Integrated forecasting model based on LSTM and TCN for short-term power load forecasting

Yangbo Yue

China-UK College, Shenzhen, 518101, China

Abstract. Power load forecasting is important to ensure the stability and reliability of regional power systems. Researchers have put forward many combined forecasting models, but most of them cannot capture the global characteristics of data well. So as to improve the accuracy of short-term power load forecasting, this paper puts forward a combined forecasting model based on long-term and short-term memory networks (LSTM) and time convolution networks (TCN). In terms of the power load data, the LSTM and TCN forecasting models are established at first, and then the output results of LSTM and TCN are weighted and combined according to the reciprocal ratio of the square error, and the LSTM-TCN combined forecasting model is obtained. Finally, an example is analyzed by using the real data of the Australian Energy Administration. The LSTM-TCN model constructed in this paper has more advanced model performance, and its error is obviously lower than that of a single forecasting model and other classical network models, indicating that the LSTM-TCN model has higher accuracy in short-term load forecasting.

Keywords: power load forecasting; long-term and short-term memory neural network; time convolution neural network; error square reciprocal ratio.

1. Introduction

Short term load forecasting (STLF) mainly refers to predicting the power load in the next few hours, one day or several days. The safe operation of the power system, the start and stop of generator units, and the scheduling and trading plans of electricity all require reliable STLF. Therefore, short-term load forecasting has become increasingly important as a research field in contemporary power system management. Various load forecasting techniques have been proposed during the past few decades, which can be roughly divided into traditional forecasting models, single network models, and hybrid network models.

Traditional forecasting models include the regression method, autoregressive moving average method, and exponential smoothing method. Milton et al. [1] made an effective short-term load forecast for the university campus through the Gaussian process regression (GRP). Through the traditional autoregressive method, Dodamani et al. [2] provided a good forecast result for the power load in the next 4 to 6 hours. Yu et al. [3] proposed an AR model based on particle swarm optimization (PSO), which significantly increases the forecast accuracy compared with the traditional AR model based on the least square method. Mazira et al. [4] used a seasonal ARIMA model to predict the future power load, which verifies the applicability of this forecasting technology. Goswami et al. [5] compared the forecasting performance of ARIMA and SARIMA and found that the error of the SARIMA model which considers the seasonality of load data is smaller. By comparing several forecasting methods of time series, Taylor [6] found that due to its ability to recognize the seasonality present in the data, the double-season Holt-Winters exponential smoothing method provides the highest predicting accuracy. These approaches work well for consistent power load forecasts and are simple to predict, but they are not appropriate for load series with significant randomness.

At present, load forecasting methods which based on artificial intelligence have also been widely applied, with single neural networks and support vector machines being the most typical. Sorab Singh et al. [7] used an artificial neural network to predict the hourly power load in NEPOOL, New England. By analyzing the working days and weekends separately, the prediction accuracy is improved. Muhammad et al. [8] developed an improved artificial neural network model with an adaptive backpropagation algorithm (ABPA). This method eliminates the behavior difference between the training dataset and the future dataset by introducing adjustment factors and further improves the forecasting formula so that the improved neural network technology can be effectively used for long-term power load forecasting. Huang et al. [9] provided a paradigm for a hybrid neural algorithm. The hybrid neural network's topological structure and time step are optimized using the enhanced differential evolution technique, which significantly boosts prediction accuracy and speed. A short-term load forecasting model and a load early warning model for charging stations based on PSO-SVM were proposed by Liao et al. [10] The support vector machine (SVM) model's parameters are optimized using particle swarm optimization (PSO), and a PSO-SVM load forecasting model with optimal kernel parameters is established in accordance with the normalized root mean square error. This model has a high level of predicting accuracy and stability. A load forecasting model for XGBoost based on similar days was proposed by Liao et al. [11]. After studying the common law of weather and daily pattern on load, a model with second-order Taylor expansion and loss function is added to the regularization term, which will successfully manage the complexity and over-fitting problem of the model. Chen et al. [12] used the gradient boosting decision tree (GBDT) for load forecasting and found that it is better than RBF neural network and kernel SVR model in short-term load forecasting. Li et al. [13] found that the prediction accuracy of SVR with the linear kernel is higher than that of artificial neural networks (ANN). Jiang et al. [14] proposed an improved Grey Wolf Optimizer Optimized Support Vector Regression (IGWO-SVR) model, which has the advantages of good stability and fast convergence. Compared with the traditional BP neural network and SVR model, the prediction accuracy is higher. Due to the increasingly complex power system structure, the nonlinearity and uncertainty of load are becoming more and more obvious, and the relationship between load and factors affecting load is difficult to be clearly expressed by a single model.

In order to further improve the forecasting performance, most researchers now put forward mixed models and combined forecasting models. Muhammad et al. [15] developed a hybrid forecasting model (FE-WNN-SAMF). The adaptive momentum factor is used to optimize the threshold and initial weight of the wavelet neural network, and the feature selection module is integrated to solve the dimension problem. This model is superior to other traditional models in stability, convergence speed, and accuracy. In order to handle the power load data and increase the model's predictive accuracy, Lv et al. [16] proposed a hybrid model based on variational modal decomposition, long-term, and short-term memory and applied the seasonal factor removal approach based on the original data features. A short-term load forecasting method based on a combined optimization of gradient boosting decision trees was proposed by Liu et al. [17]. Making full use of the data provided by each model, the prediction accuracy of the model is increased by employing the outputs of ARIMA, SVM, and back propagation neural network (BPNN) as the input of GBDT. Zhang et al. [18] proposed a load forecasting method based on VMD-SVR-PSO combined optimization model. VMD is used to decompose data, SVR is used to predict each decomposition sequence, and PSO is used to optimize parameters so that the model has high forecasting accuracy. Zhang et al. [19] put forward a combined forecasting method based on the least squares support vector machine (LSSVM) and BP neural network (BPNN), and its forecasting accuracy is better than that of a single method. Li et al. [20] put forward a short-term power load combination model based on singular spectrum analysis, neural network, and least square support vector machine, and used particle swarm optimization and simulated annealing to optimize the parameters of the combination model respectively. The combination forecasting model can significantly improve the forecasting accuracy.

Even if the previous hybrid model has achieved excellent performance in load forecasting, as the previous study [18] indicated, the VMD-SVR-PSO model cannot capture the global characteristics of data well. Therefore, this paper introduces a combined forecasting model based on a time convolution neural network (TCN) and long-term and short-term memory network (LSTM). Firstly, LSTM is used to forecast the power load. Secondly, TCN is used to predict the power load. Finally, the prediction results of the two models are weighted and combined to get the final result. Generally speaking, the contributions of this study are as follows:

1. A hybrid forecasting model based on LSTM and TCN is proposed, which can effectively combine the advantages of LSTM and TCN and has strong robustness.
2. The combined forecasting model has high forecasting accuracy, which is of great significance to ensure the reliability of power system operation.
3. Experiments on several real datasets show excellent generalization ability and prediction performance.

2. Model Framework

This section introduces all single prediction models, including long-term and short-term memory networks (LSTM), time convolution networks (TCN), weighting methods based on the reciprocal ratio of error squares, and indicators for evaluating the performance of prediction models.

2.1 Long-term and short-term memory network (LSTM)

LSTM introduces a cell state to realize long-term memory on the basis of RNN, which solves the problem of RNN. The specific calculation flow is as follows:

f_t represents a forgetting gate, which can decide which information to discard, as shown in formula 1:

$$f_t = \sigma(W_f . [h_{t-1}, x_t]) + b_f \tag{1}$$

In this formula, σ represents the sigmoid function, W_f is the weight of the forgetting gate, and $\left[h_{t-1}, x_t\right]$ represents the input value at the last moment and the output value information at that moment, and b_f is the paranoid value of the gate.

i_t represents an input gate, which can determine what information to add, as shown in formula 2:

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

In this formula, W_i is the weight of the input gate, $\left[h_{t-1}, x_t\right]$ indicates the input value at the last moment and the output value information at that moment, and b_i is the paranoid value of the gate.

According to the operating results of the above two gates, their state is updated, which is shown in formula 3:

$$\overset{\circ}{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{3}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \overset{\circ}{c}_t \tag{4}$$

In the two formulas, c_t represents the state of neurons in time t, W_c is the weight of neuron state, b_c is the paranoid value of this state, and tanh is the processing function, from which alternative update content can be generated.

o_t represents an output gate, which can determine the output of information, as shown in formulas 5 and 6:

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{5}$$

$$h_t = o_t \cdot \tanh\left(c_t\right) \tag{6}$$

In the above formulas, W_o is the weight of the output gate, b_o is the paranoid value of the gate, and h_t represents the output value of the neuron at t time.

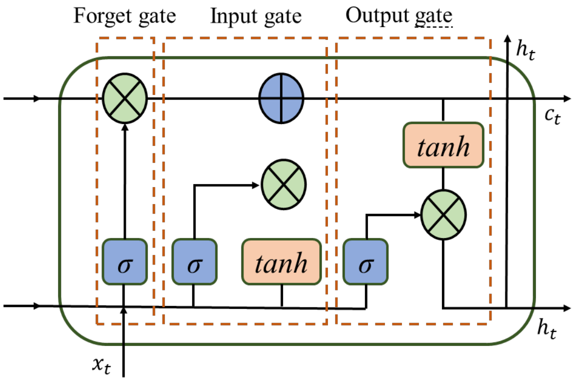

Figure 1 LSTM memory cell structure

LSTM memory cell structure mainly consists of three memory cells, namely the forgetting gate, input gate, and output gate.

2.2 Time Convolution Network (TCN)

TCN is an improved network model based on CNN, which has a unique expansive causal convolution structure and is more suitable for solving time series problems. Expansion convolution can expand the input of the upper layer and extract the features of the data with long interval. Causal convolution can ensure the causality of extracted features, that is, the output yt at time t only depends on the input before time t. TCN with a convolution kernel size of 2 and expansion coefficients of 1, 2 and 4. Its expansion causal convolution structure is shown in Figure 2:

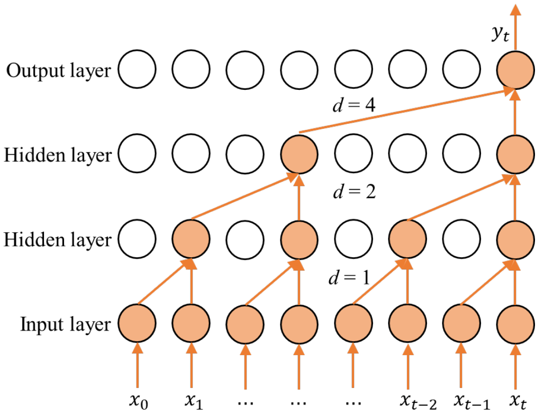

Figure 2 TCN expansion causal convolution structure

TCN expansion causal convolution structure mainly includes input layer, hidden layer and output layer.

The memory length of the TCN network is determined by convolution kernel size, expansion coefficient and convolution layers. The extended convolution operation of the sth sequence element in the network can be defined as formula 7:

$$F(S)=(x*f)(s)=\sum_{i=0}^{k-1} f(i)\cdot x_{s-z\cdot i} \tag{7}$$

In this formula, x is the input sequence; * indicates convolution operation; z is the expansion coefficient; k is the convolution kernel size; $f(i)$ is the ith element in the convolution kernel; $x_{s-z\cdot i}$ is the element corresponding to the fish convolution kernel in the input sequence.

The residual module of TCN includes the basic TCN causal expansion convolution layer, weight normalization layer, activation function Relu and Dropout layer, and the results are shown in Figure 3:

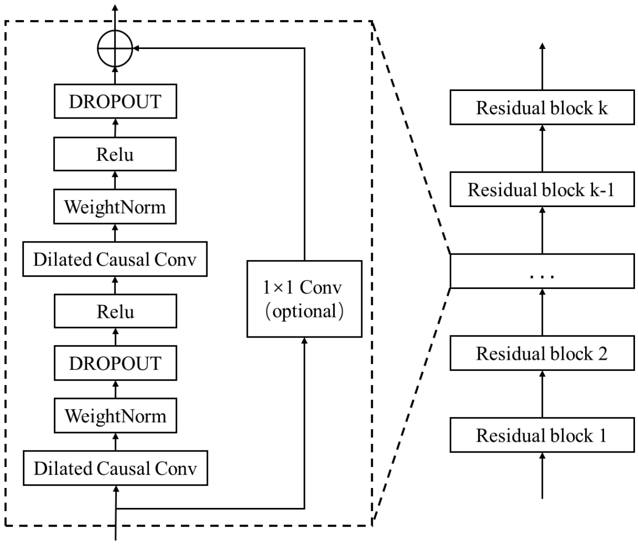

Figure 3 Residual module structure diagram

The residual module structure diagram consists of two layers of extended convolution, two layers of weight normalization, two layers of activation function ReLU and two layers of Dropout.

Normalization of weights can eliminate the problem of gradient explosion and effectively speed up the calculation. In order to make TCN network nonlinear and not too simple, Relu activation function is adopted, and Dropout layer is added after Relu activation layer to prevent over-fitting, so as to achieve regularization effect. This study adjusts the different dimensions of residual tensor by 1×1 convolution.

2.3 Combination forecasting model based on LSTM and TCN

Bates and Granger proposed the forecasting technique known as combination forecasting in 1969. The starting point is that the prediction results of different methods are different, and it is necessary to give different weights to each prediction model, so as to get a better comprehensive model. Figure 4 presents the combined forecasting model flow chart based on LSTM and TCN.

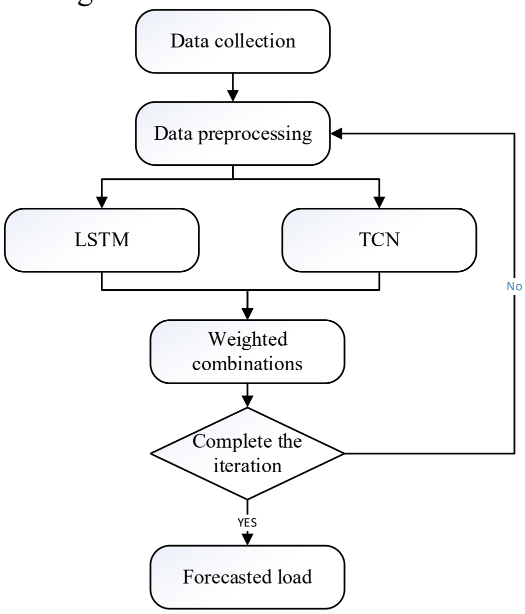

Figure 4 Combined forecasting model flow chart of LSTM and TCN

In this paper, the optimal weight method based on residual is adopted. The smaller the prediction error of a single model, the greater its weight in the combined model. Assuming that two single models predict m time points at the same time, the calculation formula of this method is shown in Formula 8-10:

$$w_i = \frac{1}{h_i \cdot \sum_{i=1}^{2} \frac{1}{h_i}} \tag{8}$$

$$h_i = \sum_{i=1}^{m} v_{it}^2 \tag{9}$$

$$v_{it} = f_{it} - f_t \tag{10}$$

In these formulas, f_{it} represents the predicted value of the ith method at time t, f_t is the actual value at time t, v_{it} represents the residual of the ith method at time t, h_i is the square sum of the residual of the ith method at an m time point, and w_i is the weight coefficient of the ith method, and its value is equal to the reciprocal ratio of the square of the prediction error of each single model.

2.4 Evaluation of model performance

In this study, the prediction outcomes of each model are assessed using the mean absolute percentage error (MAPE) and the root mean square error (RMSE). Formulas 11 and 12 provide their definitions.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{12}$$

In these formulas, y_i represents the real load value, $\hat{y}_i$ is the load value that is be predicted and n is the number of time series points.

3. Experimental Results and Discussion

3.1 Data description

The data of the present study comes from AEMO (Australian Energy Market Operator). This data is the power load collected once every half hour in Tasmania from January to March, 2014. The average value of this set of data is 1044.22 kWh, and the standard deviation is 88.322. To better see the regularity of the dataset, the present paper paper selects the data with a time span of one week and visualizes it, as shown in Figure 6.

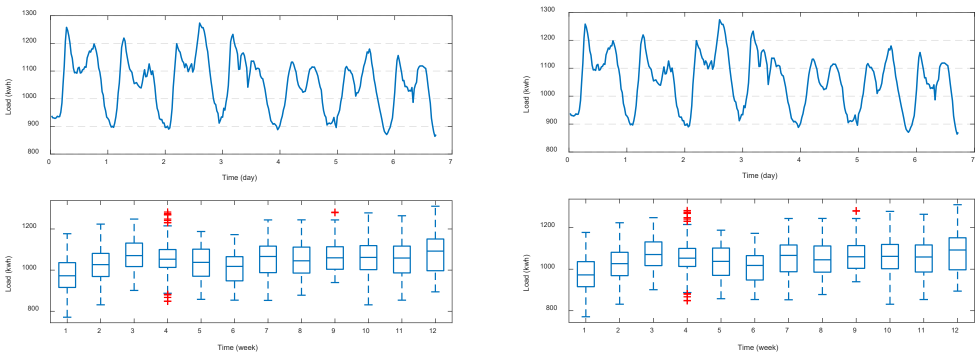

Figure 5 Dataset description [21]

According to Figure 6, the change of load is periodic, and the trend of load change is basically the same every day, that is, the peak and trough of load curve appear at the same time. From the box chart, it can be seen that in these three months, there are a lot of abnormal values in the load in the fourth week. It shows that the power grid load fluctuates greatly this week.

3.2 Results of the model

The emergence of Back-propagation (BP) algorithm is a major breakthrough in the development of neural networks, and it is also the basis of many deep learning training methods. This method will calculate the loss function gradient to each parameter in the neural network, and cooperate with the optimization method to update the parameters and reduce the loss function.

BP originally only refers to the process that the gradient of the loss function flows reversely through the network, but now it is often understood as the whole training method of neural network, which consists of two steps, that is, error propagation and parameter updating.

Recurrent neural networks (RNNs) are used to address problems where the training samples' input is a continuous sequence but the sequence’s length varies, such as time series problems. The primary distinction between RNN and a basic neural network is that while the latter only creates weighted connections between layers, the former also creates weighted connections between neurons.

In this paper, we set the number of hidden layer neurons in LSTM as 200, the convolution kernel size in TCN model is 2, and the convolution expansion factor is 2. The original data are divided into three groups by month, and each group of data is divided into training set and test set according to 7:3, and the performances of commonly used BP, RNN and LSTM-TCN used in this paper are compared, respectively on the test set. Among them, the number of hidden layers of BP and RNN are both set to 200. The comparison results are shown in Table 1.

Table 1 Comparisons of the performance of different models in different datasets

Dataset	BP		RNN		LSTM-TCN	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Jan.	1.538	21.864	1.774	24.906	1.377	19.3693
Feb.	1.653	23.267	1.917	26.130	1.449	20.6649
Mar.	1.819	24.962	2.410	33.828	1.340	18.3162
All	1.773	24.842	1.878	26.357	**1.442**	20.5893

As shown in Table 1, in the same dataset, the MAPE value of LSTM-TCN is 1.422, which is lower than the traditional BP and RNN models. In different datasets, the MAPE value of the LSTM-TCN model used in this paper is still smaller than other traditional models, which demonstrates that our model has strong generalization ability.

3.3 Ablation experiment

The term ablation research comes from the experimental psychology of the 1960s and 1970s, when researchers removed an animal's brain to examine how it affected the animal's behavior. Ablation research has been utilized in machine learning, particularly sophisticated deep neural networks, to define the process of deleting unwanted network components.

To verify the performance ability of the model, the experiment also evaluates the prediction effect of each independent model and compares them with the performance of the combined model. When LSTM acts alone, its input feature dimension is 2, and the hidden layer neurons are set to 200. When TCN acts alone, the convolution kernel size in the model is set to 2 and the convolution expansion factor is set to 2. After 1000 iterations of training, the prediction effects of different models are shown in Table 2.

Table 2 Comparisons of the performance of independent model and combined model in different datasets

Data set	LSTM		TCN		LSTM-TCN	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Jan.	2.873	40.5343	1.499	20.7559	1.377	19.3693
Feb.	2.107	28.9641	1.631	22.5112	1.449	20.6649
Mar.	2.389	32.9566	1.622	21.1821	1.340	18.3162

All	2.007	28.4452	1.515	21.0238	1.442	20.5893

As indicated in Table 2, the LSTM-TCN combined forecasting model proposed in this paper has the smallest MAPE value and RMSE value in the datasets of different months. In all datasets, the MAPE value of the LSTM-TCN model is 1.442%, and the RMSE value is 20.5893. Compared with the LSTM and TCN single model, the accuracy is improved.

4. Conclusion and Prospect

Deep learning models such as LSTM and TCN have been widely used in various fields of life. The combined model based on LSTM and TCN proposed in this paper is used for short-term load forecasting of power system. The presents study selects the real data of the Australian Energy Administration, compares the experimental results of the LSTM-TCN model with traditional model and single model, and draws the following conclusions:

1. In the LSTM-TCN model proposed in this paper, LSTM and TCN models are weighted by the reciprocal ratio of the square of error (integrated model), and the time series data with the large error of single model are corrected to reduce the error of the single model. The experimental results show that the LSTM-TCN model can effectively improve the accuracy of short-term power load forecasting.

2. In the future research, we should consider adding an attention mechanism to TCN. Moreover, meteorological factors such as temperature, humidity and weather can also be introduced into short-term load forecasting. In addition, the model can be extended to specific fields such as industry, agriculture and commerce, and more effective information can be mined in combination with specific scenarios.

References

[1] M. Madhukumar, A. Sebastian, X. Liang, M. Jamil and M. N. S. K. Shabbir, "Regression Model-Based Short-Term Load Forecasting for University Campus Load," in IEEE Access, vol. 10, pp. 8891-8905, 2022.

[2] S. N. Dodamani, V. J. Shetty and R. B. Magadum, "Short term load forecast based on time series analysis: A case study," 2015 International Conference on Technological Advancements in Power and Energy (TAP Energy), 2015, pp. 299-303.

[3] J. Yu, H. Lee, Y. Jeong and S. Kim, "Short-term hourly load forecasting using PSO-based AR model," 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS), 2014, pp. 1449-1453.

[4] H. Matsila and P. Bokoro, "Load Forecasting Using Statistical Time Series Model in a Medium Voltage Distribution Network," IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, 2018, pp. 4974-4979.

[5] K. Goswami and A. B. Kandali, "Electricity Demand Prediction using Data Driven Forecasting Scheme: ARIMA and SARIMA for Real-Time Load Data of Assam," 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 570-574.

[6] J. W. Taylor and P. E. McSharry, "Short-Term Load Forecasting Methods: An Evaluation Based on European Data," in IEEE Transactions on Power Systems, vol. 22, no. 4, pp. 2213-2219, Nov. 2007.

[7] S. Singh, S. Hussain and M. A. Bazaz, "Short term load forecasting using artificial neural network," 2017 Fourth International Conference on Image Information Processing (ICIIP), 2017, pp. 1-5.

[8] Mohammed, N.A., Al-Bazi, A. An adaptive backpropagation algorithm for long-term electricity load forecasting. Neural Comput & Applic 34, 477–491 (2022)

[9] Y. Huang, R. Zhao, Q. Zhou and Y. Xiang, "Short-Term Load Forecasting Based on a Hybrid Neural Network and Phase Space Reconstruction," in IEEE Access, vol. 10, pp. 23272-23283, 2022.

[10] X. Liao, X. Kang, M. Li and N. Cao, "Short Term Load Forecasting and Early Warning of Charging Station Based on PSO-SVM," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2019, pp. 305-308.

[11] X. Liao, N. Cao, M. Li and X. Kang, "Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2019, pp. 675-678,.

[12] B. Chen, R. Lin and H. Zou, "A Short Term Load Periodic Prediction Model Based on GBDT," 2018 IEEE 18th International Conference on Communication Technology (ICCT), 2018, pp. 1402-1406.

[13] L. W. Chong, D. Rengasamy, Y. W. Wong and R. K. Rajkumar, "Load prediction using support vector regression," TENCON 2017 - 2017 IEEE Region 10 Conference, 2017, pp. 1069-1074.

[14] F. Jiang, Z. Peng and J. He, "Short-term load forecasting based on support vector regression with improved grey wolf optimizer," 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), 2018, pp. 807-812.

[15] M. ZulfiqAr, M. Kamran, M. B. Rasheed, T. Alquthami and A. H. Milyani, "A Short-Term Load Forecasting Model Based on Self-Adaptive Momentum Factor and Wavelet Neural Network in Smart Grid," in IEEE Access, vol. 10, pp. 77587-77602, 2022.

[16] L. Lv, Z. Wu, J. Zhang, L. Zhang, Z. Tan and Z. Tian, "A VMD and LSTM Based Hybrid Model of Load Forecasting for Power Grid Security," in IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 6474-6482, Sept. 2022.

[17] S. Liu, Y. Cui, Y. Ma and P. Liu, "Short-term Load Forecasting Based on GBDT Combinatorial Optimization," 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), 2018, pp. 1-5.

[18] R. Zhang and H. Qu, "Short-Term Prediction of Power Load for Urban Residents Based on VMD-SVR-PSO Model," 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI), 2022, pp. 519-524.

[19] Zhang Ming-guang and Li Lin-rong, "Short-term load combined forecasting method based on BPNN and LS-SVM," 2011 IEEE Power Engineering and Automation Conference, 2011, pp. 319-322.

[20] Zhang, H., Yang, Y., Zhang, Y. et al. A combined model based on SSA, neural networks, and LSSVM for short-term electric load and price forecasting. Neural Comput & Applic 33, 773–788 (2021).

[21] Accessed at: https://aemo.com.au