EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

Short-Term Power Load Forecasting for Industrial Parks Using CNN-BiLSTM Network with Kernel Density Estimation-based Interval Prediction Method

Rui Hua[1], Tao Han[1], Yue Ran[1], and Qiongxiao Wu[2,*]

[1]Three Gorges Electric Energy Co., Ltd., Yichang, China
[2]Suzhou Tongyuan Soft Control Information Technology Co., Ltd., Suzhou, China

Abstract

Predicting electrical power loads in industrial parks is challenging due to the uncertainty and variance in model construction. This paper proposes an interval forecasting approach based on a residual-form CNN-BiLSTM, which effectively captures spatial patterns and bidirectional temporal dependencies in time series data. To handle uncertainty, a kernel density estimation (KDE)-based method is employed to generate probabilistic prediction intervals. Experiments using 18 days of real industrial data validate the model's performance. Results show superior robustness and accuracy compared with LSTM, BiLSTM, and GRU. The proposed model achieves perfect prediction interval coverage probability (PICP = 1.0), narrow normalized interval width (PINAW = 0.0828), and minimal CWC = 0.0828, while baseline models exhibit low coverage or excessively wide intervals. These findings confirm that the method provides both reliable and sharp uncertainty quantification, making it suitable for practical energy forecasting applications.

Keywords: load forecasting, neural network, kernel density function, probabilistic prediction interval

Received on 09 September 2025, accepted on 19 December 2025, published on 28 April 2026

doi: 10.4108/ew.12681

1. Introduction

Recent years have witnessed the tremendous development of electrical power load prediction in energy management [1]-[5], smart grid [6]-[11], and energy systems [12]-[16] due to its widespread applications. In the domain of the electrical power load, there are two classes of main problems: short-term and long-term prediction problems. Long-term electricity forecasting primarily focuses on predicting power load over extended time horizons, typically spanning several years to decades. Due to the influence of numerous uncertain factors, such as economic growth, demographic changes, policy shifts, technological advancements, and climate variability, achieving high accuracy in long-term load forecasting is inherently challenging. As a result, the objective of long-term load forecasting is not to provide exact figures, but rather to capture the general trends and directional changes in electricity demand. This helps guide power system planning, infrastructure development, and strategic decision-making. In contrast to long-term load forecasting, short-term load forecasting [17]-[20] focuses on accurately predicting the specific values of electricity demand over shorter time horizons, typically ranging from hours to a few days. The short-term aims for forecasting are to support real-time operations and industrial applications, such as

*Corresponding author. Email: catherinewus@163.com

power dispatching, unit commitment, demand response, and energy market trading, where precise load values are critical for efficient and reliable system operation.

As representative works, [21] presented a multistage artificial neural network-based short-term load forecasting system designed to support unit commitment and dispatch operations in utility systems. [22] proposed a comprehensive short-term load forecasting methodology for a U.S. power system, leveraging multiple meteorological forecasts from various commercial weather services. A semi-parametric additive modeling approach was proposed in [23] for short-term load forecasting, aiming to capture the relationship between electricity demand and influencing variables such as calendar data, lagged demand, and temperature forecasts. Then, a robust and accurate forecasting approach was proposed in [24] to integrate support vector machine and long short-term memory (LSTM) models with a dynamic weighting mechanism. Furthermore, [25] developed a novel two-stage hybrid short-term load forecasting method to address the rapidly growing electricity demand in China. [26] devised a time-series clustering-based probabilistic short-term load forecasting approach to support improved demand-side management in smart grids, accounting for weather and data noise uncertainties by outputting probabilistic load intervals for both individual and aggregated households.

Although these aforementioned studies have advanced the development of electrical power load prediction, they mainly adopt point forecasting approaches, which fail to capture the uncertainty and variability inherent in real-world electricity consumption., In real-world scenarios, the inherent uncertainty in power systems makes it challenging to rely solely on point forecasts. Consequently, interval load forecasting has gained attention as a more robust alternative. As a remedy, [27] presented a forecasting system for energy management systems that integrates uncertainty representation for both renewable generation and load forecasts. Using fuzzy prediction interval models, the approach generates some forecasts with expected targets and prediction intervals to capture future-variant bounds with specified coverage probabilities. Along this research line, [28] proposed a novel distribution-free learning model for interval load forecasting under attacks, integrating an isolation forest to detect and mitigate load anomalies caused by cyber threats, and an attention-enhanced gated recurrent unit (GRU) to capture both short- and long-term temporal dependencies. Expanding the horizon, [29]. Additionally, [30] employed reinforcement learning combined with neural networks to achieve online adaptive prediction interval generation, improving prediction interval quality under imbalanced data distributions and enhancing robustness against concept drift. However, the above studies do not consider the load variations in industrial parks, where power loads exhibit large fluctuations and irregular changes due to the movement of personnel. Moreover, due to the unique power load distribution in our industrial park, it is necessary to design a specialized load forecasting model. In consideration of this challenge, this paper develops a residual-form CNN-BiLSTM neural network structure. Then, based on the kernel function estimation method, an interval-forecasting method combined with the proposed CNN-BiLSTM is designed to generate the future prediction interval of the electrical power load. Then, the electrical power load data in our industrial park is sampled, and these data is applied to train our proposed CNN-BiLSTM, and the numerical simulation is conducted to verify the effectiveness of the present prediction method. Finally, some comparative simulations with the baselines LSTM, BiLSTM, and GRU are also conducted to illustrate the feasibility of the proposed method.

In brief, the contribution of this study is threefold:

- Develop a CNN-BiLSTM neural network structure to achieve the single-point prediction of electrical power load.
- Design a KDE-based method to generate the forecasting interval by the established CNN-BiLSTM.
- Sample the real electrical power load data in our industrial park, conduct the associated numerical simulation in our established model, and compare the results with the baselines LSTM, BiLSTM, and GRU.

The remainder of this paper is organized as follows. Section 1 introduces the necessary preliminaries and thereby formulates the electrical power load prediction problem. Section 2 develops the CNN-BiLSTM neural network structure and KDE-based interval forecasting method, and accordingly designs the KDE-based load prediction algorithm. Afterwards, numerical simulations are conducted in Section 3 to verify the effectiveness of the present interval forecasting method. Subsequently, Section 4 delves into the discussion of the numerical simulations by comparing results. Finally, a conclusion is drawn in Section 5.

2. Preliminary and Problem Formulation

2.1. Background of Power Load Forecasting in Industrial Parks

As shown in Figure 1, the total power load in industrial parks typically consists of several key components, each corresponding to different functions within the park. These components can be broadly categorized as: production equipment loads, lighting and auxiliary system loads, office and administrative loads, logistics and transportation system loads, standby and emergency system loads, and heating, ventilation, and so on. Here, production equipment loads are the dominant component due to the majority of energy consumption of the industrial parks, including heavy machinery, manufacturing lines, industrial robots, and other specialized production tools.

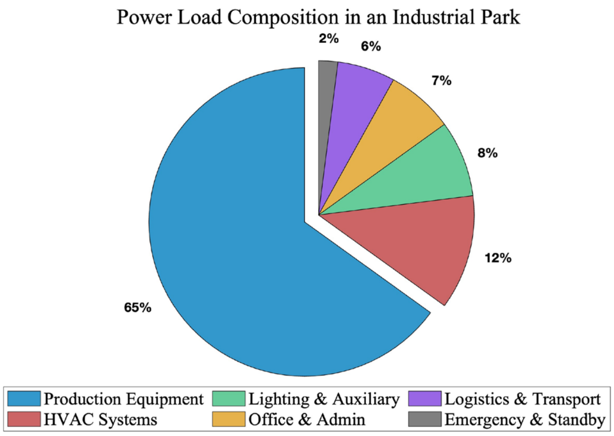

Figure 1. The component of the power loads in industrial parks

As shown in Figure 2, the temporal evolution of total power loads in our industrial parks from 0 AM – 24 AM in some days. Figure 2 depicts that the power load in the industrial parks remains consistently high during the period from 06:00 AM to 12:00 AM, followed by a noticeable decline from 13:00 PM to 22:00 PM. This diurnal load pattern indicates that the majority of industrial activities—and consequently, energy-intensive operations—are predominantly scheduled during the morning hours. Such temporal concentration of electricity demand suggests that production processes are front-loaded within the daily cycle. Moreover, the overall energy demand across the observed time intervals consistently exceeds 5000 units, underscoring the substantial and continuous energy requirements of the industrial parks. This sustained high-load profile reflects the intensive nature of industrial operations and presents significant implications for energy system planning, load forecasting, and infrastructure reliability.

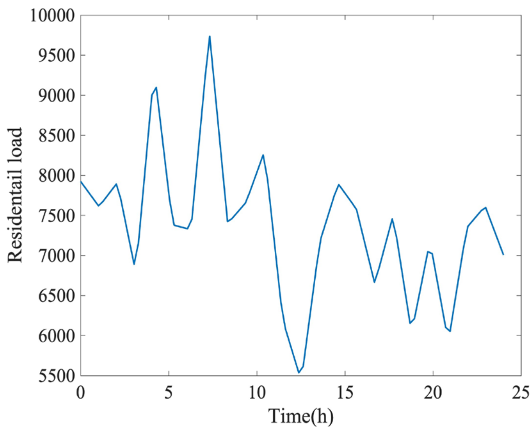

Figure 2. The evolution of the total power load in industrial parks, which is the sampled real data

2.2 Bi-LSTM network

A recurrent neural network (RNN) incorporates internal memory through recurrent connections, enabling the network to maintain and update a hidden state according to both current input and previous states, which is usually employed for prediction tasks, such as trajectory prediction, load prediction, and so on. However, standard RNN suffers from issues, mainly including gradient vanishing and exploding, which limit the effectiveness of RNNs in prediction tasks with long-term dependencies. So, the long short-term memory (LSTM) and gated recurrent unit (GRU) have emerged. Here, we employ the LSTM as the basic tool. As shown in Figure 3, its calculation is given below,

- The forget gate of LSTM: This component is used to decide how much previous information is necessary to forget, whose calculation is represented as below,
$$f_t = \sigma\left(W_f[h_{t-1}, x_t]^\top + b_f\right) \quad (1)$$
- where $\sigma(\cdot)$ stands for the sigmoid activation function, W_f the weight matrix, h_{t-1} the hidden layer state at the previous instants, x_t the current input state, b_f the bias, and the range of f_t is from 0 to 1.
- The input gate of LSTM: This component is employed to measure how much input information is necessary to enter the memory cells, whose calculation is given as below,
$$i_t = \sigma(W_i[h_{t-1}, x_t]^\top + b_i) \quad (2)$$
$$C_{i,t} = \tanh(W_C[h_{t-1}, x_t]^\top + b_C) \quad (3)$$
- where both W_i and W_C denote the weight matrices, both b_i and b_C the bias, and $\tanh(\cdot)$ the hyperbolic tangent function.
- The memory cell state of LSTM: This component aims to update the states of the memory cells, whose calculation is established as below,
$$C_t = f_t C_{t-1} + i_t C_{i,t} \quad (4)$$
- where C_t and C_{t-1} denote the cell memory states at current and previous states, respectively.
- The output gate of LSTM: This component is utilized to pick how much memory cell information is necessary to output, whose calculation is given by
$$O_t = \sigma(W_o[h_{t-1}, x_t]^\top + b_o) \quad (5)$$
- where W_o denotes the weight matrix, and b_o the bias.
- The hidden layer state of LSTM: This component calculates the final hidden layer state, whose formula is
$$h_t = \tanh(O_t) + C_t \quad (6)$$

Traditional LSTM networks process input time sequences in a single temporal direction—typically from the first element to the last. Bidirectional LSTM (BiLSTM) employs two-layer separable LSTM layers: the first layer processes the sequence from front to back (forward LSTM), and the second addresses it from back to front (backward LSTM). These two LSTM layers are then concatenated or otherwise combined at each time step to form a comprehensive representation of the input at that position. For instance, given a sequence dataset $X = \{x_1, x_2, \cdots, x_n\}$, the forward LSTM calculates the forward hidden layer states h_t^l from 1 to n, while the backward LSTM calculates the backward hidden layer states h_t^r from n to 1. The final hidden layer states are stacked as $h_t = [h_t^{l\top}, h_t^{r\top}]^\top$. As illustrated in Figure 4, the BiLSTM architecture consists of two parallel LSTM layers processing the same input sequence in opposite directions. The outputs from the forward and backward LSTMs are then merged and passed to subsequent layers, or say, fully connected layers.

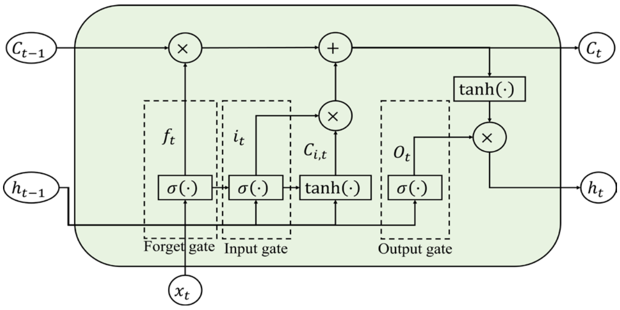

Figure 3. The calculation structure of the LSTM

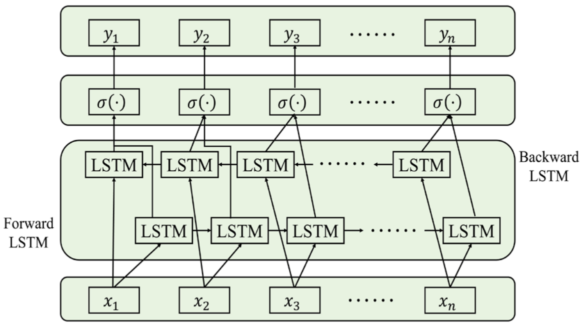

Figure 4. The calculation structure of the BiLSTM. Here, $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ denotes the n-dimensional input and output of BiLSTM, respectively

2.3 CNN Operation

A convolutional neural network (CNN) is used to extract hierarchical features by preprocessing raw input data, reducing the need for manual feature engineering. This capability has made them the backbone of modern computer vision systems and an essential component in tasks such as image classification and otherwise. In this work, CNN is employed to process the time-sequence load data to extract the data features, whose structure is shown in Figure 5.

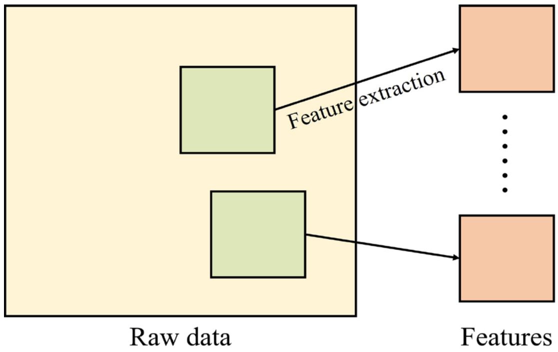

Figure 5. The illustration of the structure of the CNN

2.4 Problem Formulation

With these preliminaries about the backgrounds of industrial parks, LSTM network, and CNN in the subsections 2.1, 2.2, and 2.3, the problem addressed by this paper is naturally motivated as below,

Problem 1: Given a series of time sequence input data $\{X_1, X_2, \cdots, X_N\}$ about power load in industrial parks, design a neural network structure such that the following loss function is minimized,

$$L = \frac{1}{N}\sum_{i=1}^{N}\|X_i - \hat{X}_i\|^2 \tag{7}$$

where $\hat{X}_i$ denotes the predicted power load data.

3. Results

In this section, we are ready to establish a CNN-BiLSTM network architecture with an interval forecasting mechanism to address Problem 1. To this end, we first design a residual-form CNN-BiLSTM neural network structure, which is shown in Figure 6.

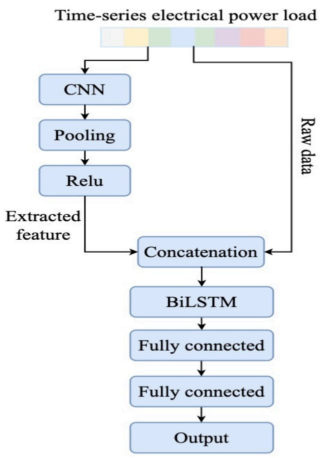

Figure 6. The structure of the designed residual-form CNN-BiLSTM

Then, we design an interval forecasting method as below. Compared with interval prediction, single-point prediction fails to capture the uncertainty inherent in many real-world forecasting problems. It provides a single deterministic estimate without reflecting the possible range of variation, making it less informative for risk-aware decision-making. Moreover, point predictions are more sensitive to noise and model errors, and offer limited metrics for evaluating predictive reliability.

Interval forecasting of electrical power load aims to predict both the upper and lower bounds of future load with a preset confidence level, thereby producing a prediction interval, which quantifies the uncertainty associated with the load forecasting. Reliable prediction intervals generated by interval forecasting offer critical insights and contribute to reducing operational risks in electrical power systems, thereby providing a more holistic representation of power load demand uncertainty. The flowchart structure of the interval forecasting is shown in Figure 7. Figure 7 exhibits that, for the time series power load from the input, we generate M interval forecasters, including a CNN-BiLSTM single-point power load forecaster and a kernel density estimation (KDE) method. Here, the KDE is used to estimate the probability density function of the input data with an unknown distribution. Since the interval forecasting method gives the range of the future electrical power load, this KDE-based interval forecasting method has some robustness for the uncertainty of the variance of the future electrical power load. Define the error of single-point forecasters for $j = 1,2,\cdots,M$ as $\mathbf{e} := [e_1, e_2, \cdots, e_M]^\top \in \mathbb{R}^M$, where $e_j := X_{j,i} - \hat{X}_{j,i}$ denotes the i-th electrical power load data training error of the j-th single-point forecaster (i.e., CNN-BiLSTM). So, according to the KDE method, one can calculate the probability distribution function as below,

$$f_k(\mathbf{e}) = \frac{1}{hM}\sum_{j=1}^{M} K\left(\frac{\mathbf{e} - e_j}{h}\right) \tag{8}$$

where $h > 0$ denotes the bandwidth parameter used to control the degree of smoothness, and $K(\cdot)$ the kernel function.

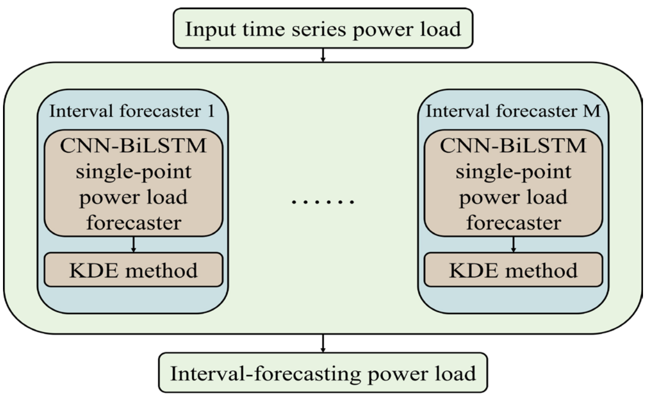

Figure 7. The illustration of electrical power load interval forecasting

Note that, the kernel function $K(\cdot)$ in the expression of $f_k(\mathbf{e})$ is necessary to be set. There are usually three types of kernel functions, i.e., triangle, normal, and Epanechnikov kernel functions. According to the different tasks, one can design the associated kernel functions as well. In this work, we mainly focus on the normal kernel function.

The calculation of the KDE-based interval forecasting is given below,

- Fit the KDE by the electrical power load training error so as to estimate the load error distribution.
- Obtain the future single-point electrical power load in the test set by the designed CNN-BiLSTM.
- Generate the lower and upper bounds of the predicted electrical power load by sampling from the KDE model.
- Evaluate the performance of the KDE-based interval forecasting by i) prediction interval coverage probability (PICP):

$$PICP = \frac{1}{M}\sum_{j=1}^{M} c_j \tag{8}$$

- where c_j is Boolean, and is 1 if the actual electrical power load is within the range of the forecasting interval, and 0 otherwise. ii) prediction interval normalized average width (PINAW):

$$PINAW = \sum_{j=1}^{M} U_j - L_j \tag{9}$$

- where U_j and L_j denote the upper and lower bounds of the forecasting interval, respectively. iii) coverage width criterion (CWC):

$$CWC = PINAW \times (1 + g\exp(-\lambda(PICP - \mu))) \tag{10}$$

- where λ denotes a tunable parameter, μ the desired confidence level, $g = 0$ if $PICP \geq \lambda$, and 0 otherwise.

Note that, if the actual electrical power load $\hat{X}_{j,i} \notin [L_j, U_j]$, then a re-forecasting mechanism begins, which implies that

$$\hat{X}_{j,i} = h(\hat{X}_{j,i}, \epsilon_{j,i}) \tag{11}$$

is running until $\hat{X}_{j,i} \in [L_j, U_j]$, where $\epsilon_{j,i} = \hat{X}_{j,i} - X_{j,i}$ and $h(\cdot)$ denotes the correction funciton. The pick of correction function $h(\cdot)$ is usually given by the meta-learner or set by experiences.

4. Numerical simulations

In this section, we present several numerical simulation examples to demonstrate the effectiveness and practicality of the proposed method. To ensure the relevance and reliability of the experiments, we utilize real-world electrical power load data collected from our industrial park.

Specifically, the dataset contains hourly power consumption records measured from January 1, 2025, to January 19, 2025. Each data point corresponds to the total electrical load observed during a one-hour interval, meaning that each day contains exactly 24 data points. The dataset, therefore, provides a detailed and high-resolution temporal profile of the energy consumption behavior within the industrial park over a period of 19 consecutive days.

In the implementation of the KDE-based prediction interval estimation, we employ the triangular kernel function due to its simplicity and effective smoothing performance in one-dimensional density estimation tasks. The overall network structure used for time series forecasting combines convolutional and bidirectional LSTM components to capture both local patterns and long-range temporal dependencies in the data. The architecture of the proposed model is illustrated in Figure 8, and the key parameter settings are summarized in Table 1.

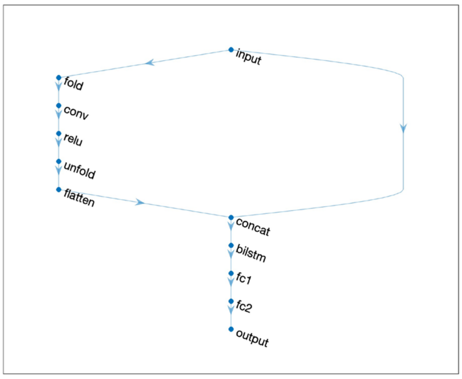

Figure 8. The structure of the CNN-BiLSTM

Table 1. The setting parameters of the proposed CNN-BiLSTM

Parameter Name	**Setting value**
Input layer size	10×1
CNN filter number	16
CNN filter size	5
BiLSTM units	200
FC layer 1 unit	200
FC layer 2 unit	3
Output layer size	3×1
Optimizer	Adam
Maximal epochs	1000
Gradient threshold	1
Initial Learning Rate	0.01
Learning rate dropping factor	0.15
Learning rate dropping period	400

Note that, as shown in Table 1, the size of the input layer is 10×1, which corresponds to the length of the input sequence used in the model. This is because we adopt a sliding window approach, where each input sample consists of 10 consecutive historical time steps of the electrical power load. The model uses these 10 time steps as context information to predict the next 3×1 values, i.e., the electrical power load for the subsequent three hours. This technique allows the model to capture short-term temporal dependencies and local patterns in the data, which are critical for accurate multi-step forecasting in time series problems.

It is worth noting that, in our model design, we intentionally avoid using a simple architecture with an input layer of size 1 and an output layer of size 1. Such a structure, while straightforward, suffers from significant limitations in the context of time series forecasting. Specifically, it fails to effectively capture the historical context of the data, as it relies solely on the most recent single observation to make predictions. This leads to poor modeling of temporal dependencies. Moreover, when this type of model is extended to multi-step forecasting, by recursively feeding the predicted outputs back into the model as new inputs, it tends to suffer from severe error accumulation. That is, even small prediction errors in the early steps can propagate and amplify over time, resulting in significant degradation of prediction accuracy for longer horizons.

The training process of the designed CNN-LSTM is shown in Figure 9. Figure 9 exhibits that the loss and RMSE of the proposed CNN-BiLSTM decrease as the iteration increases, and finally, the loss approaches 0.03 and the RMSE to 00.24, $with$the expression of RMSE as below,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{M} e_j} \tag{12}$$

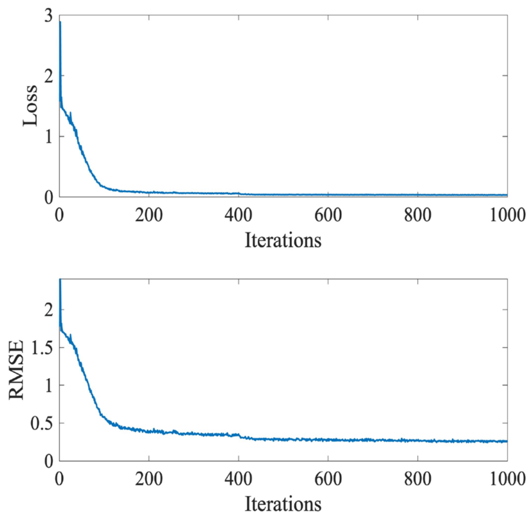

Figure 9. The training process of the CNN-BiLSTM

The single-point performance of the proposed CNN-BiLSTM model is illustrated in Figure 10. It can be observed that the training errors vary within a wide range, from as low as 0.013 to as high as 392.61. Notably, relatively large errors occur at the very beginning of the training sequence, indicating potential difficulties in model initialization or the presence of abrupt changes in the data at those time steps. However, for the majority of the training points, the prediction errors remain relatively small. As exhibited in Figure 11, most of the training errors fall below 100, suggesting that the proposed model is generally capable of capturing the temporal patterns in the data and fitting the training set effectively. This observation confirms the model's ability to learn the underlying data dynamics, with only a few outliers contributing to larger deviations in the early stage.

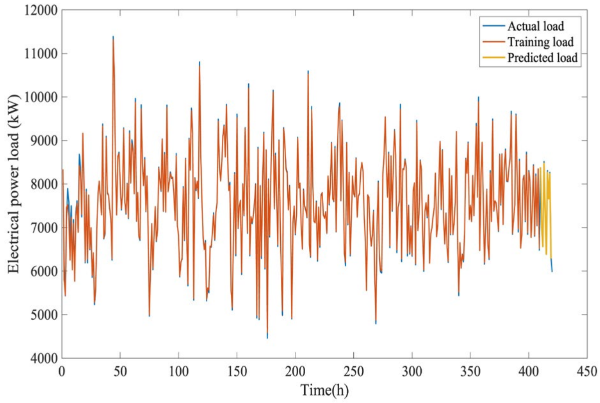

Figure 10. The single-point performance of the CNN-BiLSTM, which predicts the electrical power load of the future 10 instants

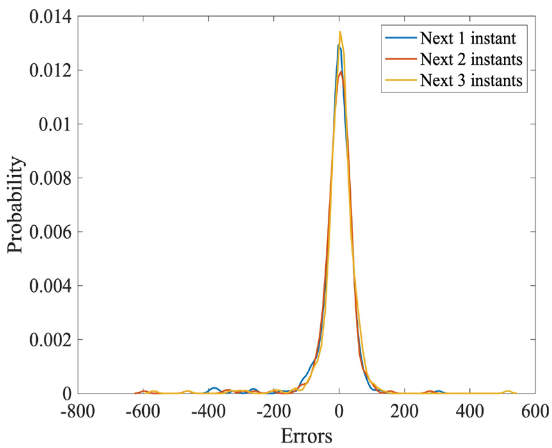

Figure 11. The distribution of the training errors of the CNN-BiLSTM

As shown in Figure 12, the proposed method demonstrates reliable interval forecasting performance. Specifically, the predicted intervals successfully capture the true load values with a confidence level of 95%, indicating that the model is capable of producing statistically meaningful uncertainty estimates. Moreover, as shown in Table 2, the width of the prediction interval remains within a reasonable range, with the upper and lower bounds deviating no more than 180 units from the predicted mean load. This suggests that the model maintains both high coverage probability and tight interval width, which are essential for practical applicability in power load forecasting. The evaluation of the PICP, PINAW, and CWC is shown in Table 3. Here, $\mu = 0.95$, $\lambda = 50$, and $g = 10$. As shown in Table 3, the proposed method achieves a perfect prediction interval coverage probability (PICP = 1), indicating that all actual load values are successfully captured within the predicted intervals. Furthermore, the average normalized width of the prediction intervals (PINAW = 0.0828) remains relatively small, suggesting that the intervals are not excessively wide. Consequently, the CWC metric also equals 0.0828, confirming that the model provides reliable and sharp uncertainty quantification without overestimation. The aforementioned results illustrate that the CNN component possesses strong feature extraction capabilities, which enable the model to capture local patterns and fluctuations in the power load data more effectively.

Table 2. The lower and upper bounds of the 95% confidence interval of the probability density functions generated by the KDE method

Predicted instant	Lower bound	Upper bound
Instant 1	-109.3	73.9
Instant 2	-97.9	81.5
Instant 3	-100.6	83.5

Table 3. The evaluation results of the PICP, PINAW, and CWC of the proposed CNN-BiLSTM

Evaluated index	Value
PICP	1
PINAW	0.0828
CWC	0.0828

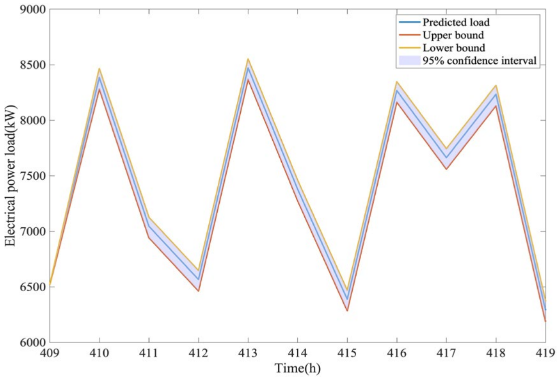

Figure 12. The interval forecasting performance of the CNN-BiLSTM

5. Discussion

To illustrate the effectiveness of the proposed CNN-BiLSTM model, we conduct a comparative study with three widely used baseline models: standard LSTM, bidirectional LSTM (BiLSTM), and gated recurrent unit (GRU). These models are selected due to their proven capabilities in time series forecasting tasks and serve as suitable benchmarks to evaluate the performance gains introduced by the convolutional feature extraction and bidirectional memory mechanisms in our proposed architecture. The comparison results are shown in Table 4.

Table 4. The comparison results of the PICP, PINAW, and CWC of LSTM, BiLSTM, GRU, and the proposed CNN-BiLSTM

Methods	PICP	PINAW	CWC
LSTM	0.6	0.5769	3.9×10^{8}
BiLSTM	0.5	0.001	5.9×10^{10}
GRU	0.4	0.3827	8.7×10^{12}
CNN-BiLSTM	**1**	**0.0828**	**0.0828**

Table 4 presents the comparison of uncertainty quantification performance among four different models, namely LSTM, BiLSTM, GRU, and the proposed CNN-BiLSTM. It is observed that the CNN-BiLSTM significantly outperforms the other models across all evaluation metrics. In particular, the proposed CNN-BiLSTM achieves a perfect prediction interval coverage probability (PICP = 1.0), meaning that all ground truth values fall within the predicted intervals. Moreover, the normalized average width of the interval (PINAW = 0.0828) remains reasonably small, indicating that the intervals are sharp and not overly conservative. As a result, the Coverage Width-based Criterion (CWC) also reaches its minimum possible value, 0.0828. In contrast, the baseline models perform poorly. While LSTM and GRU yield wider intervals (e.g., PINAW = 0.5769 and 0.3827, respectively), they fail to capture most true values (PICP = 0.6 and 0.4), leading to severely penalized CWC scores (e.g., 10^8 and above). BiLSTM produces extremely narrow intervals (PINAW = 0.001), but with a poor PICP of 0.5, showing that the predicted intervals are too tight and unreliable. These results clearly demonstrate the superior capability of the CNN-BiLSTM model in providing accurate and trustworthy interval predictions, thanks to its combination of convolutional feature extraction and bidirectional temporal learning.

6. Conclusion

To address the issue of predicting the electrical power load, this paper develops a KDE-based interval-forecasting CNN-BiLSTM time-series prediction method. Specifically, a normal kernel function is selected to generate the forecasting intervals and account for the uncertainty of the established model. Extensive simulation experiments have been conducted based on real-world power load data sampled in our industrial park. The proposed method achieves promising results in both point prediction and interval forecasting tasks. Particularly, it yields a perfect coverage probability (PICP = 1.0), a narrow average interval width (PINAW = 0.0828), and the minimum CWC value among all compared models. These results indicate that the proposed model not only makes accurate forecasts but also produces reliable and compact prediction intervals. Compared with traditional models such as LSTM, BiLSTM, and GRU, which either suffer from low coverage or overly wide intervals, the CNN-BiLSTM architecture demonstrates superior capability in capturing temporal dependencies and learning complex patterns in power load dynamics. The use of KDE further enhances its ability to quantify uncertainty in a data-driven manner, making the method more robust and practical for real-world energy management and decision-making scenarios.

Acknowledgements.
This research was funded by the Research and Demonstration Application of Intelligent Comprehensive Energy Planning and Simulation Design Platform, funded by China Yangtze Power Co., Ltd, grant number Z152302044/Z612302013.

References

[1] Quilumba FL, Lee W-J, Huang H, Wang DY, and Szabados RL. Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. IEEE Transactions on Smart Grid. 2015; 6(2):911–18. https://doi.org/10.1109/TSG.2014.2364233.

[2] Liu B, Nowotarski J, Hong T, and Weron R. Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. IEEE Transactions on Smart Grid. 2017; 8(2):730–37. https://doi.org/10.1109/TSG.2015.2437877.

[3] Liu J-X and Leu J-S. LARSI-Tpe-Xgb: Short-Term Load Forecasting by Load-Adaptive Relative Strength Index and Fusion of Tree-Structured Parzen Estimator and Xgboost. IEEE Transactions on Power Delivery. 2025; 40(3):1318–30. https://doi.org/10.1109/TPWRD.2025.3545638.

[4] Lauricella M and Fagiano L. Day-Ahead and Intra-Day Building Load Forecast with Uncertainty Bounds Using Small Data Batches. IEEE Transactions on Control Systems Technology. 2023; 31(6):2584–95. https://doi.org/10.1109/TCST.2023.3274955.

[5] Ren L, Deng J, Song W, Zio E, and Cattani C. A Fractional Brownian Motion Model for Forecasting Lost Load and Time Interval Between Power Outages. IEEE Access. 2021; 9:6623–32. https://doi.org/10.1109/ACCESS.2020.3048499.

[6] Fan S, Methaprayoon K, and Lee W-J. Multiregion Load Forecasting for System with Large Geographical Area. IEEE Transactions on Industry Applications. 2009; 45(4):1452–9. https://doi.org/10.1109/TIA.2009.2023569.

[7] Son J, Cha J, Kim H, and Wi Y-M. Day-Ahead Short-Term Load Forecasting for Holidays Based on Modification of Similar Days' Load Profiles. IEEE Access. 2022; 10:17864–80. https://doi.org/10.1109/ACCESS.2022.3150344.

[8] Azeem A, Ismail I, Jameel SM, and Harindran VR. Electrical Load Forecasting Models for Different Generation Modalities: A Review. IEEE Access. 2021; 9:142239–63. https://doi.org/10.1109/ACCESS.2021.3120731.

[9] Ranaweera DK, Karady GG, and Farmer RG. Effect of Probabilistic Inputs on Neural Network-Based Electric Load Forecasting. IEEE Transactions on Neural Networks. 1996; 7(6):1528–32. https://doi.org/10.1109/72.548183.

[10] Smyl S, Dudek G, and Pełka P. ES-dRNN: A Hybrid Exponential Smoothing and Dilated Recurrent Neural Network Model for Short-Term Load Forecasting. IEEE Transactions on Neural Networks and Learning Systems. 2024; 35(8):11346–58. https://doi.org/10.1109/TNNLS.2023.3259149.

[11] Faustine A, and Pereira L. FPSeq2Q: Fully Parameterized Sequence to Quantile Regression for Net-Load Forecasting with Uncertainty Estimates. IEEE Transactions on Smart Grid. 2022; 13(3):2440–51. https://doi.org/10.1109/TSG.2022.3148699.

[12] Park K, Yoon S, and Hwang E. Hybrid Load Forecasting for Mixed-Use Complex Based on the Characteristic Load Decomposition by Pilot Signals. IEEE Access. 2019; 7:12297–12306. https://doi.org/10.1109/ACCESS.2019.2892475.

[13] Madhukumar M, Sebastian A, Liang X, Jamil M, and Shabbir MNSK. Regression Model-Based Short-Term Load Forecasting for University Campus Load. IEEE Access. 2022; 10:8891–8905. https://doi.org/10.1109/ACCESS.2022.3144206.

[14] Khosravi A, Nahavandi S, and Creighton D. Construction of Optimal Prediction Intervals for Load Forecasting Problems. IEEE Transactions on Power Systems. 2010; 25(3):1496–1503. https://doi.org/10.1109/TPWRS.2010.2042309.

[15] Quan H, Srinivasan D, and Khosravi A. Incorporating Wind Power Forecast Uncertainties into Stochastic Unit Commitment Using Neural Network-Based Prediction Intervals. IEEE Transactions on Neural Networks and Learning Systems. 2015; 26(9):2123–35. https://doi.org/10.1109/TNNLS.2014.2376696.

[16] He Y, Zhu J, and Wang S. A Novel Neural Network-Based Multiobjective Evolution Lower Upper Bound Estimation Method for Electricity Load Interval Forecast. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2024; 54(5):3069–83. https://doi.org/10.1109/TSMC.2024.3352665.

[17] Gross G, and Galiana FD. Short-Term Load Forecasting. Proceedings of the IEEE. 1987; 75(12):1558–73. https://doi.org/10.1109/PROC.1987.13927.

[18] Christiaanse WR. Short-Term Load Forecasting Using General Exponential Smoothing. IEEE Transactions on Power Apparatus and Systems. 1971; PAS-90(2):900–911. https://doi.org/10.1109/TPAS.1971.293123.

[19] Vangipuram SR, and Giridhar AV. Short Term Residential Load Forecasting Using Temporal Weather Based Embedding Stacked Lstms. IEEE Latin America Transactions. 2025; 23(6):497–507. https://doi.org/10.1109/TLA.2025.11007195.

[20] Zhou Q, Tesfatsion L, and Liu C-C. Short-Term Congestion Forecasting in Wholesale Power Markets. IEEE Transactions on Power Systems. 2011; 26(4):2185–96. https://doi.org/10.1109/TPWRS.2011.2123118.

[21] Methaprayoon K, Lee W-J, Rasmiddatta S, Liao JR, and Ross RJ. Multistage Artificial Neural Network Short-Term Load Forecasting Engine with Front-End Weather Forecast. IEEE Transactions on Industry Applications. 2007; 43(6):1410–6. https://doi.org/10.1109/TIA.2007.908190.

[22] Fan S, Chen L, and Lee W-J. Short-Term Load Forecasting Using Comprehensive Combination Based on Multimeteorological Information. IEEE Transactions on Industry Applications. 2009; 45(4):1460–6. https://doi.org/10.1109/TIA.2009.2023571.

[23] Fan S, and Hyndman RJ. Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. IEEE Transactions on Power Systems. 2012; 27(1):134–41. https://doi.org/10.1109/TPWRS.2011.2162082.

[24] Muzumdar AA, Modi CN, M MG, and Vyjayanthi C. Designing a Robust and Accurate Model for Consumer-Centric Short-Term Load Forecasting in Microgrid Environment. IEEE Systems Journal. 2022; 16(2):2448–59. https://doi.org/10.1109/JSYST.2021.3073493.

[25] Wang Y, Xia Q, and Kang C. Secondary Forecasting Based on Deviation Analysis for Short-Term Load Forecasting. IEEE Transactions on Power Systems. 2011; 26(2):500–507. https://doi.org/10.1109/TPWRS.2010.2052638.

[26] Masood Z, Gantassi R, and Choi Y. Enhancing Short-Term Electric Load Forecasting for Households Using Quantile Lstm and Clustering-Based Probabilistic Approach. IEEE Access. 2024; 12:77257–68. https://doi.org/10.1109/ACCESS.2024.3406439.

[27] Sáez D, Ávila F, Olivares D, Cañizares C, and Marín L. Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids. IEEE Transactions on Smart Grid. 2015; 6(2):548–56. https://doi.org/10.1109/TSG.2014.2377178.

[28] Chen Z, Du C, Zhang B, Yang C, and Gui W. A Cybersecure Distribution-Free Learning Model for Interval Forecasting of Power Load Under Cyberattacks. IEEE Transactions on Industrial Informatics. 2025; 21(3):2540–9. https://doi.org/10.1109/TII.2024.3507946.

[29] Zhao C, Wan C, Song Y, and Cao Z. Optimal Nonparametric Prediction Intervals of Electricity Load. IEEE Transactions on Power Systems. 2020; 35(3):2467–70. https://doi.org/10.1109/TPWRS.2020.2965799.

[30] Zhang Y, Wen H, Wu Q, and Ai Q. Optimal Adaptive Prediction Intervals for Electricity Load Forecasting in Distribution Systems via Reinforcement Learning. IEEE Transactions on Smart Grid. 2023; 14(4):3259–70. https://doi.org/10.1109/TSG.2022.3226423.