

Article

Pre-Attention Mechanism and Convolutional Neural Network Based Multivariate Load Prediction for Demand Response

Zheyu He ^{1,*}, Rongheng Lin ¹, Budan Wu ¹, Xin Zhao ² and Hua Zou ¹

¹ State Key Laboratory of Networking and Switching Technology, School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China

² Economic & Research Institute, State Grid Shandong Electric Power Company, Jinan 250021, China

* Correspondence: rinchannow@bupt.edu.cn

Abstract: The construction of smart grids has greatly changed the power grid pattern and power supply structure. For the power system, reasonable power planning and demand response is necessary to ensure the stable operation of a society. Accurate load prediction is the basis for realizing demand response for the power system. This paper proposes a Pre-Attention-CNN-GRU model (PreAttCG) which combines a convolutional neural network (CNN) and gate recurrent unit (GRU) and applies the attention mechanism in front of the whole model. The PreAttCG model accepts historical load data and more than nine other factors (including temperature, wind speed, humidity, etc.) as input. The attention layer and CNN layer effectively extract the features and weights of each factor. Load forecasting is then performed by the prediction layer, which consists of a stacked GRU. The model is verified by industrial load data from a German dataset and a Chinese dataset from the real world. The results show that the PreAttCG model has better performance (3~5% improvement in MAPE) than both LSTM with only load input and LSTM with all factors. Additionally, the experiments also show that the attention mechanism can effectively extract the weights of relevant factors affecting the load data.

Keywords: load prediction; attention; convolutional neural network; gate recurrent unit



Citation: He, Z.; Lin, R.; Wu, B.; Zhao, X.; Zou, H. Pre-Attention Mechanism and Convolutional Neural Network Based Multivariate Load Prediction for Demand Response. *Energies* **2023**, *16*, 3446. <https://doi.org/10.3390/en16083446>

Academic Editor: Georgios Christoforidis

Received: 13 March 2023

Revised: 21 March 2023

Accepted: 11 April 2023

Published: 14 April 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

The construction of smart grids has greatly changed the power grid pattern and power supply structure. It is also a new challenge for the safe and stable operation of power systems. For power systems, reasonable power planning and demand response are necessary to ensure the stable operation of a society. Accurate load prediction is the basis for realizing demand response, economic operation, and scientific management of the power system. It is of great significance for optimizing unit combinations, power dispatching, and power market transactions.

For users on the demand side, their electricity consumption behavior has been dynamically changing, which also leads to the non-linear characteristics of user load data. At the same time, users' electricity behaviors are also very easily affected by a variety of external factors, such as climate change, holiday activities, electricity prices, etc. Electricity behavior and these external factors are uncertain and not linear. Traditional methods and simple neural network methods cannot achieve good effects. Therefore, on the premise that relevant data can be obtained, multivariate prediction models considering these external factor data and load data are becoming a very valuable research direction [1]. However, determining the correlation weight of each external factor is a big challenge.

Based on the above considerations, this paper proposes a novel multivariate load prediction model based on the pre-attention mechanism and convolution load network (Pre-Attention-CNN-GRU, or PreAttCG) for multiple data, including meteorological data, electricity price data, and load data. Putting the attention mechanism in front of the neural

network model provides practical significance in terms of data analysis to attention layer weights. We can directly use the attention layer weight values in the final training model to comprehensively analyze the weights of the time dimension effect and factor dimension effect with respect to load data. We conduct multivariate time series data using a two-dimensional matrix and take advantage of the convolutional neural network to extract features. We can then use a recurrent neural network to better capture the internal changes and further improve prediction accuracy. Experiments on Chinese and German datasets show that the PreAttCG model has more accuracy in load prediction tasks than baseline methods such as LSTM. Additionally, the PreAttCG model can effectively find out the weight of each external factor affecting the load.

2. Related Works

Many researchers have studied multivariate load forecasting.

Lang et al. [2] applied random weights and kernels into a neural network for short-term forecasting of load data with old load and temperature data. Unterluggauer et al. [3] proposed a multivariate multi-step model based on LSTM to predict short-term charging load data. Bracale et al. [4] and Xing et al. [5] used multivariate quantile regression for short-term load forecasting. Huang et al. [6] proposed a novel hybrid predictive model based on multivariate empirical mode decomposition (MEMD) and support vector regression (SVR) with parameters optimized by particle swarm optimization (PSO), which can capture precise electricity peak load. Xiao et al. [7] proposed the Multi-scale Skip Deep Long Short-Term Memory (MSD-LSTM) model for short-term load prediction with multivariate data. Khan et al. [8] applied SVR to realize multivariate time series forecasting model for load prediction.

Roy et al. [9] proposed a hybrid model based on Multivariate Adaptive Regression Splines (MARS) and an Extreme Learning Machine (ELM) to estimate heating load in buildings. Similarly, Cheng et al. [10] used Evolutionary Multivariate Adaptive Regression Splines (EMARS) to predict building energy. Fan et al. [11] used the features extracted by unsupervised deep learning as inputs for cooling load prediction. Zhang et al. [12] filtered original input data using an Unscented Kalman Filter (UKF) and then used an improved coupled generative adversarial stacked auto-encoder (ICoGASA) that consisted of three generative adversarial networks (GANs) to generate more similar errors in weather forecasting and the lifestyles of different residents for prediction analysis. Zhang et al. [13] proposed a novel asynchronous deep reinforcement learning model with an adaptive early forecasting method and reward incentive mechanism for short-term load forecasting. Hu et al. [14] proposed a multivariate regression load forecasting algorithm based on variable accuracy feedback. Gupta et al. [15] proposed a joint feature selection framework for multivariate prediction. Ouyang et al. [16] proposed a combined multivariate model through the use of different kernel functions in support vector regression models for wind power prediction.

The algorithms and models mentioned mainly used regression methods and simple structured neural networks. These methods can only accommodate data with low dimensions. They cannot take advantage of more useful factors that are strongly related to load data.

The convolutional neural network (CNN) [17] is mainly used in image processing to extract the features of pictures based on maintaining the spatial relations between the pixels. As time series data can be converted to 2-D curves, we can apply a CNN to them to extract the features efficiently. As a result, many researchers have introduced CNNs to their forecasting models. Bendaoud et al. [18] provide 2-D input to a CNN and conducted one-quarter-ahead and 24 h-ahead forecasting. Dong et al. [19] combined a CNN and K-means clustering to improve the scalability of short-term load forecasting. Deng et al. [20] used multi-scale convolutions (MS-CNN) to extract different level features for short-term load forecasting. Zhao et al. [21] built a new model based on a CNN to improve short-term heat load prediction of different buildings in residential districts. Jin et al. [22] proposed a CNN-GRU hybrid model with parameter-based transfer learning to optimize short-term

load prediction. Yu et al. [23] used a 2-D CNN to improve their bird swarm algorithm for torsional capacity evaluation of RC beams.

Alhussein et al. [24] and Rafi et al. [25] combined LSTM and a CNN for load forecasting and achieved better results than LSTM-only models. Similar to LSTM methods, Sajjad et al. [26] used a GRU instead of LSTM.

Li et al. [27], Khan et al. [28], Imani et al. [29], Tudose et al. [30], and Dong et al. [31] introduced a CNN to their models for short-term load forecasting and achieved better results in evaluation indexes.

The studies mentioned above introduced CNNs to extract load features and obtain ideal results. However, those studies did not consider multivariate factors and their structures were simple. Therefore, there is much space for improvement.

Attention is a mechanism that can help improve neural networks. It can calculate the weights of features efficiently, which can help the model to understand the data better. The mechanism is mainly used in the fields of computer vision (CV) [32] and natural language processing (NLP) [33].

The efficiency of extracting the best weight of each factor can also help to achieve better performance in load forecasting. Tang et al. [34] introduced attention to a Temporal Convolutional Network (TCN) for short-term load forecasting. Thus, to achieve better performance in load forecasting, we have proposed the Pre-Attention-CNN-GRU (PreAttCG) model.

3. Algorithm Model Design

With the continuous improvement of smart grid construction, collectable data are not only load curve data in the actual power system, but also rich regional location data, real-time electricity price data, etc. Through interaction with the meteorological system, some meteorological data can also be obtained. Users' electricity consumption behaviors are closely related to these various external factors. Fully excavating and analyzing the effect of these multiple factors on power consumption behaviors is helpful to predict electricity consumption behavior more accurately. It can help to reasonably plan power distribution, save energy, and support the sustainable development of power and other energy. When considering these various factors, users' load data form a typical multi-factor time series. The data have the same time dimension as the ordinary time series and have a multi-time data dimension with multiple factors affecting the load data in each time dimension. This paper introduces a comprehensive analysis of the effect weight of the time dimension and factor dimension on power load. We also use a convolutional neural network to extract two-dimensional multiplex time series data as input to the subsequent recurrent neural network layer. The model's structure is shown in Figure 1.

Figure 1 shows the designed model's steps for processing the original multivariate data from input to output, and the internal principles of each step are specifically described below.

3.1. Input Data of the Model

The main input is load data. Because users' electricity consumption is often related to meteorological data [1], we include meteorological data such as the temperature, rainfall, visibility, air pressure, and electricity price as external factor input data. These data are all time series data, so they have the practical significance of both time and external factors. In terms of external factors, there may be a variety of data with different factors according to the different actual data situation. The time series may be 96 points, 24 points, and so on according to the different actual acquisition frequency. Therefore, analyzing the effect weight of different aspects and different factors is conducive to a better understanding of user behavior.

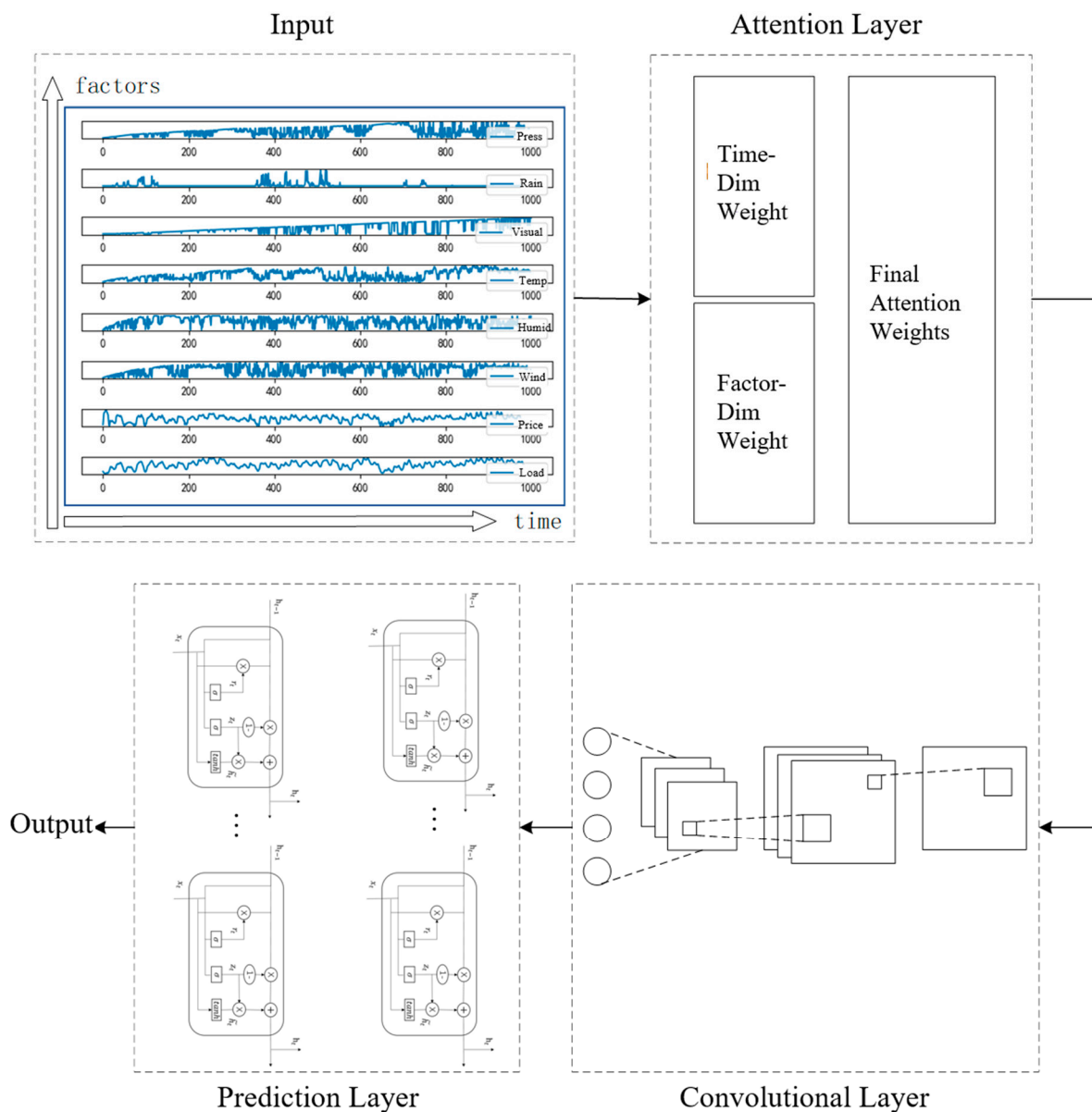


Figure 1. Multivariate load prediction model based on a pre-attention mechanism and convolutional neural network.

3.2. Attention Layer

The attention mechanism (Attention) is not a complete model, but should be a technology. It functions to focus on and fully learn from the more important parts of a dataset and can be applied to any relevant model of sequence data. Under the traditional encoder–decoder model architecture, the codec needs to be limited by a fixed-length vector in the internal structure. The emergence of the attention mechanism breaks this point. In fact, a model based on the attention mechanism can also be used as a real measure of similarity. The current input weight is proportional to the similarity of the target state, and the more similar the weight, the greater the result. Therefore, the introduction of the attention mechanism allows the model to selectively focus on the corresponding relevant information in the input when making the output. It is also widely used in many sequence prediction problems, which is why this paper uses Attention to analyze the effect weight of different external factors on electricity behavior.

The essence of the attention mechanism is to introduce a fully connected layer, but the activation function in the internal structure of the layer is set to SoftMax. Its output is a set of weights representing attention, which is then combined with the original input to obtain the “importance” of each original feature. In order to comprehensively consider the different weights of the time dimension and the external factor dimension, the specific weight calculation method of the attention layer designed in this section is shown in Figure 2.

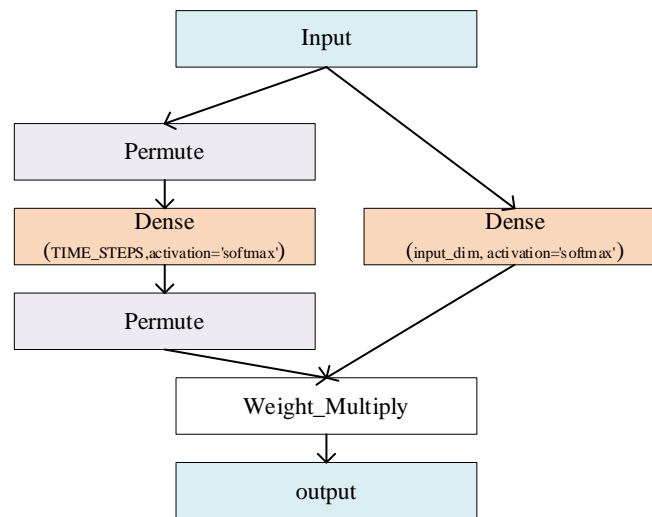


Figure 2. Attention mechanism weights are calculated in this section.

For the time dimension, the attention allocation matrix is as follows:

$$M_t = \begin{bmatrix} a_1 & a_1 & \dots & a_1 \\ a_2 & a_2 & \dots & a_2 \\ \dots & \dots & \dots & \dots \\ a_{timestep} & a_{timestep} & \dots & a_{timestep} \end{bmatrix}_{Timestep * input_dim} \quad (1)$$

For the time dimension, the attention allocation matrix is as follows:

$$M_f = \begin{bmatrix} b_1 & b_2 & \dots & b_{input_dim} \\ b_1 & a_2 & \dots & b_{input_dim} \\ \dots & \dots & \dots & \dots \\ b_1 & b_2 & \dots & b_{input_dim} \end{bmatrix}_{Timestep * input_dim} \quad (2)$$

The resulting final attention distribution matrix is therefore as follows:

$$M_{final} = \begin{bmatrix} a_1 * b_1 & a_1 * b_2 & \dots & a_1 * b_{input_dim} \\ a_2 * b_1 & a_2 * b_2 & \dots & a_2 * b_{input_dim} \\ \dots & \dots & \dots & \dots \\ a_{timestep} * b_1 & a_{timestep} * b_2 & \dots & a_{timestep} * b_{input_dim} \end{bmatrix}_{Timestep * input_dim} \quad (3)$$

The final attention allocation matrix is an element-by-element product of the time dimension and the factor dimension. Therefore, the value of each element in the matrix is the final weight obtained by considering the time dimension and the space dimension comprehensively. After model training is completed, the weight of each part is output, and the obtained value reflects the weight of the corresponding time or factor dimension. This reflects the degree of effect of the corresponding factors on electricity consumption behavior, which has a certain practical significance.

3.3. Convolutional Layer

Convolutional neural networks (CNNs) show excellent performance in target monitoring and image classification and can acquire local features from the higher level of inputs and combine them into more complex features at the lower level. A CNN is usually used for the processing of visual data, namely data formatted as a two-dimensional matrix. The multivariate load data to be analyzed in this paper are exactly this type of matrix data with the dual features of multivariate factors and time series.

The convolutional network used in the method of this paper mainly consists of multiple stacked convolution and pooling operations. Where the number of convolution kernels can determine the degree of feature extraction. The size of the convolution kernel can be adjusted according to the fixed length of the input sequence data. The pooling layer is used to filter some unimportant features.

In deep learning-related model frameworks, the stacking of multiple convolutional layers enables the initial layers to learn low-level features in the application inputs. However, the output feature map of the convolutional layer has a limitation: it will track the specific location of the input feature more accurately, that is, even a very small movement of the input feature will cause the generation of different feature maps. Therefore, a pooling layer is added to the middle of the continuous convolution layer to reduce the limitation of the invariance of the generated feature map, while the activation function is used to enhance the ability of the model to learn complex structures. The activation function used in this section is ReLU, or the Rectified Linear Unit function, as shown in Equation (4).

$$f_{relu}(x) = \text{Max}(0, x) \quad (4)$$

The ReLU function retains values greater than 0 (which are also relatively good features in the data), discarding values with features less than 0; this activation function can effectively address gradient-related problems in model training and make the network easier to train.

3.4. Prediction Layer

The prediction layer of the model designs a three-layer stacked GRU network based on GRUs. The network structure not only solves the gradient problem of RNNs itself, but also improves the training efficiency of the model due to its simple unit structure.

4. Experiment Design and Comparative Analysis

This section introduces the superiority and practical significance of the proposed method in prediction accuracy mainly through the experiments conducted on real datasets. Accuracy will be reflected by methods such as established indexes and control variables, and the practical significance of the methods will be analyzed separately for different datasets.

4.1. Dataset

This paper uses different industrial electricity consumption datasets from China and Germany to validate the proposed model. In the German datasets, the load data and real-time electricity price data come from the actual data of some regions, which have been published by an agency in Germany since October 2018. These data represent the actual electricity consumption situation of a region and the historical real-time electricity price situation of a region. The meteorological data used came from the Climate Data Center of the German Meteorological Bureau, which provided the meteorological conditions of the electricity price and load dataset in the provided dataset. Combining the two sets of data provides the multivariate load dataset used for the experiments in this section.

The Chinese dataset used was derived from ledger data provided by the relevant departments, including load data and meteorological data from some regions from January 2020 to May 2021. The details of the dataset are shown in Tables 1 and 2.

Table 1. Definition of the German dataset.

Symbol	Meaning	Instance	File Format
TI	Time	2018-10-02 01:00:00	–
LD	Load	3627.00	csv
PC	Price	56.65	csv
P0	Pressure	9.383	txt
RR	Rainfall	0.3	txt
VV	Visibility	4.893	txt
T	Temperature	17.8	txt
U	Humidity	72	txt
FF	Wind speed	6.7	txt

Table 2. Definition of the China dataset.

Symbol	Meaning	Instance	File Format
Station_Id	Station ID	54,863	txt
Province	Province	Henan	txt
City	City	Zhoukou	txt
Region	Region	Huaiyang	txt
Year	Year	2020	txt
Month	Month	3	txt
Day	Day	1	txt
Hour	Hour	0	txt
Min	Minute	0	txt
TEM	Temperature	19.8	txt
RHU	Humidity	72	txt
Pre_1H	Rainfall	0.1	txt
WIND_S_Avg_2 min	Wind speed over two minutes	1.2	txt
WIND_S_Avg_10 min	Wind speed over ten minutes	1.6	txt

For the dataset shown in Table 1, after certain processing of the time identification and region identification data, the multiple input variables include temperature, humidity, precipitation, wind speed within 2 min, and wind speed within 10 min, as well as load data.

4.2. Index Definition

The core of this paper is more accurate load prediction, so we defined RMSE, MAPE, and R2_Score to evaluate the prediction effect.

4.3. Comparison Methods

Based on the PreAttCG method presented in this paper, the following section briefly introduces the method selected from the technical selection aspects of deep learning prediction methods and the consideration of multivariate factors.

(A) LSTM network (LoadLSTM) with load data input. This method focuses on incorporating multiple external factors into the model input. Therefore, a set of comparison tests will be set to only use the load data themselves without considering the effect of external factors, using the results obtained as a benchmark.

(B) LSTM network (FullLSTM) with full amount of data input. The input of this method will use all external factor data and load data of this section's dataset, which can be used to prove the improvement in model prediction ability through the consideration of external effect factors.

4.4. Experiment Results and Comparative Analysis

After clarifying the result index and the experimental comparison method, experimental verification will be conducted on the selected dataset. The specific experimental results are as follows.

4.4.1. German Dataset

A comparison of prediction accuracy with the selected contrast method was first used on the German dataset, and the performance of each method is shown in Table 3.

Table 3. Comparison of prediction accuracy indexes (German dataset).

Index	LoadLSTM	FullLSTM	PreAttCG
MAPE (%)	10.217	6.103	4.983
RMSE	439.071	422.919	417.083
R2_SCORE	0.931	0.939	0.953

As can be seen from each index, the prediction accuracy of the method proposed in this section is better than the benchmark method. Compared with only inputting load data, inputting both meteorological and price data can obtain a better prediction effect. To show the differences between methods more intuitively, the load prediction results obtained by PreAttCG and the LSTM model only with load data are shown in Figure 3.

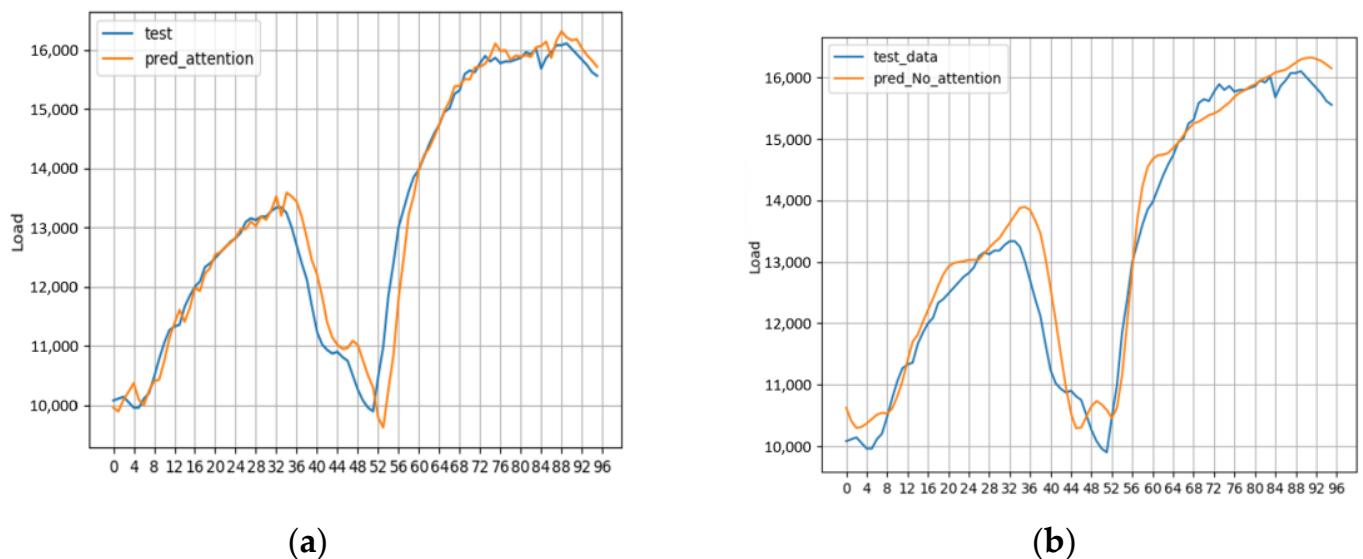


Figure 3. Comparison of the prediction results for the German dataset ((a): PreAttCG; (b): LSTM).

As can be seen from Figure 3, although the benchmark LSTM method can predict the general electricity consumption trend of users, the details of the users' electricity consumption behavior and the accuracy of the model are inferior to the proposed model because of the impact of external factors on electricity consumption behavior.

Due to the attention design of the method in our model, the attention weight output of all dimensions after model training can be further analyzed to analyze the effect of various factors on actual electricity consumption behavior. The different weights for the time and factor dimensions are shown in Figure 4.

As Figure 4 shows in the time dimension, the impact weight of 7:00 and 23:00 on future power is higher. In the dimension of external factors, user electricity behavior is affected by real-time price weight, as well as temperature, rainfall, and other meteorological data, though by not as much as price. We can speculate that the user may belong to industrial and commercial users, and real-time electricity price can be controlled by electricity suppliers; therefore, to further regulate user electricity behavior, the load curve for peak filling can be adjusted. This is one of the research topics of this paper.

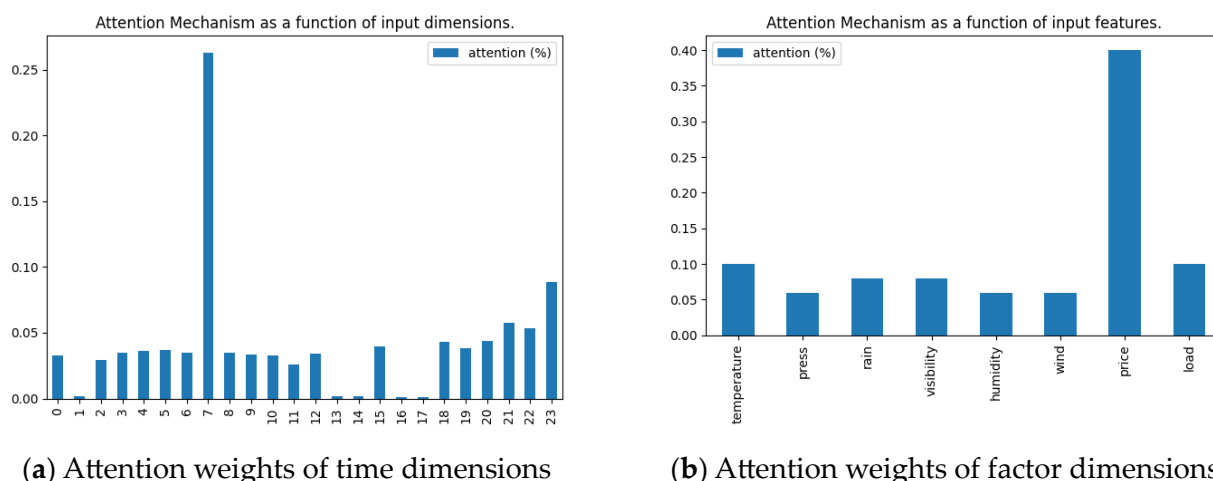


Figure 4. Time dimension and factor dimension effect on weights in the German dataset.

In order to further verify the importance of each factor on the prediction results, the input meteorological data and electricity price data are deleted from the model input one by one according to the idea of the control variable. We compared the prediction results with the results of the full data input and observed the change in each index. Taking MAPE as an example, the specific results are shown in Table 4, where the change in MAPE is the comparison of the value of this section with the first line of full input.

Table 4. Comparison of prediction results in terms of changes in MAPE (%) for different input data in the German dataset.

Input	LoadLSTM	FullLSTM	MAPE Change
Full Data	14.103	10.983	–
No Electricity Price	15.015	12.042	+1.059
No Temperature	14.782	11.392	+0.409
No Pressure	14.133	11.275	+0.292
No Rainfall	14.201	11.417	+0.434
No Visibility	14.108	11.022	+0.039
No Humidity	14.072	11.071	+0.088
No Wind Speed	14.091	10.994	+0.011
Load Data Only	15.547	12.639	+1.656

It can be seen from Table 4 that any external factor data will have a certain impact on the prediction of electricity consumption behavior. It once again confirms that when predicting users’ electricity consumption behaviors, considering more external factor data comprehensively can help to improve the prediction effect. As can be seen from the change in MAPE in the table, without the electricity price data input, the corresponding index has the largest change, and with no wind speed input, the index data change is very small. The result is consistent with the weight of the output model proposed in this paper. It illustrates the effectiveness of the attention mechanism designed in the model in this section. In addition, under our PreAttCG model, only the univariate of input load data is predicted, with the highest MAPE value and the worst prediction effect, which also shows that the inclusion of analysis of some data related to user electricity consumption behavior is beneficial for improving prediction accuracy.

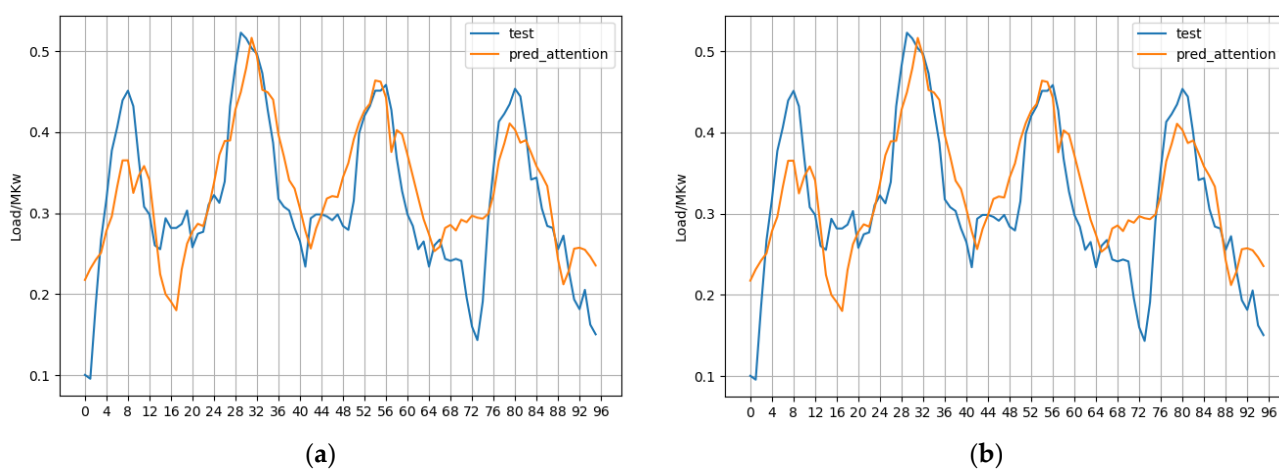
4.4.2. Chinese Dataset

As with previous dataset experiments, experiments were continued on the multivariate load dataset used in China, which compared the method proposed in this section with the selected comparison method, and the specific performance of each method is shown in Table 5.

Table 5. Comparison of prediction accuracy indexes (Chinese dataset).

Index	LoadLSTM	FullLSTM	PreAttCG
MAPE (%)	12.172	10.078	9.421
RMSE	39.262	38.402	35.657
R2_SCORE	88.097	89.171	90.391

Similarly, it can be seen from the various indexes that the PreAttCG model has better prediction accuracy than the benchmark method. Compared with only inputting load data, considering meteorological data in model input can obtain better prediction results. To show the method differences across the methods more intuitively, the load prediction results obtained by the methods in this section and by the LSTM model that only considers load data are shown in Figure 5.

**Figure 5.** Comparison of prediction results for the Chinese dataset ((a): PreAttCG; (b): LSTM).

As shown in Figure 5, the experimental results were obtained using one of the stations (Station_ID: 53982) as an example. The right panel shows the predictions of the LSTM model, and the left panel shows the predictions of the proposed model. It can also be seen that the proposed model better describes the details of user consumption behavior, and the value at the peak is closer to the true value.

Figure 6 shows the effect of the past 24 h and external factors for the current area. It can be seen in the time dimension that the effect weights of 2:00, 13:00, and 20:00 are higher; in the feature dimension, the effect of the 0th and 1st feature weights are higher. It can be said that for current user behavior, temperature and humidity factors have more effect on power load.

As with the previous dataset, the effect of these external factors on users' electricity behavior is further verified by using control variables. They can help to improve the accuracy of the model's prediction compared to deleting the meteorological data from the model input. We observed changes in the various indexes, including MAPE, with MAPE changes compared to the value of the first line of full input in Table 6.

As can be seen from Table 6, the meteorological factor data will have a certain impact on the prediction of electricity consumption behavior. It once again confirms that when predicting users' electricity consumption behavior, considering more external factor data related to electricity consumption behavior comprehensively can help to improve the prediction effect.

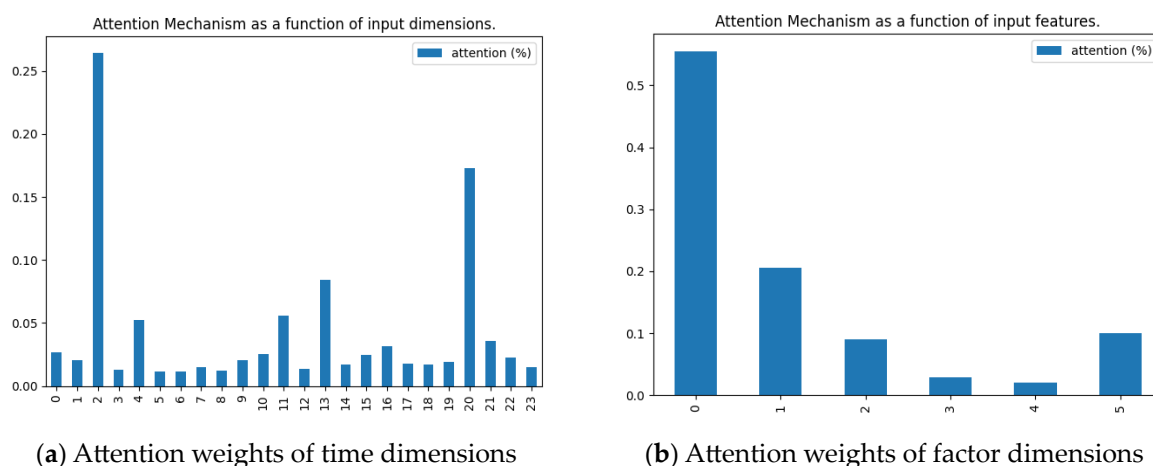


Figure 6. Time dimension and factor dimension effects on weights in the Chinese dataset.

Table 6. Comparison of prediction results in terms of changes in MAPE (%) for different input data in the Chinese dataset.

Input	LoadLSTM	FullLSTM	MAPE Change
Full Data	10.078	9.421	--
No Temperature	10.725	9.682	+0.261
No Humidity	10.515	9.584	+0.163
No Rainfall	10.139	9.497	+0.076
No Wind Speed in two minutes	10.082	9.433	+0.012
No Wind Speed in ten minutes	10.091	9.421	0.00
Load Data Only	11.023	10.048	+0.627

As can be seen from the change in MAPE in the table, the largest corresponding index change occurs in the absence of temperature data in input; in the absence of wind speed data in input, the index change is very small. The result is consistent with the weight of the model output results proposed in Figure 6. The result for the Chinese dataset also illustrates the effectiveness of the attention mechanism designed in the model in our paper. As with the German dataset experiment, prediction with univariate input load data has the worst effect under the algorithm proposed in this paper, which also shows that external factors related to user electricity behavior are beneficial for improving prediction accuracy. When data conditions permit, inputting more external factor data into the model can improve the prediction effect to a certain extent.

5. Conclusions

This paper proposes a multivariate prediction method based on a pre-attention mechanism and convolutional neural networks that considers multivariate load data, including meteorological data, electricity price data, and load data. By improving the method of calculating attention weight, we use an attention layer comprising weight values to comprehensively analyze the effect weights of power load under the time dimension and external factor dimension. The proposed model helps to intuitively understand users' electricity behaviors. This paves the way for subsequent studies on load regulation through the regulation of some human-controllable factors. The advantages of the proposed method in terms of industrial load prediction accuracy and power consumption curve characterization are proven by the experiments involving German and Chinese datasets exploring the effect of the time dimension and factor dimension on load data. We transformed multivariate time series data into a two-dimensional matrix and took advantage of a convolutional neural network to extract features. We could then use a recurrent neural network to better capture internal changes and further improve prediction accuracy.

For future work, we will focus more on how to improve the accuracy of electricity load forecasting for a certain type of user (residential, commercial, etc.) or a certain type of

specific industry. We will analyze the characteristics of each user type more accurately and establish more accurate and efficient models. Besides accuracy, running speed will also be within the scope of consideration.

Author Contributions: Conceptualization, Z.H. and R.L.; methodology, R.L.; software, Z.H.; validation, Z.H.; formal analysis, R.L.; investigation, B.W., X.Z. and H.Z.; resources, B.W., X.Z. and H.Z.; data curation, X.Z.; writing—original draft preparation, Z.H.; writing—review and editing, Z.H. and R.L.; visualization, Z.H.; supervision, R.L.; project administration, R.L.; funding acquisition, R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of State Grid Corporation of China, grant number 5108-202218280A-2-380-XG (The project's ERP system number is 520625220031), whose project title is "Research and Application of Hierarchical Precise Adjustment Technology for Industrial User Demand Response with Safety Constraints".

Data Availability Statement: The data is unavailable due to privacy.

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

References

- Habbak, H.; Mahmoud, M.; Metwally, K.; Fouda, M.M.; Ibrahim, M.I. Load Forecasting Techniques and Their Applications in Smart Grids. *Energies* **2023**, *16*, 1480. [\[CrossRef\]](#)
- Lang, K.; Zhang, M.; Yuan, Y.; Yue, X. Short-Term Load Forecasting Based on Multivariate Time Series Prediction and Weighted Neural Network with Random Weights and Kernels. *Clust. Comput.* **2019**, *22*, 12589–12597. [\[CrossRef\]](#)
- Unterluggauer, T.; Rauma, K.; Järventausta, P.; Rehtanz, C. Short-Term Load Forecasting at Electric Vehicle Charging Sites Using a Multivariate Multi-Step Long Short-Term Memory: A Case Study from Finland. *IET Electr. Syst. Transp.* **2021**, *11*, 405–419. [\[CrossRef\]](#)
- Bracale, A.; Caramia, P.; De Falco, P.; Hong, T. Multivariate Quantile Regression for Short-Term Probabilistic Load Forecasting. *IEEE Trans. Power Syst.* **2019**, *35*, 628–638. [\[CrossRef\]](#)
- Xing, Y.; Zhang, S.; Wen, P.; Shao, L.; Rouyendegh, B.D. Load Prediction in Short-Term Implementing the Multivariate Quantile Regression. *Energy* **2020**, *196*, 117035. [\[CrossRef\]](#)
- Huang, Y.; Hasan, N.; Deng, C.; Bao, Y. Multivariate Empirical Mode Decomposition Based Hybrid Model for Day-Ahead Peak Load Forecasting. *Energy* **2022**, *239*, 122245. [\[CrossRef\]](#)
- Xiao, Y.; Zheng, K.; Zheng, Z.; Qian, B.; Li, S.; Ma, Q. Multi-Scale Skip Deep Long Short-Term Memory Network for Short-Term Multivariate Load Forecasting. *J. Comput. Appl.* **2021**, *41*, 231.
- Khan, M.; Javaid, N.; Iqbal, M.N.; Bilal, M.; Zaidi, S.F.A.; Raza, R.A. Load Prediction Based on Multivariate Time Series Forecasting for Energy Consumption and Behavioral Analytics. In Proceedings of the Conference on Complex, Intelligent, and Software Intensive Systems, Matsue, Japan, 4–6 July 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 305–316.
- Roy, S.S.; Roy, R.; Balas, V.E. Estimating Heating Load in Buildings Using Multivariate Adaptive Regression Splines, Extreme Learning Machine, a Hybrid Model of MARS and ELM. *Renew. Sustain. Energy Rev.* **2018**, *82*, 4256–4268.
- Cheng, M.-Y.; Cao, M.-T. Accurately Predicting Building Energy Performance Using Evolutionary Multivariate Adaptive Regression Splines. *Appl. Soft Comput.* **2014**, *22*, 178–188. [\[CrossRef\]](#)
- Fan, C.; Xiao, F.; Zhao, Y. A Short-Term Building Cooling Load Prediction Method Using Deep Learning Algorithms. *Appl. Energy* **2017**, *195*, 222–233. [\[CrossRef\]](#)
- Zhang, G.; Guo, J. A Novel Ensemble Method for Residential Electricity Demand Forecasting Based on a Novel Sample Simulation Strategy. *Energy* **2020**, *207*, 118265. [\[CrossRef\]](#)
- Zhang, W.; Chen, Q.; Yan, J.; Zhang, S.; Xu, J. A Novel Asynchronous Deep Reinforcement Learning Model with Adaptive Early Forecasting Method and Reward Incentive Mechanism for Short-Term Load Forecasting. *Energy* **2021**, *236*, 121492. [\[CrossRef\]](#)
- Hu, Y.; Xia, X.; Fang, J.; Ding, Y.; Jiang, W.; Zhang, N. A Multivariate Regression Load Forecasting Algorithm Based on Variable Accuracy Feedback. *Energy Procedia* **2018**, *152*, 1152–1157. [\[CrossRef\]](#)
- Gupta, S.; Dileep, A.D.; Gonsalves, T.A. A Joint Feature Selection Framework for Multivariate Resource Usage Prediction in Cloud Servers Using Stability and Prediction Performance. *J. Supercomput.* **2018**, *74*, 6033–6068. [\[CrossRef\]](#)
- Ouyang, T.; Zha, X.; Qin, L. A Combined Multivariate Model for Wind Power Prediction. *Energy Convers. Manag.* **2017**, *144*, 361–373. [\[CrossRef\]](#)
- O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. *arXiv* **2015**, arXiv:1511.08458.
- Bendaoud, N.M.M.; Farah, N. Using Deep Learning for Short-Term Load Forecasting. *Neural Comput. Appl.* **2020**, *32*, 15029–15041. [\[CrossRef\]](#)
- Dong, X.; Qian, L.; Huang, L. Short-Term Load Forecasting in Smart Grid: A Combined CNN and K-Means Clustering Approach. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Republic of Korea, 13–16 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 119–125.

20. Deng, Z.; Wang, B.; Xu, Y.; Xu, T.; Liu, C.; Zhu, Z. Multi-Scale Convolutional Neural Network with Time-Cognition for Multi-Step Short-Term Load Forecasting. *IEEE Access* **2019**, *7*, 88058–88071. [[CrossRef](#)]
21. Zhao, A.; Mi, L.; Xue, X.; Xi, J.; Jiao, Y. Heating Load Prediction of Residential District Using Hybrid Model Based on CNN. *Energy Build.* **2022**, *266*, 112122. [[CrossRef](#)]
22. Jin, Y.; Acquah, M.A.; Seo, M.; Han, S. Short-Term Electric Load Prediction Using Transfer Learning with Interval Estimate Adjustment. *Energy Build.* **2022**, *258*, 111846. [[CrossRef](#)]
23. Yu, Y.; Liang, S.; Samali, B.; Nguyen, T.N.; Zhai, C.; Li, J.; Xie, X. Torsional Capacity Evaluation of RC Beams Using an Improved Bird Swarm Algorithm Optimised 2D Convolutional Neural Network. *Eng. Struct.* **2022**, *273*, 115066. [[CrossRef](#)]
24. Alhussein, M.; Aurangzeb, K.; Haider, S.I. Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting. *IEEE Access* **2020**, *8*, 180544–180557. [[CrossRef](#)]
25. Rafi, S.H.; Deeba, S.R.; Hossain, E. A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. *IEEE Access* **2021**, *9*, 32436–32448. [[CrossRef](#)]
26. Sajjad, M.; Khan, Z.A.; Ullah, A.; Hussain, T.; Ullah, W.; Lee, M.Y.; Baik, S.W. A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. *IEEE Access* **2020**, *8*, 143759–143768. [[CrossRef](#)]
27. Li, L.; Ota, K.; Dong, M. Everything Is Image: CNN-Based Short-Term Electrical Load Forecasting for Smart Grid. In Proceedings of the 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC), Exeter, UK, 21–23 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 344–351.
28. Khan, S.; Javaid, N.; Chand, A.; Khan, A.B.M.; Rashid, F.; Afridi, I.U. Electricity Load Forecasting for Each Day of Week Using Deep CNN. In Proceedings of the Workshops of the International Conference on Advanced Information Networking and Applications, Matsue, Japan, 27–29 March 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1107–1119.
29. Imani, M. Electrical Load-Temperature CNN for Residential Load Forecasting. *Energy* **2021**, *227*, 120480. [[CrossRef](#)]
30. Tudose, A.M.; Sidea, D.O.; Picioroaga, I.I.; Boicea, V.A.; Bulac, C. A CNN Based Model for Short-Term Load Forecasting: A Real Case Study on the Romanian Power System. In Proceedings of the 2020 55th International Universities Power Engineering Conference (UPEC), Torino, Italy, 1–4 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
31. Dong, X.; Qian, L.; Huang, L. A CNN Based Bagging Learning Approach to Short-Term Load Forecasting in Smart Grid. In Proceedings of the 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), San Francisco, CA, USA, 4–8 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
32. Mnih, V.; Heess, N.; Graves, A.; Kavukcuoglu, K. Recurrent Models of Visual Attention 2014. *arXiv* **2014**, arXiv:1406.6247.
33. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
34. Tang, X.; Chen, H.; Xiang, W.; Yang, J.; Zou, M. Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network. *Electr. Power Syst. Res.* **2022**, *205*, 107761. [[CrossRef](#)]

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.