International Journal of Innovative Computing, Information and Control
Volume **21**, Number **6**, December **2025**

ICIC International ©2025 ISSN 1349-4198
pp. 1599–1612

SHORT-TERM POWER LOAD FORECASTING USING A HYBRID MSCNN-BILSTM MODEL WITH ATTENTION MECHANISM

Zhihua Xu[1] and Tinglong Pan[2]

[1]School of Electrical Engineering
Yancheng Institute of Technology
No. 1, Hope Avenue Middle Road, Yancheng 224051, P. R. China
xzh@ycit.cn

[2]School of Internet of Things Engineering
Jiangnan University
No. 1800, Lihu Avenue, Wuxi 214122, P. R. China
tlpan@jiangnan.edu.cn

Received April 2025; revised July 2025

Abstract. *To address the challenge of extracting key features from power load data, this study proposes a hybrid MSCNN-BiLSTM-Attention model for short-term power load forecasting. First, the maximal information coefficient (MIC) is employed to analyze feature correlations within the power load dataset, enabling the selection of highly relevant features for dataset construction. The processed data is then fed into a multi-scale convolutional neural network (MSCNN) to extract temporal features at varying scales. Subsequently, these features are processed by a bidirectional long short-term memory network (BiLSTM) for sequential prediction, while an attention mechanism dynamically filters and weighs the temporal features. Finally, a fully connected layer integrates the outputs to generate the forecasted values. The model is validated using three years of multidimensional power load data from a region in Australia, with five comparative models established for benchmarking. Additional verification is performed using a three-year dataset from Spain. Experimental results demonstrate that the proposed MSCNN-BiLSTM-Attention model outperforms alternative approaches, achieving superior prediction accuracy and effectively resolving the difficulty of extracting critical regional-level power load features.*
Keywords: Power load forecasting, Multi-scale convolutional neural network, Bidirectional long short-term memory, Attention mechanism, Deep learning, Maximal information coefficient

1. **Introduction.** The accelerated modernization process has led to a substantial surge in electricity demand, consequently elevating the precision requirements for power load forecasting to unprecedented levels [1]. Short-term power load forecasting facilitates optimal multi-energy complementarity in power generation systems while enabling rational allocation, management, and dispatch of electricity consumption, thereby ensuring stable operation of power grids. Accurate and effective electric load forecasting is fundamental to national energy policy formulation, guiding the production of electrical equipment and infrastructure planning while ensuring power system operational efficiency [2].

Power load is predominantly influenced by key factors such as climatic conditions and economic indicators, which impart significant stochastic volatility and uncertainty to load profiles. These characteristics consequently elevate the technical requirements for load forecasting methodologies. Currently employed approaches can be categorized into conventional prediction methods and machine learning techniques [3]. Traditional forecasting

DOI: 10.24507/ijicic.21.06.1599

methods encompass various analytical approaches, including time series analysis methods [4], trend analysis methods, regression analysis methods [5], exponential smoothing techniques, and grey prediction models [6]. These methods essentially achieve prediction by establishing statistical relationships within historical data. However, they demand high-quality historical data and may produce significant deviations when external influencing factors undergo substantial changes [7]. Machine learning-based forecasting methods utilize advanced computational power to extract latent patterns from vast datasets, autonomously improving through experiential learning to develop generalized predictive models. Traditional machine learning methods include support vector machine (SVM) [8], decision tree [9], and random forest (RF) [10], which can handle nonlinear problems with relatively small amounts of data. Compared to traditional forecasting methods, machine learning techniques can improve prediction accuracy. However, they still have limitations in feature extraction. When applied to more complex power systems, their forecasting performance tends to be poor, resulting in lower accuracy in power load prediction [11].

Deep learning models, due to their adaptive learning and updating capabilities, have gradually become the mainstream research method in the field of load forecasting. The essence of these models is to construct multi-layer neural networks to extract data features and build nonlinear mapping relationships, enabling time-series forecasting. [12] proposed a new end-to-end training deep neural network DeepRidge is proposed to obtain better accuracy with smaller parameter sizes. In power load forecasting, common deep learning models include recurrent neural networks (RNN), long short-term memory (LSTM), gated recurrent units (GRU), convolutional neural networks (CNN), deep belief networks (DBN), and generative adversarial networks (GAN), among others [13]. [14] proposed an LSTM-based short-term power load forecasting method for individual residential households, effectively improving the short-term prediction accuracy of power load data. [15] introduced a CNN-BiGRU-NN hybrid model for short-term power load forecasting, which outperformed other models in terms of accuracy. [16] built a deep LSTM network and conducted hyperparameter search, achieving higher prediction accuracy compared to machine learning methods. [17] improved the DBN network, enhancing both the computational speed and prediction accuracy of the model, although the network still exhibited instability during operation. [18] proposed a load data repair method based on GRUI-GAN, where GAN is capable of handling incomplete time series, thus improving data quality, but the issue of lower fitting accuracy remains. [19] presented a hybrid prediction model based on CNN-BiGRU-Attention, which improved the accuracy of power load forecasting. However, it lacked correlation analysis of features affecting load and did not perform feature selection and filtering in the data preprocessing stage. In [20] combining the use of temporal convolutional network (TCN) and long-LSTM improves the accuracy of EV charging load forecasting, while a new method is used to identify composite similarity days, which significantly improves the accuracy of EV charging load forecasting. [21] proposed a novel architecture of MSTGCN-Transformer, which substantially improves the accuracy and stability of multi-node load prediction.

In summary, considering the diverse features of the power load dataset, this paper proposes a deep learning hybrid model (MSCNN-BiLSTM-Attention), which combines multi-scale convolutional neural networks (MSCNN), bidirectional long short-term memory networks (BiLSTM), and the attention mechanism. First, the maximum information coefficient (MIC) is used to select the features highly correlated with power load characteristics, constructing the dataset for the model. Next, MSCNN is employed to extract temporal features at different scales. Then, BiLSTM is applied for time-series forecasting, with the attention mechanism used to filter the temporal features. Finally, the predicted values are output through a fully connected layer. Experimental results show that the proposed hybrid model outperforms others, offering better forecasting results and higher prediction accuracy.

2. Principles of Deep Learning.

2.1. Multi-scale convolutional neural network (MSCNN). Convolutional neural network (CNN) is a deep learning model. The convolutional layer extracts feature information from the input data by performing convolution operations. The pooling layer then filters the results of the convolution operation.

Due to the fact that traditional convolutional neural networks (CNNs) can only use convolutional kernels of a single scale, the extracted feature information is not comprehensive enough [22]. Therefore, in order to capture multi-scale features when facing complex time series data, the MSCNN model designed in this paper consists of three convolutional layers, three maximal pooling layers, and fully connected (FC) layer. Finally, the multi-scale feature vectors extracted by the 3 pooling layers are fused. The MSCNN model is shown in Figure 1.

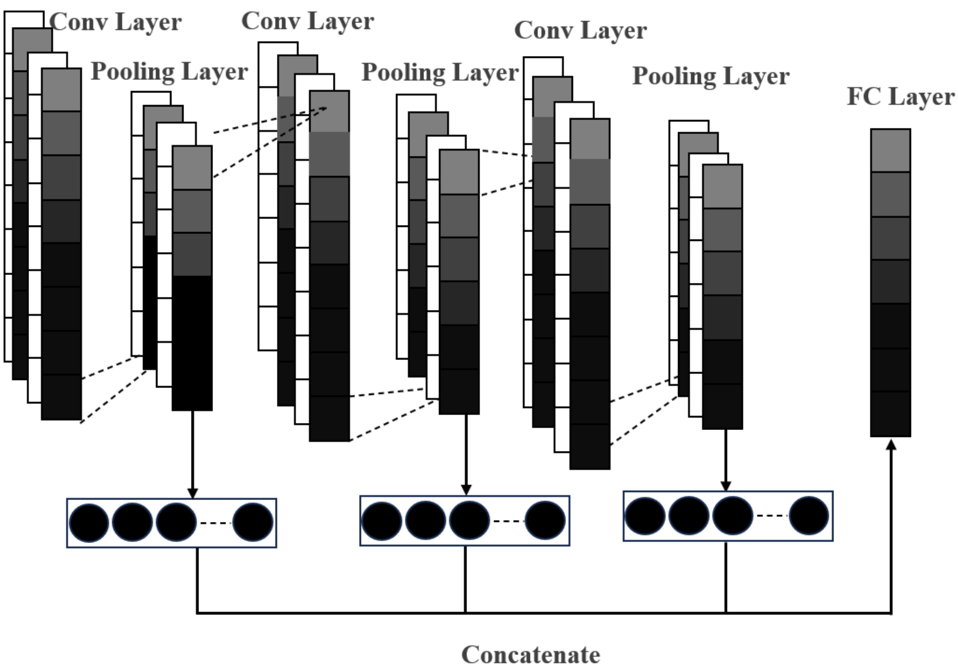

Figure 1. MSCNN structure diagram

2.2. Bidirectional long short-term memory network (BiLSTM). The power load feature information extracted and fused by MSCNN is output and then input into BiLSTM for time-series prediction.

Long short-term memory (LSTM) is a type of neural network used for processing sequential data, first introduced by Schmidhuber and Hochreiter [14]. It is designed to address the issues of gradient explosion and gradient vanishing that often occur during the training of recurrent neural networks (RNNs).

Long short-term memory (LSTM) networks read data in a unidirectional flow, from past to future. This method relies on historical data to predict the future, but it may not fully capture the features of the data, and the unidirectional information flow can lead to the loss of some information. The BiLSTM network model, on the other hand,

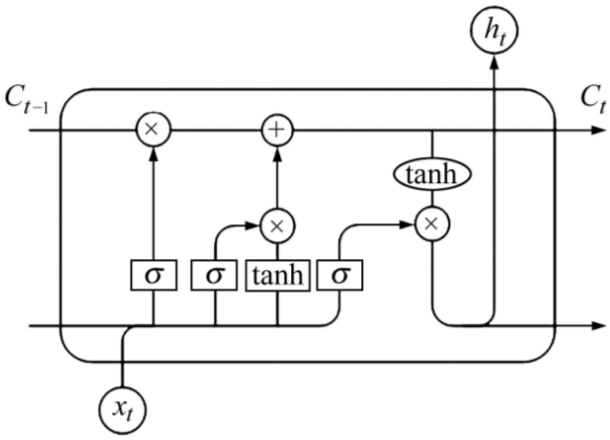

FIGURE 2. LSTM structure diagram

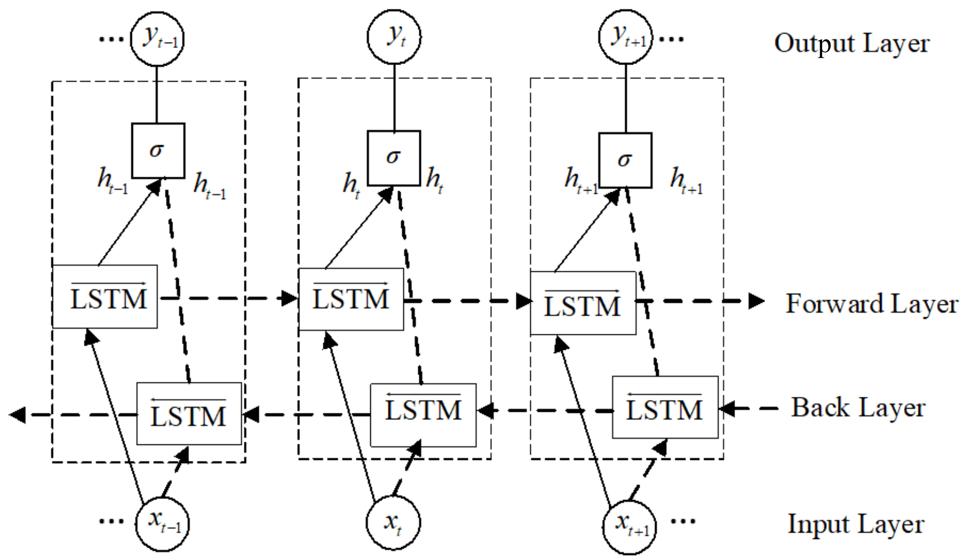

FIGURE 3. BiLSTM structure diagram

has bidirectional information flow, allowing it to simultaneously consider both past and future information [23]. For each element in the sequence, BiLSTM can provide richer contextual information. By utilizing both past and future information, BiLSTM can offer more accurate predictions in time-series forecasting. In the two sequences, one LSTM processes the input sequence in the forward direction, while the other processes it in the reverse direction. After processing, the output vectors are concatenated, completing bidirectional training and further enhancing feature extraction capabilities. Figure 3 shows the structure of the BiLSTM.

2.3. **Attention mechanism.** In the application of deep learning, the attention mechanism enhances model performance and achieves better prediction results by calculating and assigning attention weights to the input data, enabling the filtering and selection of time-series features. The computational process is shown in Figure 4. First, the feature inputs are transposed and fed into a multi-layer perceptron (MLP) to calculate the attention weights, and the output is then transposed. The attention weights are then element-wise multiplied (Hadamard product) with the feature inputs, resulting in the feature output with the attention weight distribution applied.

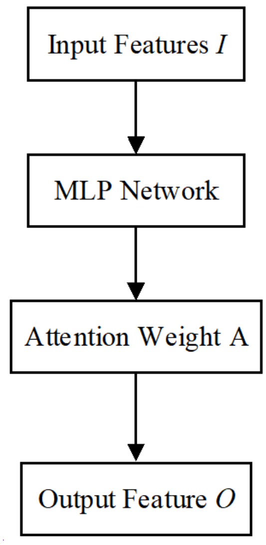

FIGURE 4. Flowchart of attention mechanism

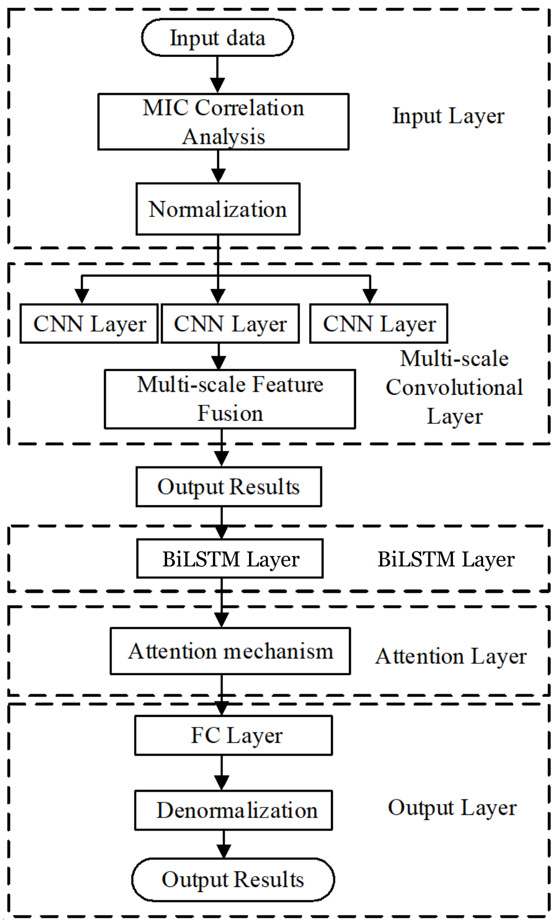

FIGURE 5. Flowchart of prediction model

3. **Construction of the MSCNN-BiLSTM-Attention Prediction Model.** The flowchart of the load forecasting model proposed in this paper is shown in Figure 5. The main structure of the model consists of five key components: the input layer, MSCNN layer, BiLSTM layer, attention mechanism layer, and output layer.

1) Input Layer: The input feature vector dimension is set to $48 * 6$, which includes 48 time points of power load data, 4 types of meteorological data (dry bulb temperature, wet bulb temperature, dew point temperature, and humidity), and electricity price data. The input data undergoes maximum information coefficient (MIC) correlation analysis [24] to select the features with higher correlations. The formula for calculating the MIC is as follows:

$$\begin{cases} Q(X,Y) = \sum P(x,y)\log\left(\dfrac{P(x,y)}{P(x)P(y)}\right) \\ M_{\text{mic}} = \max\limits_{|X||Y|<B}\left(\dfrac{\max\left(Q(X,Y)\right)}{\log_2(\min(X,Y))}\right) \end{cases} \tag{1}$$

In the formula, $Q(X,Y)$ represents the mutual information between variables X, Y; $P(x)$, $P(y)$ represent the marginal distributions of variables X, Y, respectively; $P(x,y)$ represents the joint distribution of variables X, Y; M_{mic} denotes the maximum information coefficient; B is the maximum number of grid cells used during discretization.

Next, the filtered data undergoes normalization, and the formula for normalization is

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

In the formula, X^*, X represent the normalized and original multidimensional power load data, respectively. $X_{\max}$, $X_{\min}$ represent the maximum and minimum values in the multi-dimensional power load data sequence, respectively.

2) MSCNN Layer: The MSCNN layer consists of three convolutional layers and three pooling layers. It performs multi-scale feature extraction on the input multidimensional power load data and fuses the extracted features.

3) BiLSTM Layer: The BiLSTM layer performs time-series forecasting on the multi-scale power load temporal features extracted by the MSCNN layer. It captures the temporal dependencies in the data by processing it in both forward and backward directions.

4) Attention Layer: The attention layer assigns attention weights to the input power load features, resulting in a feature sequence with attention-based weight distribution. This improves the neural network's learning capability and computational efficiency while saving computational resources.

5) Output Layer: The output layer connects the fully connected layer with the attention layer. To enhance the model's generalization ability and prevent overfitting, dropout regularization is introduced to reduce the dependencies between neurons during the training process. The output results undergo inverse normalization to restore the original and time-series predicted power load data. Accuracy evaluation metrics are then calculated to analyze the experimental results. The inverse normalization formula is

$$X = X'(X_{\max} - X_{\min}) + X_{\min} \tag{3}$$

In the formula, X, X' represent the multidimensional power load data before and after inverse normalization, respectively; $X_{\max}$, $X_{\min}$ represent the maximum and minimum values in the multi-dimensional power load data sequence, respectively.

4. **Case Simulation and Analysis.**

4.1. **Data preprocessing.**

This study uses multi-dimensional electricity load data from a region in Australia, spanning from January 1, 2008, to December 31, 2010, for a total of 3 years. The dataset includes electricity load data, four types of meteorological data (dry bulb temperature, wet bulb temperature, dew point temperature, and humidity), as well as electricity price data, all sampled at 30-minute intervals, with 48 data points collected per day. The time series data is split in a 4 : 1 ratio. The first 80% is used as the training set, while the remaining 20% is used as the test set.

This experiment is written in Python and uses the Keras framework to build the MSCNN-BiLSTM-Attention combined prediction model. The model is debugged on the Jupyter Notebook platform.

4.2. **Model evaluation metrics.** To evaluate the performance of the model, the following metrics were used: mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) [25].

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |(y_t - y_t')| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - y_t')^2} \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{y_t' - y_t}{y_t}\right| \times 100\% \tag{6}$$

In the formulas, y_t represents the true value of the original data for the ith power load; y_t' represents the predicted value of the ith electricity load data; n is the total number of samples in the electricity load test set. MAPE is used to measure the average relative error between the predicted and actual electricity load values, expressed as a percentage. MAE and RMSE are measured in MW, with smaller values indicating a smaller deviation in the electricity load prediction and higher accuracy.

4.3. **Model hyperparameter settings.** The hyperparameters of the deep learning prediction model typically include the number of filters, kernel size, batch size, and the number of training epochs. The Adam optimization algorithm is used for parameter optimization of the model. After repeated experiments and evaluations, the best hyperparameter settings for the model are shown in Table 1.

TABLE 1. Model hyperparameter configuration

Hyperparameter	Value
filters1/kernel_size1	32/8
filters2/kernel_size2	64/16
filters2/kernel_size2	64/32
epochs	300
Batch-size	48
units	64
Dropout	0.2

4.4. **Feature selection result and analysis.** The optimization and selection of multi-dimensional electricity load data were performed through the MIC analysis, and the result is shown in Figure 6. According to the MIC analysis, the electricity load has a higher correlation with dry bulb temperature (DBT), humidity, and electricity price, while the correlation with dew point temperature (DT) and wet bulb temperature (WBT) is lower.

Filter out the DT and WBT feature quantities with low correlation. DBT, humidity and electricity prices, which have high correlation, are brought into the prediction model as feature quantities.

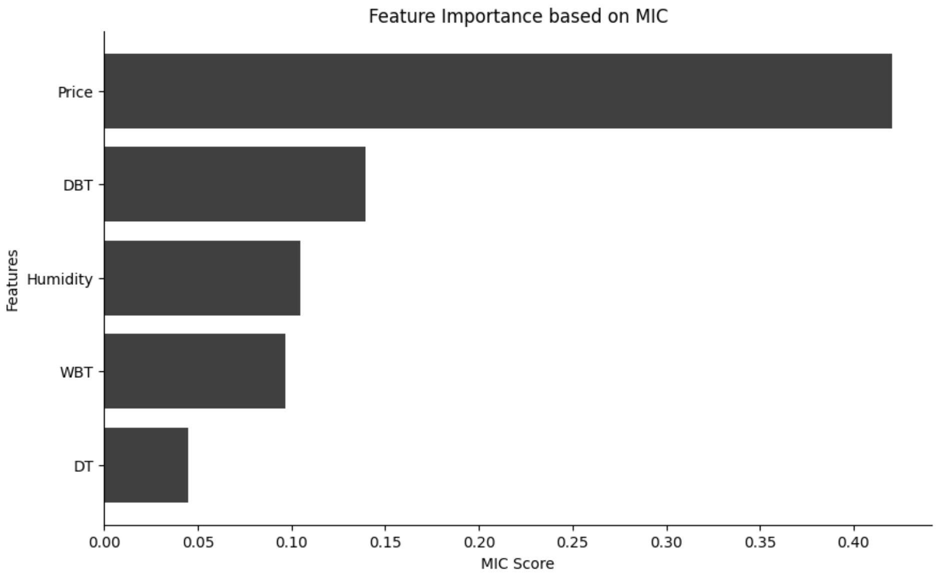

FIGURE 6. MIC correlation analysis diagram

4.5. **Comparison of different sliding windows.** The size of the sliding window can have a significant impact on the model's performance. The window size can be adjusted based on the specific requirements. As the window slides over the data stream or dataset, the data within the window is continuously updated, and the algorithm processes the data within the window accordingly.

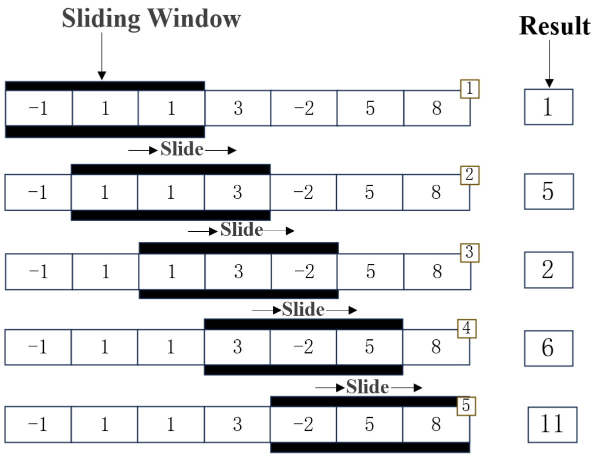

FIGURE 7. Sliding window diagram

In this study, the sliding window width starts at 24 and is doubled incrementally to find the optimal window size. The results in Table 2 and Table 3 clearly show that when the sliding window width is set to 72, the model's prediction performance is the best, with accuracy evaluation metrics of MAPE, RMSE, and MAE being 0.61%, 68.763 MW, and 53.393 MW, respectively. The R^2 was improved by 0.1%, 0.2%, 0.6%, 10.12%.

Figure 8 shows the predicted results of electricity load for the last two days of the test set with different sliding window widths. It can be observed that the choice of sliding window width has a significant impact on the model's prediction performance. When the sliding window width is small, the model predicts better in regions with smooth load changes; however, its prediction results are relatively poor in scenarios with drastic

Table 2. Prediction errors under different sliding window widths

Width \ Evaluation	MAPE/%	RMSE/MW	MAE/MW
24	0.73	85.741	64.122
48	0.84	91.504	71.977
72	0.61	68.763	53.393
96	1.03	123.681	94.435
128	1.41	160.128	127.144

Table 3. Evaluation metrics for different sliding window widths

Width \ Evaluation	R^2	Time/s
24	0.996	4506
48	0.995	12300
72	0.997	18000
96	0.991	21000
128	0.985	39000

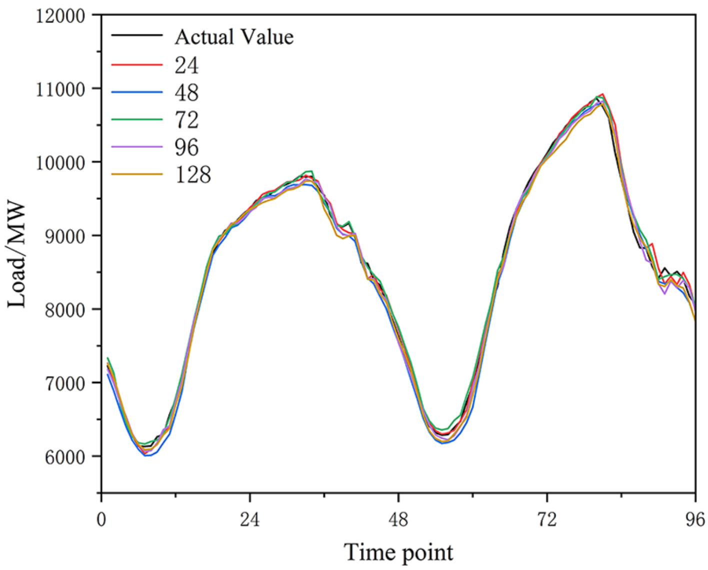

Figure 8. (color online) Prediction curves of different sliding window width

load fluctuations. Conversely, when the sliding window width is large, while the model captures the overall trend well, it struggles to accurately capture small fluctuations in the load data.

4.6. **Ablation experiment.** To verify the accuracy and reliability of the proposed model's predictions, different control groups were set up for the ablation experiment. The prediction results are shown in Table 4 and Table 5.

As shown in Table 4 and Table 5, the MSCNN-BiLSTM-Attention model outperforms the CNN-BiLSTM-Attention, CNN-BiLSTM, BiLSTM, LSTM, and GRU models, with MAPE decreasing by 0.21%, 0.46%, 0.67%, 0.75%, and 1.03%, respectively. RMSE decreases by 25.12%, 42.94%, 50.48%, 55.07%, and 58.63%, and MAE decreases by 26.02%, 43.62%, 50.55%, 54.81%, and 61.06%, respectively. This demonstrates that the model

TABLE 4. Prediction errors of different models

Evaluation / Model	MAPE/%	RMSE/MW	MAE/MW
MSCNN-BiLSTM-AM	0.61	68.723	53.393
CNN-BiLSTM-AM	0.82	91.829	72.168
CNN-BiLSTM	1.07	120.506	94.694
BiLSTM	1.28	138.858	107.983
GRU	1.36	153.057	118.167
LSTM	1.64	166.221	137.100

TABLE 5. Evaluation metrics of different models

Evaluation / Model	R^2	Time/s
MSCNN-BiLSTM-AM	0.997	18000
CNN-BiLSTM-AM	0.995	11200
CNN-BiLSTM	0.992	9000
BiLSTM	0.989	8700
GRU	0.987	8500
LSTM	0.984	8400

proposed in this paper has superior prediction accuracy. The R^2 was improved by 0.2%, 0.5%, 0.8%, 1.0%, 1.3%.

Figure 9 shows a comparison of the prediction results for the last two days of the test set. LSTM and GRU show poorer predictions for the peaks and valleys of the load. Comparing BiLSTM and LSTM, it is clear that the former has a significantly better curve fitting performance, with its predictions more closely aligning with the original data, highlighting the advantage of BiLSTM in predictions. Comparing CNN-BiLSTM and BiLSTM, since CNN introduces additional feature extraction for the electricity data, the former's curve is less volatile than the latter's, and its accuracy is improved. Comparing CNN-BiLSTM-Attention and CNN-BiLSTM, with the help of the attention mechanism, the former demonstrates better prediction performance at the peaks and valleys. Between MSCNN-BiLSTM-Attention and CNN-BiLSTM-Attention, the prediction errors of these two models are roughly the same, but the former shows better prediction accuracy in areas with more significant load fluctuations.

4.7. **Model validation.** To further validate the generalization ability of the model proposed in this paper, it is applied to the prediction of electricity load data at the regional level in a foreign area. The data comes from a region in Spain, covering a period from January 1, 2015, to December 31, 2018, for a total of 3 years. It includes multi-dimensional electricity load data, as well as five types of meteorological data (temperature, wind speed, pressure, relative humidity, and rainfall), sampled at 1-hour intervals with 24 data points collected per day. The time-series data is sequentially divided into a 4 : 1 ratio, with the first 80% as the training set and the last 20% as the test set. The experimental software and hardware environment, as well as the model hyperparameter configuration, are the same as those in Section 4.3. The prediction results are shown in Table 6 and Table 7.

As shown in Table 6 and Table 7, the model selected in this paper achieves a reduction in MAPE of 0.22%, 0.56%, 0.79%, 0.91%, and 1.09% compared to the other models. RMSE is reduced by 14.26%, 27.19%, 29.11%, 33.52%, and 37.02%, while MAE is reduced by

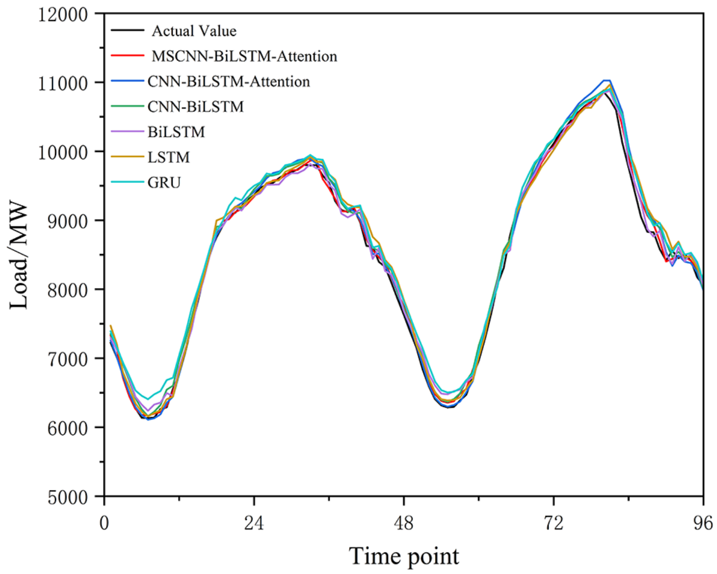

Figure 9. (color online) Prediction curves of different models

Table 6. Prediction errors of different models

Model \ Evaluation	MAPE/%	RMSE/MW	MAE/MW
MSCNN-BiLSTM-AM	1.12	476.527	316.474
CNN-BiLSTM-AM	1.34	555.757	379.195
CNN-BiLSTM	1.68	654.445	475.456
BiLSTM	1.91	672.183	526.054
GRU	2.03	706.219	538.255
LSTM	2.21	756.592	547.825

Table 7. Evaluation metrics of different models

Model \ Evaluation	R^2	Time/s
MSCNN-BiLSTM-AM	0.989	17400
CNN-BiLSTM-AM	0.984	14000
CNN-BiLSTM	0.979	12100
BiLSTM	0.978	10080
GRU	0.975	8600
LSTM	0.969	8450

16.54%, 33.44%, 39.84%, 41.20%, and 42.23%, respectively. The R^2 was improved by 0.5%, 1.0%, 1.1%, 1.4%, 2.0%.

Figure 10 shows a comparison of the prediction results for the last day of the test set. From the figure, it can be observed that the prediction curves for LSTM and BiLSTM exhibit high volatility. After introducing CNN, the curve fitting has significantly improved. Compared to CNN-BiLSTM, after incorporating the attention mechanism, CNN-BiLSTM-Attention shows a clear increase in prediction accuracy, and the smoothness of the prediction curve has also been greatly improved. Finally, MSCNN-BiLSTM-Attention uses different scales of convolution for the extraction of the feature information of the load, which gives good prediction in both peaks and troughs compared to other models.

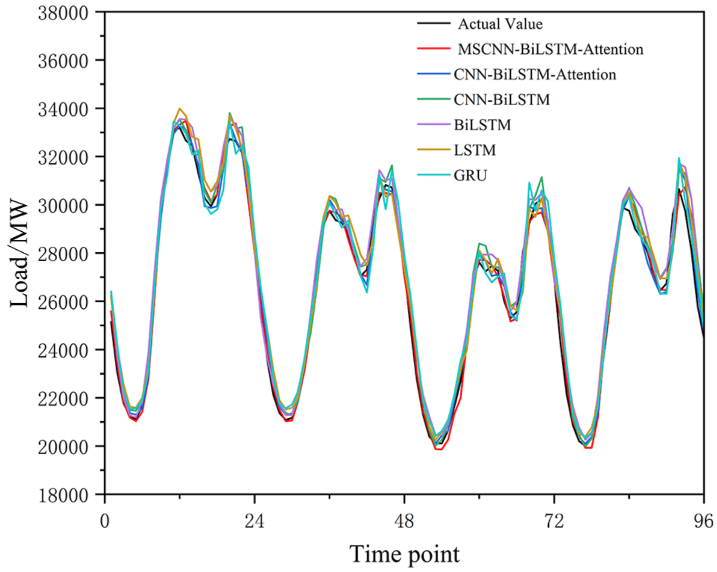

FIGURE 10. (color online) Prediction curves of different models

5. **Conclusion.** This paper first utilizes the maximum information coefficient (MIC) for feature selection on the dataset, and then employs the deep learning ensemble model MSCNN-BiLSTM-Attention for electricity load prediction on the selected dataset, effectively improving prediction accuracy. Through comparison with different models, the following conclusions are drawn.

1) MIC selects features with strong correlation to electricity load, which reduces the data dimensionality and filters out weakly correlated data, thereby improving the model's operational performance.

2) Compared to CNN with a single convolutional kernel size, MSCNN uses different scales of convolution on the load to extract feature information, which is beneficial to comprehensively extract the feature information. This enhances the model's generalization ability and robustness. The BiLSTM network compensates for the limitation of LSTM, which can only process time-series features in one direction. BiLSTM can process time-series data bidirectionally and perform time-series forecasting, achieving better prediction results. In time-series prediction tasks, the attention mechanism can accurately identify and select key temporal features. By integrating the attention mechanism, the model's accuracy in the prediction process is significantly improved.

3) The MSCNN-BiLSTM-Attention model proposed in this paper for short-term electricity load forecasting fully considers the advantages of different models. Through experiments and comparisons with other models, it significantly outperforms others in terms of prediction accuracy. In future research, intelligent optimization algorithms can be introduced to further optimize the model's hyperparameters. Additionally, considering the impact of other external factors on electricity load, other time-series data with high correlation to electricity load can be added as feature inputs to further improve the model's prediction accuracy.

Acknowledgment. This work is supported by Basic Research Program of Jiangsu Province Natural Science Foundation – Surface Project (BK20211235).

REFERENCES

[1] P. Xu, F. Zhang and J. Du, Application of deep learning methods in energy load forecasting, *Gas & Heat*, vol.43, no.4, pp.1-14, 2023.

[2] X. Kong, Y. Ma, Q. Ai et al., Review on electricity consumption characteristic modeling and load forecasting for diverse users in new power system, *Automation of Electric Power Systems*, vol.47, no.13, pp.2-17, 2023.

[3] D. Jiang, T. Li and W. Liu, Short-term power load forecasting using similar day and SAE-DBiLSTM model, *Journal of Electrical Engineering*, vol.17, no.4, pp.240-249, 2022.

[4] K. Li, L. Li and Y. Xiao, Research on power consumption forecasting and peak shaving during load period based on time series analysis, *Optics & Optoelectronics Technology*, vol.20, no.4, pp.153-159, 2022.

[5] M. Madhukumar, A. Sebastian et al., Regression model-based short-term load forecasting for university campus load, *IEEE Access*, vol.10, pp.8891-8905, 2022.

[6] Y. Wang, X. Shen, Q. Li et al., Forecasting of medium and long-term maximum power load for offshore oilfields based on PCA-GRD-LWR model, *Journal of China University of Petroleum (Edition of Natural Science)*, vol.47, no.2, pp.129-135, 2023.

[7] M. Zeng, X. Xiao and F. Xu, Short-term load forecasting method based on wavelet transform and BiGRU-NN model, *Electrical Measurement & Instrumentation*, vol.60, no.6, pp.103-109, 2023.

[8] B. Wei, G. Bao and Z. Li, Short-term electricity load forecasting based on support vector regression forecasting model considering weather factors and time-of-use tariff factors, *Power System and Clean Energy*, vol.39, no.11, pp.9-19, 2023.

[9] Y. Cai, Y. Zhang, S. Cao et al., Short-term power load big data forecasting model based on decision tree algorithm, *Manufacturing Automation*, vol.44, no.6, pp.152-155+182, 2022.

[10] Q. Shen, L. Mo, G. Liu et al., Short-term load forecasting based on multi-scale ensemble deep learning neural network, *IEEE Access*, vol.11, pp.111963-111975, 2023.

[11] Y. Qian, Y. Kong and C. Huang, Review of power load forecasting, *Sichuan Electric Power Technology*, vol.46, no.4, pp.37-43, 2023.

[12] Y. Meng, H. Zhang and X. Zhang, DeepRidge: A hop-layer deep learning model for detecting energy ridges in hoist operation noise spectrograms, *International Journal of Innovative Computing, Information and Control*, vol.21, no.2, pp.307-322, 2025.

[13] Z. Li, Y. Li, Y. Liu et al., Deep learning based densely connected network for load forecasting, *IEEE Transactions on Power Systems: A Publication of the Power Engineering Society*, vol.36, no.4, pp.2829-2840, 2021.

[14] W. C. Kong, Z. Y. Dong, Y. W. Jia et al., Short-term residential load forecasting based on LSTM recurrent neural network, *IEEE Transactions on Smart Grid*, vol.10, no.1, pp.841-851, 2019.

[15] N. Zeng, X. Xiao, F. Xu et al., A short-term load forecasting method based on CNN-BiGRU-NN model, *Electric Power*, vol.54, no.9, pp.17-23, 2021.

[16] Y. Zhang, Q. Ai, L. Lin et al., A very short-term load forecasting method based on deep LSTM RNN at zone level, *Power Grid Technology*, vol.43, no.6, pp.1884-1891, 2019.

[17] L. Guan, Modeling and simulation of power load forecasting based on improved DBN, *Microcomputer Applications*, vol.38, no.2, pp.42-45, 2022.

[18] H. Zhao, X. Shen, L. Lv et al., Load data restoration based on generative adversarial network and its application in short-term load forecasting of electric, *Automation of Electric Power Systems*, vol.45, no.16, pp.143-151, 2021.

[19] S. Ren, K. Yang, J. Shang et al., Short-term power load forecasting based on CNN-BiGRU-Attention, *Journal of Electrical Engineering*, vol.19, no.1, pp.344-350, 2024.

[20] H. Meng, T. Zhang, J. Wang, J. Zhang, D. Li and G. Shi, Multi-node short-term power load forecasting method based on multi-scale spatiotemporal graph convolution network and transformer, *Power System Technology*, vol.48, no.10, pp.4297-4305, 2024.

[21] J. Tian, H. Liu, W. Gan et al., Short-term electric vehicle charging load forecasting based on TCN-LSTM network with comprehensive similar day identification, *Applied Energy*, vol.381, 125174, 2025.

[22] K. Li, T. Pan and D. Xu, Short-term power load forecasting based on MSCNN-BiGRU-Attention, *Electric Power*, pp.1-9, https://link.cnki.net/urlid/11.3265.tm.20250102.1824.004, 2025.

[23] J. Ren, H. Wei and Z. Zou, Ultra-short-term power load forecasting based on CNN-BiLSTM-Attention, *Power System Protection and Control*, vol.50, no.8, pp.108-116, 2022.

[24] Y. Wang, Z. Liu, H. Dong et al., Multivariate load forecasting of integrated energy system based on CEEMDAN-CSO-LSTM-MTL, *Electric Power Construction*, vol.46, no.1, pp.72-85, 2025.

[25] B. Shao and D. Ji, Multi-feature short-term prediction of power load components based on VMD-SE, *Electric Power*, vol.57, no.4, pp.162-170, 2024.

Author Biography

Zhihua Xu received the B.Eng. degree in Electronic Information Engineering from Jiangnan University, China, in 2003, and the M.Eng. degree in Control Theory and Control Engineering with Jiangnan University, China, in 2011. He is currently a Lecturer with Yancheng Institute of Technology, China. His research interests include new energy generation technology, energy storage technology, electronic technology design and application and signal analysis and processing.

Tinglong Pan received his B.Eng. degree in Industrial Automation from China University of Mining and Technology, China, in 1999, and the Ph.D. degree in Power Electronics and Power Drive from China University of Mining and Technology, China, in 2004. He is currently a Professor at Jiangnan University, China. His research interests include microgrid control technology, power conversion technology, power drive system and its intelligent control technology.