

Article

Long-Term Power Load Forecasting Using LSTM-Informer with Ensemble Learning

Kun Wang ^{1,2}, Junlong Zhang ^{1,2}, Xiwang Li ^{1,2,*} and Yaxin Zhang ³¹ Department of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China² Shenyang Institute of Computing Technology of Chinese Academy of Sciences, Shenyang 110168, China³ School of Mathematics, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

* Correspondence: lixw@sict.ac.cn

Abstract: Accurate power load forecasting can facilitate effective distribution of power and avoid wasting power so as to reduce costs. Power load is affected by many factors, so accurate forecasting is more difficult, and the current methods are mostly aimed at short-term power load forecasting problems. There is no good method for long-term power load forecasting problems. Aiming at this problem, this paper proposes an LSTM-Informer model based on ensemble learning to solve the long-term load forecasting problem. The bottom layer of the model uses the long short-term memory network (LSTM) model as a learner to capture the short-term time correlation of power load, and the top layer uses the Informer model to solve the long-term dependence problem of power load forecasting. In this way, the LSTM-Informer model can not only capture short-term time correlation but can also accurately predict long-term power load. In this paper, a one-year dataset of the distribution network in the city of Tetouan in northern Morocco was used for experiments, and the mean square error (MSE) and mean absolute error (MAE) were used as evaluation criteria. The long-term prediction of this model is 0.58 and 0.38 higher than that of the lstm model based on MSE and MAE. The experimental results show that the LSTM-Informer model based on ensemble learning has more advantages in long-term power load forecasting than the advanced baseline method.

Keywords: ensemble learning; energy consumption forecasting; neural networks; long-term load forecasting



Citation: Wang, K.; Zhang, J.; Li, X.; Zhang, Y. Long-Term Power Load Forecasting Using LSTM-Informer with Ensemble Learning. *Electronics* **2023**, *12*, 2175. <https://doi.org/10.3390/electronics12102175>

Academic Editor: Ahmed Abu-Siada

Received: 25 February 2023

Revised: 7 April 2023

Accepted: 15 April 2023

Published: 10 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background and Literature Review

In 2021, the share of electricity in global final consumption increased by 0.2 points, reaching 20.4% [1]. Electricity consumption has been increasing in recent years. It can be seen that electricity is becoming more and more important in our daily life. With the increasing demand for electricity, the country needs to build more power stations to meet the needs of human production. The purpose of the establishment of the national power system is to meet the power demand [2]. If there is no accurate prediction of long-term power load, it will lead to too many power generation facilities or insufficient power generation facilities. Excessive establishment of power generation facilities will lead to a waste of electricity and affect economic decision-making. The lack of power generation facilities is more serious, which may lead to insufficient power supply and affect people's daily life. Nowadays, the main task of power companies is to predict the power load so as to adjust the power supply and study the expansion planning of power generation facilities. This paper studies the long-term power forecasting, aiming to solve the problem of expansion planning and transformation of the power system.

Because the electric energy in the power grid system cannot easily be stored in large quantities and the power demand changes all the time, the power company needs the system to generate electricity and charge changes to achieve dynamic balance. In order

to realize the dynamic balance between power generation and charge change, researchers mostly study the power load based on the time series [3] for power load forecasting. Until now, there have been traditional models [4] and artificial intelligence (AI) models [5] for prediction.

Traditional modeling is based on statistical analysis and has good interpretability. The autoregressive moving average model (ARMA) is a classical statistical modeling method [6]. Nowicka-Zagrajek et al. [7] applied the ARMA model to California's short-term power forecasting and achieved good results. However, ARMA is only suitable for stationary stochastic processes [7]. Most sequences in nature are non-stationary. Therefore, the researchers proposed an autoregressive integrated moving average (ARIMA) model [8], which converts non-stationary sequences into stationary sequences based on differential operations. Valipour et al. [9] used the ARIMA model to predict the monthly inflow of Deziba Reservoir, and the prediction results are better than ARMA model. In addition, traditional modeling can achieve good results in solving linear problems [10], but it cannot solve nonlinear problems well and cannot deal with multivariate time series problems.

AI modeling is data-driven, which has been widely used in power load forecasting, such as back propagation (BP) [11] and artificial neural network (ANN) [12]. In general, power load forecasting needs to be abstracted into time series forecasting, while traditional BP neural networks cannot deal with time series problems well. A.S. Carpinteiro et al. [13] proposed the use of ANN models to make long-term predictions of future power loads using data obtained from North American Electric Power Corporation. The recurrent neural network (RNN) is a special kind of ANN [14], which retains a small amount of previous information through a self-connected structure, so as to establish the relationship between the past and the present. Tomonobu et al. [15] applied RNNs to long-term forecasting of wind power generation. This model is more accurate than the feed-forward neural network (FNN) model. However, because the RNN model is prone to gradient disappearance or gradient explosion problems, it is not effective in solving long-term dependence problems and can only deal with short-term dependence problems. However, in the LSTM model, the use of gating unit design alleviates the problem to a certain extent. Therefore, the LSTM model has the ability to deal with both short-term dependence and long-term dependence [16]. Jian et al. [17] proposed a periodic long-term power load forecasting model based on the LSTM network. The performance of this model is better than that of the ARIMA model. With the introduction of the Transformer model, the Attention mechanism has become a useful tool for solving problems in the natural language processing (NLP) field. Vaswani et al. [18] proposed an attention mechanism to solve machine translation problems, and looked forward to the future, believing that the attention mechanism can be applied to other fields. The attention mechanism can solve the long-term dependence problem of input and output, which makes researchers try to apply the Transformer model in the long-term prediction of time series. Wu et al. [19] proposed the Autoformer model. This model added the seasonality of data and other factors on the basis of the Transformer, carried out LTLF for the power consumption of 321 customers, and achieved good performance.

Due to the complexity of power load data, the researchers found that the accuracy of a single model is difficult to improve, so they began to explore ensemble learning. Divina et al. [10] elaborated the concept of the Stacking Ensemble Scheme and divided ensemble learning into three categories: Bagging, Boosting, and Stacking. The model constructed by the author in this paper is based on Evolutionary Algorithms (EAs), Random Forests (RF), ANN [20], and Generalized Boosted Regression Models (GBM), which improves the accuracy of power prediction. Kaur et al. [21] used the RNN-LSTM integrated learning model to manage the smart grid, which improves the prediction accuracy. Jung et al. [22] proposed using the LSTM-RNN model to predict long-term photovoltaic power generation for 63 months of data collected from 167 PV sites, and achieved good performance.

From the time range of prediction, the prediction of power load is divided into short-term power forecasting (STLF), medium-term power forecasting (MTLF), and long-

term power forecasting (LTLF) [23]. LTLF allows people to find and evaluate suitable photovoltaic power generation locations on a large scale. In the above papers, the authors found that the ensemble learning model has better performance after comparing the effects of the ensemble model and the single model. In addition, economic, environmental, and other factors also affect the consumption of power load [24]. Fan et al. enhanced the precision of power load by adding weather multivariate variables [25]. It can be seen that multivariate prediction has improved power prediction.

1.2. Research Gap

Through the background investigation of Section 1.1, the advantages of the traditional model are simple and interpretable. However, the performance on nonlinear and non-stationary sequences is not ideal. Even if the ARIMA model enhances the ability to solve non-stationary sequences, it still does not have the ability to solve nonlinear sequences well. Although the ARIMA model can solve the non-stationary problem, its parameters are difficult to adjust. At the same time, traditional algorithms cannot handle multivariate time series problems. In AI modeling, the BP neural network [26] model cannot capture time domain information. The RNN [27] model can capture time-domain information, but due to the problem of gradient vanishing and gradient exploding, the RNN model does not perform well in long-term prediction. The LSTM model alleviates the problem of gradient vanishing and gradient exploding to a certain extent by adding gating units, and has the ability of long-term prediction, but there are also large errors. The Attention mechanism solves the long-term dependence of input and output by calculating the correlation of all data and has good long-term prediction ability. However, it is not sensitive to short-term sequence features, resulting in large errors in the final results.

In summary, a single model always has its limitations, and the ensemble models can compensate for the shortcomings of a single model by integrating the advantages of multiple individuals. In previous studies, most of the work focused on the use of a single model for tuning and failed to combine the advantages of each model to improve performance.

1.3. Contribution

In the past, AI modeling was applied to power load forecasting research. Due to the inability to solve the gradient problem, most studies focused on STL. LTLF articles only accounted for 5% [28], but long-term power forecasting has practical production significance. Therefore, this paper proposes the LSTM-Informer model to solve the LTLF problem. This paper introduces the Transformer model which has achieved excellent results in other fields and uses the Attention mechanism to solve the long-term dependence of input and output data. At the same time, the integrated LSTM model captures short-term features, making up for the insensitivity of the Transformer model to short-term dependencies. Finally, the LSTM model and the Informer model were integrated to form the LSTM-Informer model, and the experiment proves that the LSTM-Informer model has excellent performance in long-term power load forecasting.

- The Pearson model and the RF model were used to process multivariate data to find the variables most relevant to electricity forecasting.
- A relatively new integrated learning LSTM-Informer model was constructed. The bottom layer of this model uses the LSTM model as a learner to capture the short-term time correlation of power load, and the top layer uses the Informer model to solve the long-term dependence of input and output. The model can accurately predict long-term power load while capturing short-term time correlation.
- For the analyzed data, the model proposed in this paper has good performance in accuracy and fitting degree.

2. Problem Statement

In this work, we assume that we can obtain the weather data (\vec{W}) of the power load during N years and the corresponding power load data (\vec{P}). The available weather data are temperature (\vec{Temp}), humidity (\vec{Hum}), wind speed (\vec{Ws}), etc.

This work proposes a new ensemble learning method for the accurate forecast of power load. The time stamp (in ten minutes) was used to deal with the correlation of time. Suppose that the power load at t minute on day d is $P(t, d)$. The influencing factors of power load forecasting weather variables $\vec{W}(t - i, d) = [Temp(t - i, d), Hum(t - i, d), Ws(t - i, d)]$, Following that, the vector input by the ensemble method is: $\vec{x}(t - i, d) = [P(t - i, d), \vec{W}(t - i, d)]$, and the power load to be predicted is: $y(t, d) = \hat{P}(t, d)$.

3. Deep Learning Forecast Model

3.1. Long Short-Term Memory Networks

RNN is a type of ANN that takes care of dependencies among data nodes [29]. The concept of the hidden state is introduced into the RNN model. The hidden state can extract the eigenvalues of the data and output after transformation. RNNs excel on short-term dependencies issues. However, the model can not deal with the long-term dependence problem well [30]. LSTM network was invented with the goal of addressing the vanishing gradients problem [31]. The LSTM model introduces a gate control mechanism [32]. The forgetting gate determines the information of cell state loss, the input gate determines the cell state to store new information, and the output gate determines the information to be output [33]. Figure 1 shows the structure of LSTM. c_t is the state of the cell at t, c_{t-1} the state of the cell at t - 1. f_t denotes the forget gate. i_t denotes the input gate. o_t denotes the output gate. Its mathematical formula is:

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [x_{l_{s_t}} \cdot h_{l_{s_{t-1}}}] + b_{l_{s_f}}) \\
 i_t &= \sigma(W_{l_{s_i}} \cdot [h_{l_{s_{t-1}}} \cdot x_{l_{s_t}}] + b_{l_{s_i}}) \\
 \tilde{c}_t &= \tanh(W_{l_{s_c}} \cdot [h_{l_{s_{t-1}}} \cdot x_{l_{s_t}}] + b_{l_{s_c}}) \\
 c_t &= f_t * c_{t-1} + i_t * \tilde{c}_t \\
 o_t &= \sigma(W_{l_{s_o}} \cdot [h_{l_{s_{t-1}}} \cdot x_t] + b_{l_{s_o}}) \\
 h_t &= o_t * \tanh(c_t)
 \end{aligned}
 \tag{1}$$

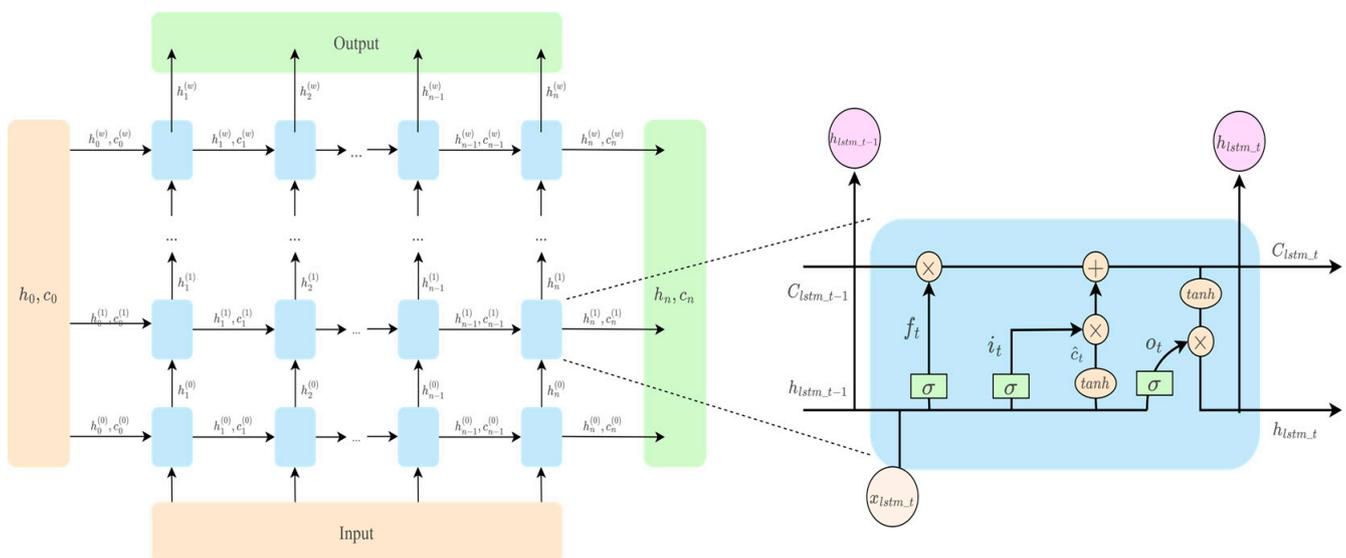


Figure 1. The structure of LSTM.

3.2. Informer

The Informer structure is shown in Figure 2. It consists of an encoder and a decoder. When the Informer model inputs the vector, it will add position encoding (timestamp) information to mark the position relationship at different times to deal with the correlation of time. Compared with the Transformer model, the multi-head attention mechanism of the Informer model focuses more on data with more obvious degradation trends so as to better solve the long-term dependence problem. The input of the decoder consists of two parts—one from the hidden intermediate data features of the output of the encoder and the other from the original input vector. The value to be predicted is assigned to 0, so as to prevent the previous position and pay attention to the information of the position to be predicted in the future. Following that, the data were connected to the multi-head attention mechanism. Finally, the full connection was performed to obtain the final output.

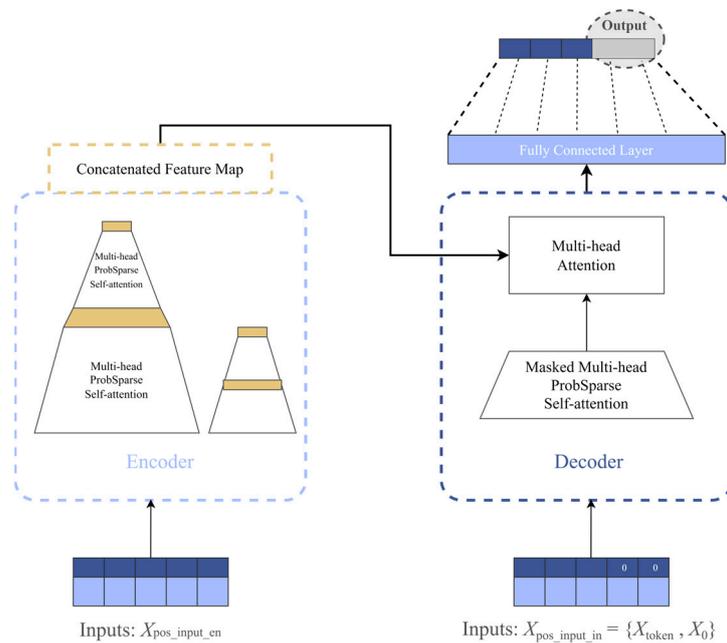


Figure 2. Structure of Informer.

Q, K, V vector constitutes the attention mechanism. Suppose q_i, k_i, v_i is the i th row of matrix QKV . The line i of the final output z can be expressed as:

$$A(q_i, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = E_{p(k_j|q_i)}[v_j] \tag{2}$$

The evaluation of the i th query sparsity is:

$$M(q_i, K) = \ln \sum_{j=1}^{L_k} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \tag{3}$$

The final attention mechanism is:

$$A(Q, K, V) = \text{Softmax}\left(\frac{\overline{Q}K^T}{\sqrt{d}}\right)V \tag{4}$$

The Informer model encodes the data and adds position coding through the timestamp so that the data have time series dependence. EncoderStack is composed of multiple

Encoder layers and distilling layers. The EncoderStack structure is shown in Figure 3. The Layer Normalization [34] formula is expressed as:

$$o = \text{LayerNorm}(x + \text{Sublayer}(x)) \quad (5)$$

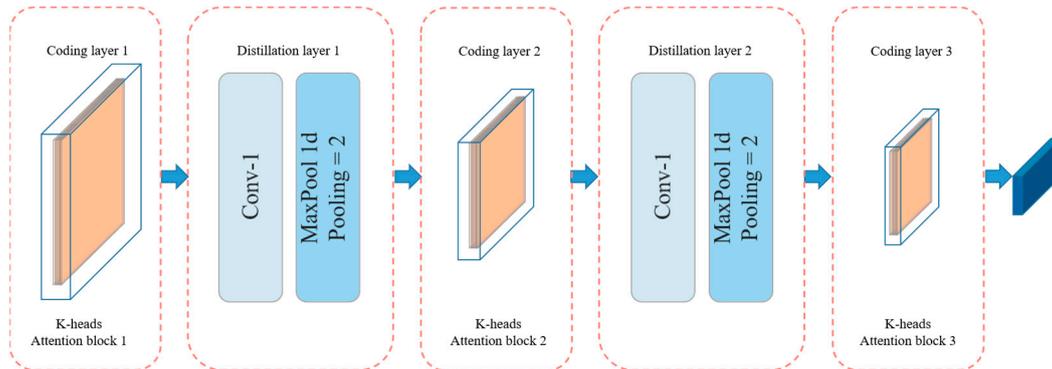


Figure 3. Structure of EncoderStack.

In the formula, LayerNorm is a Layer Normalization function [35]. The distilling mechanism improves robustness and reduces the use of network memory by a one-dimensional Convolution Layer, Activation Layer, and Pooling Layer in the time dimension. The distilling function formula is:

$$X_{j+1}^t = \text{MaxPool}(\text{ELU}(\text{Conv1d}([X_j^t]_{AB}))) \quad (6)$$

3.3. LSTM-Informer

Up until now, ensemble learning has gained an increasingly important position in regression, classification, and time series problems. For example, Jamali et al. [36] proposed a new hybrid model for solar heating. Mishra et al. [37] proposed using LSTM and wavelet transform to predict photovoltaic power generation, and the accuracy was improved compared with a single model. These methods fuse different models to improve the results of a single model. Ensemble learning transforms multiple weak learners into a strong learner. The process of ensemble learning is usually as follows: new feature data were generated through the training of the first-layer base learner, and then these new feature data were used as the input of the second layer and trained with the original data in the second layer learner. By synthesizing multiple single learners, a composite learner will have better performance than any single learner constructed. Its prediction level will be improved accordingly.

LSTM can effectively capture the correlation of data time. Since the LSTM model introduces the design of the door, the gradient vanishing problem is slowed down. Compared with other models, it also has a good effect on long-term dependence problems. Therefore, the underlying learner of this article is the LSTM model. The Informer model can effectively solve the long-term dependence problem. Therefore, the top-level learner in this article is Informer. The framework is used to preprocess data with the underlying LSTM model to solve the short-term dependence problem, and then the top-level Informer model is added to solve the long-term dependence problem.

The final model of this paper is shown in Figure 4. The original data were preprocessed by DataHandle data, and then data standardization [38]. After standardization, the data were divided into historical data vector and power load data to be predicted. The length of historical data is h , and the length of power load to be predicted is w . The whole data were divided into three parts according to the proportions. The training set and the validation set were put into the LSTM model for training to capture the short-term time correlation of power, and new eigenvalues were generated. The new eigenvalues were added to

the Informer model, and the overall data were trained to generate the final power load forecasting. Table 1 shows the architecture of the LSTM in this article.

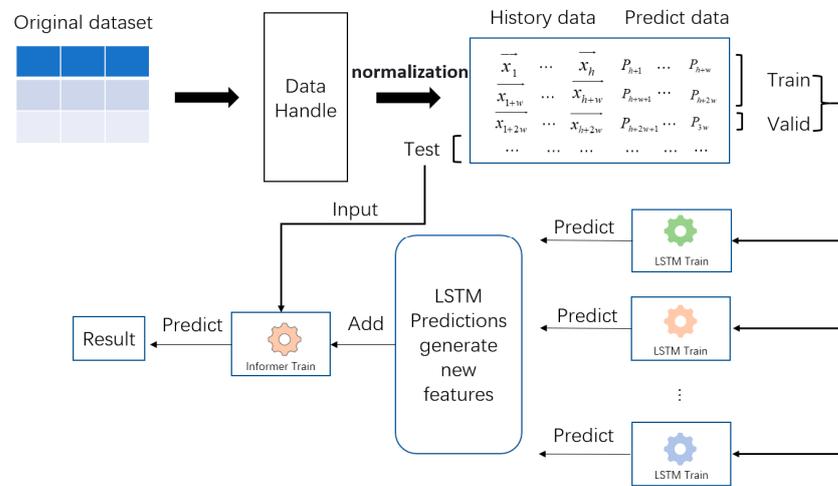


Figure 4. Power load forecasting architecture based on LSTM-Informer.

Table 1. LSTM architecture.

| Hidden_Size | Num_Layers | Batch_Size |
|-------------|------------|------------|
| 15 | 2 | 32 |

4. Experiment

4.1. Datasets

The data set is the power consumption data of the power grid in the three regions of Quads, Smir, and Boussafou in the city of Tetouan, Morocco. The data are made up of 52,416 pieces of data in a 10 min window from 2017. Each datapoint contains nine feature values. In the work of this paper, we used weather and other factors, the power generation of region one and region two as input variables to predict the power generation of region three.

Data sets are susceptible to various factors [29]. These data will affect the prediction results of the model. Therefore, the data need to be processed first. The model will be able to predict more accurately using the processed data. As shown in Figure 5, the process of data processing is in this paper.

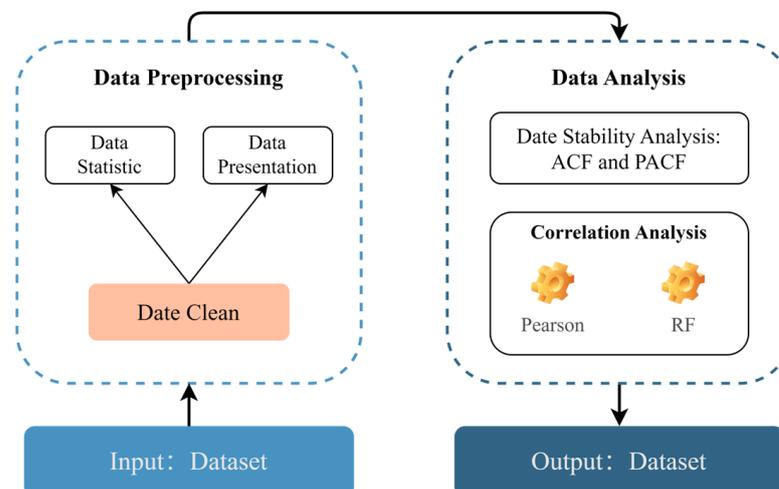


Figure 5. Structure of data handle.

Table 2 shows the statistical results of the data. For the nine variables, the count is completely consistent, indicating that there is no missing value. The mean value reflects the mean level of the number of variables with symmetrical normal distribution.

Table 2. Data statistics.

| | Temperature (°C) | Humidity (%) | Wind Speed (m/s) | General DiffuseFlows (w/m ²) | DiffuseFlows (w/m ²) | PowerC_Zone1 (kw) | PowerC_Zone2 (kw) | PowerC_Zone3 (kw) |
|-------|------------------|--------------|------------------|--|----------------------------------|-------------------|-------------------|-------------------|
| count | 52,416.00 | 52,416.00 | 52,416.00 | 52,416.00 | 52,416.00 | 52,416.00 | 52,416.00 | 52,416.00 |
| mean | 18.81 | 68.26 | 1.96 | 182.70 | 75.03 | 32,344.97 | 21,042.51 | 17,835.41 |
| Std | 5.82 | 15.55 | 2.35 | 264.40 | 124.21 | 7130.56 | 5201.471 | 6622.17 |
| min | 3.25 | 11.34 | 0.05 | 0.00 | 0.01 | 13,895.70 | 8560.08 | 5935.17 |
| max | 40.01 | 94.80 | 6.48 | 1163.00 | 936.00 | 52,204.40 | 37,408.86 | 47,598.33 |

As shown in Figure 6, the first five pictures show the data on weather factors during the year, while the next three are statistical maps of power load consumption in the three regions. It can be seen that temperature and power load consumption are related.

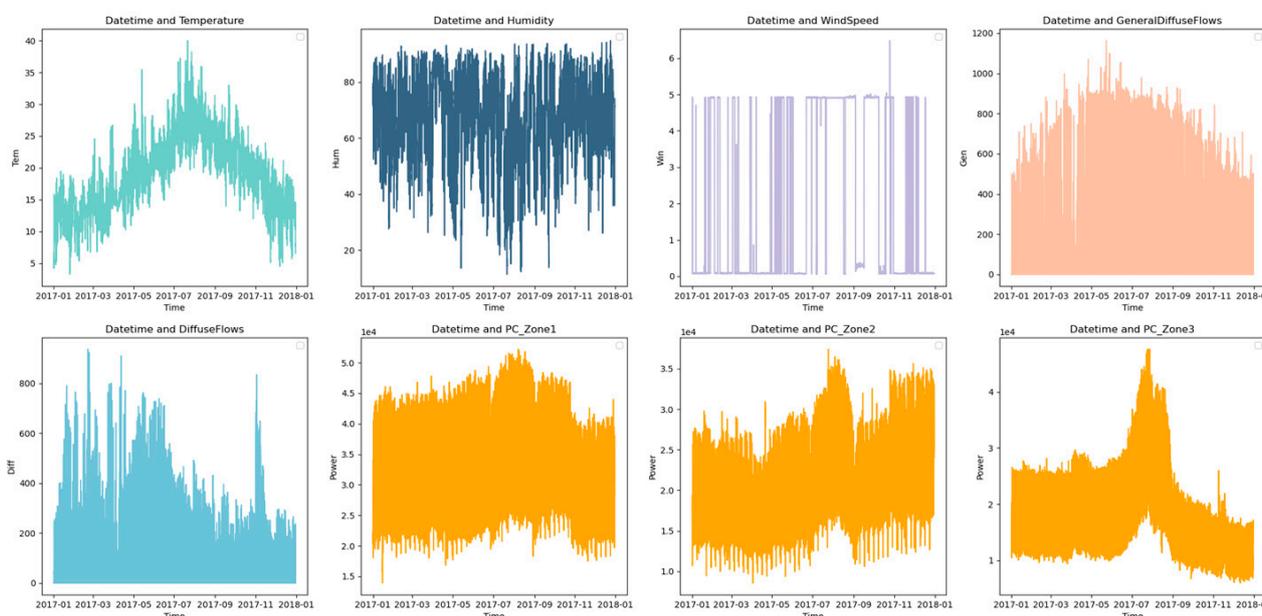


Figure 6. Statistics of weather factors and electricity consumption in three regions.

Data preprocessing processes the outliers of the data and allows us to visually see the basic characteristics of the data set. However, seeing only the basic features is not enough to better analyze the correlation between the data set and the model used, so we will enter the data analysis stage.

4.2. Variable Selection

In this paper, the Pearson coefficient and RF algorithm were used to analyze the relationship between this data set and the power load forecasting of area three that needs to be predicted. Select the associated variables to predict the power load consumption.

Pearson correlation coefficient [39] can measure the degree of non-linear correlation and correlation. Therefore, the Pearson coefficient can be used to calculate the correlation between power load of region 3 and other characteristics. If the power load of region 3 is X and the other characteristics are Y , the coefficient equation is:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \tag{7}$$

In the formula $\rho_{X,Y} \in (-1, 1)$, denotes the correlation strength of X , and Y , A positive value indicates that the relationship between this feature and power load is positively correlated, and the closer to 1, the stronger the correlation. A negative value indicates that the feature is negatively correlated with the power load. The closer the value is to 0 indicates that the selected eigenvalue is independent of the power load.

As shown in Table 3, temperature, wind speed, GeneralDiffuseFlows, and charge consumption in regions 1 and 2 are positively correlated with charge consumption in region 3. Humidity is negatively correlated with the charge consumption of DiffuseFlows and Region 3.

Table 3. Correlation analysis of power load value from region 3.

| Feature | ρ | Correlation |
|------------------------|--------|-------------|
| Temperature | 0.49 | + |
| Humidity | -0.23 | - |
| WindSpeed | 0.28 | + |
| GeneralDiffuseFlows | 0.063 | + |
| DiffuseFlows | -0.039 | - |
| PowerConsumption_Zone1 | 0.75 | + |
| PowerConsumption_Zone2 | 0.57 | - |

Figure 7 shows the relationship between the power load of region 3 and other characteristics more obviously. It can be seen that the temperature, region 1, and region 2 are related to the charge consumption of region 3.

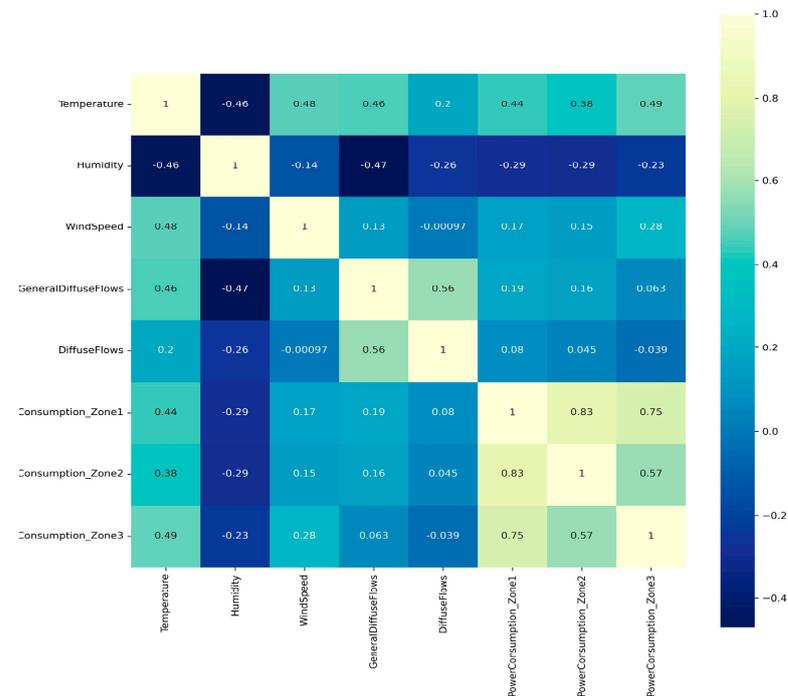


Figure 7. On the regional three power consumption correlation coefficient heat map.

RF correlation coefficient analysis:

Breiman (2001) proposed random forests, which add an additional layer of randomness to bagging [40]. RF is a forest formed by multiple decision trees in a random way. Different trees in a random forest may classify new incoming samples into different belongs [41]. We counted which category was judged the most and predicted this sample for this category. The structure is shown in Figure 8:

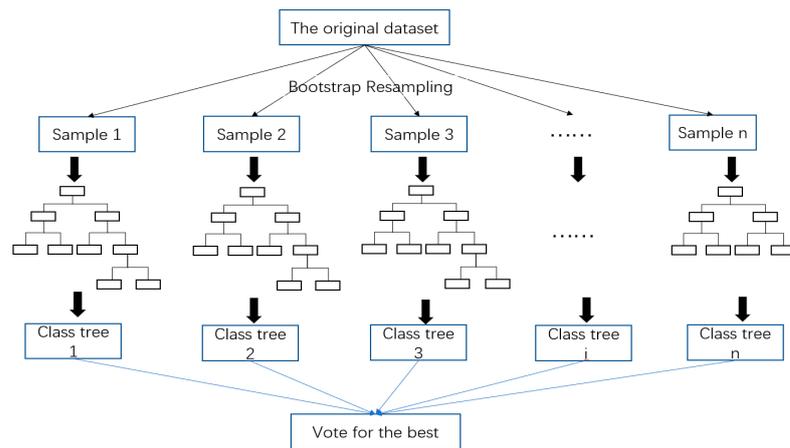


Figure 8. Random forest structure diagram.

RF mainly selects feature principle: the OOB (out of bag) principle [42]. Adding random noise to important features will have a great impact on the precision of the algorithm during RF training [43]. Out-of-bag data were selected in each decision tree and the error is calculated, and record it as: $OOB1_{err}$ [44]. Randomly add noise interference to all the sample features X obtained from the out-of-bag data, and calculate the out-of-bag data error again, recorded as: $OOB2_{err}$ [45]. Supposing that there are N trees, the X formula is as follows:

$$X_{importance} = \frac{1}{N} \sum_{i=1}^N |OOB2_{err} - OOB1_{err}| \tag{8}$$

Through the analysis of the Pearson coefficient and RF (Table 4), it can be known that the power load consumption of region 3 is related to the power consumption of the other two regions, that is, the power consumption of people’s daily habits is more relevant. The second is the effect of temperature. It shows that the weather factor is also an important factor affecting power consumption, which provides reference conditions for feature selection of power load forecasting.

Table 4. Correlation analysis of power load value from region 3.

| Feature | The Importance of Feature X |
|------------------------|-----------------------------|
| PowerConsumption_Zone1 | 0.580088 |
| PowerConsumption_Zone2 | 0.137847 |
| Temperature | 0.135731 |
| GeneralDiffuseFlows | 0.046054 |
| DiffuseFlows | 0.035704 |
| WindSpeed | 0.033177 |
| Humidity | 0.031397 |

4.3. Data Normalization

The results of data partitioning are shown in Figure 9. First, the data were standardized by Z-Score. Convert different orders of magnitude into the same order of magnitude. Following that, the data can be compared [46]. Its formula is:

$$x' = \frac{x - \mu}{\sigma} \tag{9}$$

After normalization, the data were divided into training set, validation set, and test set, accounting for 70%, 10%, and 20%, respectively. The data of a timestamp is \vec{x} . The length of the required historical data is h . The length of the power load to be predicted is w .

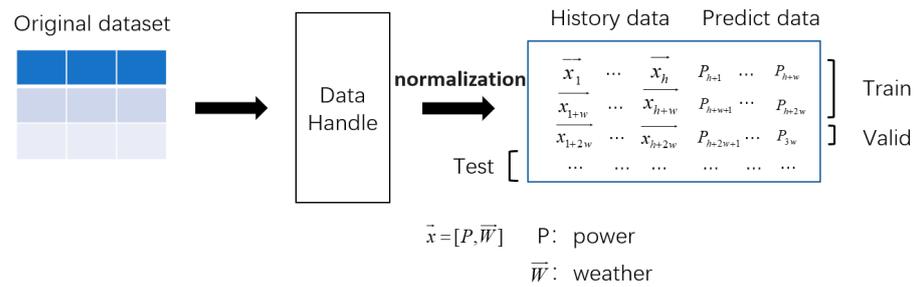


Figure 9. Data set preprocessing.

4.4. Methods for Comparison

This paper selects the more advanced models for comparison.

Informer [47]: The model improves on the basic Transformer model and is more suitable for long-term prediction problems.

Transformer [18]: At the beginning, it was a very famous text processing model. It first introduced the attention mechanism and was later applied to various fields, thus promoting the development of various fields. For the time series problem, the Transformer model shows superior performance in capturing the relationship of the gas dependence problem. The transformer model shows great potential in solving the LSTF problem [35].

Autoformer [19]: The model based on Transformer introduces seasonal and periodic terms to better solve the long-term prediction problem.

Reformer [48]: The use of locality-sensitive hashing instead of the original dot product Attention reduces the time complexity of Transformer, but the performance is comparable. It has higher memory efficiency and speed in long sequences.

LSTM [49]: A classical algorithm for dealing with time series problems which slows down the vanishing gradient and exploding gradient problems so that it can better capture long-term dependence problems.

4.5. Evaluation Metrics

In this chapter, we used the LSTM-Informer model to test and summarize the data set in Section 4.1. This paper selects MAE and MSE to contrast the precision of different algorithms. MAE refers to the average absolute value of the absolute deviation of each measurement. MSE is a measure of the difference between the estimator and the estimated quantity [34]. The formula of the above indicators is:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{10}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \tag{11}$$

In the above equation: n represents the time points of predictions, \hat{y}_i represents the result of the i th model prediction, and y_i represents the true value of the.

4.6. Method Comparison and Analysis

4.6.1. Experimental Result

In this paper, the proposed LSTM-Informer model was compared with advanced single models for prediction accuracy, including Informer model, Transformer model, Autoformer model, Reformer model, and LSTM model. The data used in this paper were sampled once every ten minutes. In order to comprehensively analyze the model performance, two comparison schemes were used here. The first comparison scheme is that the input length is shorter than the output length. The second comparison scheme is that the output length is longer than the input length. In the first comparison strategy, we used 288 pieces of data (48 h) to respectively predict the power load in Region3 of 24 pieces of data (4 h), 48 pieces

of data (8 h), 72 pieces of data (12 h), 96 pieces of data (16 h), and 120 pieces of data (20 h) that can evaluate the performance of the model’s short-term dependence problem. The results are shown in Table 5.

Table 5. Using 48 h forecast 4 h, 8 h, and other short-term power load forecasting.

| Methods | LSTM-Informer | | Informer | | Autoformer | | Transformer | | Reformer | | LSTM | |
|---------|---------------|--------|----------|--------|------------|--------|-------------|--------|----------|--------|--------|--------|
| Metric | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 24 | 0.0617 | 0.1988 | 0.0597 | 0.1929 | 0.2464 | 0.3997 | 0.1153 | 0.2943 | 0.2983 | 0.4738 | 0.0787 | 0.2245 |
| 48 | 0.0846 | 0.2431 | 0.0910 | 0.2424 | 1.2121 | 0.9072 | 0.1855 | 0.3766 | 0.3196 | 0.4922 | 0.1057 | 0.2533 |
| 72 | 0.0856 | 0.2415 | 0.0964 | 0.2584 | 0.2771 | 0.4115 | 0.1866 | 0.3779 | 0.4423 | 0.5895 | 0.1299 | 0.2988 |
| 96 | 0.1064 | 0.2755 | 0.1165 | 0.2879 | 0.8902 | 0.7659 | 0.2354 | 0.4396 | 0.4228 | 0.5731 | 0.1233 | 0.2879 |
| 120 | 0.1098 | 0.2802 | 0.1159 | 0.2848 | 0.4739 | 0.5371 | 0.2106 | 0.4043 | 0.4412 | 0.5858 | 0.1372 | 0.3157 |
| Count | 7 | | 3 | | 0 | | 0 | | 0 | | 0 | |

From the data in Table 5, we can notice the characteristics of each model in short-term power load forecasting. The autocorrelation mechanism of the Autoformer’s model is based on Period-based dependencies and Time delay aggregation. In STLFL, it is not easy to find its cycle, so the performance of the Autoformer model is not stable. Because the LSTM model can capture the time correlation of short-term power load, its performance is relatively stable. As the time to be predicted increases, the performance of the model in STLFL decreases slightly. In fact, the LSTM-Informer model performs better than other models in predicting short-term dependent power loads. On the issue of STLFL, the Informer model is the closest to the model performance proposed in this paper.

In order to analyze these methods intuitively, we drew six models of STLFL by ranking. They were used to better observe the effect of each model on short-term power load. The MSE and MAE results of the six STLFL models are shown in Figure 10. Through the ranking of MSE index, we can see that LSTM-Informer model and Informer model have the best performance in STLFL. The LSTM model and the Transformer model rank relatively stable in STLFL. Through the ranking of MAE indicators, we found that the Informer model and the LSTM-Informer model have similar performance in STLFL. Through the ranking of MAE indicators, we found that the Informer model and the LSTM-Informer model have similar performance in STLFL. Through two graphs, we found that the time complexity of the model optimized based on the Transformer model framework is reduced. However, half of the short-term power load forecasting problem was due to the basic model, and the other half of the effect ranking lagged behind the Transformer model. Since then, we have analyzed the model and the Informer model in detail on the issue of STLFL.

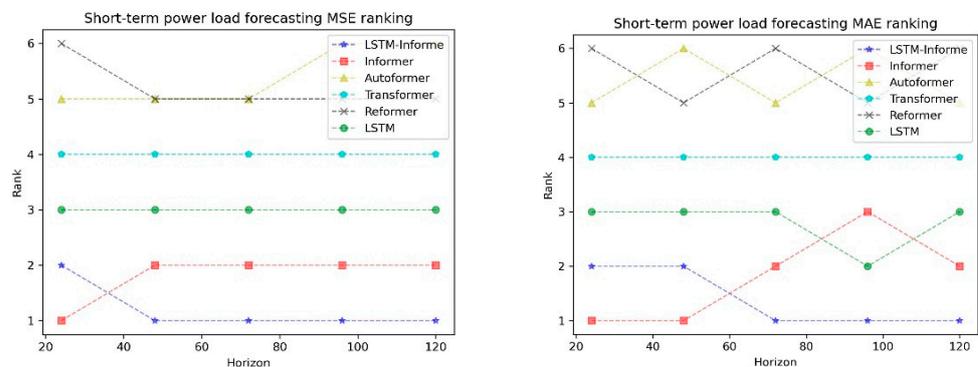


Figure 10. Comparison results of six short-term power load models via MSE and MAE metric.

In the second comparison strategy, we used 48 pieces of data (8 h) to respectively predict the power load in Region3 of 144 pieces of data (24 h), 192 pieces of data (32 h), 240 pieces of data (40 h), 288 pieces of data (48 h), and 432 pieces of data (72 h) that can evaluate the performance of the model’s short-term dependence problem. This is shown in Table 6.

Table 6. Using 8 h forecast 24 h, 32 h, and other long-term power load forecasting.

| Methods | LSTM-Informer | | Informer | | Autoformer | | Transformer | | Reformer | | LSTM | |
|---------|---------------|--------|----------|--------|------------|--------|-------------|--------|----------|--------|--------|--------|
| Metric | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 144 | 0.2085 | 0.3963 | 0.2690 | 0.4561 | 0.7061 | 0.6181 | 0.2692 | 0.4517 | 0.8949 | 0.8672 | 0.7509 | 0.7675 |
| 192 | 0.2644 | 0.4477 | 0.2979 | 0.4813 | 1.0563 | 0.7284 | 0.2736 | 0.4566 | 0.8762 | 0.8449 | 0.8128 | 0.8054 |
| 240 | 0.2407 | 0.4253 | 0.2691 | 0.4542 | 0.4664 | 0.5451 | 0.2510 | 0.4353 | 0.8868 | 0.8447 | 0.5970 | 0.6731 |
| 288 | 0.2490 | 0.4232 | 0.2544 | 0.4305 | 1.3293 | 0.7713 | 0.2111 | 0.3920 | 0.9092 | 0.8578 | 0.7085 | 0.7285 |
| 432 | 0.2030 | 0.3783 | 0.2786 | 0.4440 | 0.4519 | 0.5148 | 0.2357 | 0.4173 | 0.9253 | 0.8592 | 0.7808 | 0.7629 |
| Count | 8 | | 0 | | 0 | | 2 | | 0 | | 0 | |

Table 6 shows that the precision of Transformer architecture model generally tends to be stable in the long-term forecast of power load, but LSTM’s precision is gradually deteriorating by the increase of the forecast time. It shows that the LSTM model alleviates the gradient disappearance problem but does not completely solve the gradient disappearance problem, while the Transformer architecture model can generally solve the long-term dependence problem of power load prediction. The LSTM-Informer model can still be superior to other models in the long-term power load forecasting problem. At this time, the Transformer model is the closest to the performance of the LSTM-Informer model. After that, we will compare the precision of the two methods of LTLF.

The MSE and MAE results of the six LTLF models are shown in Figure 11. Through ranking, we can observe that the performance of LSTM model in LTLF problems has declined. Transformer’s model has improved the performance of long-term power load forecasting problems. It is close to the performance of the LSTM-Informer model, while the LSTM model drops seriously. It shows that the model architecture based on Transformer has better performance for LTLF. At this time, the closest performance to the LSTM-Informer model is the Transformer model. After that, we will compare the precision of the two methods concerning LTLF.

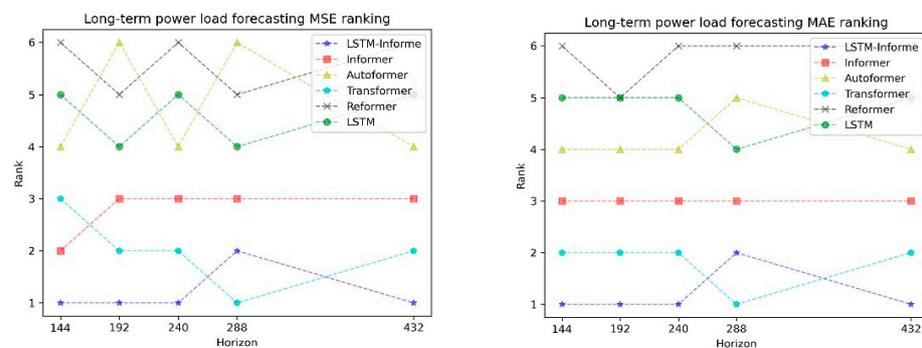


Figure 11. Comparison results of six long-term power load models via MSE and MAE metric.

Combining the results of the two tables, we can conclude that the LSTM-Informer model has improved the performance of the base learner on the short-term dependence problem. It has relatively good performance compared with other more advanced models with similar architectures. On the issue of long-term dependence, in most cases, it is superior to other single models. Although the Transformer model performs well in long-term prediction, its accuracy in short-term prediction is less than 50% of the LSTM-Informer performance. If a model is needed for power load forecasting, the LSTM-Informer model has the best performance. It is optimal in both STLf and LTLF.

4.6.2. Results Analysis

In order to further compare the differences between the LSTM-Informer model, Informer model, Autoformer model, Transformer model, Reformer model, and LSTM model, we will analyze the results in detail.

The experimental results of MAE of our selected model are shown in Figures 12 and 13. Figure 12 shows the MAE index of the selected model using 48 h of data to predict the power load of 4 h, 8 h, 12 h, 16 h, and 20 h, respectively. We can see that in the issues of STLTF, the LSTM model is superior to most of the Transformer structure models, indicating that preserving past information through hidden states is more important for STLTF. The LSTM-Informer model uses the LSTM model at the bottom to better find the time correlation between data, so that it pays attention to the time correlation of short-term data when training in the Informer model. Therefore, in short-term power load forecasting, the LSTM-Informer model has achieved good results. Figure 13 shows the MAE index of the selected model using 8 h of data to predict the power load of 24 h, 32 h, 40 h, 48 h and 72 h respectively. In the comparison models, the Informer model, Autoformer model, and Reformer model are all improved models based on the Transformer model. The transformer model is mainly composed of an encoder and decoder. Its internal self-attention mechanism enables data at each time point to pay attention to the data at other time points. In solving the time series problem, it embeds the time-tamp into the data through position encoding. The self-attention mechanism and position coding make it independent of the past hidden state to capture the dependence on the previous data, so it does not produce the gradient vanishing problem. At this time, the improved model based on Transformer has achieved good results, and as the time of power load to be predicted increases, the performance is also in a good state. However, as the time of the power load to be predicted increases, the accuracy of the LSTM model gradually decreases, indicating that the LSTM model has a gradient vanishing problem. However, the accuracy of the LSTM-Informer model proposed does not show a significant trend that the performance gradually deteriorates with the increase of the prediction length, indicating that when the LSTM model is capturing short-term event correlation, the Informer model successfully solves the long-term dependence problem. Therefore, in long-term power load forecasting, the LSTM-Informer model can still achieve the greatest results.

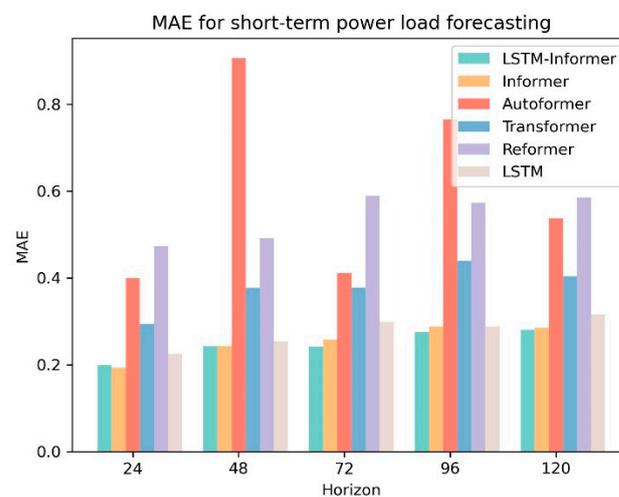


Figure 12. The MAE index of the selected model using 48 h of data to predict the power load of 4 h, 8 h, 12 h, 16 h, and 20 h.

In the above experiments, we can conclude that the performance of the STLTF Informer model is close to the LSTM-Informer model proposed in this paper. Therefore, we will further test the performance of the two models in STLTF. In the evaluation model performance in this paper, the two models are similar in numerical value. Therefore, we further compared the differences between the two models from the degree of fitting. As shown in Figure 14, the Informer model predicts the power load consumption value of 8 h with 48 h of power load value and fits the predicted power load value of 8 h with the true value. As shown in Figure 15, it shows that the LSTM-Informer model predicts the 8-h power load consumption value with 48-h power consumption and fits the 8-h power load consumption

prediction value with the true value. As shown in Figures 14 and 15, for STLFL, the predicted value of the LSTM-Informer model is smoother than the predicted value of the Informer model. The predicted value of the Informer model for the power load consumption trend is roughly the same, but compared to the real value, its predicted value is not so smooth, and the prominent value is higher. The power load consumption value predicted by the LSTM-Informer model is relatively smooth, and the trend of the real value is well-fitted. It shows that the LSTM-Informer model for STLFL is not only superior to other models in the performance index of the judgment model in this paper but is also superior to other models in terms of the real value.

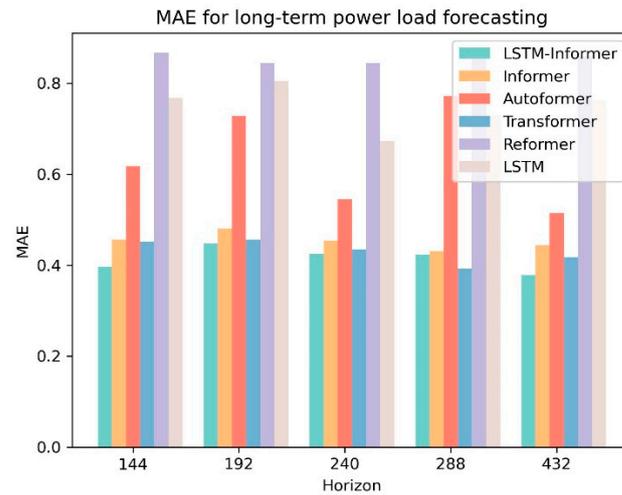


Figure 13. The MAE index of the selected model uses 8 h of data to predict the power load of 24 h, 32 h, 40 h, 48 h, and 72 h.

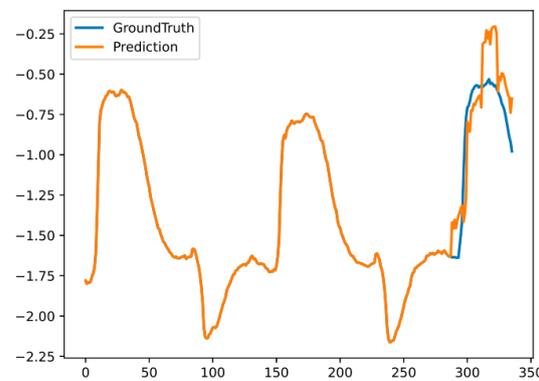


Figure 14. Informer model 48 h forecast 8 h power load consumption.

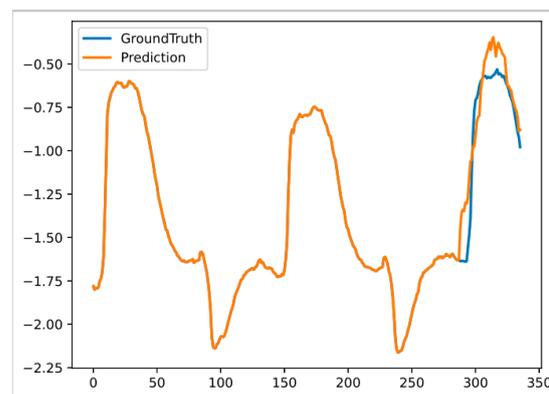


Figure 15. The LSTM-Informer model uses 48 h to predict 8 h of power load consumption.

From the above experiments, the Transformer model has the closest performance to the LSTM-Informer model proposed in this paper in long-term power load forecasting. Therefore, we will compare the performance of the two models in LTLF on the other hand. In the evaluation model performance in this paper, the two models are similar in numerical value. Therefore, we further compared the differences between the two models from the degree of fitting. Figure 16 shows the fitting degree of the LSTM-Informer model using 8-h power load consumption to predict 20-h power load. Figure 17 shows the fitting degree of the Transformer model using 8-h power load consumption to predict 20-h power load. Through the comparison of the two graphs, we can find that when the Transformer model predicts long-term power load, and its peak easily exceeds the real value. The slight fluctuation of the real value will cause the predicted value to fluctuate greatly. Therefore, the fitting degree of the predicted value relative to the real value is not very good. Compared with the Transformer model, the LSTM-Informer model has a better fitting degree for the peak when predicting long-term power load. In the red circle in Figure 16, it can better fit the real value at the peak, and basically does not exceed the real value. Even if some areas do not reach the peak, the error is still smaller than the error in Transformer model. Its predicted value does not change greatly with the slight change of the real value, but it fits the fluctuation trend of the real value well. This is due to the use of hidden states in the LSTM model, so that the LSTM is good at extracting short-term temporal correlations. Thus, we can better capture the trend of power load consumption and then capture the long-term power load consumption trend through the Informer model. The combination of the two models makes the LSTM-Informer model better predict the value of real power load consumption.

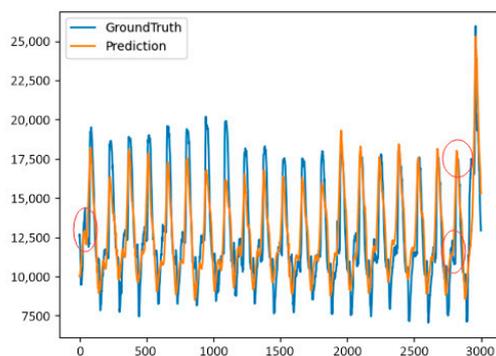


Figure 16. The LSTM-Informer model uses 8 h to predict the fitting degree of 20 h of power load.

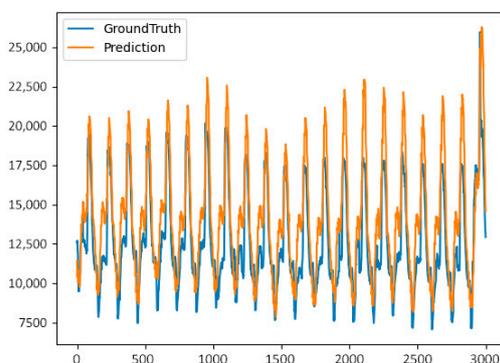


Figure 17. The Transformer model uses 8 h to predict the fitting degree of 20 h of power load.

5. Conclusions

For the nonlinear power load forecasting problem, the model based on artificial intelligence is one of the effective methods to solve the problem. This paper proposes a new ensemble learning (LSTM-Informer) model. The method is based on a two-layer structure. Its underlying learner first learns the training and validation sets based on the

LSTM model to capture the time correlation of short-term power load consumption and generate new features. Following that, the new features were added to the next layer of the learner Informer model. While obtaining the time correlation of short-term power load consumption, the Informer model does not depend on the past hidden state to capture the dependence on the previous data, so it will not produce the gradient disappearance problem so as to better predict the long-term power load consumption. Comparing the results of the LSTM-Informer model with several more advanced single models, it was found that the LSTM-Informer model outperforms other models on short-term and long-term power forecasting problems.

Author Contributions: X.L.: Project administration, Funding acquisition, Supervision. K.W.: Data curation, Writing—Original draft preparation, Conceptualization, Methodology. J.Z.: Visualization, Investigation. Y.Z.: Visualization. All authors have read and agreed to the published version of the manuscript.

Funding: In 2022: China Liaoning Province “unveiled the leader” key special project of science and technology, virtual power plant resource control and coordination optimization of key technologies and platform research and development, project number: 2022JH1/10800031.

Data Availability Statement: The data in this article is only used for research and has no permission to shares.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Share of Electricity in Total Final Energy Consumption. Available online: <https://yearbook.enerdata.net/electricity/share-electricity-final-consumption.html> (accessed on 1 January 2022).
2. Tang, L.; Wang, X.; Wang, X.; Shao, C.; Liu, S.; Tian, S. Long-term electricity consumption forecasting based on expert prediction and fuzzy Bayesian theory. *Energy* **2019**, *167*, 1144–1154. [CrossRef]
3. Kim, T.Y.; Cho, S.B. Predicting the household power consumption using CNN-LSTM hybrid networks. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Madrid, Spain, 21–23 November 2018; Springer: Cham, Switzerland, 2018; pp. 481–490.
4. Fan, D.; Sun, H.; Yao, J.; Zhang, K.; Yan, X.; Sun, Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy* **2021**, *220*, 119708. [CrossRef]
5. Ahmad, T.; Zhang, D.; Huang, C.; Zhang, H.; Dai, N.; Song, Y.; Chen, H. Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *J. Clean. Prod.* **2021**, *289*, 125834. [CrossRef]
6. Yu, C.; Li, Y.; Chen, Q.; Lai, X.; Zhao, L. Matrix-based wavelet transformation embedded in recurrent neural networks for wind speed prediction. *Appl. Energy* **2022**, *324*, 119692. [CrossRef]
7. Nowicka-Zagrajek, J.; Weron, R. Modeling electricity loads in California: ARMA models with hyperbolic noise. *Signal Process.* **2002**, *82*, 1903–1915. [CrossRef]
8. Chen, X.; Jia, S.; Ding, L.; Xiang, Y. Reasoning over temporal knowledge graph with temporal consistency constraints. *J. Intell. Fuzzy Syst.* **2021**, *40*, 11941–11950. [CrossRef]
9. Valipour, M.; Banihabib, M.E.; Behbahani, S.M.R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* **2013**, *476*, 433–441. [CrossRef]
10. Divina, F.; Gilson, A.; Gómez-Vela, F.; García Torres, M.; Torres, J.F. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* **2018**, *11*, 949. [CrossRef]
11. Gu, B.; Shen, H.; Lei, X.; Hu, H.; Liu, X. Forecasting and uncertainty analysis of day-ahead photovoltaic power using a novel forecasting method. *Appl. Energy* **2021**, *299*, 117291. [CrossRef]
12. Wang, J.; Gao, J.; Wei, D. Electric load prediction based on a novel combined interval forecasting system. *Appl. Energy* **2022**, *322*, 119420. [CrossRef]
13. Carpinteiro, O.A.; Leme, R.C.; de Souza, A.C.Z.; Pinheiro, C.A.; Moreira, E.M. Long-term load forecasting via a hierarchical neural model with time integrators. *Electr. Power Syst. Res.* **2007**, *77*, 371–378. [CrossRef]
14. Eldeeb, E. Traffic Classification and Prediction, and Fast Uplink Grant Allocation for Machine Type Communications via Support Vector Machines and Long Short-Term Memory. Master’s Thesis, University of Oulu, Oulu, Finland, 2020.
15. Senjyu, T.; Yona, A.; Urasaki, N.; Funabashi, T. Application of recurrent neural network to long-term-ahead generating power forecasting for wind power generator. In Proceedings of the 2006 IEEE PES Power Systems Conference and Exposition, Atlanta, GA, USA, 29 October–1 November 2006; pp. 1260–1265.
16. Shewalkar, A.; Nyavanandi, D.; Ludwig, S.A. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. *J. Artif. Intell. Soft Comput. Res.* **2019**, *9*, 235–245. [CrossRef]

17. Wang, J.Q.; Du, Y.; Wang, J. LSTM based long-term energy consumption prediction with periodicity. *Energy* **2020**, *197*, 117197. [CrossRef]
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed on 1 January 2022).
19. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
20. Khodayar, M.; Liu, G.; Wang, J.; Khodayar, M.E. Deep learning in power systems research: A review. *CSEE J. Power Energy Syst.* **2020**, *7*, 209–220.
21. Kaur, D.; Kumar, R.; Kumar, N.; Guizani, M. Smart grid energy management using rnn-lstm: A deep learning-based approach. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
22. Jung, Y.; Jung, J.; Kim, B.; Han, S. Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea. *J. Clean. Prod.* **2020**, *250*, 119476. [CrossRef]
23. Pallonetto, F.; Jin, C.; Mangina, E. Forecast electricity demand in commercial building with machine learning models to enable demand response programs. *Energy AI* **2022**, *7*, 100121. [CrossRef]
24. Ahmad, N.; Ghadi, Y.; Adnan, M.; Ali, M. Load forecasting techniques for power system: Research challenges and survey. *IEEE Access* **2022**, *10*, 71054–71090. [CrossRef]
25. Fan, S.; Mao, C.; Chen, L. Peak load forecasting using the self-organizing map. In Proceedings of the International Symposium on Neural Networks, Chongqing, China, 30 May–1 June 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 640–647.
26. Mishra, M.; Nayak, J.; Naik, B.; Abraham, A. Deep learning in electrical utility industry: A comprehensive review of a decade of research. *Eng. Appl. Artif. Intell.* **2020**, *96*, 104000. [CrossRef]
27. Ozcanli, A.K.; Yaprakdal, F.; Baysal, M. Deep learning methods and applications for electrical power systems: A comprehensive review. *Int. J. Energy Res.* **2020**, *44*, 7136–7157. [CrossRef]
28. Nti, I.K.; Teimeh, M.; Nyarko-Boateng, O.; Adekoya, A.F. Electricity load forecasting: A systematic review. *J. Electr. Syst. Inf. Technol.* **2020**, *7*, 13. [CrossRef]
29. Bedi, J.; Toshniwal, D. Deep learning framework to forecast electricity demand. *Appl. Energy* **2019**, *238*, 1312–1326. [CrossRef]
30. Bandara, K.; Bergmeir, C.; Smyl, S. Forecasting across time series databases using long short-term memory networks on groups of similar series. *arXiv* **2017**, *8*, 805–815. [CrossRef]
31. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]
32. Tian, Z.; Chen, H. A novel decomposition-ensemble prediction model for ultra-short-term wind speed. *Energy Convers. Manag.* **2021**, *248*, 114775. [CrossRef]
33. Takyi-Aninakwa, P.; Wang, S.; Zhang, H.; Yang, X.; Fernandez, C. An optimized long short-term memory-weighted fading extended Kalman filtering model with wide temperature adaptation for the state of charge estimation of lithium-ion batteries. *Appl. Energy* **2022**, *326*, 120043. [CrossRef]
34. Liu, B.; He, X.; Song, M.; Li, J.; Qu, G.; Lang, J.; Gu, R. A Method for Mining Granger Causality Relationship on Atmospheric Visibility. *ACM Trans. Knowl. Discov. Data (TKDD)* **2021**, *15*, 1–16. [CrossRef]
35. You, Z.; Congbo, L.; Lihong, L. Centrifugal blower Fault trend prediction method of centrifugal blower based on Informer under incomplete data. *Comput. -Integr. Manuf. Syst.* **2023**, *29*, 133–145. [CrossRef]
36. Jamali, B.; Rasekh, M.; Jamadi, F.; Gandomkar, R.; Makiabadi, F. Using PSO-GA algorithm for training artificial neural network to forecast solar space heating system parameters. *Appl. Therm. Eng.* **2019**, *147*, 647–660. [CrossRef]
37. Mishra, M.; Dash, P.B.; Nayak, J.; Naik, B.; Swain, S.K. Deep learning and wavelet transform integrated approach for short-term solar PV power prediction. *Measurement* **2020**, *166*, 108250. [CrossRef]
38. Ying, R.K.; Shou, Y.; Liu, C. Prediction Model of Dow Jones Index Based on LSTM-Adaboost. In Proceedings of the 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), Beijing, China, 14–16 May 2021; pp. 808–812.
39. Sibbo, W.; Zhigang, C.; Rui, H. ID 3 optimization algorithm based on correlation coefficient. *Comput. Eng. Sci.* **2016**, *38*.
40. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
41. Lu, L.; Zeng, B.; Yang, S.; Chen, M.; Yu, Y. Continuous Monitoring Analysis of Rice Quality in Southern China Based on Random Forest. *J. Food Qual.* **2022**, *2022*, 7730427. [CrossRef]
42. Huang, S.; Tan, E.; Jimin, R. Analog Circuit Fault Diagnosis based on Optimization Matrix Random Forest Algorithm. In Proceedings of the 2021 International Symposium on Computer Technology and Information Science (ISCTIS), Guilin, China, 4–6 June 2021; pp. 63–67.
43. Du, H.; Ni, Y.; Wang, Z. An Improved Algorithm Based on Fast Search and Find of Density Peak Clustering for High-Dimensional Data. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 9977884. [CrossRef]
44. Lv, B.; Wang, G.; Li, S.; Wu, Y.; Wang, G. A weight recognition method for movable objects in sealed cavity based on supervised learning. *Measurement* **2022**, *205*, 112149. [CrossRef]

45. Zhou, J.; Hu, L.; Jiang, Y.; Liu, L. A correlation analysis between SNPs and ROIs of Alzheimer's disease based on deep learning. *BioMed Res. Int.* **2021**, *2021*, 8890513. [[CrossRef](#)]
46. Huang, X.; Xiang, Y.; Li, K.C. Green, pervasive, and cloud computing'. In Proceedings of the 11th International Conference, GPC 2016, Xi'an, China, 6–8 May 2016.
47. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 11106–11115.
48. Kitaev, N.; Kaiser, L.; Levskaya, A. Reformer: The efficient transformer. *arXiv* **2020**, arXiv:2001.04451.
49. Graves, A.; Graves, A. Long short-term memory. *Supervised Seq. Label. Recurr. Neural Netw.* **2012**, *385*, 37–45.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.