IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 March 2023, accepted 9 April 2023, date of publication 13 April 2023, date of current version 19 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3266783

RESEARCH ARTICLE

Construction and Application of Short-Term and Mid-Term Power System Load Forecasting Model Based on Hybrid Deep Learning

HONGSHENG XU[1,2], GANGLONG FAN[1,2], GUOFANG KUANG[3], AND YANPING SONG[1,2]

[1]College of Electronic Commerce, Luoyang Normal University, Luoyang, Henan 471934, China
[2]Henan Key Laboratory for Big Data Processing and Analytics of Electronic Commerce, Luoyang Normal University, Luoyang, Henan 471934, China
[3]School of Information Technology, Luoyang Normal University, Luoyang, Henan 471934, China

Corresponding author: Hongsheng Xu (xhsls@lynu.edu.cn)

This work was supported in part by the National Natural Science Funds of China under Grant 61272015, and in part by the 2022 Henan Province Key Research and Development and Promotion Projects (Science and Technology) under Grant 222102320342.

ABSTRACT Power system load forecasting model plays an important role in all aspects of power system planning, operation and control. Therefore, accurate power load forecasting provides an important guarantee for the stable operation of the power grid system. This paper first analyzes the current status of power system load forecasting, and finds that there are still some deficiencies in the existing forecasting models. In order to make up for these shortcomings, this paper proposes construction of short-term and mid-term power system load forecasting models based on hybrid deep learning. In the data preprocessing part, this paper proposes to use the exponential weight moving average method to process missing values. The detection method of abnormal value is the GeneralizedESDTestAD(GESD). This paper analyzes the historical load data of a regional power grid and four industries, and proposes a short-term power system load forecasting model based on Bi-directional Long Short-Term Memory(BiLSTM); For mid-term load forecasting, this paper first uses random forest and Pearson correlation coefficient to select features. Then a hybrid deep learning model is constructed based on BiLSTM and random forest. After optimizing the parameters of the model, a mid-term power system load forecasting model based on hybrid deep learning is constructed. Finally, the benchmark models are selected for comparative experiments. The experimental results show that the MAPE of the proposed model is 2.36%, which is better than the benchmark models. This proves that the proposed hybrid model can effectively improve the accuracy of power system load forecasting.

INDEX TERMS Hybrid deep learning, long short-term memory (LSTM), power load forecasting, short-term, mid-term, random forest, bi-directional long short-term memory (BiLSTM).

I. INTRODUCTION

Power system load forecasting is to make a forecast of the system load for a future period by fully considering the influence of historical system load, economic conditions, meteorological conditions and social events. Traditional power system load forecasting methods cannot handle such a huge amount of data. The introduction of data mining technology to complete the load forecasting of the power system can effectively increase the forecasting accuracy, improve the stability and economy, and at the same time facilitate the overall planning of the power grid.

By forecasting the duration of the load, the power system load forecast can be divided into three types: short-term load, medium-term load and long-term load. Short-term load forecasting is the basis for the start-up, shut-down, dispatch and operation planning of the power grid internal units. Short-term load forecasting is the basis for unit start-up and shut-down, scheduling and operation planning within the grid. The medium-term forecast supports the decision to protect the

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Raza.

power consumption of enterprise production and social life, and rationalize the operation and maintenance of the grid [1]. With the diversification of load structures in power systems, the effectiveness of load forecasting model applications has decreased, so it becomes important to optimize load forecasting models and improve forecasting accuracy.

This paper first analyzes the current status of domestic and international research. Foreign countries started earlier than China in load forecasting, and most of them are devoted to the application of artificial intelligence algorithms in electric load, and methods such as neural networks and support vector machines are popular forecasting methods.

The paper uses the LSTM algorithm to predict the electricity load of non-residential customers by using multiple correlation sequence information. The K-means method is also used to cluster the daily load of non-residential users, and the experimental results are better than other load prediction methods [2]. The authors use trained LSTM networks and developed models to make predictions, comparing the predictions generated by LSTM with the results of the traditional methods of RMSE and MAPE in all prediction ranges [3]. The experimental results show that LSTM-based prediction is superior to other traditional methods and has the potential to further improve the prediction accuracy. The literature [4] used improved deep residual networks for short-term electric load forecasting modeling and used a two-stage integration strategy to improve the generalization capability of the model and obtained good results.

To improve the prediction accuracy, the authors first built the LSTM prediction model and eXtreme Gradient Boosting (XGBoost) prediction model separately, and then combined the LSTM and XGBoost for load prediction using the error inverse method. Experiments show that the model predicts better than a single model [5]. Peng et al. [6] proposed a short-term load forecasting model based on Attention and Long Short-Term Memory (LSTM). The model takes full advantage of the time-series nature of the load and uses an attention mechanism to highlight the input features that play a key role in load prediction. Rafi et al. developed a combined model based on a convolutional neural network and a long and short-term memory network, and the simulation results verified the validity of the proposed model [7]. Cai et al. [8] proposed a two-tier short-term load prediction architecture based on transfer learning to improve the prediction accuracy of load in the target area.

Aguilar Madrid et al [9]. present short-term electricity load forecasting based on a set of machine learning models. Since the XGBoost model is based on an ensemble of decision trees, it improves the interpretability of the model. Through comparative experiments, it is found that the XGBoost model shows the best prediction results. This paper proposes an intelligent power load forecasting method based on deep learning. The proposed deep learning model is an integration of Deep Convolutional LSTM and Stacked GRU. Finally, the experimental results on the competitive power load data set show excellent results [10]. The authors propose short term forecast of electricity load based on DNNs [11]. The proposed method is evaluated on a real dataset collected over a 4-year period and forecasts the electricity load for the coming days and weeks. The main contribution of the paper is the time-frequency (TF) feature selection process from the actual dataset, which contributes to the regression process initiated by DNNs.

Through the analysis of current research status, the existing power load forecasting models have not been able to effectively consider the time-series dependence of the load itself. The BiLSTM network can make up for this defect. BiLSTM can replace the neurons in the original recurrent neural network (RNN), so that the original RNN has the ability of temporal memory to realize the modeling of the temporal features of the load and the relationship between the load and other influencing factors. Therefore, the BiLSTM model is used to construct the power system load forecasting model in this paper.

In this paper, the BiLSTM model is used for time-series forecasting to create a short-term power load forecasting model based on the characteristics of power data. For mid-term load forecasting, the data are first resampled for the daily load maximum and minimum values of the data. The features are then found by the plausibility of the data. The importance of the constructed features is ranked by the random forest model, and redundant features are eliminated by combining the correlation coefficient method. This paper proposes a hybrid deep learning model based on BiLSTM and random forest. Then the parameters of the model are optimized, and the mid-term power system load forecasting model is constructed. In order to verify the performance of the proposed hybrid model, we select the benchmark models for comparative experiments. The error indicators are used as the evaluation methods. The experimental results show that the forecasting accuracy of the proposed model is higher than other benchmark models. The model makes full use of historical data and characteristics to predict the observed values. This paper uses the model to forecast the maximum and minimum daily load of each industry in the next three months. Experiments show that the proposed model can effectively improve the accuracy of power system load forecasting.

II. RELATED TECHNICAL ANALYSIS

A. BILSTM

LSTM is a deep learning framework based on cyclic neural networks. RNN can only have short-term memory due to the disappearance of gradients. The LSTM network is able to capture historical data information at long intervals and solves the problem of gradient disappearance to a certain extent [12].

The key to LSTM is a conveyor belt-like cell state, which can run directly over the entire chain with only a small number of linear interactions, on which information flows relatively easily and remains constant. The structures called gates in the LSTM are used to remove or add an information to the cell state [13]. A gate is a method that contains a sigmoid neural network layer and a pointwise multiplication operation, which serves to selectively let information pass through.

(1) The forgetting gate is controlling whether to forget or not, deciding with a certain probability what information to discard from the cellular state. The gate reads the hidden state of the previous sequence h_{t-1} and the data of this sequence x_t. An activation function outputs a value between 0 and 1, f_t representing the probability of forgetting the state of a hidden cell in the previous layer. 1 means "completely reserved", 0 means "completely discarded". Its mathematical formula is as follows [14].

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

(2) Memory gates determine what new information is stored in the cell state and are responsible for processing the input of the current sequence location. There are two parts: the sigmoid layer determines the output of the information i_t to be remembered by the network, and the tanh layer is the output of a normal RNN propagation C_t. The two results are multiplied to update the cell status. Its mathematical formula is as follows.

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

$$\tilde{C}_t = tanh\left(W_C\left[h_{t-1}, x_t\right] + b_C\right) \tag{3}$$

(3) The update gate is used to update the old cell state, updating C_{t-1} to C_t. The forgetting gate determines what will be forgotten and the memory gate determines what will be remembered. Before studying the output gate of the LSTM, the cell state of the LSTM needs to be determined, and the update gate performs this step. Its mathematical equation is shown below.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

(4) The output gate determines what the network outputs at this time step, and this output will be based on the current cell state [15]. The gate consists of two parts, the first part o(t) is obtained from the hidden state of the previous sequence h(t-1) and the present sequence data x(t), and the activation function sigmoid. The second part consists of the hidden state c(t) and the tanh activation function, which is given by the following equation.

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

$$h_t = O_t * tanh\left(C_t\right) \tag{6}$$

Although these doors are functionally different, they are the same in terms of the operation of the tasks performed. The LSTM model uses the sigmoid function as a selection tool and the tanh function as a transformation tool, combining these two functions to achieve the gate function.

The BiLSTM model is actually a variant of the LSTM network. This model combines forward LSTM and backward LSTM to extract semantic features from sentences, and forward and backward information can be obtained at each time. BiLSTM model is good at processing long-distance text information, and CRF model obtains the most likely prediction results through the adjacent relationship between labels. Because the one-way structure of LSTM cannot deal with the problem of information encoding in the opposite direction, it cannot fully learn the two-way features in a specific period of time. Therefore, this paper uses BiLSTM to extract the contextual feature information of the text and captures the information using bi-directional dependency, so that the features at each moment have the backward and forward dependency relationship.

This module combines the state vectors h and h of forward and reverse LSTM output into an h vector output with complete context information. The combined feature vectors are calculated by the activation function to obtain the result with the following equation.

$$O_t = \tanh(w_h \cdot \overleftrightarrow{h} + b_o) \tag{7}$$

In the above formula, $\overleftrightarrow{h}$ represents the output vector after the combination of $\vec{h}$ and $\overleftarrow{h}$, and $w_h \in R^{|n| \times 2|n|}$ represents the weight corresponding to this vector. $b_o \in R^{|n|}$ represents the offset vector, and O_t is the output result and the input of CRF module.

B. ANALYSIS OF RANDOM FOREST MODEL

Random forest is a supervised machine learning algorithm and an integrated algorithm, which is a classifier containing multiple decision trees. The decision tree classifies the data step by step based on the principle of inductive algorithm. It can be represented by a top-down tree structure, which is easy to extract rules and understand, and has obvious advantages in high-dimensional data [16].

Random forest is an integrated algorithm that trains and predicts samples through a number of independent and unrelated decision trees [17]. It takes into account the modeling results of multiple classifiers and summarizes them to produce a comprehensive result. Random forests generally perform better than single decision trees in theory because they are generally determined by voting on average or by majority of the results of multiple decision trees.

According to the principle and basic idea of random forest, the generation of random forest mainly consists of three steps:

Step1: K training sample sets are extracted from the original sample set S by Bootstrap method. Generally, the sample size of each training set is the same as that of S.

Step2: K training sets are learned to generate K decision tree models. In the decision tree generation process, assume there are M input variables, F variables are randomly selected from M variables. Each internal node splits by using the optimal splitting method on the F feature variable, and the F value is a constant during the formation of the random forest model.

Step3: Finally, the results of K decision trees are combined to form the final result. For classification problems, the combination method is simple majority voting. For regression problems, the combination method is the simple average method.

In the above steps, S represents the original sample set, K represents the training sample set, and both M and F represent variables.

III. DATA PREPROCESSING

The data source analyzed in this paper is the load data of the power system in a region within a certain time interval of 15 minutes. Additionally, the data analyzed include the maximum and minimum daily load data for four industries, namely, large industrial power, non-general industry, general industry and commerce, in the same time period. According to the statistics of the dataset, there are 388 missing values in the whole data source, with missing values in all industries, and non-popular industrial data are also missing in certain time periods.

A. MISSING VALUE PROCESSING

1) PROCESSING METHOD ANALYSIS

Time series missing value processing methods are divided into three main categories:

The first type is the direct deletion method. Such methods are not applicable for time series data, which will destroy the integrity of the series and are not conducive to model building;

The second is based on statistics. The common methods are mean filling, median filling, common value filling and so on, but they ignore the time series information of the data;

The third is machine learning-based filling methods. Common methods are missing value filling algorithms based on KNN, RNN, EM (Expectation-Maximization) and Matrix Factorization [18]. However, this kind of method rarely takes into account the time series information between two adjacent data.

2) ANALYSIS OF MISSING VALUE FILLING METHODS

In order to maintain the continuity and integrity of the time series, as well as to consider the temporal information between adjacent data, we first consider the forward-filling approach. The core idea of this method is to fill in the current missing values with the values from the last point in time before the missing values. By analyzing the data, it is found that a single data point has a large fluctuation, so the filling method is not effective. Therefore, we use the Moving average method for filling in order to reduce the large impact of a single mutation point on the filling data.

The core idea of the moving average method is to take out the values in a rolling time before the missing values occur and calculate their mean or median to fill in the missing values. Considering the different effects of time on our data filling, we use the Exponential Weighted Moving Average (EWMA) method to fill in missing values.

The method has two main features: First, the exponential weighted moving average method can calculate the weighted average, the value of the nearest time has the highest weight, and the weighted exponential order of the previous time values decreases; Second, the EWMA implementation of pandas can effectively handle missing values. It is observed that the missing values are continuous missing in a day, and only forward/backward filling can destroy the periodicity of the data, thus causing data noise. The exponential weighted moving average method is more suitable for processing. Therefore, this paper uses the exponential weighted moving average method to deal with missing values. The comparison before and after the above treatment is shown in Fig. 1 and Fig. 2. The experiment shows that the method used in this paper is feasible.

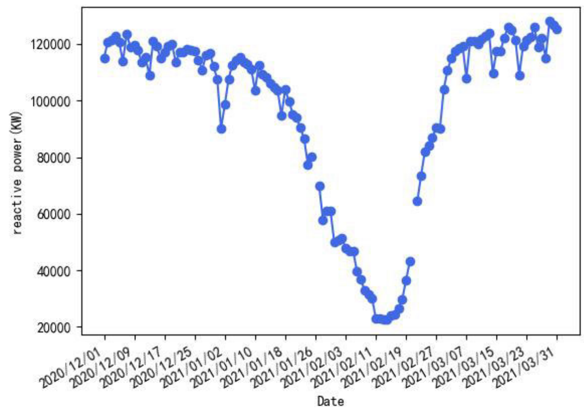

FIGURE 1. Missing value map of load data for power systems in 2021.

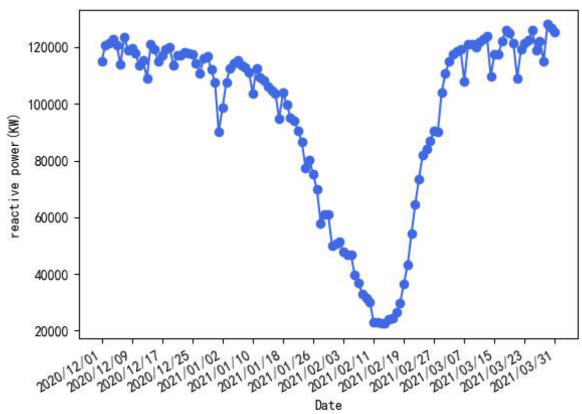

FIGURE 2. Data map with missing values processed by exponential weighted moving average method.

B. ANALYSIS OF DETECTION AND PROCESSING METHODS FOR OUTLIERS

1) EXAMINING OUTLIERS

Traditional outlier detection methods are 3σ Principle outlier detection and box plot outlier detection [19]. Since the collected data sources fluctuate more frequently over time, and the 3σ principle outlier detection is more applicable to

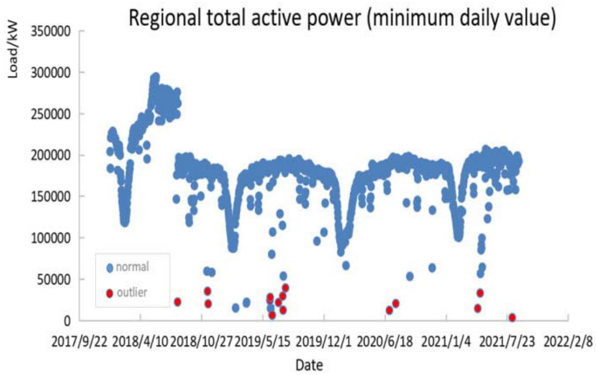

FIGURE 3. Regional daily load minimum anomaly detection.

the data as a whole, which is more strict and obeys normal distribution, the detection method is not used. The data used in this paper fluctuates periodically, and there is a large gap between the maximum and minimum values. It is not only possible to detect the maximum value to determine that it is an outlier. Therefore, the box plot anomaly detection method is not suitable for the outlier detection in this paper.

The normal test GESD is a simple statistical method used to detect one or more outliers in a univariate dataset that follows an approximate normal distribution. This statistical method assume that regular data follows a statistical model (or distribution) and data that does not follow a model (or distribution) is an outlier. This method is not only suitable for time series data, but also better for power load and business volume prediction to improve the quality of data analysis. The main idea of this method is that outliers are often multiple rather than single in real datasets. In order to extend Grubbs'Test to k outlier detection, the maximum (maximum or minimum) deviations from the mean values need to be removed from the dataset step by step, and the corresponding t-distribution critical values updated synchronously to verify the validity of the original hypothesis. Since power system load forecasting is based on time series, this paper uses the normal test GESD to detect outliers [20]. And through the analysis of the collected data sources, it is found that some of the outliers are negative.

Each industry has a large difference in power consumption, so outliers can be identified by adjusting significant levels for different industries. This paper takes the minimum daily load of a region as an example, as shown in Fig. 3.

The red part shown in Fig. 3 is the detected outlier, which is much lower than the average or even contains negative values, and will affect subsequent predictions if not handled.

2) ANALYSIS OF HANDLING OUTLIERS

There are several common outlier handling methods [21].

(1) Delete: There are usually two strategies for deleting outlier data directly: whole delete and pair delete. This method is the easiest to operate, but it also has two drawbacks that cannot be ignored: First, in the case of few original data, direct deletion will result in insufficient sample size, which will affect subsequent operations. Second, direct deletion is likely to have an impact on the original distribution of the data, resulting in statistical model instability, so this method is not applicable to power data.

(2) Mean Correction: If the sample size of the data is small, the outlier can also be corrected by the mean of the first and last two observations. This method is a relative compromise method. The advantage is that it can overcome the shortcomings of missing samples, while the disadvantage is that it loses the characteristics of the samples themselves, so it is not suitable for the data collected.

(3) Regression Interpolation: This method finds patterns of change between two related variables and smooths the data by fitting it to a function. If there is a dependency between variables, that is $y = f(x)$, try to find the dependency f and predict y based on x. The above approach is the essence of the regression problem. A common assumption in practical problems is $p(y) = N(f(x))$ that N is normally distributed. Assuming y is an observation value and there is noise data, according to the dependence between X and y, and then updating y value according to x, random noise can be removed, which is the principle of regression denoising.

(4) Treat as missing values: The advantage of this method is that it can use the information of the original data to fill in the outliers. Treating outliers as missing values needs to be based on the characteristics of the outliers (missing values). The outliers are handled differently in different situations, whether they are completely random, random, or non-random.

Because the power load data depends on its original characteristics (such as time, weather, temperature, wind direction, etc.) and has a strong dependence on time series, the fourth method is most applicable, that is, outliers are treated as missing values. Sample data after outlier treatment is shown in Table 1.

IV. CONSTRUCTION AND EVALUTION ANALYSIS OF SHORT-TERM AND MID-TERM POWER SYSTEM LOAD FORECASTING MODELS

A. CONSTRUCTION OF SHORT-TERM POWER SYSTEM LOAD FORECASTING MODEL BASED ON BILSTM

1) BUILDING MODEL

After data analysis, this paper finds that when making short-term predictions, weather conditions, temperature, and the influence of wind direction on the predicted results can be ignored. In order to improve the accuracy of prediction data, we make full use of all the data on the basis of the existing features to perform a sliding window operation on the original dataset. Therefore, this paper establishes a sliding window for predicting future data from multiple historical data.

The sample data is processed in a format recognized by the BiLSTM model, and the input format requirements of the BiLSTM model are maintained [22]. The data set is divided into training and test sets. Because the model needs to learn the time factor, the original sequence of the data cannot be

TABLE 1. Sample data after outlier treatment.

Sample description	Explain	Number of original samples	Number of samples processed
Data Time	Total power (kw)	128156	128544
Industry type	Power for Large Industry	973	974
	Non-general Industry	691	692
	General industry	973	974
	business	973	974
Date, weather conditions, etc.	-----	1346	1340

destroyed in the division of the training sets, which can be cut by time [23].

The purpose of this paper is to make short-term load forecasting for a regional power network with 15-minute intervals over the next 10 days. This paper presents a time series prediction model based on a long-term and short-term memory network for power system load prediction. BiLSTM models are created by importing Sequential models from Keras and adding layers to the model using the add method. The steps are as follows:

Step1: A BiLSTM with one input time step and 50 features are added, and we construct 256 memory cells in the BiLSTM hidden layer;

Step2: We added a fully connected layer for output prediction, constructed a neuron, and added a sigmoid activation function;

Step3: For the compilation part of the model, we use Mean Squared Error as the loss function and Adam as the optimization algorithm. The advantage of Adam is that after bias correction, the learning rate of each iteration has a certain range, which makes the parameters more stable. The power load forecasting model uses 50 epochs, and each batch size is set to 100. The model is trained, and the shuffle needs to be set to False.

Based on the established short-term power system load forecasting model, the error curves of the test set and the training set are shown in Fig. 4.

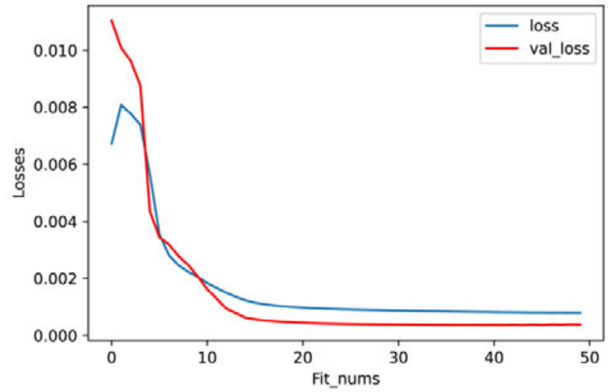

FIGURE 4. Error plot of test set and training set.

2) EVALUATION ANALYSIS OF POWER SYSTEM LOAD FORECASTING MODEL

Due to the complexity of the time series model, this paper uses the following four methods to evaluate the model in order to test the fitting degree of the model [24].

(1) r2_score function

The r2_ scope function calculates computes R^2. That is the coefficient of determination. This factor provides an estimate that future samples may be predicted by the model. The best score is 1.0, which can be negative (because the model may be worse). The purpose of this coefficient is to predict the expected value of y. In addition, the value of R^2 is set to 0.0 in a constant model that does not take into account the input characteristics. The formula is shown below.

$$R^2\left(y,\hat{y}\right)=1-\frac{\sum_{i=0}^{n_{samples}-1}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=0}^{n_{samples}-1}\left(y_i-\bar{y}\right)^2} \tag{8}$$

In above formula, R^2 Mean values can be used as error benchmarks. Its function is whether the prediction error is greater or less than the mean reference error. If R^2_score = 1, the predicted and true values in the sample are exactly the same, and there is no error, indicating that the independent variables in the regression analysis have a better interpretation of the dependent variables. If R^2_score = 0, Then the molecule is equal to the denominator and each predicted value of the sample is equal to the mean.

(2) Interpretation variance score

This function of Explained_Variance_Score calculates the explained variance regression score. If $\hat{y}$ is the predicted target output, y is the corresponding target output, and Var is the variance, which is the square of the standard deviation. Then the explanatory variance is estimated by the following formula:

$$explained\_variance\left(y,\hat{y}\right)=1-\frac{Var\left\{y-\hat{y}\right\}}{Var\left\{y\right\}} \tag{9}$$

Formula(9) calculates that the best score is 1.0 and the lower the value, the worse it is.

(3) Median absolute error

The function of median_absolute_error has strong outliers and calculates the loss from the median of all absolute differences between the target and the prediction. If $\hat{y}_i$ is the predicted value for the first sample and y_i is the corresponding true value. The definition of the median absolute error (MedAE) estimate is shown below.

$$MedAE\left(y,\hat{y}\right)=median\left(\left|y_1-\hat{y}_1\right|,\cdots,\left|y_n-\hat{y}_n\right|\right) \tag{10}$$

(4) Mean Squared Log Error

The formula is as follows:

$$E\left(f;D\right)=\frac{1}{m}\sum_{i=1}^{m}\left(\log\left(1+f\left(x_i\right)\right)-\log\left(1+y_i\right)\right)^2 \tag{11}$$

In formula(11), the mean square logarithmic error is robust to noise and continuous over the entire interval.

Based on the above evaluation indexes, the evaluation index values of the power system load forecasting model are obtained as shown in Table 2.

TABLE 2. Effect evaluation index value of LSTM.

Evaluating indicator	Value
R^2 score	0.951
Interpretation variance score	0.959
Median absolute error	0.01
Mean Squared Log Error	0.0002

In order to observe the fitting effect more intuitively, the model building experiment is carried out in this paper. Fig. 5 shows the prediction result of the test set and the true data graph of the test set.

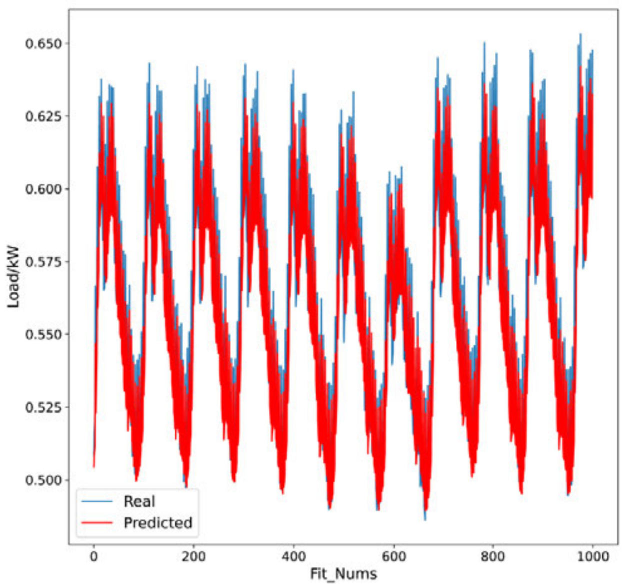

FIGURE 5. Fitting effect diagram of test results and test data.

Experiments are conducted here, first, the index values obtained from the evaluation index. Then a graph of the fit effect of the test results to the test data is obtained. The purpose of this experiment is to predict the power load in this area every fifteen minutes for the next fifteen days. The following predictions can be obtained from the model we have established, as shown in Fig. 6. The experimental results show that the model has little error in prediction effect on the test set.

The predicted trend in Fig. 6 shows that it is roughly the same as the true load trend, so the predicted results are in line with the expected results. The results of the prediction experiments show that the model can effectively improve the accuracy of power system load prediction.

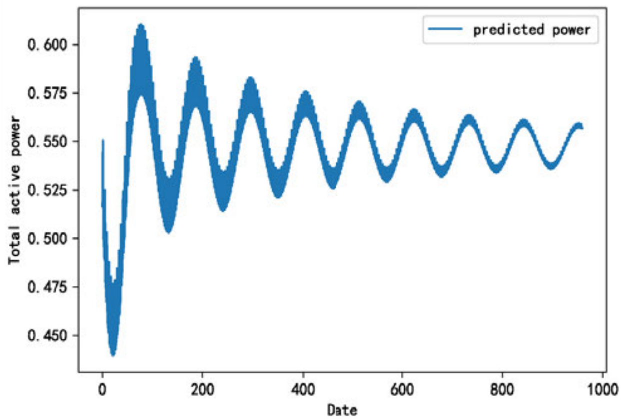

FIGURE 6. Trend chart of forecast effect at intervals of 15 minutes in the next 15 days.

B. MID-TERM POWER SYSTEM LOAD FORECASTING MODEL BASED ON HYBRID DEEP LEARNING MODEL

1) STANDARDIZATION AND NORMALIZATION OF DATA

The data collected in this paper are the maximum and minimum loads of the regional power grid in the next three months, which belong to the mid-term forecast. Since the given data has only one time feature, in order to improve the accuracy of prediction, this paper constructs the feature based on the given data.

First we get the resampled data. Fig. 7 is a trend map of the minimum and maximum values from 2018 to 2021.

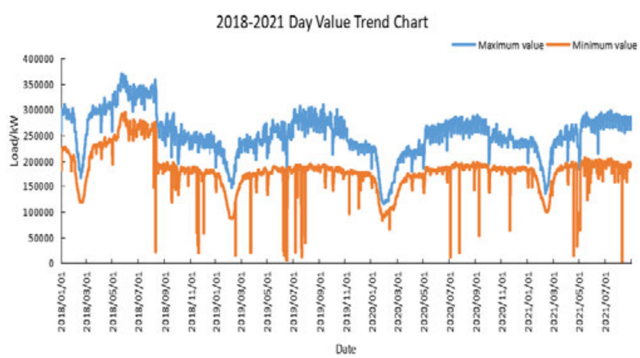

FIGURE 7. 2018-2021 maximum and minimum daily load trend chart.

Through the analysis of the sample data, we can see that the load of the daily maximum and minimum values in this area changes regularly every quarter. The frequency of fluctuations is constant and seasonal with a fixed length of time. So we construct the features of year, month, day and week of year according to time.

In order to eliminate the differences between different evaluation indexes, it is easy to compare the data and other operations. To ensure the reliability of the results, it is necessary to standardize the original index data. The main idea of data standardization or normalization is to scale multiple evaluation indexes of different attributes, orders of magnitude and units to the same data interval and extent to reduce the impact of size, characteristics, distribution differences on the model.

Common data standardization or normalization methods include: maximum-minimum normalization, mean variance standardization, decimal scaling, quantitative feature binarization, and so on. Maximum and Minimum Normalization transforms the original data linearly so that the results are mapped to the range of [0,1] and scales the original data equally. Then the normalization formula is as follows:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

where in equation(12), $x_{\max}$ is the maximum value of the sample data. $x_{\min}$ is the minimum value of the sample data.

This article can use class of sklearn. preprocessing. MinMaxScaler, which has the advantage of saving the parameters (maximum and minimum) in the training set and converting the test set data directly using its objects. In addition to standardization, we also normalize test datasets. The test dataset cannot be normalized directly by a formula, but rather by using the maximum and minimum values of the training dataset. Because the following may occur in a real environment: (1) The data will be continuously exported into the model, unable to get the maximum and minimum values; (2) The training dataset is the data simulated in the real environment and cannot use its own maximum and minimum values directly; (3) In real-world environments, it is not possible to normalize individual data. The process for Standardization in this paper is shown in the following figure.

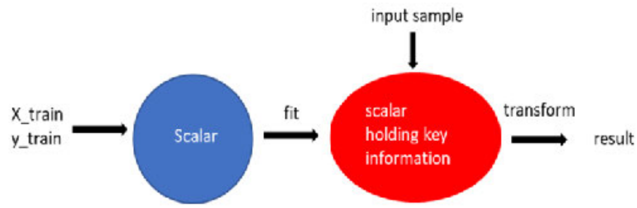

FIGURE 8. Standardized flowchart.

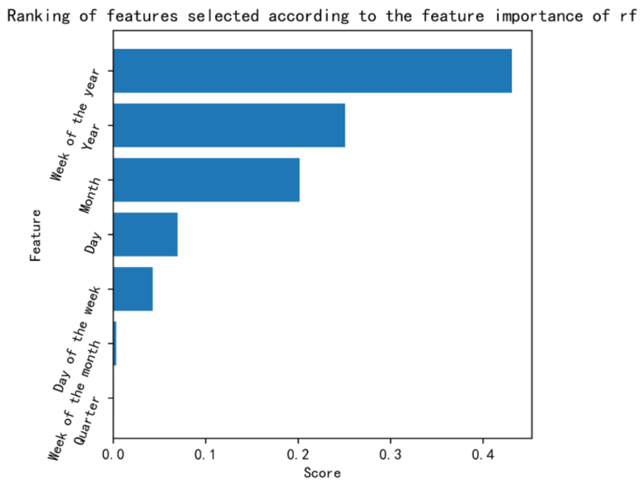

FIGURE 9. Feature ordering based on feature importance of random forest.

2) FEATURE SELECTION BASED ON RANDOM FOREST MODEL

The importance of a characteristic X in random forest is calculated as follows:

(1) For each decision tree in random forest, the error of out-of-pocket data is calculated using the corresponding OOB (out-of-pocket data), which is recorded as errOOB1.

(2) Noise interference is randomly added to the characteristic X of all samples of out-of-pocket data OOB, and the out-of-pocket data error is calculated again, which is recorded as errOOB2.

(3) Assuming that there are Ntree trees in a random forest, the importance for Feature X is $\sum$(errOOB2-errOOB1)/Ntree. The reasons why this expression is used as a measure of the importance of the corresponding feature are as follows: When noise is randomly added to a feature, the accuracy outside the bag is greatly reduced. This indicates that this feature has a great influence on the classification results of the samples, that is, it is more important.

An experiment was carried out to verify the signature importance algorithm. The result of the experiment is that the signature we constructed and its corresponding signature importance, as shown in the following figure.

3) AUTOCORRELATION ANALYSIS

Due to the intrinsic properties of the time series features themselves, there is bound to be some degree of autocorrelation between the features we construct. Therefore, before constructing the model, the autocorrelation analysis of the built features should be carried out in order to exclude some variables with high autocorrelation.

Coefficient of correlation is a statistical index method that investigates the degree of linear correlation between variables [25]. The letter r is usually used. Because of the different subjects, there are many ways to define the correlation coefficient, Pearson correlation coefficient is more commonly used. The Pearson coefficient method is used to calculate the autocorrelation of the constructed features [26]. The corresponding formula is:

$$\rho\,(X \cdot Y) = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - \left(\sum X\right)^2}\sqrt{N\sum Y^2 - \left(\sum Y\right)^2}} \tag{13}$$

The absolute value of correlation coefficient is generally above 0.8. X and Y can be considered to have strong correlation. A weak correlation can be considered between 0.3 and 0.8. If the absolute value of the correlation coefficient is below 0.3, then there is no correlation. The autocorrelation thermogram of the built features is shown in Fig. 10.

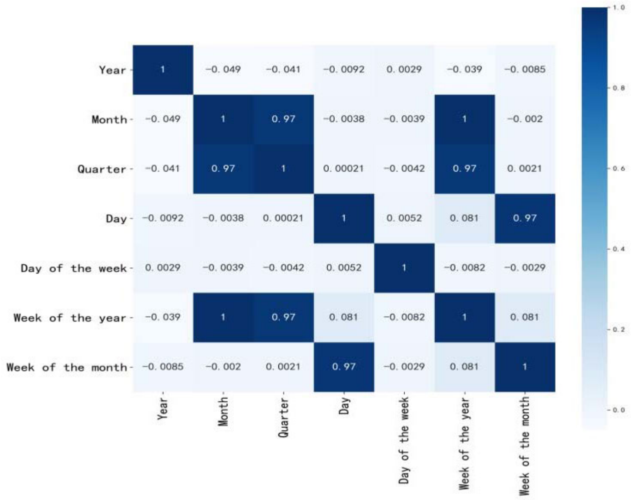

FIGURE 10. Autocorrelation heatmap of constructed features.

TABLE 3. Final selected indicators.

Num	Feature
1	YEAR
2	DAY
3	DAY OF THE WEEK
4	WEEK OF THE YEAR

Through the autocorrelation analysis in Fig. 10 of the experimental results, in order to construct the load prediction model effectively, the features with large autocorrelation and small contribution of features are removed. Finally, the features we selected are shown in Table 3.

4) CONSTRUCTION OF MID-TERM POWER SYSTEM LOAD FORECASTING MODEL

This paper constructs a hybrid deep learning model based on BiLSTM and random forest. The flowchart of the whole forecasting system is Fig.11.

The whole process of the power system load forecasting model based on the hybrid deep learning model is follow: The first step is data preprocessing; The second step is feature selection. We first use Random Forest model to select features. In this paper, the autocorrelation analysis method is used to remove the features with large autocorrelation and small contribution; The third step: the BiLSTM model and the random forest model are respectively used in this paper to forecast the power load after parameter adjustment. Finally, this paper uses the BiLSTM model and the random forest model respectively, and constructs the power system load forecasting model after parameter adjustment. On this basis, we use the error reciprocal method to weighted combine the prediction data of BiLSTM and random forest, so as to obtain the hybrid model forecasting model.

The network architecture of the proposed hybrid Deep learning network is shown by figure12. In Fig. 12, the feature selection of power load based on random forest model

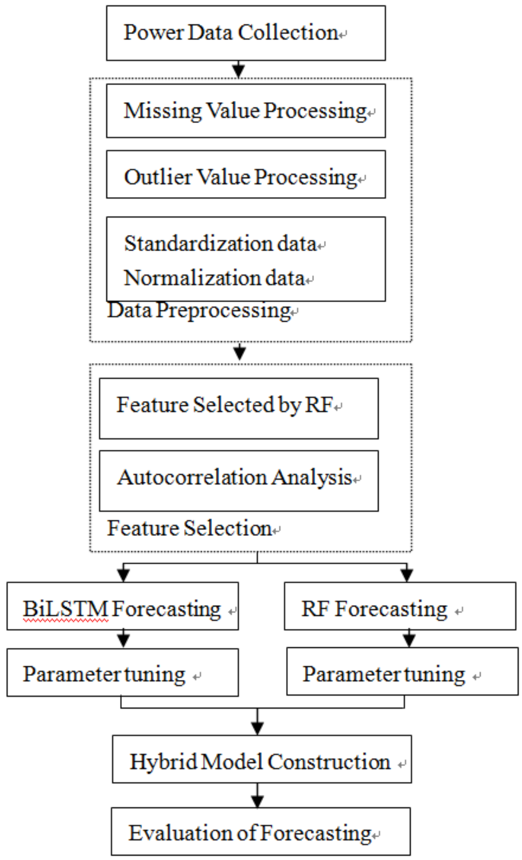

FIGURE 11. The whole forecasting system based on the hybrid deep learning model.

is shown. Then, after autocorrelation analysis, the features with large autocorrelation and small feature contribution are removed, and the final feature set is obtained. Next, it integrated and learned with the BiLSTM model, and finally constructed a hybrid deep learning model.

Then, this experiment uses this hybrid deep learning to build a mid-term power system load forecasting model. First, the GridSearchCV method is used for superparametric tuning to find the best BiLSTM model parameters. This method can adjust parameters automatically and give the optimal results and parameters as long as the parameters are input. At the same time, an estimator with the best parameter is instantiated. For the purpose of parameter tuning experiments, Table 4 illustrates the name and meaning of the parameters in the BiLSTM model.

This paper uses grid search to optimize the parameters. After experimentation, the final parameter results are shown in Table 5.

After the test, the number layers of this BiLSTM multivariate time series prediction model is 3. The number of hidden neurons is 80 in the first layer and 100 in the second layer. The number of output neurons is 2. The number of iterations of the model is 600. The learning rate is 0.01.

In this paper, random forest model is used for load prediction, and the final optimized parameter values are as follows:

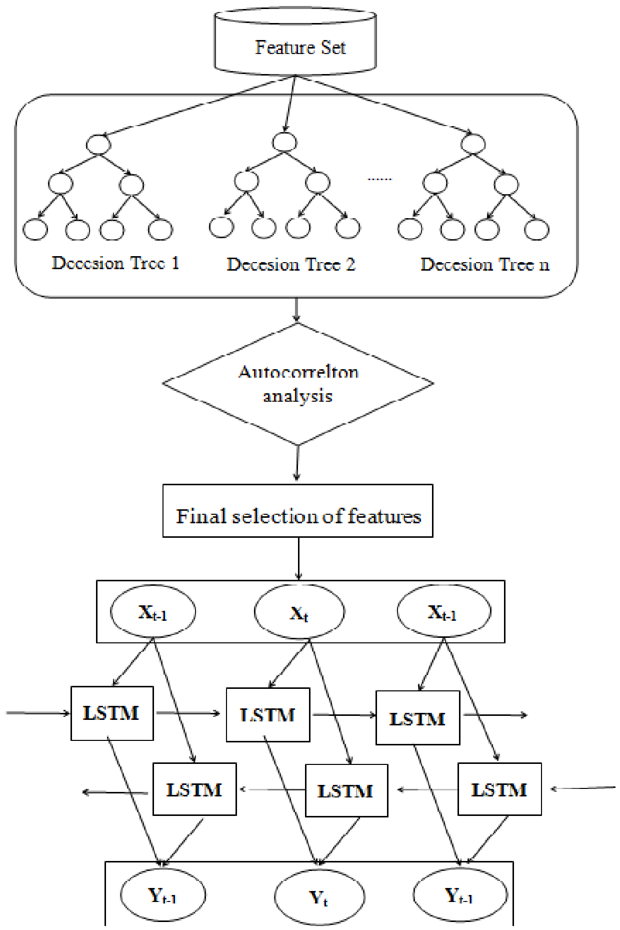

FIGURE 12. The network architecture of the proposed hybrid Deep learning network.

TABLE 4. Parameter name and meaning.

No.	Parameter Name	Parameter Meaning
1	batch_size	This parameter manages and updates the values of the learnable parameters in the model, which can make the model output closer to the real label. It is a learnable parameter for the sampling gradient update model, which reduces the loss.
2	epochs	This parameter is defined as a single training iteration for all batches in forward and backward propagation. This parameter is how many times the data will be rotated during training.
3	optimizer	Select a set of samples in the training set to update the weights.

max_features is 3, max_depth is 10, min_samples_split is 6, min_samples_leaf is 80.

We added an BiLSTM loop layer consisting of memory units, which is BiLSTM In this paper, when stacking BiLSTM layers, it is necessary to output a sequence of values for each input instead of a single value in order for subsequent BiLSTM layers to have the required 3D input

TABLE 5. Parameter adjustment range and results.

Parameter Name	Adjusting Parameter Range	Adjustment Result
batch_size	[16,20]	20
epochs	[8,10]	10
optimizer	['adam','Adadelta']	adam

This experiment uses return_Sequence true to implement this functionality. Dropout layer is also added in this experiment. Dropout regularization is used between the BiLSTM looping layer and the fully connected layer so that the BiLSTM layer is the input to the fully connected layer. This paper presents a hybrid deep learning technology based on BiLSTM and random forest to build the load forecasting model. The predictive effect of the load forecasting model on the test set is shown in the Fig. 13.

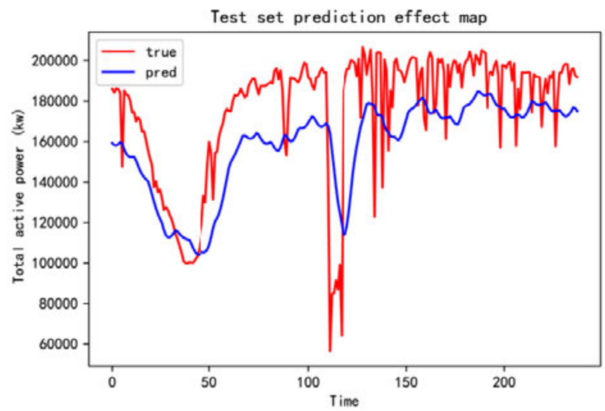

FIGURE 13. The predictive effect of load forecasting model on the test set.

This paper selects the current popular benchmark models for comparative analysis, such as: DBN, BiLSTM, RF, CNN-LSTM. To ensure the validity and reliability of the experiments, the experimental conditions were the same for all models.

The experiment is a comparative analysis based on error indicators. The error indicators use Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) with the following equations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2} \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{y}_i - y_i}{y_i}\right| \times 100\% \tag{15}$$

The experimental results show that the MAPE of the proposed model is 2.36%, which is better than other models. The MAPE of the proposed hybrid model is 1.93% lower than that of CNN-LSTM and 2.96% lower than that of BiLSTM. In addition, it is 3.76% lower than RF and 3.53% lower than

TABLE 6. Comparison of accuracy forecasting of different models.

Model	MAPE/%	RMSE/kW
BiLSTM-RF	2.36	91.96
CNN-LSTM	4.29	115.82
BiLSTM	5.32	123.85
RF	6.12	136.75
DBN	5.89	128.59

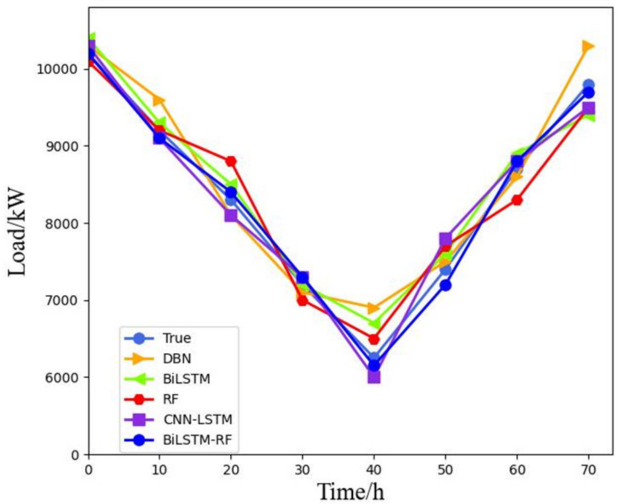

FIGURE 14. The load curves of each model.

TABLE 7. Evaluation index value of power load mid-term forecast model effect.

Model	R² score	Mean_Squared_Log_Error
BiLSTM-RF	0.437	0.028
CNN-LSTM	0.86	0.083
BiLSTM	1.13	0.12
RF	0.96	0.13
DBN	1.86	0.21

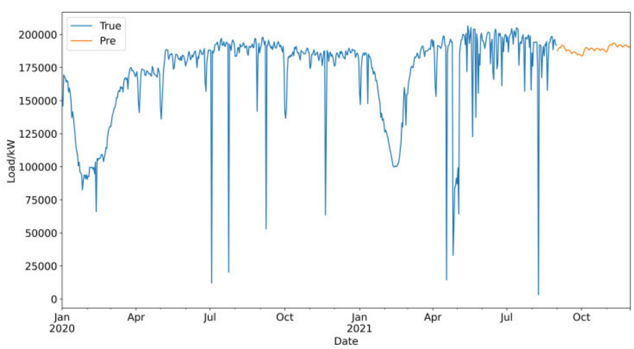

FIGURE 15. The forecast results of the minimum daily load for the next three months.

DBN. Therefore, the experimental results prove the accuracy of the hybrid model proposed in this paper. The experimental results are shown in Table 6.

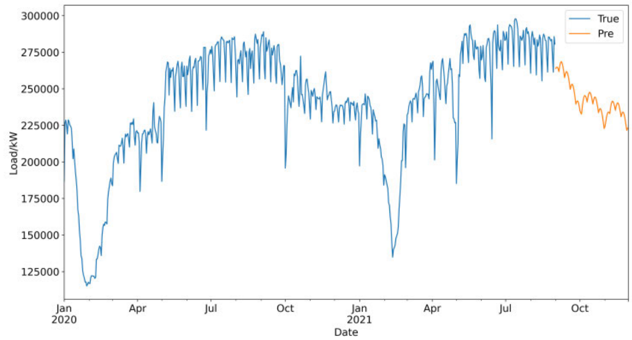

FIGURE 16. The forecast results of the maximum daily load for the next three months.

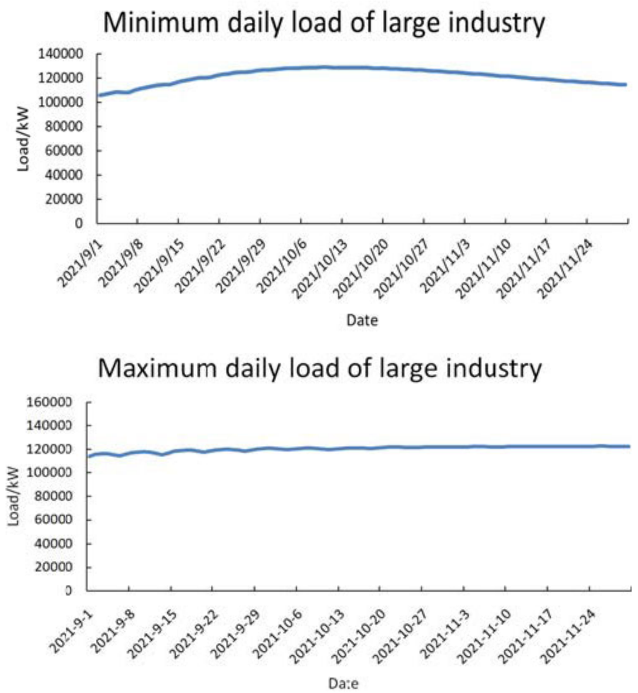

FIGURE 17. The extreme value forecast trend of large industry in the next three months.

In order to show the accuracy of the hybrid model more intuitively, Fig. 14 shows the power system load forecasting curves of each model. As can be seen from the figure, each model can forecast the trend of the curve. But over time, the difference manifests itself at the lowest point of the load. Traditional models such as DBN, BiLSTM, RF have large errors. Although CNN-LSTM also works well, the proposed hybrid model has less error and higher accuracy. Comprehensive analysis shows that the hybrid model in this paper has the best performance.

5) EVALUATION ANALYSIS OF MID-TERM POWER SYSTEM LOAD FORECASTING MODEL

This paper selects two evaluation indicators including: r2_ score function and mean squared logarithmic error. The

TABLE 8. Effect of the model on the maximum and minimum test sets for each industrial day.

Daily load / Test Effect Industry	Maximum	Minimum
Big Industry	0.004	0.011
General industry	0.005	0.006
Non-general Industry	0.003	0.010
Business	0.002	0.008

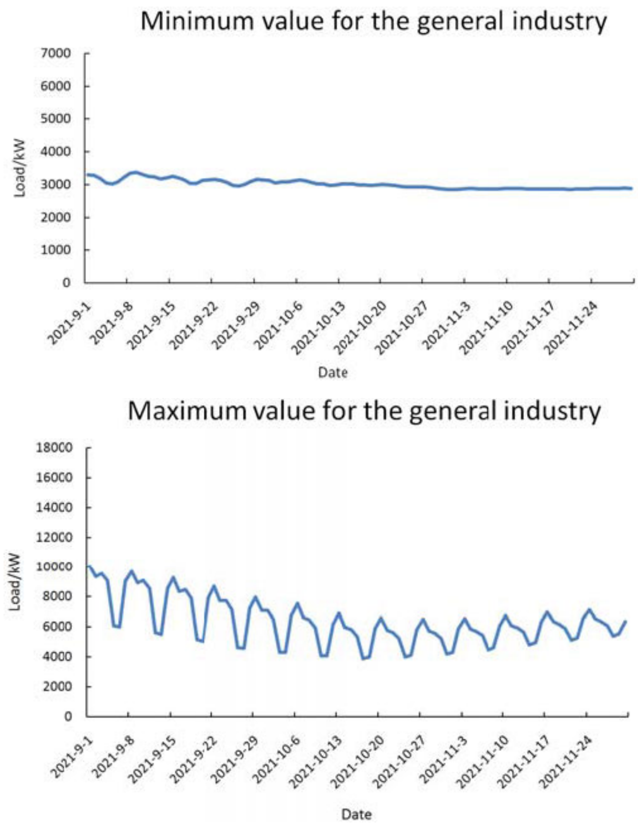

FIGURE 18. The extreme value forecast trend of general industry in the next three months.

purpose of this experiment is to verify the effect of the model on the test set. The experiment is conducted on these two indicators compared with other corresponding traditional models, such as: DBN, BiLSTM, RF, CNN-LSTM. These two indicators values are derived from the predicted results of the model on the predicted data, as shown in Table 7.

The final test results are the trend of minimum and maximum daily load forecasting for the next three months, as are shown in Fig.15 and Fig.16. In these two figures, the blue color line represents the real power load trend from January 2020 to August 2021. The yellow line represents the power load forecast trend for the next three months, that is, from September 2021 to November 2021. The two trend figures show the minimum and maximum daily power loads respectively.

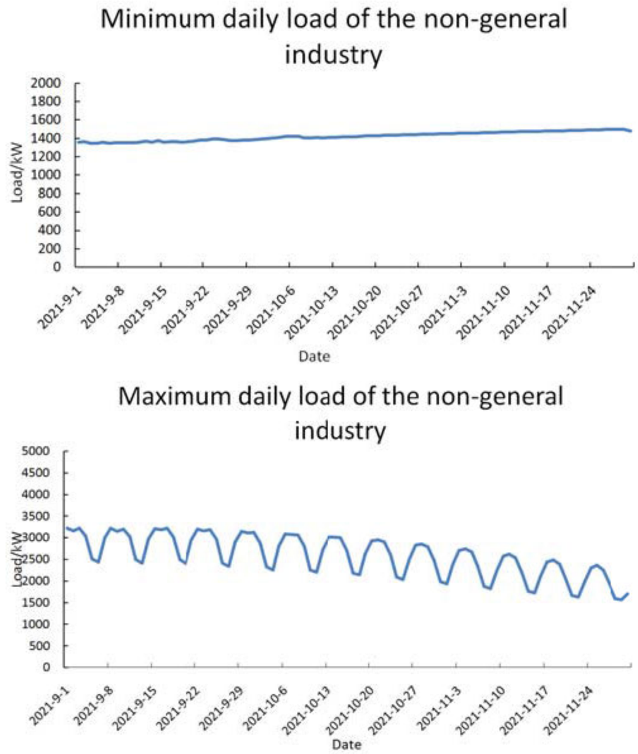

FIGURE 19. The extreme value forecast trend of non-general industry in the next three months.

The experimental results in Fig. 15 and 16 show that the trend predicted by the proposed model in this paper is basically consistent with the real power load trend in the previous period. Through mid-term power load forecasting analysis, the results show that the power load forecasting model built by BiLSTM and random forest hybrid deep learning can effectively improve the forecasting accuracy and achieve the forecasting effect.

V. FORECAST RESULTS AND ACCURACY ANALYSIS OF THE EXTREME VALUE BASED ON THE LOAD FORECASTING MODEL

This paper uses hybrid deep learning to build load forecasting models. Through experiments, this model is validated to forecast the maximum and minimum load values for each industry in the next three months. In this paper, the mean square logarithmic error is used to get the model effect evaluation index values. Table 8 shows the effect of the model on the maximum and minimum test sets of each industrial day.

The experiment in this paper is the index value obtained from the evaluation index. The experimental results show that the error of the prediction effect of the model on the test set is very small. In this paper, the maximum and minimum daily load of each industry in the next three months are predicted by the established load forecasting model. Finally, according to the upper and lower limits of the active power change curve

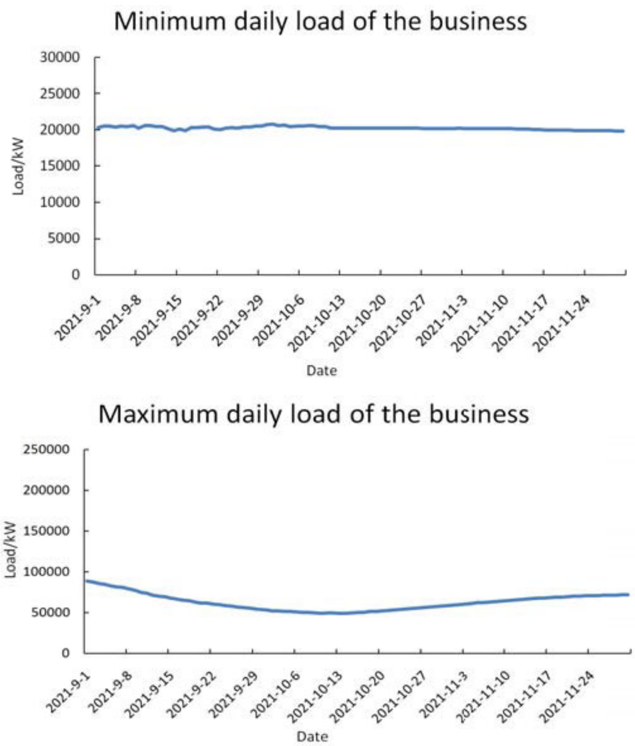

FIGURE 20. The extreme value forecast trend of business industry in the next three months.

of each industry, the forecast trend chart is constructed as follows:

(1) The extreme value forecast trend of large industry in the next three months is shown in Fig. 17.

(2) The extreme value forecast trend of general industry in the next three months is shown in Fig. 18.

(3) The extreme value forecast trend of non-general industry in the next three months is shown in Fig. 19.

(4) The extreme value forecast trend of business industry in the next three months is shown in Fig. 20.

From the experimental results in Fig. 17 to Fig. 20, it can be seen that for each industry, the annual change trend of power load is the same. According to the comparison between our prediction trend chart and the historical power load change chart, the results show that the prediction effect of the model conforms to the historical change trend. The experimental results show that the proposed model can effectively improve the accuracy of power system load forecasting.

VI. CONCLUSION

The accuracy of power load forecasting is one of the important guarantees for the economic and safe operation of power grids. It is an important basis for power enterprises to plan power structure, arrange generation plan and transaction plan. Therefore, it is very important to improve the accuracy of power load forecasting model.

This paper analyzes the historical load data of a regional power grid and four industries. First, the data is preprocessed. For the continuous missing values, this paper uses the exponential weight moving average method to fill in the missing values. Then, we delete duplicate values with the same indicators and abnormal values of data. Experiments show that these methods can effectively preprocess data. In this paper, a short-term power system load forecasting model based on BiLSTM is proposed, and the effect of the model is evaluated and analyzed. For mid-term load forecasting, this paper uses random forest and Pearson correlation coefficient to select features. A hybrid deep learning model based on BiLSTM and random forest is constructed, and then the parameters of the model are optimized. Finally, the mid-term power system load forecasting model is constructed based on the hybrid deep learning. In this paper, the benchmark models are selected for comparative experiments. The experimental results show that the MAPE of the proposed model is 2.36%, which is better than the benchmark models. This proves that the hybrid model is effective in prediction accuracy. The forecast model is applied to forecast the maximum and minimum daily load of each industry in the next three months. The experimental results show that the model can effectively improve the accuracy of power system load forecasting.

The limitation of this paper is that LSTM network model only considers one hidden layer, and the number of hidden layers can be increased to improve the accuracy of power forecasting model. In the future research, in order to improve the accuracy of power load model feature extraction, we can try to classify the data and use different feature extraction algorithms for different data. In addition, the load data can be divided into different components to extract their own characteristic information. Therefore, the calculation speed and prediction accuracy of the prediction model can be improved by reducing the calculation scale of the prediction model.

REFERENCES

[1] L. Lv, Z. Wu, L. Zhang, B. B. Gupta, and Z. Tian, "An edge-AI based forecasting approach for improving smart microgrid efficiency," *IEEE Trans. Ind. Informat.*, vol. 18, no. 11, pp. 7946–7954, Nov. 2022.
[2] R. Jiao, T. Zhang, Y. Jiang, and H. He, "Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network," *IEEE Access*, vol. 6, pp. 59438–59448, 2018.
[3] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Proc.*, vol. 158, pp. 2922–2927, Feb. 2019.
[4] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.
[5] Z. Chen, J. Liu, C. Li, X. Ji, D. Li, Y. Huang, F. Di, X. Gao, and L. Xu, "Ultra short-term power load forecasting based on LSTM and XGBoost combined model," *Power Syst. Technol.*, vol. 44, no. 2, pp. 614–620, 2020.
[6] W. Peng, J. R. Wang, and S. Q. Yin, "Short-term load forecasting model based on attention-LSTM in electricity market," *Power Syst. Technol.*, vol. 43, no. 5, pp. 1745–1751, 2019.
[7] S. H. Rafi, N.-A. Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated CNN and LSTM network," *IEEE Access*, vol. 9, pp. 32436–32448, 2021.
[8] L. Cai, J. Gu, and Z. Jin, "Two-layer transfer-learning-based architecture for short-term load forecasting," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1722–1732, Mar. 2020.
[9] E. A. Madrid and N. Antonio, "Short-term electricity load forecasting with machine learning," *Information*, vol. 12, no. 2, p. 50, Jan. 2021.

[10] F. U. M. Ullah, A. Ullah, N. Khan, M. Y. Lee, S. Rho, and S. W. Baik, "Deep learning-assisted short-term power load forecasting using deep convolutional LSTM and stacked GRU," *Complexity*, vol. 2022, pp. 1–15, Aug. 2022.
[11] G. M. U. Din and A. K. Marnerides, "Short term power load forecasting using deep neural networks," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Silicon Valley, CA, USA, Jan. 2017, pp. 594–598.
[12] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
[13] L. Zhang, C. Xu, Y. Gao, Y. Han, X. Du, and Z. Tian, "Improved Dota2 lineup recommendation model based on a bidirectional LSTM," *Tsinghua Sci. Technol.*, vol. 25, no. 6, pp. 712–720, Dec. 2020.
[14] Y. Jin, H. Guo, J. Wang, and A. Song, "A hybrid system based on LSTM for short-term power load forecasting," *Energies*, vol. 13, no. 23, p. 6241, Nov. 2020.
[15] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting," *IEEE Access*, vol. 8, pp. 180544–180557, 2020.
[16] H. Gamal, A. Alsaihati, S. Elkatatny, S. Haidary, and A. Abdulraheem, "Rock strength prediction in real-time while drilling employing random forest and functional network techniques," *J. Energy Resour. Technol.*, vol. 143, no. 9, pp. 1–21, Sep. 2021.
[17] Q. Wu, Y. Ye, H. Zhang, M. K. Ng, and S.-S. Ho, "ForesTexter: An efficient random forest algorithm for imbalanced text categorization," *Knowl.-Based Syst.*, vol. 67, pp. 105–116, Sep. 2014.
[18] L. Lv, J. Chen, Z. Zhang, B. Wang, and L. Zhang, "A numerical solution of a class of periodic coupled matrix equations," *J. Franklin Inst.*, vol. 358, no. 3, pp. 2039–2059, Feb. 2021.
[19] N. Mughees, S. A. Mohsin, A. Mughees, and A. Mughees, "Deep sequence to sequence bi-LSTM neural networks for day-ahead peak load forecasting," *Expert Syst. Appl.*, vol. 175, Aug. 2021, Art. no. 114844.
[20] H. Zang, R. Xu, L. Cheng, T. Ding, L. Liu, Z. Wei, and G. Sun, "Residential load forecasting based on LSTM fusing self-attention mechanism with pooling," *Energy*, vol. 229, Aug. 2021, Art. no. 1206682.
[21] L. Zhang, Y. Huo, Q. Ge, Y. Ma, Q. Liu, and W. Ouyang, "A privacy protection scheme for IoT big data based on time and frequency limitation," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–10, Jul. 2021.
[22] L. Zhang, Z. Huang, W. Liu, Z. Guo, and Z. Zhang, "Weather radar echo prediction method based on convolution neural network and long short-term memory networks for sustainable e-agriculture," *J. Cleaner Prod.*, vol. 298, May 2021, Art. no. 126776.
[23] J. Li, D. Deng, J. Zhao, D. Cai, W. Hu, M. Zhang, and Q. Huang, "A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2443–2452, Apr. 2021.
[24] L. Lv, Z. Wu, J. Zhang, L. Zhang, Z. Tan, and Z. Tian, "A VMD and LSTM based hybrid model of load forecasting for power grid security," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6474–6482, Sep. 2022.
[25] H. Xu, G. Fan, and Y. Song, "Novel key indicators selection method of financial fraud prediction model based on machine learning hybrid mode," *Mobile Inf. Syst.*, vol. 2022, pp. 1–12, Mar. 2022.
[26] H. Xu, G. Fan, and Y. Song, "Application analysis of the machine learning fusion model in building a financial fraud prediction model," *Secur. Commun. Netw.*, vol. 2022, pp. 1–13, Mar. 2022.

HONGSHENG XU received the M.S. degree from Henan University, China, in 2007. He is currently pursuing the degree with the Henan Key Laboratory for Big Data Processing and Analytics of Electronic Commerce, Luoyang Normal University. He was an Associate Professor with the College of Electronic Commerce, Luoyang Normal University. His research interests include artificial intelligence, deep learning, power load forecasting, knowledge graph, and machine learning.

GANGLONG FAN received the M.S. degree from the Huazhong University of Science and Technology, China. He was a Professor with the College of Electronic Commerce, Luoyang Normal University, where he is currently with the Henan Key Laboratory for Big Data Processing and Analytics of Electronic Commerce. His research interests include deep learning, machine learning, and electronic commerce.

GUOFANG KUANG received the M.S. degree from Harbin Engineering University, China, in 2007. He was an Associate Professor with the School of Information Technology, Luoyang Normal University. His research interests include artificial intelligence, machine learning, knowledge graph, and big data analysis.

YANPING SONG received the B.S. degree from Luoyang Normal University, China, in 2023. She is currently pursuing the degree with the College of Electronic Commerce, Luoyang Normal University, and the Henan Key Laboratory for Big Data Processing and Analytics of Electronic Commerce, Luoyang Normal University. Her research interests include deep learning, knowledge graph, and power load forecasting.