International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7650-7662

https,//doi.org/10.70517/ijhsa463656

Research on Short-term Load Forecasting Techniques for Distribution Networks Based on Time Series Analysis

Zhou Wang[1,*], Junqiang Mu[2], Tianting Li[1] anf 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
Corresponding authors: (e-mail: ZhouWang2024@126.com).

Abstract Power system load forecasting is a key aspect of grid scheduling and operation, which is affected by factors such as national policies, population growth, seasonal changes, and weather changes. In this paper, a prediction model based on improved AlexNet-GRU is proposed to address the problem of short-term load prediction accuracy in distribution networks. Firstly, the basic principles and characteristics of power system load forecasting are analyzed, and the load data processing methods, including abnormal data correction and missing data completion, are studied. Then the AlexNet network in the field of image recognition is improved into a one-dimensional convolutional structure, and combined with the GRU network to construct the prediction model, fully utilizing AlexNet's ability to extract complex features and GRU's advantage of processing time-series data. The analyses of the algorithms show that the model reduces the average absolute percentage error MAPE by 1.082%, 1.314%, 1.939%, and 2.323%, and improves the average prediction accuracy by 1.085%, 1.236%, 1.876%, and 2.223%, respectively, when compared with the CNN-GRU, GRU, LSTM, and RNN models, during the consecutive six-month test in a province of Southwest China, 2.223%. In the validation of a regional dataset in Australia, the mean absolute error MAE is reduced by 22.77 MW and the root mean square error RMSE is reduced by 18.48 MW compared with the CNN-GRU model.The experimental results show that the proposed model can effectively improve the accuracy and stability of short-term load forecasting of the distribution network, and provide decision-making support for the safe and economic operation of the power system.

Index Terms Distribution network, Short-term load forecasting, Improved AlexNet-GRU, Feature extraction, Time series data, Prediction accuracy

I. Introduction

The most important task of the power system is to provide power users with high-quality, stable power [1]. The prerequisite for the realization of this task is that the power operation, scheduling and planning departments should understand the users' demand for electricity, the law and the trend of change, so as to formulate a reasonable marketing plan and development strategy, realize the optimal allocation of social resources, and improve the social and economic benefits [2]-[5]. Short-term load forecasting of distribution networks is an important technology for the economic and efficient operation of distribution networks. Short-term load forecasting is mainly based on the distribution network historical operation data, external meteorology, and time and other holidays and other information, mining the distribution network load changes in the law and influencing factors, and inferring the trend of the load changes in the future short-term period of time [6]. Accurate prediction of distribution network short-term load can provide effective auxiliary decision-making for distribution system planning and design, scheduling and operation, energy efficiency management and demand response, which is of great significance for reducing power generation cost and improving the level of fine management and operation of distribution network [7]-[9].

Due to the complexity and variety of load types included in the distribution network, the short-term fluctuation and change rules of each type of load are different [10]. In addition, influenced by seasonal changes, meteorological changes and holidays and other characteristic factors, each type of load fluctuation also has uncertainty [11], [12]. With the increase in the types of distribution system loads and the introduction of bi-directional flexible load equipment such as electric vehicles into the grid, the nonlinear complexity of distribution system loads has gradually increased, and the difficulty of short-term load forecasting has also been increasing [13]-[15]. Traditional short-term load forecasting methods mainly use trend analysis, multiple linear regression method and autoregressive sliding average method, etc. Such forecasting methods are more effective for smoothly changing short-term loads, but the forecasting effect is not good in the case of frequent load fluctuations [16]-[18]. Therefore, it has become the focus and difficulty of short-term load forecasting research to comprehensively consider the characteristics of the law affecting load changes in the time dimension so as to improve the forecasting accuracy and effect [19], [20].

Power system load forecasting is the fundamental work of power system planning and operation, which has an important impact on the economy, security and reliability of the power system. Accurate short-term load forecasting can guide the power grid to reasonably arrange the power generation plan, optimize the scheduling strategy, reduce the waste of electric energy, and lower the system operating costs. Under the current power market environment, the accuracy of load forecasting is directly related to the economic benefits and market competitiveness of power enterprises. However, power load is affected by many factors, such as weather changes, customer behavior, economic activities, holiday effects, etc., showing complex nonlinear and variable characteristics, which brings challenges to accurate forecasting. Traditional load forecasting methods mainly include statistical methods and artificial intelligence methods, but these methods often suffer from insufficient prediction accuracy and limited generalization ability when facing the complex and variable actual power system. Especially for the distribution network, a system with diverse user behaviors and large load fluctuations, it is difficult for traditional forecasting methods to meet the demand for high-precision forecasting. In recent years, deep learning technology has achieved remarkable results in the field of time series prediction, and its powerful feature extraction and pattern recognition capabilities provide new ideas for load prediction. Convolutional neural networks (CNNs) are excellent in feature extraction, while gated recurrent units (GRUs) have a unique advantage in processing time series data, but it is still difficult for a single model to fully explore the complex spatio-temporal dependencies in load data.

In this study, we propose a novel improved AlexNet-GRU hybrid prediction model, which transforms the AlexNet network originally used for image recognition into a one-dimensional convolutional structure suitable for load time-series data, reduces parameter redundancy by reducing the number of output nodes in the convolutional layer, adjusting the size of convolutional kernel, optimizing the fully-connected layer, etc., and adopts Adam's optimizer instead of the SGD algorithm to improve the computational efficiency. The model combines the improvement of AlexNet and the SGD algorithm. The model combines the powerful feature extraction capability of AlexNet and the advantage of GRU in capturing timing dependencies, forming a two-layer architecture of “feature extraction and timing analysis”. The model is validated with provincial power data in Southwest China and public datasets in Australia to evaluate the prediction effect of the model in different time periods and seasons, and compared with mainstream prediction models such as CNN-GRU, GRU, LSTM, RNN, etc., to validate the prediction accuracy and stability of the proposed model in terms of multiple evaluation indexes.

II. Technical basis for load forecasting in power systems

II. A. Power System Load Forecasting

Power system load forecasting is a complex and delicate work, in the process of power system load forecasting, it will not only be affected by the system load's own changing condition, but also affected by many factors such as national policy, population growth, seasonal change, weather change, economic development speed and so on [21], [22].

Power system load forecasting steps are:

(1) Collect historical load data of the power system.

(2) Analyze the history of power load and find out the development pattern.

(3) Establish a forecasting model to forecast the future load.

II. A. 1) Basic principles

The accuracy of short-term load forecasting is an important basis for arranging the starting and stopping of units. Under the current environment of power market intelligence, short-term load forecasting enables power generation companies to formulate reasonable maintenance plans and power generation plans according to the power demand of users. At the same time, short-term load forecasting can also enable the power supply company to fully understand the power demand of users and make reasonable power purchase plan. For transmission companies, short-term load forecasting can be used to make reasonable power generation plans and arrange the operation mode of the system. In short-term load forecasting, the following basic principles should be observed:

(1) Principle of Knowability

(2) Principle of probability

(3) Principle of inertia

(4) Principle of similarity

(5) Systematic principle

(6) The principle of feedback

II. A. 2) Characterization

The characteristics of load forecasting are specified below:

(1) Inaccuracy

Power system load is easily affected by many factors (e.g., temperature, season) in the process of forecasting, and the influencing factors related to it are also developing and changing. In addition, it may also be affected by unexpected events, so that the prediction results have inaccuracy.

(2) Conditionality

Conditionality refers to the fact that when forecasting future power load data, the influence of various factors must be taken into account. And the said conditions, through can be divided into can reliably and directly affect the load prediction results of the inevitable conditions and due to the development of the future load, change regularity is difficult to control, in order to get more accurate prediction, need to add some of the assumption conditions.

(3) Temporality

Temporality refers to the power system load forecasting is completed within a certain time frame, that is, the power system load forecasting is required to have real-time. And when doing power system load forecasting, it should specifically indicate the forecast time corresponding to the forecast value and the time range of this forecast.

(4) Multi-program

Because of the inaccuracy and conditionality of load forecasting, under the condition of comprehensive consideration of various influencing factors, different forecasting methods should be adopted for loads with different characteristics and different mathematical models should be established to make load forecasting have a high degree of accuracy.

II. A. 3) Forecasting characteristics

Based on the previous analysis of load characteristics, the total power system load forecasting model can generally be described by the following equation:

$$y(t)=b(t)+w(t)+s(t)+v(t) \tag{1}$$

where $y(t)$ is the total load at moment t. $b(t)$ is the classical load, which refers to the basic normal load component at moment t. $w(t)$ is the weather-influenced load component at moment t, and $s(t)$ is the load component at moment t influenced by the day's date type. $v(t)$ is the load component increased by the influence of a special event at moment t.

In general, the occurrence of special events is random and unpredictable, so in this paper, only typical loads, loads increased by weather changes and date types are considered for load forecasting.

II. B.Requirements and steps for short-term load forecasting

II. B. 1) Steps in short-term load forecasting

Short-term load forecasting is based on the historical data of the load, the establishment of a mathematical model of load forecasting as the core, and the purpose of obtaining the optimal forecast results. The basic process of short-term load forecasting for power systems is shown in Figure 1.

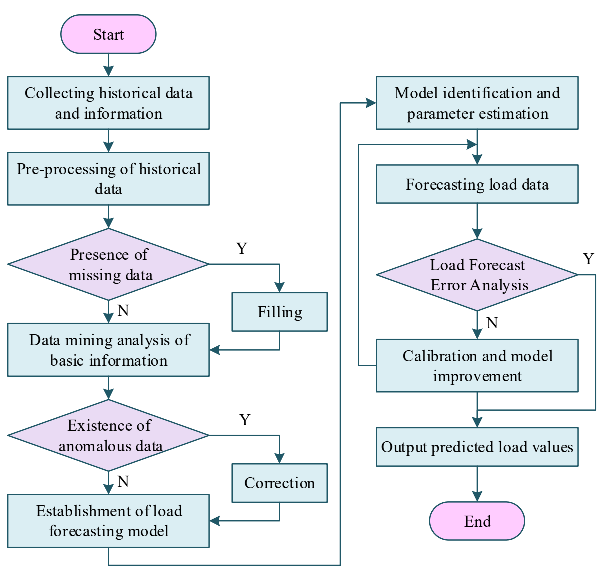


Figure 1: Basic process of short-term load prediction of power system

II. B. 2) Evaluation indicators for forecasting models

The deviation between the forecasted and actual load values is collectively referred to as the forecast error. The ratio between the predicted and actual values is collectively referred to as the load prediction accuracy. Usually, it is desired that the accuracy rate is as high as possible, and the higher the accuracy rate. Then it means the closer the prediction value is to the actual value. A reasonable analysis of the prediction error can improve the accuracy of the prediction, and at the same time, it can also provide an effective reference. Therefore, some form of error is often used as an index to evaluate the effectiveness of the prediction model, mainly as follows:

(1) Absolute error (AE) is the absolute value of the prediction error. That is:

$$AE = |y - \hat{y}| \tag{2}$$

(2) Relative error (RE), also known as the error rate, is the ratio of the absolute error to the measured value. That is:

$$RE = \frac{|y - \hat{y}|}{y} \times 100\% \tag{3}$$

(3) Mean Absolute Error (MAE) is an average of the sum of the absolute values of the errors between the predicted and measured values. To wit:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

(4) The mean relative error (MAPE) is the average of the absolute value of the prediction error between the predicted and measured values divided by the sum of the measured values. I.e:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \tag{5}$$

(5) Mean Square Error (MSE) is the average of the sum of squared prediction errors. To wit:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

(6) Root Mean Square Error (RMSE) is the positive square root of the mean square error. To wit:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \times 100\% \tag{7}$$

II. C.Load data processing

Data is the raw material of neural network modeling, the original data is diverse, the data quality is uneven, it will seriously affect the network's perception of the law of change, affecting the training effect. So before inputting the data into the neural network for training, the data need to be preprocessed, and data preprocessing is the most troublesome and important part of the modeling process. The data preprocessing process includes removing attribute data, bad data correction, data standardization, adding influencing factors and data reorganization.

Attribute data refers to the unique characteristic data of the target station area, such as the station ID, the belonging prefecture and city, the station capacity and other data that indicate the characteristics of the station area. These data are indispensable for operations such as identifying station areas, classifying station areas, and extracting load data, but they are not the characteristic data that affect short-term load forecasting, and they have no direct effect on training the model. So it is sufficient to delete these data directly.

Undesirable data mainly includes 2 categories of abnormal data and missing data, and the following will introduce the processing methods of abnormal data and missing data appearing in the power load data respectively.

In the historical power load data, the influence of some random factors will produce some deviations from the larger values, these values are mixed in the normal load data, increasing the randomness of the load data, which will have an impact on the prediction accuracy of the model, and need to deal with abnormal data. Define data w as abnormal data, the criterion formula is as follows equation:

$$\begin{cases} I_{QR} = Q_3 - Q_1 \\ w > Q_3 + 1.5I_{QR} \\ w < Q_3 - 1.5I_{QR} \end{cases} \tag{8}$$

where Q_1 and Q_3 are the first and third quartiles, respectively. I_{QR} is the interquartile range.

Correction of outliers is a mathematical method to correct the outliers in the raw load data so that they can be used for model training. The correction methods include horizontal and vertical processing methods. Because the load randomness of the distribution substation area is strong and the load mutation is more frequent, the horizontal processing method is not applicable to the abnormal value processing with the distribution substation area, and this paper adopts the vertical processing method.

Vertical processing method: the load data of the distribution station area has a strong periodicity, so the load data of the same date type should have a high similarity, and calculate the difference between the load values of similar day types. If the difference between similar day type data exceeds a certain range, this load data is abnormal data, abnormal data can be modified according to the following rules, the calculation formula is as follows:

$$|Y(d,t) - M(t)| > \varepsilon(t) \tag{9}$$

$$Y(d,t) = Y(d,t) \pm \varepsilon(t) \tag{10}$$

where $\varepsilon(t)$ is the set threshold and $M(t)$ is the average value of the last few days in the historical data.

Due to various reasons in the process of data collection will occur in the case of missing data, and even there will be a continuous multi-day missing data, at this time if the above horizontal or vertical processing can not be used to make up the data, but will interfere with the model training. At this time, the normal data of recent days or the same day type of the missing day should be selected for curve fitting, and then the missing data should be replaced by the fitted curve to make up the data. Assume that the fitted curve is in the following equation:

$$f = g(y, a_1, a_2, \cdots a_n) = a_1 g_1(y) + a_2 g_2(y) + \cdots + a_n g_n(y) \tag{11}$$

Let $g_1(y) = 1$, $g_2(y) = y$, $g_3(y) = y^2$, and so on.

First bring the normal data y_k for the last few days or the same day type into the following equation:

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} a_i g(y_1) \\ \sum_{i=1}^{n} a_i g(y_2) \\ \vdots \\ \sum_{i=1}^{n} a_i g(y_n) \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{12}$$

Apply the principle of least squares as in the following equation:

$$\sum_{k=1}^{m} \delta_k^2 = \Delta_{\min} \tag{13}$$

and the concept of extreme value is given in the following equation:

$$\sum_{k=1}^{m} \delta_k \frac{\partial \delta_k}{\partial a_i} = 0 \tag{14}$$

Solving for the coefficient $a_1, a_2 \cdots a_n$ is a reasonable fit to the curve.

Data normalization is the scaling of data to a specified range. The input data in the load forecasting modeling process have different scales and ranges, resulting in different orders of magnitude and no comparability of the indicators. For data that cannot be used directly because of different data types, the data needs to be quantized into machine-recognizable data first. Data with different scales need to be standardized in order to unify the scales.

III. Presentation and testing of load forecasting methods

III. A. Short-term load forecasting methods for distribution networks

III. A. 1) Gated Recirculation Unit (GRU)

The gated recurrent unit (GRU) network can not only overcome the problems of “gradient vanishing” and “gradient explosion” encountered by traditional recurrent neural networks (RNNs) when dealing with long time series information, but also solve the computational inefficiency of long short-term memory networks (LSTMs) due to the large number of parameters.

The training and prediction efficiency of GRU model is significantly improved in the case of large sample size, and thus it is widely used in the prediction field. Although the basic principles of the GRU model and the LSTM model are similar, the GRU model is optimized on the basis of the LSTM model to simplify the model structure.

GRU uses only two gating structures, the reset gate and the update gate.By reducing the number of gating mechanisms, the GRU model results in fewer model parameters and higher computational efficiency.The GRU model not only retains the advantages of the LSTM model in capturing the long- and short-term dependencies of the time series, but also improves the performance of the model in dealing with the long time series data.

The inputs of the reset gate are the hidden layer h_{t-1} of the previous moment and the input x_i of the current moment, and the input value is obtained by the sigmoid function as r_t, whose value is determined to be in the range of $[0,1]$, and the smaller the value is indicates that less information is passed to the hidden layer of the current moment. The reset gate function expression is:

$$r_t = \sigma\{W_r(h_{t-1}, x_t) + b_r\} \tag{15}$$

where $\sigma(\cdot)$ is the activation function sigmoid. W_r is the weight matrix of the update gate. b_r represents the reset bias term.

The input of the update gate is the hidden layer h_{t-1} at the previous moment and the input x_i at the current moment. The input value of z_t is obtained through the sigmoid function, which takes the value of $[0,1]$, and the smaller the value is, it means that it is difficult to pass the information in the hidden layer of the previous moment to the hidden layer of the current moment.

Update the gate function expression as:

$$z_t = \sigma\{W_z(h_{t-1}, x_t) + b_z\} \tag{16}$$

where W_z is the weight matrix of the update gate. b_z represents the bias term of the update gate.

III. A. 2) CNN-GRU model construction

In this paper, we propose a forecasting model that combines CNN model and GRU model, aiming to improve the accuracy and reliability of short-term load forecasting.The CNN-GRU model integrates the advantages of both CNN model and GRU model, where the CNN model is used to extract the important features from the time-series data, and the GRU model handles the long- and short-term temporal dependencies, which enhances the model's ability to complex deformation pattern capturing ability [23].

The operation process of the CNN-GRU short-term load forecasting model for distribution networks starts with data input, which is processed through data segmentation and normalization to ensure data consistency. Next, feature extraction is performed by the CNN model to fully capture the load information in the data. Then, temporal analysis is carried out using GRU model to dig deeper into the temporal dependencies in the data. Finally, training and prediction are carried out through the fully connected layer in order to obtain the prediction results of short-term loads in distribution networks.

III. A. 3) AlexNet Neural Network Main Components

(1) Convolutional layer

The convolutional layer is an essential component of a convolutional neural network, whose input x is convolved with some trainable multidimensional convolutional kernel f_k, and then the resulting result is added to the deviation b_k. Assuming that there are K convolutional kernels, the k th output of the layer can be expressed by the following equation:

$$y_k(i,j) = \sum_{c=1}^{C} f_k^c * x^c(i,j) + b_k \tag{17}$$

where the capital letter C denotes the total number of channels in the input, and $x^c(i,j)$ denotes the region in the i th row and j th column of the c th channel in the input.

(2) Activation layer

The activation layer, also known as the modified linear unit, serves to increase the learnability of the network and diversify the internal structure of the network.All the activation functions used in AlexNet are ReLU functions with the following formulas:

$$y_k(i,j)=\begin{cases} x_k(i,j), & x_k(i,j)\geq 0 \\ 0, & x_k(i,j)<0 \end{cases} \tag{18}$$

(3) Pooling layer

The pooling layer is used to reduce the size of the feature matrix, if pooling is not taken, the problem of learning parameters too large will occur in the training process. The pooling selected in this paper are all maximum pooling, the formula is as follows:

$$y_k(I,J)=\max\left(x_k\left(I+i-1,J+j-1\right)\right) \tag{19}$$

where I and J denote the i th row and j th column after output.

(4) Linear layer

Linear layer, also known as fully connected layer, the main purpose is to flatten the multi-dimensional array left by the previous convolution into one-dimensional data, playing a role in reducing the dimensionality of the data.

III. A. 4) Improvement of the AlexNet-GRU prediction model

Combining AlexNet network with GRU makes full use of AlexNet network's ability to extract complex features and GRU's expertise in processing time-series data. However, the AlexNet network is proposed to be used in the field of image recognition and is a two-dimensional convolution (2D-CNN), which cannot be directly used in load prediction models, and thus needs to be improved.

The classical AlexNet convolutional neural network has a strong ability to discriminate and recognize images, but this paper is to apply the migration of AlexNet in the field of image recognition to the field of short-term load forecasting, therefore, targeted improvements are to be made to AlexNet, mainly to enhance the ability of the forecasting model to extract features from complex data, and also to combine it with the GRU network. In order to adapt to the characteristics of the samples, 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

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 two-dimensional convolutional layer (Conv2d) in the network is replaced with a one-dimensional convolutional layer (Conv1d) to facilitate the processing of time series data.

Through a large number of experiments, the network structure is improved as follows: reduce the number of output nodes of the convolutional layer, and change the size of some convolutional kernels, so that the output nodes of all one-dimensional convolutional layers are reduced to 32, the size of the convolutional kernel is set to 3, and the pooling operation adopts the form of maximal pooling, and at the same time, the size of the pooling operation's sliding window is set to 2, and the next output nodes of the first layer of the fully-connected layer are reduced to 192, the second fully-connected layer of the output nodes is reduced to 192. The number of output nodes in the first fully connected layer is reduced to 192, the number of output nodes in the second fully connected layer is reduced to 96, and the number of output nodes in the third fully connected layer (output layer) is reduced to 96. By improving the structure of the network, the redundancy of the parameters is reduced, which accelerates the convergence speed of the model and shortens the training time.

(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., the natural language and computer vision problems) and alleviate the problem of gradient oscillations.

Let the parameters to be optimized be w, the loss function be $loss$, the learning rate be lr, the decay rates of the first- and second-order momentum be β_1 and β_2, and take the moment of t as the current moment, and the moment of $t+1$ as the next, and the specific training process for Adam is described as follows:

Find the gradient value g_t of the loss function with respect to the covariates at the current moment at this point in time:

$$g_t = \nabla loss = \frac{\partial loss}{\partial w_t} \tag{20}$$

Derive the first-order momentum m_t and the second-order momentum V_t at the current moment:

$$m_t = \beta_1 \cdot m_{t-1} + (1-\beta_1) \cdot g_t \tag{21}$$

$$V_t = \beta_2 \cdot V_{t-1} + (1-\beta_2) \cdot g_t^2 \tag{22}$$

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

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{23}$$

$$\hat{V}_t = \frac{V_t}{1-\beta_2^t} \tag{24}$$

Calculate the gradient of descent at that moment η_t:

$$\eta_t = lr \cdot \frac{\frac{m_t}{1-\beta_1^t}}{\sqrt{\frac{V_t}{1-\beta_2^t}}} \tag{25}$$

Calculates parameter w_{t+1} for the next moment, thus completing the update:

$$w_{t+1} = w_t - \eta_t = w_t - lr \cdot \frac{\frac{m_t}{1-\beta_1^t}}{\sqrt{\frac{V_t}{1-\beta_2^t}}} \tag{26}$$

After replacing the SGD algorithm with the Adam optimizer algorithm, the computational efficiency is higher, the memory requirement is less, and after bias correction, the learning rate has a definite range for each iteration, making the parameters smoother.

Based on the above theoretical knowledge of AlexNet and GRU, a short-term load forecasting model based on improved AlexNet-GRU combination is proposed.

The model load data, meteorological data, and different clustering identifiers are jointly used as inputs, and AlexNet1d is firstly constructed by combining the AlexNet network structure to extract the complex features hidden in the multidimensional sample data. The extracted features are then fed into the GRU layer to study the dynamic change laws within the time-series data. The temporal dependence between the data is learned to reduce the loss of historical information, and an arithmetic example is given to analyze it with the data of the actual power grid system. Meanwhile, it compares with other methods from multiple perspectives, evaluates the forecasting effect of the model more objectively, and finally completes the load forecasting.

III. B. Example analysis

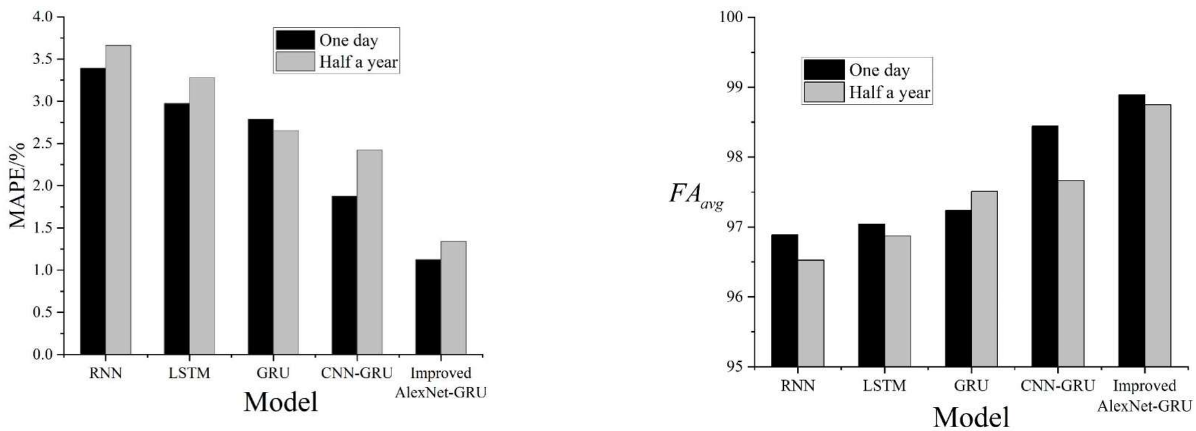
III. B. 1) Analysis of forecast results for different time periods

In order to verify the accuracy of the improved AlexNet-GRU network model prediction method proposed in this paper, an arithmetic case analysis is performed. The test samples use a total of three years of electricity load data from a province in Southwest China, with a collection period of 20 min and 72 points collected in one day, and the prediction results are compared with CNN-GRU, GRU, LSTM and RNN models to verify the reliability of the proposed model.

The evaluation metrics of the improved AlexNet-GRU short-term load forecasting method proposed in this paper for the load forecasting results in the test set data for an individual day and within a consecutive half-year period are shown in Fig. 2, with the MAPE/% value and FA_{avg} value in Fig. (a) and Fig. (b), respectively, FA_{avg} denoting the average of the forecasting accuracies.

The analysis shows that the improved AlexNet-GRU model has a higher prediction accuracy than the other comparison models, both in the day-ahead short-term load prediction on a single day and in six consecutive months.

In the consecutive six-month test, the improved AlexNet-GRU method reduces the MAPE metrics by 1.082%, 1.314%, 1.939%, and 2.323%, respectively, compared to the CNN-GRU, GRU, LSTM, and RNN4 class prediction models. The FA_{avg} metrics improved by 1.085%, 1.236%, 1.876%, and 2.223%, respectively. It shows that the GRU model has a better prediction effect in dealing with time series problems and reflects the necessity of optimizing its parameters.



(a) MAPE index (b) Mean of prediction accuracy

Figure 2: The evaluation index of the load prediction results

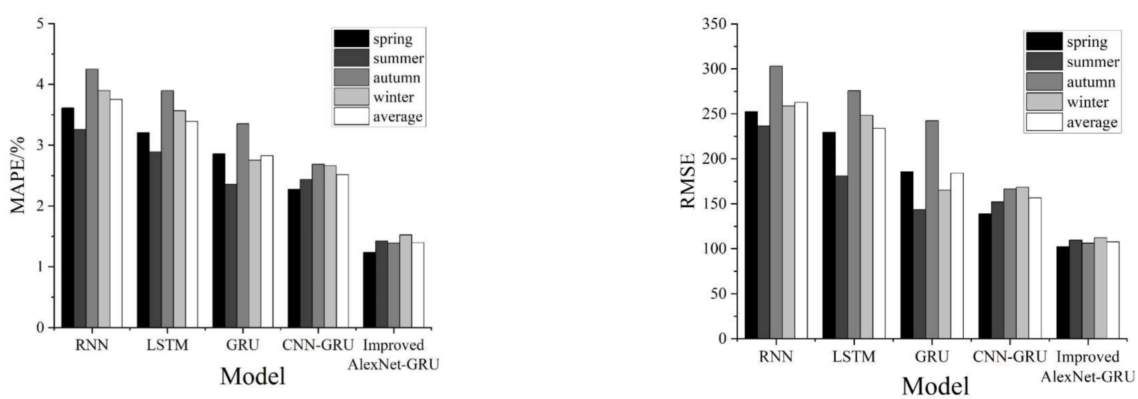
III. B. 2) Comparison of forecasting results by season

Comparison of the forecasting results of different models for different seasons is shown in Fig. 3, Fig. (a) and Fig. (b) show the MAPE/% values and RMSE values of different models, respectively.

The figure summarizes the average error of the improved AlexNet-GRU forecasting model proposed in this paper with other methods for load forecasting in four seasons in the test set data.

It can be seen that the method proposed in this paper has the lowest prediction error in different seasons of the year, and the average prediction accuracy of the four seasons reaches more than 98%.

Compared with the RNN, LSTM, GRU, and CNN-GRU models, the MAPE indexes are reduced by 2.36%, 1.996%, 1.435%, and 1.121%, respectively, and the RMSEs are reduced by 155.047, 126.055, 76.568, and 48.925, respectively.A comprehensive analysis of the experimental data demonstrates that the proposed model has a higher performance in short-term load forecasting for the grid. The comprehensive analysis of the experimental data shows that the model proposed in this paper has higher prediction accuracy and stability in the short-term load forecasting of the power grid.



(a) MAPE value (b) RMSE value

Figure 3: Comparison of different seasonal forecasts of different models

III. B. 3) Actual and predicted values and absolute percentage error

The actual and model predicted values are shown in Fig. 4. The figure shows the actual values, predicted values of monitoring load at the first 20min of each hour of the day for the improved AlexNet-GRU algorithm model. It can be seen that the predicted results of the improved AlexNet-GRU algorithm are in line with the trend of the actual values, and the actual values are similar to the model predicted values.

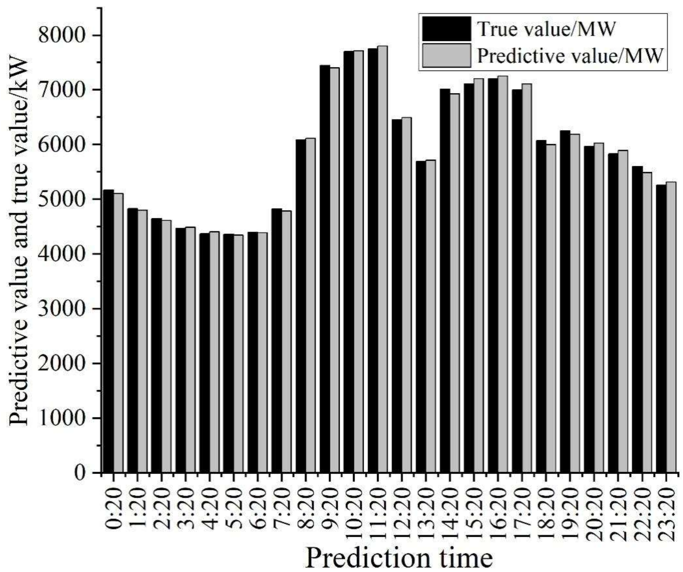


Figure 4: Actual value and model predictive value

The absolute percentage errors between the actual values and the predicted values of the model are shown in Fig. 5, and the results show that the average absolute percentage error (MAPE) and the maximum absolute percentage error (MAPE) of the prediction results of the proposed model are 1.239% and 3.266%, respectively, for the 24 monitoring points.

Compared with the CNN-GRU model, the MAPE is reduced by 1.192%, which proves that the proposed improved AlexNet-GRU prediction model has a higher prediction accuracy than the CNN-GRU model in terms of both the average error and the maximum error.

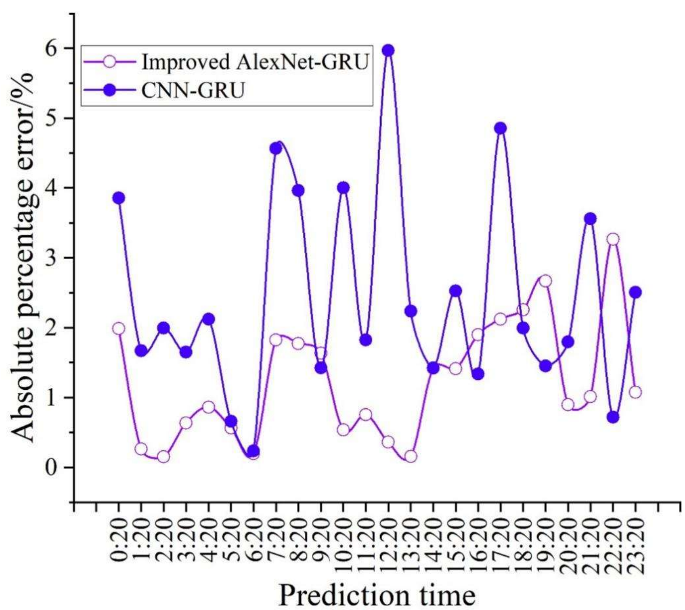


Figure 5: The absolute percentage error between the actual value and the predicted value

III. B. 4) Validation of a dataset for a region in Australia

In order to verify the effectiveness of the improved AlexNet-GRU prediction method, the more common SVR prediction models, BP neural network, LSTM deep neural network, GRU deep neural network, and CNN-GRU neural network are selected for comparison. In order to more accurately reflect the prediction performance between the models, all model parameters follow the principle of the control variable method, and the model parameters of each prediction method are determined separately in the following.

SVR is an extension of SVM application in the field of prediction, which is a classical machine learning regression model, and the SVR model parameters are selected: the kernel function is Gaussian kernel function, the kernel parameter is 0.001, the penalty factor is 0.001, and the training period is 50.

BP neural network as a traditional neural network algorithm, although its structure contains only two parts of the forward propagation of data and the backward propagation of error, but its function is very powerful, and it is widely used in the field of prediction and recognition.The parameter selection of BP neural network model is shown in Table 1.

Table 1: Parameter selection of bp neural network model

Parameter name	Set value
Training cycle	50
Hidden layer unit	50
Training algorithm	Trainlm algorithm
Loss function	MSE
Learning rate	0.001

LSTM and GRU neural networks are variant models of the RNN model, and they are both deep learning models that have their unique advantages in handling large amounts of data as well as time series prediction.

The parameter selection of LSTM and GRU models is shown in Table 2.

Table 2: Parameter selection of LSTM andGRU model

Parameter name	Set value
Training cycle	50
Hidden layer unit	50
Training algorithm	Adam
Loss function	MSE
Learning rate	0.001

The CNN-GRU model is a one-dimensional CNN applied to a GRU network, where a one-dimensional CNN is a convolutional neural network that does not include a pooling layer, and the parameters of the CNN-GRU model are chosen as shown in Table 3.

Table 3: Parameter selection of the model

Parameter name	Set value
Training cycle	50
Hidden layer unit	50
Training algorithm	Adam
Loss function	MSE
Learning rate	0.001
CNN filter number	50
Convolution kernel dimension	3

In order to validate the effectiveness and superiority of the improved AlexNet-GRU forecasting method, the training and forecasting of electricity loads are performed on a publicly available dataset for the Australian region.

The sampling frequency of this dataset is 30 minutes, and the data from January 2022 to June 2023 is used as the training set to forecast the load in July 2024, and comparison experiments with SVR, BP, LSTM, GRU, and CNN-GRU models are conducted.The results of the comparison of load forecasting from July 5-7 are shown in Fig. 6.

From the figure, it can be seen that the prediction results of the improved AlexNet-GRU prediction model are closest to the actual load data, indicating that the method has the best prediction effect.

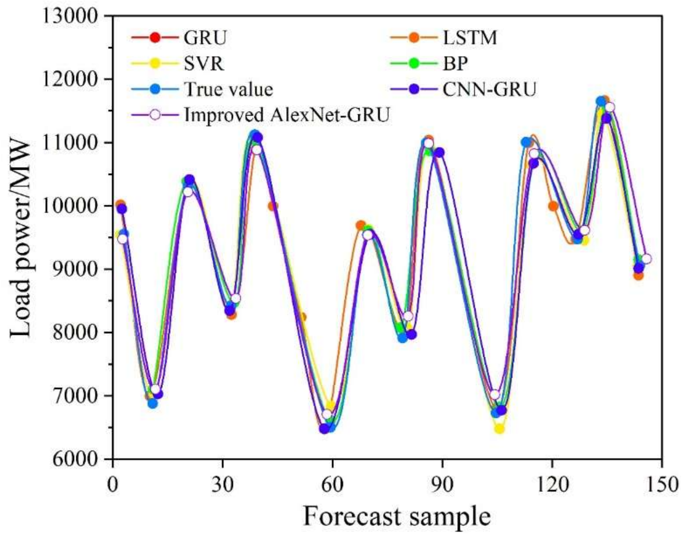


Figure 6: July 5-7 load forecast comparison results

The comparison results of the average absolute error MSE, average relative error MAPE and root mean square error RMSE of the six models for short-term electricity load forecasting in the July test set are shown in Table 4.

The traditional SVR forecasting method has the largest relative error, and the average relative errors of the other five forecasting models are all below 3%. The improved AlexNet-GRU prediction method has the lowest RMSE and MAPE, which are 80.15 MW and 0.73%, respectively.

Compared with SVR, BP, LSTM, GRU, and CNN-GRU models, where the MAPE is reduced by 2.49%, 2.02%, 0.91%, 0.51%, and 0.25%, respectively.The RMSE is reduced by 172.21 MW, 161.02 MW, 50.37 MW, 29.53 MW, and 18.48 MW, respectively.

Comprehensive analysis shows that the proposed method has a significant decrease in MAE, MAPE and RMSE indexes, which shows that the improved AlexNet-GRU forecasting model has a strong ability to recognize the characteristic information of load data. It is able to extract key information from short-term load data information more accurately in the forecasting process, reduce the model training error, and improve the accuracy of the forecasting model [24]-[26].

Table 4: Prediction error of different models

Prediction model	MAE/MW	MAPE/%	RMSE/MW
SVR	193.52	3.22	252.36
BP	160.71	2.75	241.17
LSTM	145.26	1.64	130.52
GRU	123.65	1.24	109.68
CNN-GRU	85.22	0.98	98.63
Improved AlexNet-GRU	62.45	0.73	80.15
Minimum value	62.45	0.73	80.15
Maximum value	193.52	3.22	252.36

IV. Conclusion

The improved AlexNet-GRU forecasting model significantly improves the short-term load forecasting accuracy of distribution networks by combining the feature extraction capability of convolutional neural networks and the timing processing advantage of gated recurrent units. Example analysis shows that the model achieves an average prediction accuracy of more than 98% in consecutive six-month load forecasts. Comparative experiments demonstrate that the proposed model reduces the MAPE metrics by 2.36%, 1.996%, 1.435%, 1.121%, and the RMSE by 155.047, 126.055, 76.568, and 48.925 in four-season average forecasts compared to RNN, LSTM, GRU, and CNN-GRU models, respectively. In the Australian dataset validation, the MAPE of the improved AlexNet-GRU prediction method is 0.73% and the RMSE is 80.15 MW, which are reduced by 0.25% and 18.48 MW compared with the CNN-GRU model, respectively. The model optimization process, by replacing the 2D convolutional layer with the 1D convolutional layer, reducing the number of output nodes of the convolutional layer, adjusting the size of the convolutional kernel, and using the Adam optimizer, effectively The redundancy of parameters is reduced and the computational efficiency is improved. The actual prediction results show that the proposed model can accurately capture the load change trend, and the average absolute percentage error and the maximum absolute percentage error of the 24 monitoring points are 1.239% and 3.266%, respectively. In summary, the improved AlexNet-GRU model performs well in improving prediction accuracy and computational efficiency, and provides an effective method for short-term load forecasting in distribution networks.

References

[1] Tajer, A. (2016). Load forecasting via diversified state prediction in multi-area power networks. IEEE Transactions on Smart Grid, 8(6), 2675-2684.

[2] 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.

[3] 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.

[4] Habbak, H., Mahmoud, M., Metwally, K., Fouda, M. M., & Ibrahem, M. I. (2023). Load forecasting techniques and their applications in smart grids. Energies, 16(3), 1480.

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

[6] Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. Sustainable cities and society, 35, 257-270.

[7] Azeem, A., Ismail, I., Jameel, S. M., & Harindran, V. R. (2021). Electrical load forecasting models for different generation modalities: a review. IEEE Access, 9, 142239-142263.

[8] Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020). Methods and models for electric load forecasting: a comprehensive review. Logistics, Supply Chain, Sustainability and Global Challenges, 11(1), 51-76.

[9] Zaboli, A., Tuyet-Doan, V. N., Kim, Y. H., Hong, J., & Su, W. (2023). An lstm-sae-based behind-the-meter load forecasting method. IEEE Access, 11, 49378-49392.

[10] Fahiman, F., Erfani, S. M., Rajasegarar, S., Palaniswami, M., & Leckie, C. (2017, May). Improving load forecasting based on deep learning and K-shape clustering. In 2017 international joint conference on neural networks (IJCNN) (pp. 4134-4141). IEEE.

[11] Bian, H., Zhong, Y., Sun, J., & Shi, F. (2020). Study on power consumption load forecast based on K-means clustering and FCM–BP model. Energy Reports, 6, 693-700.

[12] Chen, K., Chen, K., Wang, Q., He, Z., Hu, J., & He, J. (2018). Short-term load forecasting with deep residual networks. IEEE Transactions on Smart Grid, 10(4), 3943-3952.

[13] Huang, N., He, Q., Qi, J., Hu, Q., Wang, R., Cai, G., & Yang, D. (2022). Multinodes interval electric vehicle day-ahead charging load forecasting based on joint adversarial generation. International Journal of Electrical Power & Energy Systems, 143, 108404.

[14] Ali, M., Adnan, M., Tariq, M., & Poor, H. V. (2020). Load forecasting through estimated parametrized based fuzzy inference system in smart grids. IEEE Transactions on Fuzzy Systems, 29(1), 156-165.

[15] López, J. C., Rider, M. J., & Wu, Q. (2018). Parsimonious short-term load forecasting for optimal operation planning of electrical distribution systems. IEEE transactions on power systems, 34(2), 1427-1437.

[16] 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.

[17] 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.

[18] Saviozzi, M., Massucco, S., & Silvestro, F. (2019). Implementation of advanced functionalities for Distribution Management Systems: Load forecasting and modeling through Artificial Neural Networks ensembles. Electric Power Systems Research, 167, 230-239.

[19] Dong, M., & Grumbach, L. (2019). A hybrid distribution feeder long-term load forecasting method based on sequence prediction. IEEE Transactions on Smart Grid, 11(1), 470-482.

[20] Din, G. M. U., & Marnerides, A. K. (2017, January). Short term power load forecasting using deep neural networks. In 2017 International conference on computing, networking and communications (ICNC) (pp. 594-598). IEEE.

[21] Shuangnan Gu,Yayu Yang,Wenchao Shi & Jijian Tong. (2024). Shore Power System Load Forecasting Model based on IDBO algorithm and PCA-BiLSTM Network. Journal of Physics: Conference Series,2758(1).

[22] Lang Xuhan. (2023). Analysis of power system load forecasting based on neural networks. Journal of Physics: Conference Series,2664(1).

[23] Yuan Sun,Yong Liu,Jiangtao Wang,Chao Xiang,Sai Yin & Xiaobin Li. (2024). Intelligent monitoring method for kick risk based on CNN-GRU hybrid network model. Journal of Physics: Conference Series,2901(1),012007-012007.

[24] H. Wang, J. Wang, L.C. Li, et al. Misalignment tolerance improvement of loosely coupled transformer with ferromagnetic materials via genetic algorithm, Electrical Materials and Applications, 1 (2) e70003, 2024.

[25] Y.T. Wang, X. Li, Z.T. Gao, et al. Analysis and calculation of no-load leakage flux coefficient for arc motors, Electrical Materials and Applications, 2 (1) e70013, 2025.

[26] C.Y. Liu, F.Y. Yang, Y. Han, et al. Advances in high magnetic induction and low loss Fe-based nanocrystalline alloys, Electrical Materials and Applications, 2 (2) e70012, 2025.