Eng Int Syst (2024) 2: 149–158


**Engineering**
***Intelligent Systems***

# LSTM Power Load Peak Prediction Method Based on Bayesian Network

**Liyuan Sun*, Yuan Ai, Yiming Zhang and Jianyu Ren**

*Yunnan Power Grid Co., Ltd, China Southern Power Grid Co., Ltd, Kunming 650200, China*

In recent years, the development of the power sector has progressed rapidly, the peak power load is increasing, and the imbalance between power supply and demand is becoming more and more serious. Therefore, this research was conducted to predict the peak power load in order to provide effective response and solution measures. The current study used big data technology, Long Short-Term Memory (LSTM) network and Bayesian network as its theoretical basis, regularizes the design of loss function, optimizes the network weights, and finally obtains the LSTM electric load peak prediction model based on Bayesian network. Experiments were conducted to compare the performance of the proposed model with that of the traditional LSTM model. The model proposed in this study has a training time of 44 s for the integrated station data, and the prediction accuracy reaches 96.54% for the load peak percentage of 95.5% data. The performance results of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Relative Error (MRE), and Mean Absolute Error (MAE) are 115.947, 10.161, 0.027, and 9.656, respectively. The experimental results indicated that the LSTM peak power load prediction model based on the Bayesian network has high prediction accuracy, which confirms the validity and feasibility of the proposed model.

Keywords: Bayesian network; forbidden search algorithm; great likelihood estimation; LSTM; load peaks; prediction

## INTRODUCTION

As the demand for electricity continues to increase, accurate forecasting of peak electricity loads is vital to ensure the stability and security of electricity supply. However, there are relatively few studies on peak load forecasting, and there are also problems of poor accuracy and efficiency [1]. The traditional peak load prediction method also has obvious areas requiring improvement, one of which is the lack of statistics on the electricity consumption pattern of users in each region, without considering the specificity of electricity consumption in each region; the other is that when making peak load prediction, only the data of previous years have been considered, and there is a certain lag [2]. With the rapid development of technology for big data analysis, the electric load data can be extracted and transformed into effective information by various methods. Therefore, the current study constructed a Bayesian network based on the forbidden search algorithm and the great likelihood estimation method, and regularized the loss function design in the LSTM network, selected the RMSprop optimization function to adjust and optimize the network weights, and finally proposed a kind of LSTM electric load peak prediction model based on Bayesian network [3–4]. From the theoretical analysis, the prediction accuracy of this model can be improved to a certain extent, helping the load peak prediction to be more accurate and scientific. The aim of this research is to provide a new direction of inquiry to promote the development of electric power enterprises and solve various electric power problems.

## 1. RELATED WORK

Neural networks as an arithmetic model have more mature applications in pattern recognition, intelligent control, power

*Email of corresponding author: sunliyuan2006@126.com

management, biology, medicine and other fields, and can help solve many practical problems. Fu et al. [5] believed that signal interference or attenuation can affect the integrity and correctness of data information transmission, so the Bayesian network was applied to the development of fault diagnosis and troubleshooting for the probe card, which effectively improved the success rate of fault. Noor et al. [6] found that when assessing the condition of a pipeline, the traditional qualitative method has more serious subjectivity, which affects the calculation of pipeline fault probability. Hence, they proposed an improved Bayesian network model based on the table of probability conditions, and successfully constructed a pseudo-quantitative. Liu et al. [7] concluded that there is a significant non-linear relationship between lean adoption and organizational performance, which, if not explored accurately and in depth, could lead to the unsatisfactory application of lean tools, thereby negatively affecting corporate performance. By integrating the explanatory structural model and Bayesian network, the study analyzed and compared the trends of operational performance and environmental performance, and finally succeeded in proving a non-linear relationship between lean tools and operational and environmental performance, providing some reference for further exploration of their S-curve relationship. Aziz et al. [8] conducted an in-depth analysis of the flaws in the monitoring of complex healthcare processes, taking into account the different maternal health stages. The Bayesian network was the basis for the construction of a monitoring model that took into account the interaction between different maternal health care stages. The researchers used logistic regression models to estimate the Bayesian network parameters for different variables, which eventually succeeded in determining the importance of tracing the early health care stages and effectively improving maternal health care in a hospital setting. Kumari et al. [9] concluded that in industrial chemical processing, the key to effective troubleshooting is the accurate and unambiguous diagnosis of the root cause of the failure. This study used recurrent loops as an improvement based on Bayesian networks for achieving effective optimization of the Bayesian network probabilistic model, and finally obtained a reliable causal network structure that enables efficient decomposition and processing of recurrent networks within the time frame, substantially improving the accuracy and effectiveness of causality identification.

Sun et al. [10] innovatively applied Bayesian networks to the analysis of the severity of highway collisions involving vehicles carrying hazardous material, and the influencing factors, firstly using random forest algorithm to obtain the ranking of various traffic safety risk factors in order of importance, and then using Bayesian network algorithm to infer the probability of collision. The experimental results showed that during the road transportation of hazardous material driver fatigue and erratic driving, road features such as ramps or arch bridges, poor lighting conditions, and bad weather conditions can increase the severity of road accidents involving hazardous materials, verifying the reliability and accuracy of networks that use the Bayesian algorithm. Zhang et al. [11] improved and optimized convolutional neural networks to solve the difficult problem of fault diagnosis in the monitoring of the condition of machinery. Ghorbel and Souissi [12] concluded that in the digital advertising industry, the accurate prediction of click-through rate is crucial and can play a powerful role in increasing advertising profit and enhancing user experience. Based on this, scholars proposed a long- and short-term memory network prediction model incorporating genetic algorithm and compared it with other network hybrid models. He et al. [13] used the long and short-term memory network as the main means to evaluate the unloading relaxation estimation during the excavation of dams and the construction of embankments, and found that the unloading relaxation phenomenon and time effect were more significant in columnar jointed basalt compared to massive basalt. The researchers successfully verified the reliability and accuracy of the long and short-term memory network in the evaluation of unloading relaxation degree.

In summary, Bayesian networks and long- and short-term memory networks can be applied effectively in different fields. Therefore, this study integrates them organically and applies them to the prediction of electric load peaks to improve the accuracy of these peaks and provide reliable support for the stability and security of electric power supply.

## 2. BAYESIAN NETWORK-BASED LSTM POWER LOAD PEAK PREDICTION MODEL CONSTRUCTION

### 2.1 Construction of Bayesian Networks

The essence of the Bayesian network structure is a Directed Acyclic Graph (DAG) in which random variables are referred to as nodes, and the network can be called a Gaussian Bayesian Network (GBN) when and only when the node variables are continuous. Each variable contains a probability density distribution, and the probability density distribution of the root node variable is its edge probability density distribution. Bayesian networks are divided into structure construction and parameter learning. The network structure is used to express the qualitative relationship between the variable nodes, and the network parameters can express the quantitative relationship between the variables and their parents. Let the random variable be $X$, the probability density function of the variable is expressed by $f(x)$, and the joint probability density among the variables is shown by Equation (1) through Bayesian principle.

$$f(x) = \prod_{i=1}^{n} f_i(x_i|x_1, \cdots x_{i-1}) \tag{1}$$

In Equation (1), $f_i(x_i|x_1, \cdots x_{i-1})$ denotes the conditional probability density of the accompanying variable $X$ at each case. The probability density has some correlation with the parent node, so the joint probability density function is expressed as Equation (2).

$$f(x) = \prod_{i=1}^{n} f_i(x_i|x_{pa(i)}) \tag{2}$$

In Equation (2), $pa(i)$ denotes the parent node. This type of computational equation can deal with discrete variables and can also solve problems with continuous variables.

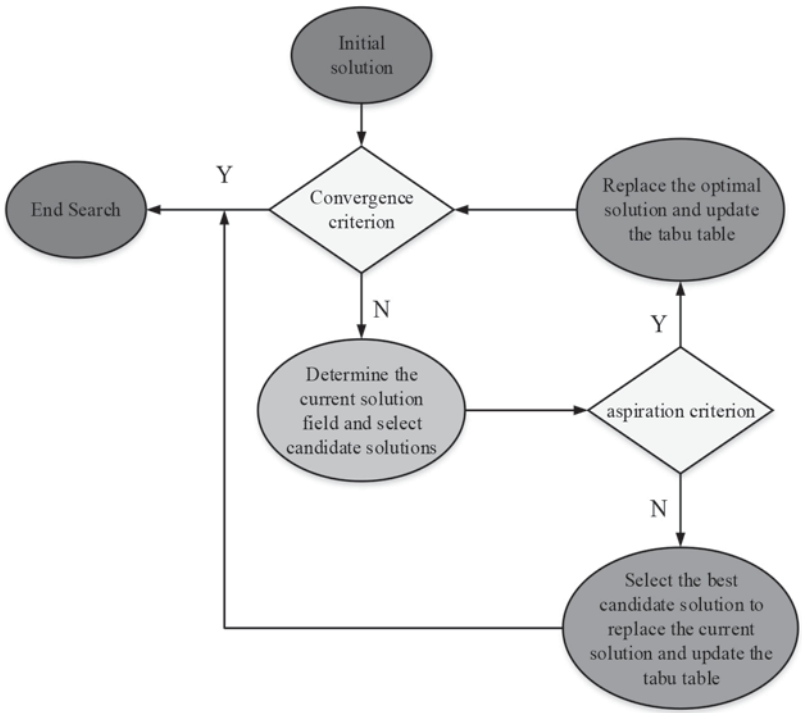


**Figure 1** Taboo search algorithm flowchart.

The electric load addressed in the study is a continuous variable problem, so GBN can be chosen for modeling. If the continuous variables of the load satisfy the Gaussian distribution, the conditional and joint densities of the variables can be represented by Equation (3) using Bayes' theorem.

$$\begin{cases} p(X|Y) = \frac{p(X,Y)}{p(Y)} \\ f(x|y) = \frac{f(x,y)}{f(y)} \end{cases} \tag{3}$$

The conditional probability density of the multidimensional normal distribution can be solved simultaneously by the probability density and distribution formulas, and the problem of solving the posterior probability density of the multidimensional normal distribution can be transformed into the problem of finding the probability density of the joint nodes. The joint probability density formula is shown in Equation (4).

$$f(X) = (2\pi)^{\frac{-n}{2}} |C|^{\frac{-1}{2}} \exp\{-(X-\mu)^T C^{-1}(X-\mu)\} \tag{4}$$

In Equation (4), $\mu$ denotes the mean vector of variables; $C$ denotes the $n * n$ dimensional covariance matrix. When the node and the parent node are known, the joint probability density can likewise be decomposed by the conditional probability density of the parent node, as expressed in Equation (5).

$$f(x_i|pa(x_i)) \sim N(x_i|\mu_i \sum_{i=1}^{i-1} \beta_{ij}(x_j - \mu_j), v_i) \tag{5}$$

In Equation (5), $\beta_{ij}$ denotes the regression coefficient between the variable and the relative parent node variable; $v$ denotes the conditional variance of the variable in the case of parent determination. The expression of the covariance matrix is shown in Equation (6).

$$C = \left[(I-B)^{-1}\right]^T D(I-B)^{-1} \tag{6}$$

In Equation (6), $D$ denotes the diagonal array of conditional variances; $B$ denotes an upper triangular array in $\beta_{ij}$; and $I$ denotes the unit matrix. The qualitative relationships of the above variables are then the Bayesian network structure.

The algorithm for Bayesian network structure learning can be a taboo search algorithm based on Bayesian Information Criterion (BIC) scoring, which is shown in the flowchart in Figure 1.

In Figure 1, the forbidden search algorithm is used to establish the parameters and the initial solution first, and determine whether the initial solution satisfies the convergence criterion. If it does not satisfy the convergence criterion, the algorithm stops searching; if the initial solution satisfies the convergence criterion, the algorithm screens the initial solution to determine the candidate solutions. After the algorithm produces the candidate solutions, it determines whether each candidate solution satisfies the contempt criterion. And the solution that satisfies the criterion becomes the new optimal solution and is assessed according to the convergence criterion. For the candidate solutions that do not satisfy the contempt criterion, the algorithm selects the current optimal solution from the candidate solutions, updates the contraindication table, and repeats the above steps. The study interprets the BIC criterion with Equation (7).

$$\begin{cases} f(X|Y) = \left[(2\pi)^{\frac{N}{2}} |S|^{\frac{1}{2}}\right]^{-1} \exp\left\{-\frac{(X-WY-U)^{T+1}}{2S}\right\} \\ W = S_{XY} S_{XY}^{-1} \\ S = \begin{bmatrix} S_{XX} & S_{XY} \\ S_{XY} & S_{YY} \end{bmatrix} \\ U = [U_X - WU_Y] \end{cases} \tag{7}$$

In Equation (7), $f(X|Y)$ denotes the Gaussian probability density function of $X$ conditional on the variable $Y$; $X$ and $Y$ denote the variables in the function; $S$ denotes the covariance

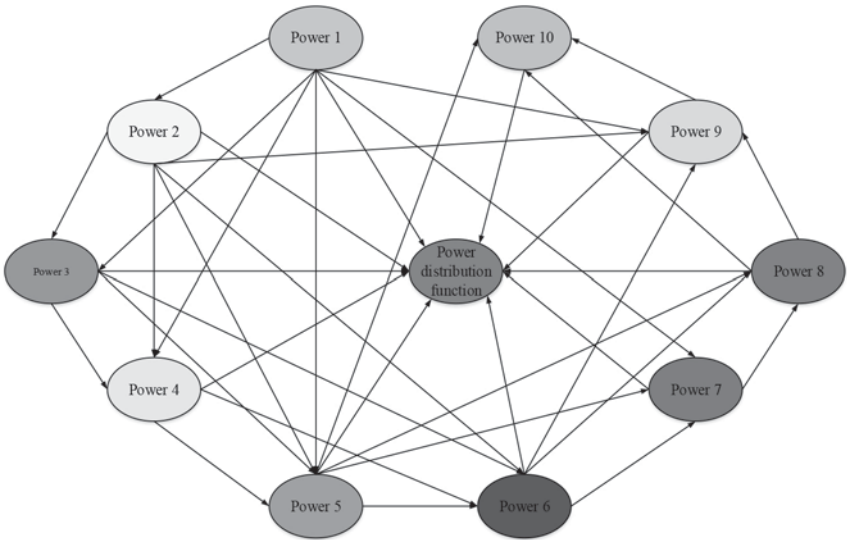


**Figure 2** Bayesian network structure based on Tabu search algorithm and BIC criterion.

matrix between the variables; $U$ denotes the mean; and $N$ denotes the length. Assuming that there are independent and identically distributed training samples $m$, the training sample log-likelihood function is expressed in Equation (8).

$$L = \frac{N}{2}\log|S| - \frac{1}{2}\sum_{i=1}^{N}\left[-\frac{1}{2}(X - WY - U)^T \times S^{-1}(X - WY - U)\right] \tag{8}$$

With the data mean, covariance and weights determined, the Gaussian distribution parameters are obtained and the Bayesian network score is obtained with Equation (9).

$$Score = L - \frac{M}{2}\ln(p) \tag{9}$$

In Equation (9), $M$ denotes the nodes associated with the probability density function; $p$ denotes the nodes of the Bayesian network. The structure of the above Bayesian network constructed based on the forbidden search algorithm and the BIC criterion is shown in Figure 2.

The parameter learning of Bayesian networks is carried out after determining the network structure. The methods of parameter learning commonly include maximum likelihood estimation and Bayesian estimation. In this study, the maximum likelihood estimation method was selected for network parameter learning.

## 2.2 Construction of long- and short-term memory neural networks

Long Short-Term Memory (LSTM) is an optimization model proposed to overcome the defects of Recurrent Neural Network (RNN) model [14–16]. The LSTM model is optimized by including a memory unit and a gating unit to preserve the information based on the RNN model. The structure of the LSTM is depicted in Figure 3.

In Figure 3, there is a separate memory unit in the LSTM neuron that follows the time series to record the valid information during the operation of the model, while being able to cascade the computation of the whole structure [17–18]. There are specific gates in neurons: Input Gate (IG), Forget Gate (FG) and Output Gate (OG) [19–20]. The expression of gating is shown in Equation (10).

$$\begin{cases} g(x) = \sigma(\theta x + b) \\ \sigma(x) = \frac{1}{1+\exp(-x)} \end{cases} \tag{10}$$

In Equation (10), $\sigma$ denotes the Sigmoid function, which maps any number in the interval [0, 1]; $x$ denotes the input of the structure; $\theta$ denotes the network matrix in the network structure; $b$ denotes the bias vector. When the information is passed from the memory unit, the information goes from the current neuron to the new neuron, the FG will control the network to forget part of the information of the neuron. And the IG will filter the processed information and add to that memory unit, and finally the OG filters the information again to obtain the output value of the hidden layer of the network, and the output result obtained by the current neuron is the input value of the neuron at the next iteration. The LSTM and the forward computation process can be described by a mathematical expression, as shown in Equation (11).

$$\begin{cases} f_t = \sigma(\theta_f[h_{t-1}, x_t] + b_f) \\ i_t = \sigma(\theta_i[h_{t-1}, x_t] + b_i) \\ o_t = \sigma(\theta_o[h_{t-1}, x_t] + b_o) \end{cases} \tag{11}$$

In Equation (11), $f_t$ denotes FG; $h_{t-1}$ denotes the output of the neuron at the moment of $t-1$; $i_t$ denotes IG; $o_t$ denotes OG; and $c_t$ denotes the control memory unit. FG is the control of the information flowing from the memory unit to the current memory unit at the previous iteration; IG is the filtering and saving of the input information; and OG is the control of the information output in the time step. The input value of the information added to the memory cell is calculated with Equation (12).

$$\tilde{c}_t = \tanh(\theta_c[h_{t-1}, x_t] + b_c) \tag{12}$$

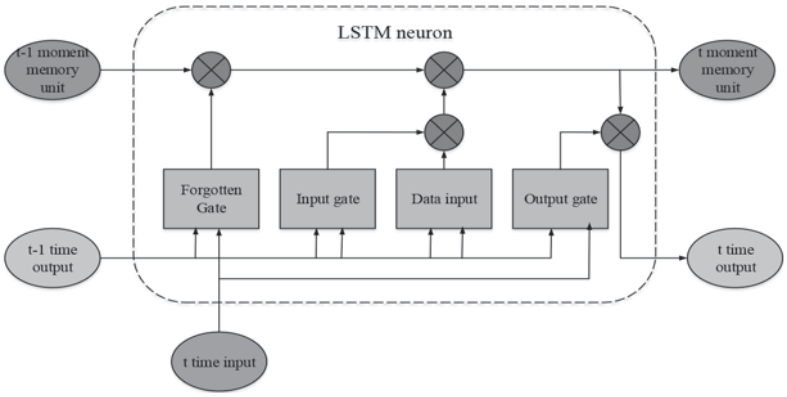


**Figure 3** Specific structure of LSTM and diagram of internal structure

In Equation (12), tanh denotes the activation function for normalization. The memory cell is updated as shown in Equation (13).

$$\tilde{c}_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \tag{13}$$

In Equation (13), $\tilde{c}_t$ denotes the updated memory unit; $\otimes$ denotes the corresponding element multiplication. The output of the neuron hidden layer is shown in Equation (14).

$$h_t = o_t \otimes \tanh(c_t) \tag{14}$$

The weights and bias parameters of the LSTM neurons need to be optimized cyclically by the loss function and the optimization function, and are generally optimized using the Back Propagation Through Time (BPTT) algorithm, which derives the gradient vertically while propagating the gradient horizontally along the time axis. Since the weights are shared at each iteration, the gradients of the weight parameters are also updated multiple times, and the total gradient after multiple updates is finally taken as the result of one weight update. Let the training samples be $N(x^\sigma, Y^\sigma)$; the input values are represented by $x^\sigma = (x_1^\sigma, x_2^\sigma, \cdots, x_m^\sigma)^T$; the labels are represented by $y^\sigma = (y_1^\sigma, y_2^\sigma, \cdots, y_m^\sigma)^T$; and the predicted output values are represented by $o^\sigma = (o_1^\sigma, o_2^\sigma, \cdots, o_m^\sigma)^T$. The study uses Euclidean distance to calculate the loss function of the output and the label. In the constructed LSTM network, the stacked linear cascade is the main connection structure. If the loss function in the structure meets the requirements, the model parameters will not produce large changes, which in turn will prevent fluctuations in the prediction results and will not converge the model with a large amount of composite data. To ensure that overfitting of the model occurs, and that the model has better adaptability and balance, regularized probabilities are included in the loss function. The loss function of the model is expressed by Equation (15).

$$Loss = \frac{1}{2}\sum_{i=1}^{m}(o_i^\sigma - y_i^\sigma)^2 + \frac{1}{2}\alpha\sum_{i=1}^{m}\left(\frac{1}{\theta}\right)^2 \tag{15}$$

In Equation (15), $\alpha$ represents the introduced regularization factor, which takes the value of 0.0001; $\theta$ represents the weight. The model adjusts the weight of each layer accordingly, and the layer with too high a weight needs to be reduced appropriately, while the layer with low weight needs to be increased appropriately, so as to ensure that the weight of each layer can contribute to the optimization. The optimization function of the LSTM network model was chosen as the forward root mean square gradient descent algorithm (RootMean Square Propagation, RMSprop). This algorithm introduces decay coefficients and gradient accumulation which can be used to solve the problems of gradient oscillation, slow convergence and local optimality in stochastic gradient descent algorithms. The forward root-mean-square gradient descent algorithm is good for training convolutional neural networks and includes the concept of mini-batch in the algorithm. The result of each update of the weights is the average of the sum of the batch gradients, which not only reduces the gradient oscillation, but also optimizes the gradient descent direction. The gradient size is calculated using Equation (16).

$$g = \frac{1}{m}\nabla_\theta \sum_{i=1}^{m} L(x_i, \theta, y_i) \tag{16}$$

In Equation (16), $L(x_i, \theta, y_i)$ denotes the loss function; $\nabla_\theta$ denotes the derivative of the weights; $m$ denotes the training samples. In the calculation of the gradient accumulation, the decay coefficient is used to decay the previous gradient accumulation to complete the update of the gradient accumulation. The update is shown in Equation (17).

$$r = \rho r(1-\rho)g \otimes g \tag{17}$$

In Equation (17), $\rho$ denotes the decay coefficient; $r$ denotes the gradient accumulation. The updated values of the calculated weights are calculated with the updated weights as shown in Equation (18).

$$\begin{cases} \Delta\theta = -\frac{\eta}{\sqrt{\delta + r}} \otimes g \\ \theta = \theta + \Delta\theta \end{cases} \tag{18}$$

In Equation (18), $\eta$ denotes the learning rate; $\delta$ denotes the small constant, which guarantees the stability of the value. The above step-by-step approach then provides a more comprehensive introduction to the LSTM power load peak prediction model based on Bayesian networks.

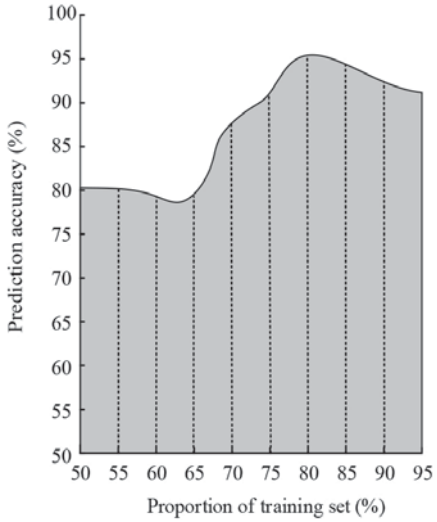


**Figure 4** Model sample division prediction results.

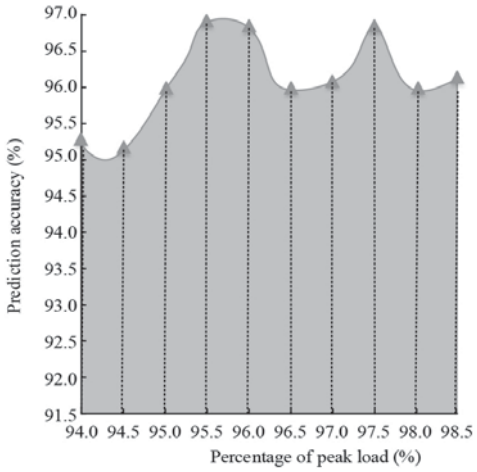


**Figure 5** Forecast results of different load peak percentage.

## 3. BAYESIAN NETWORK-BASED LSTM POWER LOAD PEAK PREDICTION ANALYSIS

### 3.1 Electricity Load Peak Sample Selection

The purpose of determining the model sample's division ratio is to find the best allocation ratio between the training sample set and the test sample set. Due to the limited number of data samples, it is necessary to achieve a high prediction accuracy while ensuring sufficient training samples. The annual peak load data of each type of station in a city from 2002 to 2021 were selected, and the proportion of the training sample set was adjusted to 50%, 60%, 70%, 80%, and 90% of the total sample, while the remaining proportion was the test sample set. The study was conducted to explore the effects of different proportions on the accuracy of the prediction model. The annual peak load prediction of distribution transformers in the station area is based on the peak load data of the previous year to predict the peak load of the next year whose prediction results with different sample division proportions are shown in Figure 4.

In Figure 4, the prediction accuracy of the Bayes-LSTM load prediction model is related to the percentage of training samples, and the model prediction accuracy is positively correlated with the percentage value in the interval of 50%–80% of training samples; the model accuracy shows a certain volatility in the interval of 80%–90% of training samples. The highest model prediction accuracy was achieved when the training sample share was around 80% and the prediction accuracy reached 96.20%. Therefore, the study fixed the partition ratio between training and testing samples at 8:2, and conducted predictive analysis on different peak load percentages to explore appropriate sample peak ratios.

In Figure 5, the prediction results of the model for different load peak percentages have certain volatility, where the highest value of prediction accuracy is 96.54%, and the load peak percentage is 95.5% at this time. For the training of the Bayes-LSTM load prediction model, the loss function was selected as the Euclidean distance after regularization, and the optimization model was selected as the RMSprop algorithm as it does not require too much load data for training, and the model is close to convergence after 50 times of batch training.

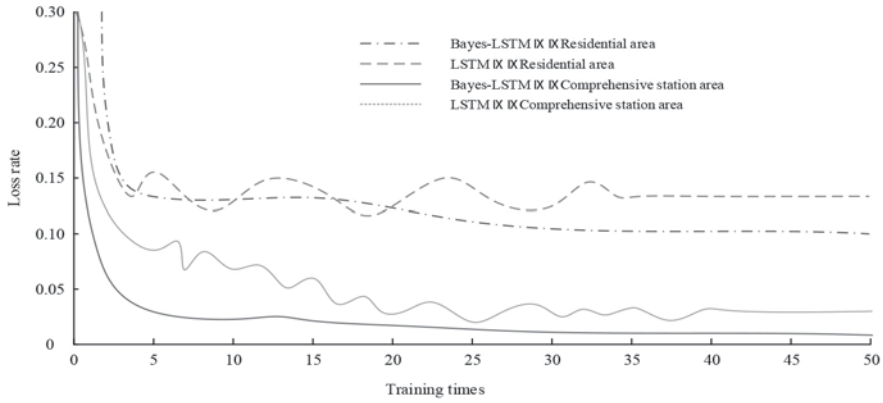


**Figure 6** Curve change diagram of loss function during model training.

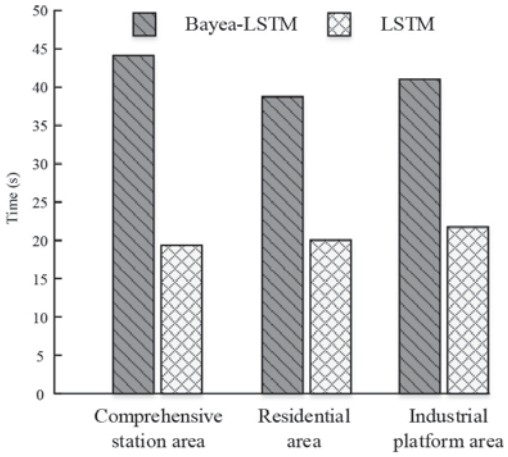


**Figure 7** Comparison of model training time results.

Figure 6 shows the variation of the loss function curve of the model during training. The Bayes-LSTM model was selected for the experiments to compare its performance with that of the traditional LSTM temporal prediction model. The loss function curves of both models converge in the interval of [0.10, 0.15] for the residential station data and in the interval of [0.00, 0.05] for the integrated station data. However, the Bayes-LSTM model has a lower loss rate and is more stable. The results indicate that the proposed model has more advantages in breaking through local optima.

For the experiments conducted to obtain a more comprehensive evaluation of the results, the TIME metric was selected to compare the Bayes-LSTM load prediction model with the traditional LSTM model. To ensure the reliability of the results, all model runs were implemented on the PyTorch platform.

In Figure 7, the training time of the Bayes-LSTM model is about 44 s for the integrated station area; about 38 s for the residential station area; and about 41 s for the industrial station area. The training time of the LSTM model is about 22 s for the integrated station area; about 19 s for the residential station area; and about 20 s for the industrial station area. The results of a comprehensive analysis show that the Bayes-LSTM model requires more training time because the network is more complex.

## 3.2 Bayes-LSTM Model Prediction Performance Analysis

The prediction accuracy of the Bayes-LSTM model proposed in the study is assessed by Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Relative Error (MRE), and Mean Absolute Error (MRE). Absolute Error (MRE) and Mean Relative Error (MRE) were used as judgment criteria. The experiments were conducted to validate the data from the integrated, residential and industrial stations respectively.

Figure 8 shows the comparison of the metrics of the models in the integrated station. In Figure 8(a), the Bayes-LSTM model gradually stabilizes and finally settles at 115.947 after performing the fifth MSE calculation; the LSTM model stabilizes at 207.337 after performing the seventh MSE calculation. In Figure 8(b), the Bayes-LSTM model gradually stabilizes and finally settles at 10.161 after performing the sixth RMSE calculation; the LSTM model stabilizes at 13.671 after performing the seventh MSE calculation. In Figure 8(c), the Bayes-LSTM model gradually stabilizes and finally settles at 10.161 after performing the sixth RMSE calculation. At 10.161, the LSTM model stabilizes at 13.671 after performing the seventh MSE calculation. In Figure 8(c), the Bayes-LSTM model gradually stabilizes and finally stabilizes at 0.027 after performing the third MRE calculation; the LSTM model stabilizes at 0.038 after performing the third MRE calculation.

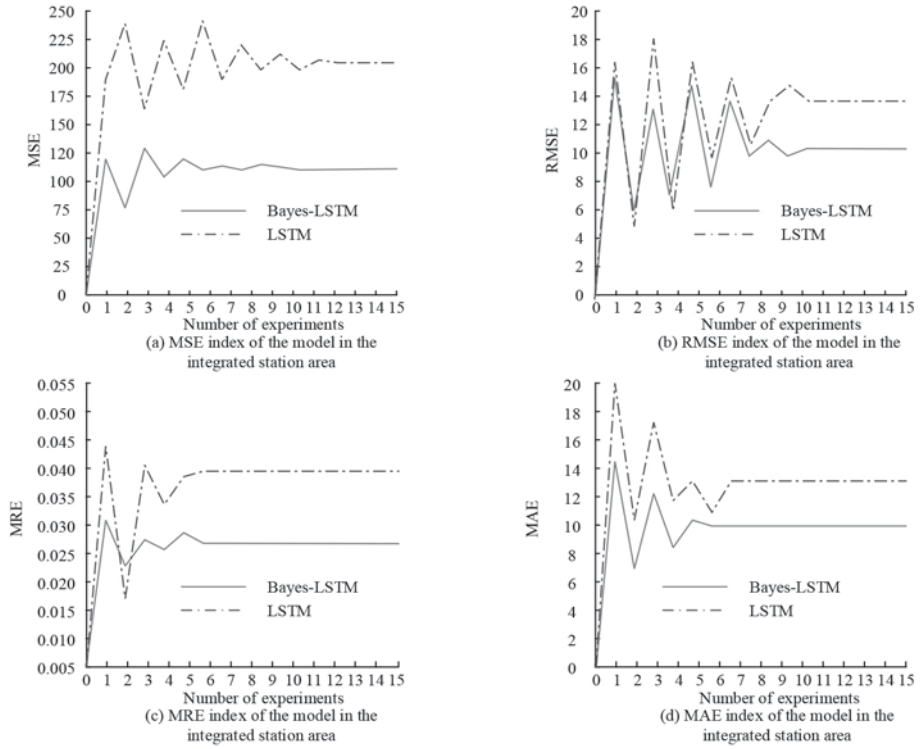


**Figure 8** Index comparison of the model in the comprehensive station area.

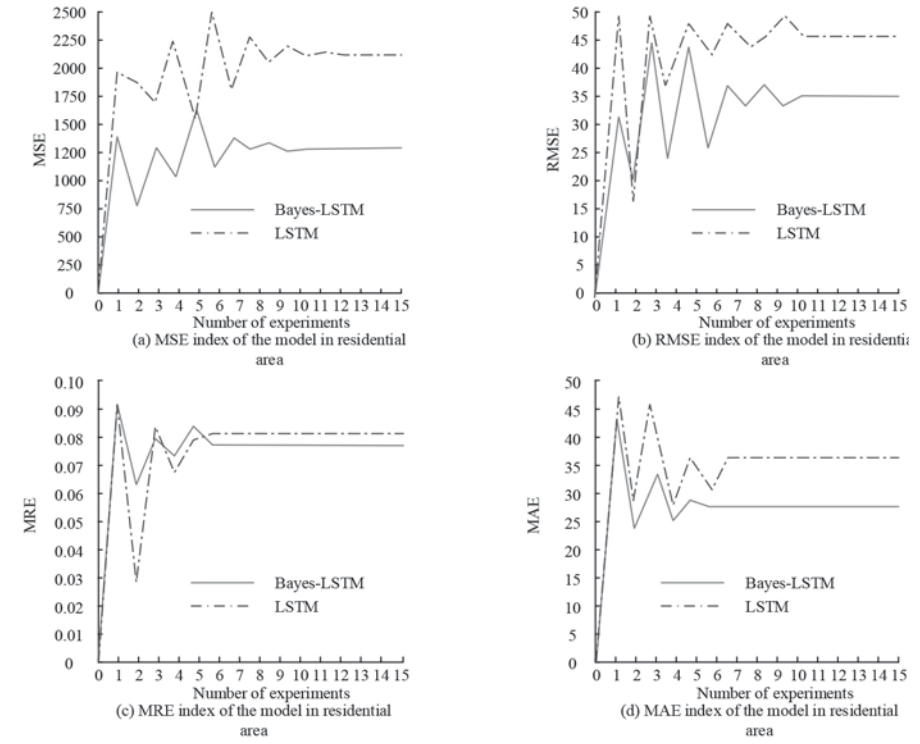


**Figure 9** Comparison of indicators of the model in residential areas.

In Figure 8(d), the Bayes-LSTM model gradually stabilizes its value after the fourth MAE calculation and finally stabilizes at 9.656; the LSTM model stabilizes at 12.088 after the fifth MSE calculation. The results show that the Bayes-LSTM model has a higher prediction accuracy for the integrated station data.

Figure 9 shows the comparison of the metrics of the models in the residential stations. In Figure 9 (a), the Bayes-LSTM model gradually stabilizes its value after the seventh MSE calculation and finally stabilizes at 1167.703; the LSTM model stabilizes at 2006.361 after the tenth MSE calculation. In Figure 9 (b), the Bayes-LSTM model gradually stabilizes its value after the sixth RMSE calculation and finally In Figure 9(c), the Bayes-LSTM model gradually stabilizes after the third MRE calculation and finally stabilizes at 0.0767; the LSTM model stabilizes at 0.0818 after the third MRE calculation. In Figure 9(d), the Bayes-LSTM model gradually stabilizes its value after the fifth MAE calculation and finally

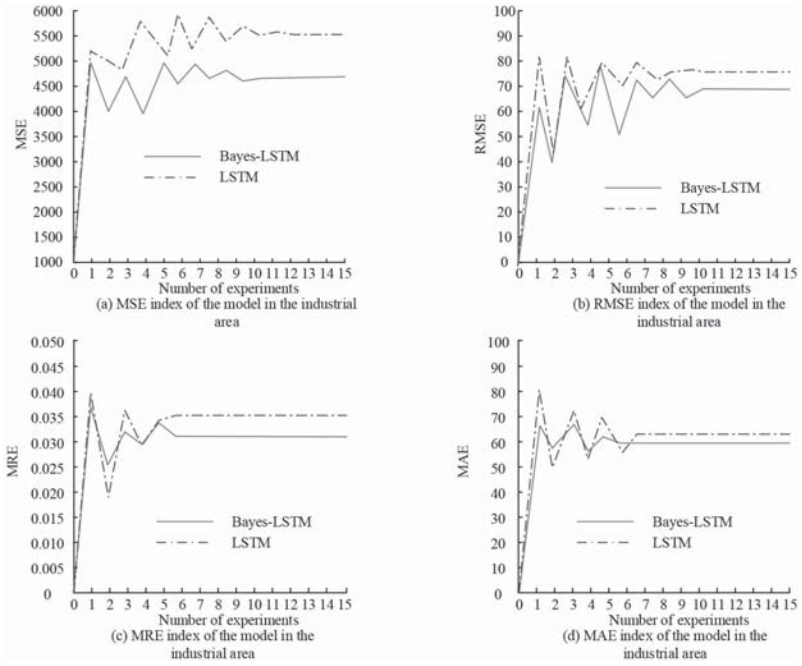


**Figure 10** Index comparison of the model in the industrial area.

stabilizes at 28.424; the LSTM model stabilizes at 34.185 after the sixth MSE calculation. The results show that the Bayes-LSTM model has a higher prediction accuracy for the residential station data.

Figure 10 shows the comparison of the metrics of the models in the industrial station area. In Figure 10(a), the Bayes-LSTM model gradually stabilizes its value after the tenth MSE calculation and finally stabilizes at 4697.656; the LSTM model stabilizes at 5515.839 after the twelfth MSE calculation. In Figure 10(b), the Bayes-LSTM model gradually stabilizes its value after the seventh RMSE calculation and finally stabilizes at 68.191; the LSTM model stabilizes at 73.902 after the seventh MSE calculation. In Figure 10(c), the Bayes-LSTM model gradually stabilizes after the third MRE calculation and finally stabilizes at 0.0304; the LSTM model stabilizes at 0.0339 after the third MRE calculation. 10(d), the Bayes-LSTM model gradually stabilizes its value after the fifth MAE calculation and finally stabilizes at 58.731; the LSTM model stabilizes at 62.625 after the sixth MSE calculation. The results show that the Bayes-LSTM model has better prediction accuracy for the industrial station data. From the analysis of the above three types of station forecasting results, the prediction error of Bayes-LSTM load forecasting model is generally better than the traditional LSTM time-series forecasting model. The Bayes-LSTM load forecasting model has more accurate prediction accuracy, smoother prediction results, and is more consistent in terms of the actual load change pattern, although it does not have better training time.

## 4. CONCLUSION

China's electricity sector is developing rapidly. The annual load peaks keep breaking new records, and the problem of supply-demand imbalance caused by peak periods is gradually becoming more pronounced. This study introduces Bayesian network and LSTM network in more detail, and proposes the LSTM electric load peak prediction model based on Bayesian network. A stacked cascade of Bayesian networks and LSTM networks was constructed, and comparative experiments were conducted to verify the performance of the proposed model. For the experiments, the data set was divided into a training set and a test set in the ratio of 8:2, and the data with no less than 95.5% of the peak load in the station area was used as the sample. The experimental results show that the Bayes-LSTM model has four performance results of 115.947, 10.161, 0.027, and 9.656 for MSE, RMSE, MRE, and MAE respectively for the data of the integrated station area; 1167.703, 34.153, 0.0767, and 28.424 for the data of industrial stations, and the results of the four performance indicators are 4697.656, 68.191, 0.0304, 58.731, respectively. The Bayes-LSTM model proposed in the study is slightly inferior to Trangrong's LSTM method in terms of training time, but improves the prediction accuracy significantly. Experimental evidence demonstrates the feasibility and effectiveness of the Bayes-LSTM load prediction model proposed in the study. However, further experiments should be conducted since the model constructed in this study could be extended by incorporating more indicators so that the new model can be open to capacity measurement and provide a range of solutions for various problems in the power system.

## REFERENCES

1. Xu L, Wang S, Tang R. 2019. Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Applied Energy*, 237:180–195.
2. Yu Z, Niu Z, Tang W, Wu Q. 2019. Deep learning for daily peak load forecasting–a novel gated recurrent neural network combining dynamic time warping. *IEEE Access*, 7:17184–17194.
3. Terki A, Boubertakh H. 2021. A new hybrid binary-real coded cuckoo search and Tabu search algorithm for

solving the unit-commitment problem. *International Journal of Energy Optimization and Engineering (IJEOE), 10*(2): 104–119.
4. Nettasinghe B, Krishnamurthy V. 2021. Maximum likelihood estimation of power-law degree distributions via friendship paradox-based sampling. *ACM Transactions on Knowledge Discovery from Data, 15*(6):1–28.
5. Fu W, Chien C F, Tang L. Bayesian network for integrated circuit testing probe card fault diagnosis and troubleshooting to empower Industry 3.5 smart production and an empirical study. *Journal of Intelligent Manufacturing, 33*(3):785–798.
6. Noor N M, Sulaiman N S, Tan H, Zardasti L. 2022. Integration of Bayesian network with fuzzy analytical hierarchy process for determining the pipeline conditions. Process Safety Progress, 2022, 41:S187-S196.
7. Liu C, Niu Z, Li Q. Relationship between lean tools and operational and environmental performance by integrated ISM–Bayesian network approach. TQM Journal, 2021, 34(4):807–828.
8. Aziz R A, Symum H, Mahbub N, Azeem A. 2021. Monitoring maternal healthcare performance in hospital setting using a multi-stage Bayesian network approach. *American Journal of Industrial Engineering, 8*(1):9–17.
9. Kumari P, Bhadriraju B, Wang Q, Kwon S I. 2022. A modified Bayesian network to handle cyclic loops in root cause diagnosis of process faults in the chemical process industry. *Journal of Process Control*, 110:84–98.
10. Sun M, Zhou R, Jiao C, Sun X. 2022. Severity analysis of hazardous material road transportation crashes with a Bayesian network using highway safety information system data. *IJERPH, 19*(7):1–22.
11. Zhang C, Wang W Z, Zhang C, Fan B, Wang JG, Gu F, Yu X. 2022. Extraction of local and global features by a convolutional neural network–long short-term memory network for diagnosing bearing faults. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 236*(3):1877–1887.
12. Ghorbel A, Souissi B. 2022. Upper confidence bound integrated genetic algorithm-optimized long short-term memory network for click-through rate prediction. *Applied Stochastic Models in Business and Industry, 38*(3):475–496.
13. He M J, Li H, Xu J R, Wang H L, Chen SZ. 2021. Estimation of unloading relaxation depth of Baihetan Arch Dam foundation using long-short term memory network. *Water Science and Engineering, 14*(2):149–158.
14. Abdelsalam A A, Hassanin A M, Hasanien H M. 2021. Categorisation of power quality problems using long short-term memory networks. I*ET Generation, Transmission & Distribution, 15*(10):1626–1639.
15. Zhang Y, Wang S, Sun G, Mao J. 2022. Aerodynamic surrogate model based on deep long short-term memory network: An application on high-lift device control. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 236(*6):1081–1097.
16. Sun Y, Kang J, Sun L, Jin P, Bai X. 2022. Condition-based maintenance for the offshore wind turbine based on long short-term memory network. *Journal of Risk and Reliability, 236*(4):542–553.
17. Geng B. 2022. Text segmentation for patent claim simplification via bidirectional long-short term memory and conditional random field. *Computational Intelligence, 38*(1):205–215.
18. Rajamoorthy R, Saraswathi H V, Devaraj J, Kasinathan P, Elavarasan R M, Arunachalam G, Mostafa T M, Mihet-Popa L. 2022. A hybrid sailfish whale optimization and deep long short-term memory (SWO-DLSTM) model for energy efficient autonomy in India by 2048. *Sustainability, 14*(3):1–35.
19. ArunKumar K E, Kalaga D, Mohan Sai Kumar C, Kawaji M, Brenza T M. 2022. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal, 61*(10):7585–7603.
20. Heddam S, Kim S, Elbeltagi A, Kisi O. 2022. Bidirectional long short-term memory-based empirical wavelet transform: A new hybrid artificial intelligence model for robust prediction of soil moisture content. *Current Directions in Water Scarcity Research*, 7:37–56.