of not excessive loss of information. On this basis, this chapter further proposes an improved short-term load forecasting method for AlexNet-GRU deep learning network, which takes the factors that have a greater influence on load changes as part of the input to the improved AlexNet-GRU forecasting model to realize the accuracy of short-term load forecasting for distribution networks.

### *3.1. Improved AlexNet-GRU prediction modeling*

#### 3.1.1. AlexNet network:

1) Convolutional Neural Network CNN [21]. CNN is similar to other NNs in that it does not need to be modeled, but uses a deep network for the initial (input> to expected (output) transfer process, it is supervised similarly to BP networks, which can regulate the network weights by backpropagating the control of the objective function. Before training, the network weights and bias values must also be set in advance, usually between $\left[-1,1\right]$, and then trained to adjust them to the optimal values. This learning method has a very good synergistic effect on the deep model, which can make it get the global optimal solution as much as possible.

(1) Forward propagation. After inputting sample values to the network, the semantic information of the upper layer is extracted layer by layer from the initial data layer through a series of operations such as convolution operation, pooling operation, etc., and then abstracted layer by layer, and finally the information passed through the full connectivity layer is gathered into a most comprehensive message. This process is called "forward propagation". Let $X$ be the input of the neural network and $Y$ be the output, the process of feed forward operation is shown in equation (8):

$$Y = f_n\left(\ldots f_3\left(f_2\left(f_1(X)\right)\right)\ldots\right) \tag{8}$$

(2) Back propagation. The operation of backpropagation in convolutional neural networks is shown in equation (9), where data flows into the network through the input layer, and the information that flows out of the output layer after a series of forward propagation is compared with the expected given output results, and then fed back forward one level at a time, thus updating the parameters of each convolutional layer and each fully connected layer:

$$E = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right) \tag{9}$$

The process is repeated until the model is collected as and the training of the neural network is completed.

2) AlexNet structure. AlexNet network is a variant of CNN and also one of the classical deep learning networks [22]. AlexNet network has mainly 8 layers structure including 5 convolutional layers and 3 fully connected layers. The AlexNet network uses the linear rectifier (ReLU) function instead of the Sigmoid function as the activation function for the convolutional and fully connected layers, which solves the problem of the neural network's gradient vanishing at deeper layers. Previously, the average pooling layer was commonly used in CNNs, while AlexNet all uses the maximum pooling layer to avoid the blurring effect of average pooling, and AlexNet is proposed to let the length of the pooling kernel be smaller than the size of the pooling kernel, so that there may be overlapping and overlapping in the output of the pooling layer, increasing the richness of the extracted features. Meanwhile, AlexNet network uses CUDA to play a role in accelerating the training of the deep network.

3) GRU structure. GRU is evolved from LSTM which is developed from RNN [23]. LSTM uses forgetting gate, input gate and output gate to regulate the flow of information in the LSTM cell. The gated loop unit mainly contains update gates, reset gates. The update gates in GRU are similar to the forgetting gates and input gates in LSTM. The update gates are used to decide what information to discard and what new information to add. The reset gate is another gate used to decide how much past information to forget. GRU is an evolved form of RNN, but it has a more concise neuronal infrastructure than LSTM and has a smaller number of parameters, thus converging faster:

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{10}$$

where $z_t$ denotes the update gate, $x_t$ is the input of the current moment, $h_{t-1}$ is the hidden state of the previous moment, $W_z$ is the weight parameter of the update gate, is the matrix dot product, and $\sigma$ is the sigmoid activation function [24]. The role of the update gate is to decide how much information from the previous and current moments needs to be passed on:

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{11}$$

where $r_t$ denotes the reset gate and $W_r$ is the weight parameter of the reset gate. The role of the reset gate is to decide how much of the information from the hidden state of the previous moment needs to be forgotten:

$$\tilde{h}_t = \tanh\left(W \cdot \left[r_t * h_{t-1}, x_t\right]\right) \tag{12}$$

where $W$ is the weight parameter of the candidate set and * is the matrix product. $\tilde{h}_t$ denotes the memory content of the current moment, which consists of the important information of the past stored in the reset gate and the information entered at the current moment:

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h}_t \tag{13}$$

where $h_t$ will retain the information from the GRU unit at the current moment and pass it to the next unit. $1 - z_t$ and $z_t$ determine the information collected from the hidden state at the current moment and the previous moment.

3.1.2. Improvement of AlexNet-GRU. The classical AlexNet convolutional neural network has a strong ability to discriminate and recognize images, but this paper is to apply the AlexNet migration in the field of image recognition to the field of short-term load forecasting, so to make targeted improvements to AlexNet, mainly to enhance the ability of the forecasting model to extract features from complex data, but also to combine with the GRU network. In order to adapt to the characteristics of the sample, obtain better training results and improve the efficiency of model training, this paper makes the following improvements after repeated debugging and optimization.

1) Improvement of network structure

(1) The use of LRN layers for normalization in AlexNet creates a competitive mechanism for local neuron activity to prevent data overfitting. In this paper, the LRN layer of AlexNet network is omitted in order to make the network easier to parallelize, and the LRN network layer has a weak ability to generalize the load characteristics.

(2) The input of classical AlexNet is an image of size 227*227*3, and its convolutional layer uses two-dimensional convolution (2D-CNN), while one-dimensional convolution (1D-CNN) is commonly used in the field of sequence modeling and natural language processing. Therefore, in this paper, the 2D convolutional layer in the network is replaced with a 1D convolutional layer to facilitate the processing of time series data.

2) Selection of optimizer. In this paper, we choose to use the Adam optimizer algorithm instead of the SGD algorithm. Adam absorbs the advantages of Adagrad and momentum gradient descent algorithms, which can adapt to sparse gradients (i.e., natural language and computer vision problems) and also alleviate the problem of gradient oscillation.

Let the parameter to be optimized be $w$, the loss function be loss, the learning rate be $lr$, the decay rate of the first and second order momentum be $\beta_1$ and $\beta_2$, take the moment $t$ as the current moment, and the moment $t+1$ as the next moment, and describe the specific training process of Adam as follows.

(1) Find the gradient value of the loss function with respect to the parameter at the current moment at this point in time $g_t$:

$$g_t = \nabla loss = \frac{\partial loss}{\partial w_t} \tag{14}$$

(2) Derive the first-order momentum $m_t$ and the second-order momentum $V_t$ at the current moment:

$$m_t = \beta_1 \cdot m_{t-1} + \left(1-\beta_1\right) \cdot g_t \tag{15}$$

$$V_t = \beta_2 \cdot V_{t-1} + \left(1-\beta_2\right) \cdot g_t^2 \tag{16}$$

(3) Calculate the corrected first and second order momentum deviations $\hat{m}_t$ and $\hat{V}_t$ for that moment:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{17}$$

$$\hat{V}_t = \frac{V_t}{1-\beta_2^t} \tag{18}$$

(4) Calculate the gradient of descent at that moment $\eta_t$:

$$\eta_t = lr \cdot \frac{\dfrac{m_t}{1-\beta_1^t}}{\sqrt{\dfrac{V_t}{1-\beta_2^t}}} \tag{19}$$

(5) Calculate the parameter $w_{t+1}$ for the next moment, thus completing the update:

$$w_{t+1} = w_t - \eta_t = w_t - lr \cdot \frac{\frac{m_t}{1-\beta_1^t}}{\sqrt{\frac{V_t}{1-\beta_2^t}}} \tag{20}$$

Replacing the SGD algorithm with the Adam optimizer algorithm is more computationally efficient, requires less memory, and is bias-corrected to have a defined range of learning rates for each iteration, making the parameters smoother.

## *3.2. Short-term load forecasting process for distribution networks*

3.2.1. Short-term charge prediction process. The process of forecasting short-term loads in distribution networks is as follows.

1) Collect data. Collect the real historical data in the forecast area and carry out the preprocessing work on the collected historical data.

2) Analyze data. Screening out the factors affecting the power load and establishing a comprehensive and accurate input feature set can effectively improve the prediction accuracy of short-term power load.

3) Modeling tuning. After completing the data preprocessing and the construction of input feature set, select the appropriate prediction method to establish the short-term power load prediction model. The prediction results are analyzed for errors, problems are identified, and then the model is optimized.

4) Forecast analysis. The optimized model is used to forecast the power load and the results are analyzed. In order to make the process more intuitive, it is drawn as a flowchart as shown in Fig. 1.

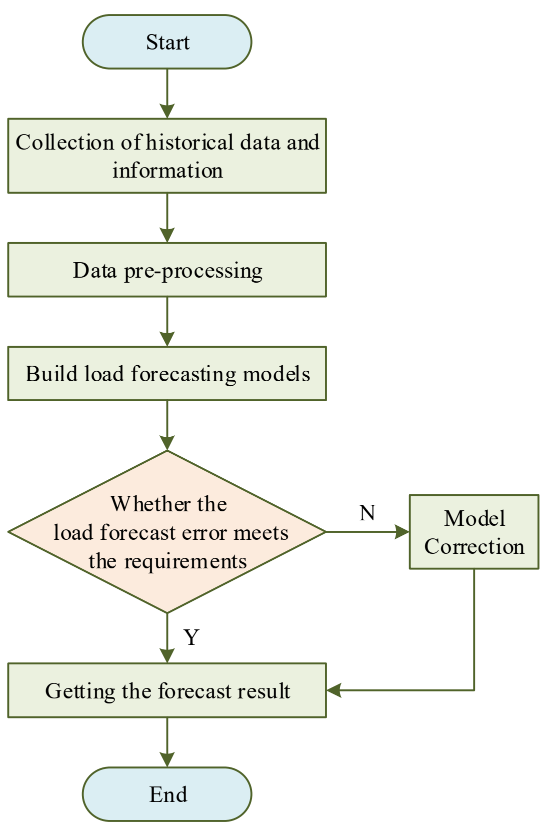



**Fig. 1.** Electricity load forecasting flow chart

3.2.2. Short-term power load data preprocessing and evaluation indicators. The quality of the data provides a strong guarantee for the forecasting accuracy, but it is easy to cause anomalies and missing data due to the problems of human factors or machine and equipment failures in the process of data collection. To address such problems, this section introduces a targeted data preprocessing approach to ensure the integrity of the collected electric load data. In addition, there is variability in the magnitude of power load data and external disturbances such as maximum and minimum temperatures, and artificial intelligence algorithms are sensitive to data types, so it is necessary to preprocess the numerical type variables in the sample data.

1) Missing data processing. For a large number of missing data, the workload to make up the data will be very large and heavy, so the deletion method is often used. When part of the missing data on the prediction model has a greater impact, more interpolation method to deal with the missing values to the effective historical data, commonly used interpolation methods are Lagrangian interpolation, three times the Elmit interpolation, three times the spline interpolation and so on. In the practical application of the project, three times Elmit interpolation and three times spline interpolation effect is more ideal, the corresponding function in MATLAB are $pchip()$ and $spline()$, in order to further reduce the error of filling data, make it closer to the real value, this paper takes the average value of three times Elmit interpolation and three times spline interpolation to fill the missing data.

2) Abnormal Data Handling. Anomalous values are usually set as a large deviation from the overall data set, and the current moment point is called the mutation point. If not processed, the model will directly fit the mutation value, while ignoring the learning of the intrinsic law of change of the real power load value, reducing the accuracy of the model prediction. Commonly used methods to identify outliers are $3\sigma$ principle, box plot, etc. Among them, the $3\sigma$ principle requires the data to obey the normal distribution or approximately obey the normal distribution, and the length coefficient of the control interval of the box plot is selected by experience and more outliers are identified. So both methods are highly subjective. In this paper, an unsupervised learning algorithm-iForest is used for outlier identification. This algorithm has better ROC performance and accuracy than most of the other outlier detection algorithms on a wide range of datasets and requires only a few parameters. In Python call function $1solation\ Forest()$ to initialize an isolated forest object with default parameters and call function $decision\_function()$ with the identified outliers to print out the outlier score for each data point.

3) Data normalization and quantification. The power load and various types of influencing factors have different values in their units or magnitudes, which, if left unprocessed and used directly as input variables for the prediction model, will increase the convergence time of the grid model and affect the model prediction performance. Considering the sensitivity of the neural network to the input data, data normalization is introduced for numerical variables to compress the data into the range of $[0,1]$ or $[-1,1]$, thus accelerating the training speed and reducing the prediction error.

(1) Data normalization:

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{21}$$

where $x$ is the sample data, $x_{\max}$ and $x_{\min}$ are the $x$ maximum and minimum values, respectively, and $\hat{x}$ is the normalized data.

(2) Data inverse normalization:

$$x_{pre} = x_n \left(x_{\max} - x_{\min}\right) + x_{\min} \tag{22}$$

where, $x_{pre}$ denotes the predicted value of the load with the same magnitude as the input load, and $x_n$ denotes the predicted value of the load after normalized load is predicted.

4) Model Evaluation Criteria. In this paper, the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to evaluate the model, which are calculated as follows:

$$MAE\left(y_i, \hat{y}\right) = \frac{1}{m}\sum_{1}^{m}\left|y_i - \hat{y}\right| \tag{23}$$

$$MAPE\left(y_i, \hat{y}\right) = \frac{1}{m}\sum_{1}^{m}\frac{\left|y_i - \hat{y}\right|}{\left|y_i\right|} \tag{24}$$

$$RMSE\left(y_i, \hat{y}\right) = \sqrt{\frac{1}{m}\sum_{1}^{m}\left(y_i - \hat{y}\right)^2} \tag{25}$$

where, $y_i$ is the real value after preprocessing, $\hat{y}$ is the predicted value and $m$ is the number of data to be predicted. Where MAE is used to reflect the real situation of the error, MAPE is used to assess the accuracy of the model, and RMSE is used to assess the prediction accuracy.

## 4. Analysis of factors affecting short-term load changes in the distribution network of village N

Due to the load in different areas has a strong randomness, so in order to comprehensively analyze and study the load characteristics of a region, through the local load influencing factors and their degree of influence of the study, you can grasp the load change law and future development trend, for the power sector power supply decision-making to provide an accurate reference basis.

N Village is located in Panyu District, Guangzhou City, China, with a total area of about 35,000 square kilometers, and there are 12 village groups under the jurisdiction of N Village, with a total of 1,266 households in the village, and a total population of about 21,000 people, including a household population of 4,388 and a foreign population of about 16,000, which makes it densely populated, and it is one of the major settlements of foreigners in the Panyu District, and the GDP of N Village in 2023 was as high as 179 million yuan, with a per capita income of 95,813 yuan, belonging to the secondary economic organization of urban development.

In the next research of this paper, the factor analysis method proposed in this paper will be used to analyze the influencing factors from the short-term load changes of the distribution network, and to explore the relationship between them and the load demand changes by using quantified data of specific sub-system indicators. The local load data were collected through the research, and the various influencing factors affecting the load changes in N village were obtained after the collation, and the influencing factors were integrated with the local load situation, as shown in Table 1.

**Table 1.** Load factor

| Load factor | |
|---|---|
| 1 | Consecutive days |
| 2 | Maximum temperature |
| 3 | Mean temperature |
| 4 | Minimum temperature |
| 5 | Mean relative humidity |
| 6 | Minimum relative humidity |
| 7 | Mean pressure |

| | |
|---|---|
| 8 | Maximum pressure |
| 9 | Minimum pressure |
| 10 | precipitation |
| 11 | Maximum wind speed |
| 12 | Mean wind speed |
| 13 | Maximum wind speed |

### *4.1. Correlation analysis of influencing factors*

This paper uses the data of N village during the period of 2023-2024 to correlate the loads and factors affecting each station area in N village, and the specific results are shown in Table 2. From the correlation analysis results, it can be seen initially that the factors that have a greater impact on the load of each station area are temperature, wind speed, air pressure, and humidity also has an impact on some stations.

**Table 2.** Correlation analysis

| - | No.9 utility transformer | 18A public transformer | No.10 Integrated room transformer | No.23Common box transformer | 16A public transformer | 14A public transforme r |
|---|---|---|---|---|---|---|
| Maximum temperature | 0.762** | 0.465** | 0.868** | 0.496** | 0.438 | 0.438** |
| Mean temperature | 0.768** | 0.493** | 0.881** | 0.501** | 0.458** | 0.457** |
| Minimum temperature | 0.762** | 0.505** | 0.864** | 0.484** | 0.464** | 0.456** |
| Maximum wind speed | 0.115* | 0.032 | 0.115* | 0.121* | 0.027 | 0.08 |
| Maximum wind speed | 0.058 | 0.024 | 0.058 | 0.083 | 0.022 | 0.083 |
| Mean wind speed | -0.181** | -0.187** | -0.232** | -0.091 | -0.078 | -0.089 |
| precipitation | 0.032 | 0.015 | 0.06 | -0.013 | 0.015 | -0.029 |
| precipitation | 0.027 | 0.012 | 0.066 | -0.005 | -0.018 | -0.007 |
| Precipitation rate | 0.035 | 0.017 | 0.071 | -0.006 | -0.006 | -0.03 |
| Maximum pressure | -0.667** | -0.524** | -0.753** | -0.462** | -.453** | -0.457** |
| Minimum pressure | -0.662** | -0.525** | -0.732** | -0.454** | -.441** | -0.462** |
| Mean pressure | -0.666** | -0.518** | -0.746** | -0.457** | -.442** | -0.456** |
| Minimum relative humidity | -0.02 | 0.083 | 0.048 | -0.067 | 0.038 | 0.018 |
| Mean relative humidity | 0.113* | 0.181** | 0.184** | -0.006 | 0.032 | 0.033 |

### *4.2. Factor commonality analysis*

The correlation test was performed on each influential factor, they were imported into SPSS data analysis software, and the results were outputted, and the value of KMO was obtained to be 0.711>0.7, and the P value in the Bartlett's test of sphericity was <0.001, which indicated that it was suitable and could be carried out for the next step of the factor analysis. The results of factor commonality analysis are specifically shown in Table 3. The second column of the table is the common degree of the variables under the initial solution of factor analysis, which indicates that for the original 10 variables if all 10 characteristic roots are extracted by principal component analysis, then all the variance of the original variables can be explained and the common degree of the variables is 1. In fact, the number of the factors is smaller than that of the original variables which is the purpose of the factor analysis so it is not possible to extract all the characteristic roots. In the third column of the display result data, they are the common degree when extracting the characteristic roots according to the specific extraction conditions, the vast majority of the information of all the variables can be explained by the factors, and the information of these variables is lost less, so the overall effect of this factor extraction is ideal.

**Table 3.** Factor common analysis

| Factor | Initial | Extraction |
|---|---|---|
| Maximum temperature | 1 | 0.937 |
| Mean temperature | 1 | 0.958 |
| Minimum temperature | 1 | 0.938 |
| Mean wind speed | 1 | 0.705 |
| Precipitation rate | 1 | 0.417 |
| Maximum pressure | 1 | 0.945 |
| Minimum pressure | 1 | 0.926 |
| Mean pressure | 1 | 0.943 |
| Minimum relative humidity | 1 | 0.864 |
| Mean relative humidity | 1 | 0.888 |

### *4.3. Factor Analysis Total ANOVA*

The total variance interpretation of the factor analysis is shown in Table 4, which includes the eigenvalues, variance contribution and cumulative variance contribution results of the calculated correlation matrix. The first set of data items describes the initial factor solution, and from the initial eigenvalues in column 2, it can be found that the eigenvalues of the first three factors are 6.058, 2.548, and 1.722, respectively, and the eigenvalues of the rest of the factors are less than 1. Moreover, the cumulative variance contribution ratio explained by the first three factors is 86.25%, so extracting the three factors as the main factors is appropriate. In addition, from the two sets of data, extracted square loading and rotated square loading, it can be seen that the factor variance explained by both factor extraction and rotation support the selection of the three common factors.

**Table 4.** Factor analysis variance

| Component | Total | Initial eigenvalue | | Total | Extract the sum of squares and load | |
|---|---|---|---|---|---|---|
| | | Variance (%) | Cumulation (%) | | Variance (%) | Cumulation (%) |
| 1 | 6.058 | 50.528 | 50.528 | 6.058 | 50.528 | 50.528 |
| 2 | 2.548 | 21.301 | 71.829 | 2.548 | 21.301 | 71.829 |
| 3 | 1.722 | 14.421 | 86.25 | 1.722 | 14.421 | 86.25 |
| 4 | 0.276 | 12.012 | 98.262 | - | - | - |
| 5 | 0.118 | 0.995 | 99.257 | - | - | - |
| 6 | 0.044 | 0.356 | 99.613 | - | - | - |
| 7 | 0.035 | 0.285 | 99.898 | - | - | - |
| 8 | 0.012 | 0.09 | 99.988 | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | 0.001 | 0.012 | 100 | - | - | - |
| 10 | 1.24E-05 | 0 | 100 | - | - | - |

In the process of data analysis, a graph explaining the magnitude of the contribution of the original variables can be generated at the same time, i.e., factor gravel plot, as shown in Figure 2. The factor analysis gravel plot is based on the number of factors as the horizontal coordinate and the eigenvalues as the vertical coordinate. As can be seen from the above graph, the eigenvalues of the first 3 factors are very high and contribute the most to explaining the original variables, while the eigenvalues of the factors from the 4th onwards are less than 1, indicating that their contribution to explaining the original variables is very small and almost negligible, so it is appropriate to extract the first 3 factors.

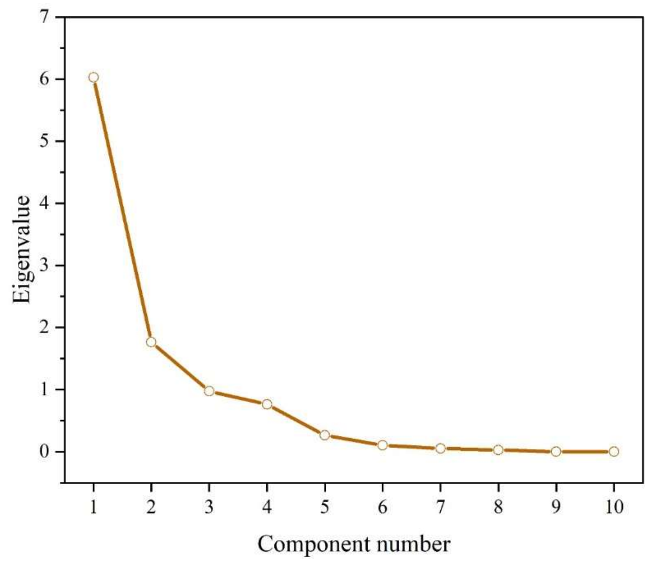



**Fig. 2.** Factor analysis lithotrripsy

### 4.4. *Component matrix analysis*

The output is the component matrix as shown in Table 5. As most of the factors are better interpreted, but there are still a small number of indicators with poorer explanatory ability, such as the "rainfall" indicator in the three factors of the loading coefficient is not much difference, so through the factor rotation method, so that the factor loading coefficients in the different factors in the load size can be differentiated, so that the results are more interpretable.

**Table 5.** Component matrix

| Component | 1 | 2 | 3 |
|---|---|---|---|
| Maximum pressure | -0.966 | -0.078 | 0.048 |
| Mean pressure | -0.965 | -0.09 | 0.052 |
| Minimum pressure | -0.956 | -0.102 | 0.068 |
| Minimum temperature | 0.947 | -0.083 | -0.184 |
| Mean temperature | 0.927 | -0.115 | -0.292 |
| Maximum temperature | 0.879 | -0.144 | -0.384 |
| Minimum relative humidity | 0.414 | 0.232 | 0.795 |
| Mean wind speed | -0.373 | 0.733 | -0.176 |
| Mean relative humidity | 0.578 | -0.037 | 0.742 |
| Precipitation rate | 0.314 | 0.346 | 0.447 |

As shown in Table 6, it can be seen that according to the 0.5 principle, the first common factor has a larger load on the variables "average temperature", "maximum temperature", and "minimum temperature", indicating that these variables mainly reflect the greater influence of temperature on the load, indicating that factor 1 comprehensively reflects their changes. The second common factor has a large load on the variables "maximum pressure", "average pressure", "minimum temperature" and "mean wind speed", which indicates that these variables mainly reflect the greater influence of air pressure on the load, indicating that factor 2 comprehensively reflects their changes. The third common factor has a large load on the variables "minimum relative humidity", "average relative humidity", and "hourly rainfall", indicating that these variables mainly reflect the greater influence of temperature on the load, indicating that factor 3 comprehensively reflects their changes. From the above analysis, it can be shown that factor 1 can be named temperature factor, and in the same way, for the second and third factors, they can be named air pressure factor and humidity factor respectively.

**Table 6.** Component matrix after rotation

| Component | 1 | 2 | 3 |
|---|---|---|---|
| Maximum pressure | 0.975 | -0.086 | 0.042 |
| Mean pressure | 0.963 | -0.092 | -0.068 |
| Minimum pressure | 0.955 | -0.085 | 0.152 |
| Minimum temperature | -0.042 | -0.916 | -0.318 |
| Mean temperature | -0.048 | -0.918 | -0.316 |
| Maximum temperature | -0.072 | -0.913 | -0.298 |
| Minimum relative humidity | -0.314 | -0.762 | -0.128 |
| Mean wind speed | 0.095 | 0.035 | 0.924 |
| Mean relative humidity | 0.276 | -0.225 | 0.875 |
| Precipitation rate | 0.124 | 0.226 | 0.896 |

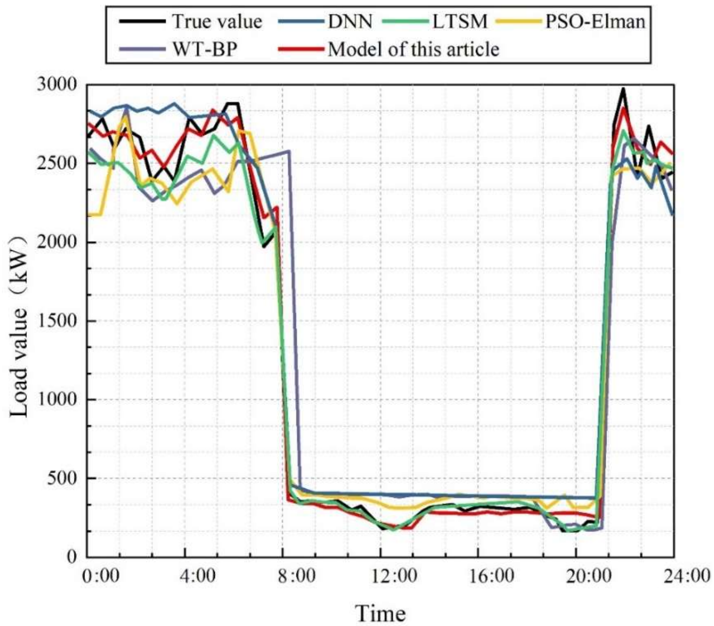
## 5. Simulation Experiment on Short-term Load Forecasting in Distribution Networks

In this chapter, simulation experiments will still be conducted based on the real data collected from the distribution network in the N-village area to further validate the practicality of the short-term load forecasting model of the distribution network proposed in this paper. The temperature factor, barometric pressure factor and humidity factor analyzed above are inputted into the short-term load prediction model of distribution network constructed in this paper to make preliminary short-term prediction, and the prediction results are compared with those of two single models, DNN and LSTM, as well as common combined prediction models, PSO-Elman and WT-BP, and the corresponding analyses are made according to the curves of the prediction results and the values of the evaluation indexes.

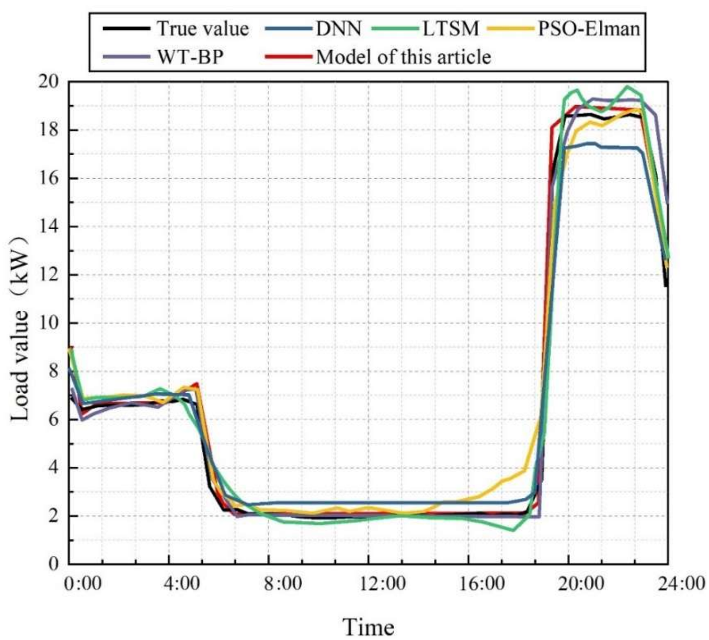
### *5.1. Short-term load forecast curve analysis*

In this paper, a large number of users' historical load data are randomly extracted from the simulation validation set, and the daily average load curve of each user in August 2024 is taken, and its distance from the clustering centers of the load classification dataset is calculated respectively to determine the classification of each user. Due to the influence of space, this paper selects four cases of typical users to show the results. The four selected typical users are NO420, NO705, NO716, and NO981, and after extracting the average daily load curves of each typical user in August, and calculating the Euclidean distance from the clustering centers of the load classification dataset, we obtain the types of the users as Category 1, Category 2, Category 3, and Category 4, respectively, and the loads of the four users on

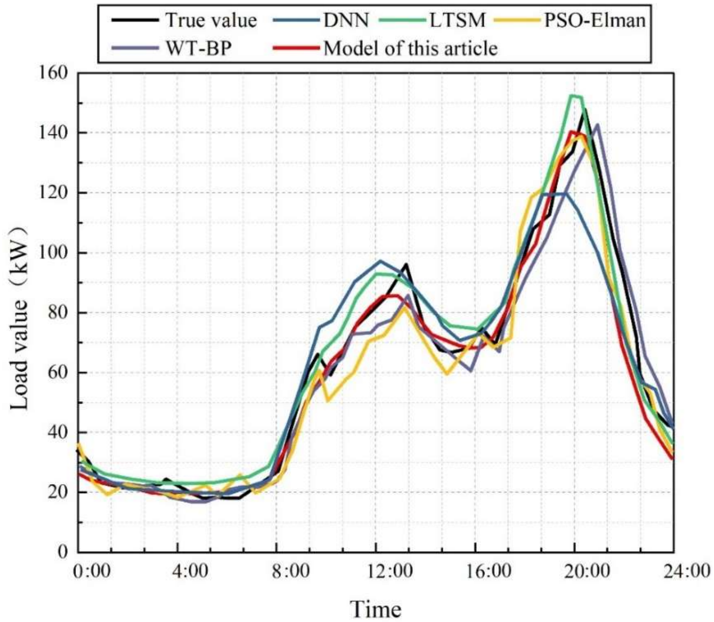
August 25 are predicted by using the algorithmic models mentioned above. Short-term prediction, the prediction results are specifically shown in Fig. 3. As can be seen from the graphs of the prediction results of the four cases of users, the three combined models, compared with a single model such as DNN and LSTM, have a stronger learning ability for the load trend, a higher degree of curve fitting, and a better control of the load mutation point. The model in this paper is less affected and has excellent stability performance. The other models have different degrees of problems in predicting different types of users, and the prediction performance needs to be improved.



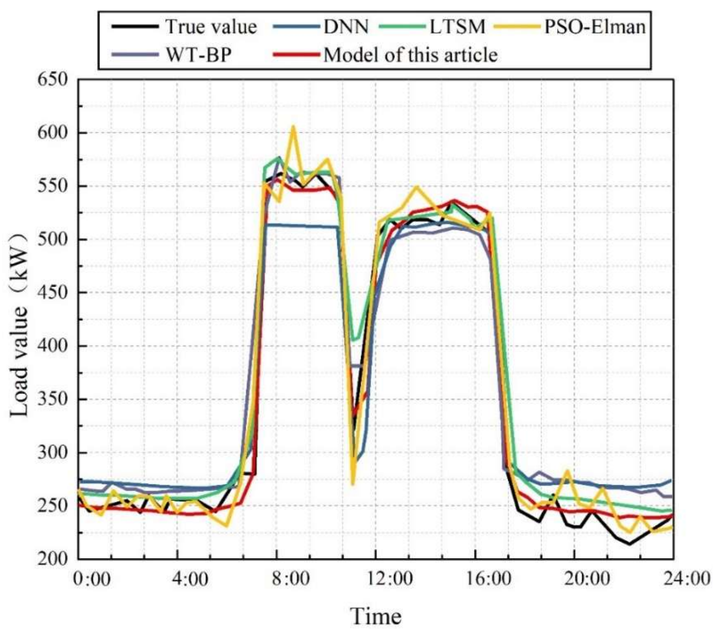

(a)NO420




(b)NO705




(c)NO716




(d)NO981

**Fig. 3.** Load short-term prediction

### *5.2. Analysis of evaluation indicators for short-term load forecasting*

Existing short-term load forecasting methods are endless, relying only on the human eye to observe the forecasting curve can not accurately derive the advantages and disadvantages of various methods, the judgment of various methods need to have certain indicators to visualize. In the above paper, this paper proposes the evaluation indexes of short-term electric load data forecasting, using the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) to evaluate. In order to highlight the superiority of this paper's model more intuitively, MRE, RMSE and MAE are selected to evaluate the prediction results, and the calculation results are specifically shown in Table 7. From the data in the table, it can be seen that the DNN model is more effective in predicting the urban residential loads such as user NO706, but it is not effective in predicting other users with high electricity consumption.The LSTM model has higher prediction accuracy for large loads, but its prediction performance is a little bit poorer for the common household loads.The PSO-Elman combination prediction model has good performance when used for short-term prediction for different types of users, but its prediction performance is not good, but its prediction performance is not good. The PSO-Elman combined prediction model has a good performance when used for short-term prediction of different types of users, but its prediction accuracy is not good enough, and seldom reaches more than 95% accuracy.The WT-BP combined prediction model converts the time-domain signals into frequency-domain signals, which effectively solves the problem of the sudden change of user loads of the distribution network, but its performance when used for the high load users is poor, and it can not comprehensively cover all types of users. The model in this paper has excellent performance when applied to the prediction of various types of user loads, and its MRE, RMSE, and MAE values are the smallest among the five methods, and the prediction accuracy can basically be maintained above 90%, which shows that the model in this paper has a certain degree of feasibility in making preliminary predictions.

**Table 7.** MRE, RMSE and MAE

| User | Model | MRE | RMSE | MAE |
|---|---|---|---|---|
| NO420 | DNN | 0.317 | 203.061 | 159.085 |
| | LSTM | 0.12 | 147.203 | 101.165 |
| | PSO-Elman | 0.246 | 225.865 | 161.748 |
| | WT-BP | 0.343 | 395.769 | 216.339 |
| | Model of this article | 0.099 | 104.69 | 79.477 |
| NO705 | DNN | 0.067 | 0.792 | 0.681 |
| | LSTM | 0.122 | 0.835 | 0.508 |
| | PSO-Elman | 0.162 | 0.911 | 0.525 |
| | WT-BP | 0.039 | 0.696 | 0.341 |
| | Model of this article | 0.045 | 0.507 | 0.219 |
| NO716 | DNN | 0.124 | 10.583 | 7.24 |
| | LSTM | 0.14 | 7.449 | 6.318 |
| | PSO-Elman | 0.106 | 7.322 | 5.806 |
| | WT-BP | 0.088 | 6.554 | 5.307 |
| | Model of this article | 0.08 | 6.712 | 4.616 |
| NO981 | DNN | 0.093 | 34.085 | 27.436 |
| | LSTM | 0.056 | 21.763 | 16.129 |
| | PSO-Elman | 0.041 | 20.58 | 15.492 |
| | WT-BP | 0.08 | 26.866 | 22.497 |
| | Model of this article | 0.027 | 17.237 | 11.468 |

Further, based on the load profiles of the above four users in October, this paper uses the load data of the first 21 days as a training set to predict the load profile on October 27, in order to verify the generalization ability of the model. In this paper, PSO-Elman and WT-BP combination models are selected for comparison experiments, and the specific simulation results are shown in Table 8 below. As the change of time and season will affect the trend of consumer electricity load, which will affect the performance of the prediction model, and the BSA combination model chosen in this paper is the best stability performance among the three combination models, the prediction errors are all below 10%, and the prediction accuracy is higher than the other two models, which verifies the high generalization ability and high accuracy of the model.

**Table 8.** Simulation results

| User | Evaluation index | Model | | |
|---|---|---|---|---|
| | | PSO-Elman | WT-BP | Model of this article |
| NO420 | MRE | 0.157 | 0.242 | 0.102 |
| | RMSE | 177.927 | 200.016 | 113.755 |
| | MAE | 90.805 | 98.724 | 68.161 |
| NO705 | MRE | 0.13 | 0.083 | 0.058 |
| | RMSE | 0.845 | 0.795 | 0.687 |
| | MAE | 0.698 | 0.382 | 0.396 |
| NO716 | MRE | 0.124 | 0.076 | 0.057 |
| | RMSE | 7.251 | 6.193 | 3.757 |
| | MAE | 4.808 | 4.31 | 3.011 |
| NO981 | MRE | 0.058 | 0.074 | 0.05 |
| | RMSE | 21.419 | 15.98 | 16.222 |
| | MAE | 14.333 | 19.659 | 10.824 |

## 6. Conclusion

Facing the high accuracy requirement of short-term load forecasting of distribution networks for the development of power marketization, this paper obtains the main influencing factors that have a greater impact on the load changes of distribution networks, and uses them as the input part of the AlexNet-GRU model to construct a short-term load forecasting model of distribution networks based on the improved AlexNet-GRU deep learning network. Influence factors of short-term load change of distribution network are analyzed for N village in Panyu District, Guangzhou City, China, and the factors that have a greater influence on load are temperature, wind speed, air pressure, and humidity through correlation analysis, and the main influence factors affecting the short-term load of the distribution network in N village are finally concluded through the analysis of the component matrix to be the temperature, air pressure, and humidity factors. The three main influencing factors are used as part of the input to the prediction model, and the simulation experiments are carried out using this paper's model with the data related to the load of the distribution network in N village as the research sample. This paper's model for NO420, NO705, NO716, NO981 four cases of users of the prediction results on the curve of the degree of fit is higher, the prediction results are more closely related to the real load values, the control of the load mutation point performance is better. The load forecasting results of this paper's model are evaluated using the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). The MRE, RMSE, and MAE values of this paper's model are lower than those of other comparative models, and it basically maintains more than 90% prediction accuracy in load forecasting, and the prediction error is always lower than 10%, which shows a good short-term load forecasting performance for distribution networks.

## References

[1] Motepe, S., Hasan, A. N., & Stopforth, R. (2019). Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms. IEEE Access, 7, 82584-82598.

[2] Van Ruijven, B. J., De Cian, E., & Sue Wing, I. (2019). Amplification of future energy demand growth due to climate change. Nature communications, 10(1), 2762.

[3] Evangelopoulos, V. A., & Georgilakis, P. S. (2022). Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks. Electric Power Systems Research, 207, 107847.

[4] Ahmad, N., Ghadi, Y., Adnan, M., & Ali, M. (2022). Load forecasting techniques for power system: Research challenges and survey. IEEE Access, 10, 71054-71090.

[5] Nti, I. K., Teimeh, M., Nyarko-Boateng, O., & Adekoya, A. F. (2020). Electricity load forecasting: a systematic review. Journal of Electrical Systems and Information Technology, 7, 1-19.

[6] Li, K., Huang, W., Hu, G., & Li, J. (2023). Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. Energy and Buildings, 279, 112666.

[7] Tang, X., Dai, Y., Wang, T., & Chen, Y. (2019). Short-term power load forecasting based on multi-layer bidirectional recurrent neural network. IET Generation, Transmission & Distribution, 13(17), 3847-3854.

[8] Motepe, S., Hasan, A. N., Twala, B., & Stopforth, R. (2019, April). Power distribution networks load forecasting using deep belief networks: The south african case. In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) (pp. 507-512). IEEE.

[9] Wang, B., Mazhari, M., & Chung, C. Y. (2022). A novel hybrid method for short-term probabilistic load forecasting in distribution networks. IEEE Transactions on Smart Grid, 13(5), 3650-3661.

[10] Grabner, M., Wang, Y., Wen, Q., Blažič, B., & Štruc, V. (2023). A global modeling framework for load forecasting in distribution networks. IEEE Transactions on Smart Grid, 14(6), 4927-4941.

[11] Jiang, H., Ding, F., & Zhang, Y. (2017, July). Short-term load forecasting based automatic distribution network reconfiguration. In 2017 IEEE Power & Energy Society General Meeting (pp. 1-5). IEEE.

[12] Wang, Y., Guo, P., Ma, N., & Liu, G. (2022). Robust wavelet transform neural-network-based short-term load forecasting for power distribution networks. Sustainability, 15(1), 296.

[13] Li, J., Feng, Y., & Hu, Y. (2021, October). Load forecasting of 5G base station in urban distribution network. In 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2) (pp. 1308-1313). IEEE.

[14] Ge, L., Xian, Y., Wang, Z., Gao, B., Chi, F., & Sun, K. (2020). Short-term load forecasting of regional distribution network based on generalized regression neural network optimized by grey wolf optimization algorithm. CSEE journal of power and energy systems, 7(5), 1093-1101.

[15] Fan, M., Hu, Y., Zhang, X., Yin, H., Yang, Q., & Fan, L. (2019, November). Short-term load forecasting for distribution network using decomposition with ensemble prediction. In 2019 Chinese Automation Congress (CAC) (pp. 152-157). IEEE.

[16] Motepe, S., Hasan, A. N., Twala, B., Stopforth, R., & Alajarmeh, N. (2019, August). South African power distribution network load forecasting using hybrid AI techniques: ANFIS and OP-ELM. In 2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) (pp. 557-562). IEEE.

[17] Zhu, D., Chen, B., Yang, S., Li, G., Feng, H., & Yao, Z. (2023, November). A load forecasting method for distribution networks considering the influence of photovoltaics and traditional load. In 2023 6th International Conference on Power and Energy Applications (ICPEA) (pp. 195-199). IEEE.

[18] Mahto, D. K., Saini, V. K., Mathur, A., & Kumar, R. (2022, August). Deep recurrent mixer models for load forecasting in distribution network. In 2022 IEEE 2nd International Conference on Sustainable Energy and Future Electric Transportation (SeFeT) (pp. 1-6). IEEE.

[19] Gong, Z., Zhang, Z., Wu, M., & Ni, H. (2019, October). Distribution Network Load Forecasting Based on Improved BP Neural Network. In 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM) (pp. 144-147). IEEE.

[20] Meng, X. F., Zhang, S. X., Wang, J., & Wu, S. H. (2016). Load forecasting method considering temperature effect for distribution network. In MATEC Web of Conferences (Vol. 63, p. 05028). EDP Sciences.

[21] E. Yu. Myasnikov,V. I. Antonov,A. V. Soldatov & R. V. Razumov. (2024). Confirmation of Electrical Network Configuration from Telemetry Data Based on Convolutional Neural Networks. Russian Electrical Engineering(8),617-624.

[22] Guo Yong & Yang Li Dong. (2022). Radar Moving Target Detection Method Based on SET2 and AlexNet. Mathematical Problems in Engineering.

[23] Meng Xiangfei,Zhu Tanbo & Li Changcheng. (2022). Construction of perfect dispatch learning model based on adaptive GRU. Energy Reports(S5),668-677.

[24] Maheshwari Sudhanshu. (2023). Analog Compatible Logic Functions Using EXCCIIs and an Extended Application for Sigmoid Activation Function. Journal of Circuits, Systems and Computers(09).