


Article

Short-Term Power Load Forecasting: An Integrated Approach Utilizing Variational Mode Decomposition and TCN–BiGRU

Zhuoqun Zou ¹, Jing Wang ^{1,2} , Ning E ¹, Can Zhang ¹, Zhaocai Wang ¹ and Enyu Jiang ^{3,*}

¹ College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; 2052231@st.shou.edu.cn (Z.Z.); wangjing@shou.edu.cn (J.W.); 15333768199@163.com (N.E.); czhang0515@163.com (C.Z.); zcwang@shou.edu.cn (Z.W.)

² Key Laboratory of Fisheries Information, Ministry of Agriculture, Shanghai 201306, China

³ Electric Power Engineering, Shanghai University of Electric Power, Shanghai 200090, China

* Correspondence: enyu_1981@163.com

Abstract: Accurate short-term power load forecasting is crucial to maintaining a balance between energy supply and demand, thus minimizing operational costs. However, the intrinsic uncertainty and non-linearity of load data substantially impact the accuracy of forecasting results. To mitigate the influence of these uncertainties and non-linearity in electric load data on the forecasting results, we propose a hybrid network that integrates variational mode decomposition with a temporal convolutional network (TCN) and a bidirectional gated recurrent unit (BiGRU). This integrated approach aims to enhance the accuracy of short-term power load forecasting. The method was validated on load datasets from Singapore and Australia. The MAPE of this paper's model on the two datasets reached 0.42% and 1.79%, far less than other models, and the R^2 reached 98.27% and 97.98, higher than other models. The experimental results show that the proposed network exhibits a better performance compared to other methods, and could improve the accuracy of short-term electricity load forecasting.

Keywords: short-term load forecasting; power systems; variational mode decomposition; TCN; BiGRU



Citation: Zou, Z.; Wang, J.; E, N.; Zhang, C.; Wang, Z.; Jiang, E. Short-Term Power Load Forecasting: An Integrated Approach Utilizing Variational Mode Decomposition and TCN–BiGRU. *Energies* **2023**, *16*, 6625. <https://doi.org/10.3390/en16186625>

Academic Editors: Sohrab Zendejboudi and Yankai Cao

Received: 10 July 2023

Revised: 4 September 2023

Accepted: 13 September 2023

Published: 14 September 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

Electricity resources are integral to the functioning of modern society, facilitating the smooth operation of key sectors such as industry and commerce. As global economic development accelerates and the population increases, the demand for electric power resources correspondingly intensifies [1–3]. Considering the difficulties inherent in electricity storage and the lagging response of electricity suppliers [4], power companies often increase generation capacity by an excess of 20% to accommodate a potential peak electricity consumption of 5% [5]. This overcapacity can lead to significant economic waste. For instance, in a medium-sized Chinese city with an annual electricity consumption of 29 billion kilowatts, a 1% reduction in forecast error can result in savings of CNY 145 million [6]. Therefore, precise power load forecasting is critical to maintaining a balance between power supply and demand, ensuring grid stability, and promoting carbon savings and emission reductions. It is the basis for guaranteeing the safe operation of electrical power. At the same time, it can formulate a reasonable plan for unit maintenance, scientifically manage and control the cost of power supply, and maximize economic and social benefits [7].

However, the characteristics of uncertainty [8] and non-linearity [9] inherent in short-term load data complicate the task of accurate prediction. The increasing incorporation of renewable energy sources necessitates the consideration of various external factors, such as weather, holidays, and electricity prices, in addition to the intrinsic time-series characteristics of load forecasting. The uncertainty associated with these factors exacerbates

the challenge of electric load forecasting. The current methodologies employed for short-term electric load forecasting can be classified into three main categories: classical statistical learning methods [10,11], traditional machine learning methods [12,13], and contemporary deep learning methods [14–18].

Classical statistical models include autoregression, sliding average, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) models, etc. In 2022, Sun et al. proposed the threshold ARMA model considering the influence of temperature, and the experimental results on a dataset of a prefecture-level city in southwest Zhejiang Province, China, demonstrated that integrating temperature factors can improve prediction accuracy [10]. The results showed that the mean absolute percentage error (MAPE) was 4.167%. In 2023, Wang et al. proposed the hybrid ARIMA and convolutional neural network (CNN) model employing wavelet transform, and conducted the experiments tested on a dataset from Tai'an City in China, which demonstrated that prior signal decomposition of load data and the use of a hybrid model could improve prediction accuracy [11]. While classical statistical models offer simplicity, easy comprehension, and rapid computation, they struggle to accommodate the influence of nonlinear factors on load data. Additionally, they demonstrate limited robustness and a weak capacity to consider complex factors. Given the rapid development of today's electricity market and the expanding utilization of various renewable energy types [19], the factors affecting load data have grown increasingly complex, posing challenges to statistical models in terms of accurately predicting load data.

As artificial intelligence evolves, machine learning and deep learning are increasingly being applied to load data forecasting. Common traditional machine learning models include support vector machines (SVMs), extreme gradient boosting (XGBoost), random forests (RF), etc. In 2022, Su et al. proposed the cuckoo search (CS)-SVM model considering demand price elasticity and conducted tests on the PJM power market datasets in the United States. The experimental results showed that MAPE achieved 13.43%, and they verified that integrating the price factor can improve the prediction performance [12]. In the same year, Dudek employed the RF model and improved the prediction accuracy by changing the methods of three input patterns and seven training patterns [13]. Although this method achieved an average MAPE of 1.53% with four datasets, it required up to 21 attempts to choose the suitable input pattern for the model training. Traditional machine learning models can process load data relatively quickly. However, in real-world load forecasting, it is difficult for traditional machine learning to extract suitable features deeply in non-linear time series data [20].

Recently, deep-learning-based methods have begun to gain traction in short-term load forecasting. In 2022, Li et al. established the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)-sample entropy (SE)-LSTM model for forecasting ultra-short term (less than/within XX minutes or seconds) power load [14]. They first introduced SE and used CEEMDAN to reduce the complexity of information, then utilized LSTM to predict each component. Experimental results on electric load data in Changsha, China, showed that decomposing the electric load data first was more accurate than using the LSTM model alone. Meanwhile, decomposition is used not only in the field of electrical load forecasting, but also in other fields where it reduces the instability of the raw data and enhances predictability. For example, Wang et al. used variational mode decomposition (VMD) to reduce the instability of raw data in water quality prediction [21]. In 2023, Wang et al. proposed a long short-term memory (LSTM) informer model based on ensemble learning [15]. This hybrid model used LSTM to capture the short-term time correlation of power load and the informer model to solve the long-term dependence problem of power load forecasting. The method was validated on a dataset from the city of Tetouan in the north of Morocco. The results reached a mean square error (MSE) of 0.2085% and a mean absolute error (MAE) of 0.3963%, which shows that capturing the long and short-term dependencies separately using the combined model could improve the accuracy of the forecast. Despite the ability of LSTM to handle non-linear time series, the

subsequent gated recursive unit (GRU) introduced a simpler gating structure [22], which enhanced the computational performance of the overall structure and improved the speed and accuracy of the iterations. In 2023, Abumohsen et al. compared the capabilities of recurrent neural networks (RNNs), LSTM, and GRU to forecast electricity load data in the Tubas area, Palestine [16]. The GRU performed best, achieving a 90.228% of goodness of fit (R^2) and a root mean square error (RMSE) of 0.04647. However, a single model is never sufficient to achieve the optimal prediction results. In 2020, Sajjad et al. proposed a novel CNN-GRU-based hybrid approach for short-term residential load forecasting [17]. This approach first learns the spatial features using CNN, then feeds them into the GRU model, which enhances the sequence learning. The proposed model was tested on the public appliances energy prediction (AEP) and individual household electric power consumption (IHEPC) datasets. The results showed that the CNN-GRU model performed better than base models such as XGBoost and RNN. Although CNN can enhance the learning capability of GRU, the temporal convolutional network (TCN) proposed in 2018 has a more flexible sense field to adapt different sequence requirements [23]. In 2023, Hong et al. proposed the CEEMDAN-TCN-GRU-Attention model [18]. They utilized CEEMDAN to reduce the nonlinearity and complexity of the sequences, and the TCN and Attention modules to enhance the ability of GRU to capture feature information. The validation was carried out on power load data in Quanzhou City, Fujian Province, China, and the results show that the CEEMDAN-TCN-GRU-Attention method is well-structured and has better accuracy than a general combination model such as GRU-TCN.

In summary, decomposing the electric load data first can reduce the nonlinearity and complexity of the data, which is conducive to improving the prediction performance. Meanwhile, the use of combinatorial models can further improve the accuracy of the electric load data prediction. From the review of relevant works, it has been determined that price, climate, and other factors impact the prediction results, which is more applicable to the current complex power consumption environment.

Building upon previous research, as shown in Table A1, this study proposes a method that incorporates factors such as temperature, electricity prices, and holiday effects to address the uncertainty inherent in load data. Further, VMD is employed to decompose the original load data, reducing model training and computation complexity, enhancing model stability and accuracy, and addressing the pronounced non-linearity of load data. In this study, we formulate a hybrid regression model integrating a bidirectional GRU (BiGRU) and a TCN. BiGRU, by merging forward and backward GRUs, allows for a more effective capture of time-series information and facilitates deeper feature extraction and analysis. The coupled TCN can discern spatiotemporal relationships in the sequence data, manage long-term dependencies, and improve the model's execution speed. In summary, this study introduces a method for short-term electricity load forecasting, integrating VMD with a TCN-BiGRU hybrid model. This proposed method was validated through experiments using electricity market data from Singapore and Australia, and its performance was compared with various standalone and integrated models. The experimental results substantiate the feasibility and superiority of the proposed model for addressing short-term electricity load forecasting challenges.

After a broad review of previous work on electrical load forecasting, we propose a combining VMD-TCN-BiGRU regression model for the uncertainty and non-linearity of electric load data. The major contributions of the paper are as follows.

- First, we use VMD to decompose the original signal in order to obtain several simpler signal components. This helps to reduce the complexity of the electrical load data and solve the nonlinearities within the load data.
- We highlight the necessity of considering economic, social, and climatic multidimensional characteristics in addition to considering the data's own characteristics.
- Subsequently, we trained the combined neural network GRU-TCN by using the fused data to obtain the long- and short-term dependencies of the data, as well as the prediction results of the testing datasets.

- We tested the new model on two open datasets from Singapore and Australia in comparison with a variety of recent and significant models applied in this field. We also demonstrate the rationality of the components in the combined model through ablation experiments.

The rest of the paper is organized as follows. Section 2 describes the materials and methods of the paper. Then, Section 3 describes the validation experiments conducted in this paper and the related discussion of the results. Finally, the paper concludes with a perspective on future work and open research challenges in Section 4.

2. Materials and Methods

The short-term load forecasting method proposed herein integrates VMD and TCN-BiGRU. Initially, VMD decomposes the load data into several signal components. The decomposed data are then combined with temperature, electricity prices, and calendar features to generate new data. These new data are segmented into datasets and normalized independently. Subsequently, these normalized data are employed to train the integrated model after a data window panning operation. Ultimately, the model is tested on a validation set, and the real data are obtained after inverse normalization of the prediction results. This method capitalizes on the inherent time-series characteristics of load data while considering external influences such as temperature, electricity prices, and calendar effects. This dual approach enhances the model's accuracy and stability and diminishes the impacts of uncertainty and non-linearity in load data. The specific process is depicted in Figure 1.

2.1. Variational Modal Decomposition (VMD)

VMD is a nonrecursive, fully adaptive variational method proposed by Dragomiretskiy et al. [24]. The primary objective of VMD is to decompose the actual input signal into multiple discrete subsignals, termed intrinsic mode functions (IMFs). By regulating bandwidth, VMD is capable of effectively suppressing the modal overlap phenomenon [25].

To determine the bandwidth of each mode, VMD involves the following key processes:

- 1 The correlation analysis signal of each mode is calculated by means of the Hilbert transform to derive the one-sided frequency;
- 2 Each mode is combined with the exponential term and modulated to the base-band frequency;
- 3 Based on Gaussian smoothness and the squared parametric of the gradient, the center frequency of each mode is determined by demodulating the signal. The resulting constrained variational problem is expressed as follows:

$$\min_{\{u_k\}\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad (1)$$

$$\text{s.t.} \sum_k u_k = f, \quad (2)$$

where $\{u_k\} := \{u_1, \dots, u_K\}$ represents the set of all modes, and $\{\omega_k\} := \{\omega_1, \dots, \omega_K\}$ denotes the set of the center frequency of each mode. Equation (2) indicates that the sum of all modes is equivalent to the actual input signal.

The VMD method chiefly uses quadratic penalty terms α and Lagrange multipliers λ to render the problem unconstrained. Following the addition of α and λ , the sum of the equation is as follows:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \quad (3)$$

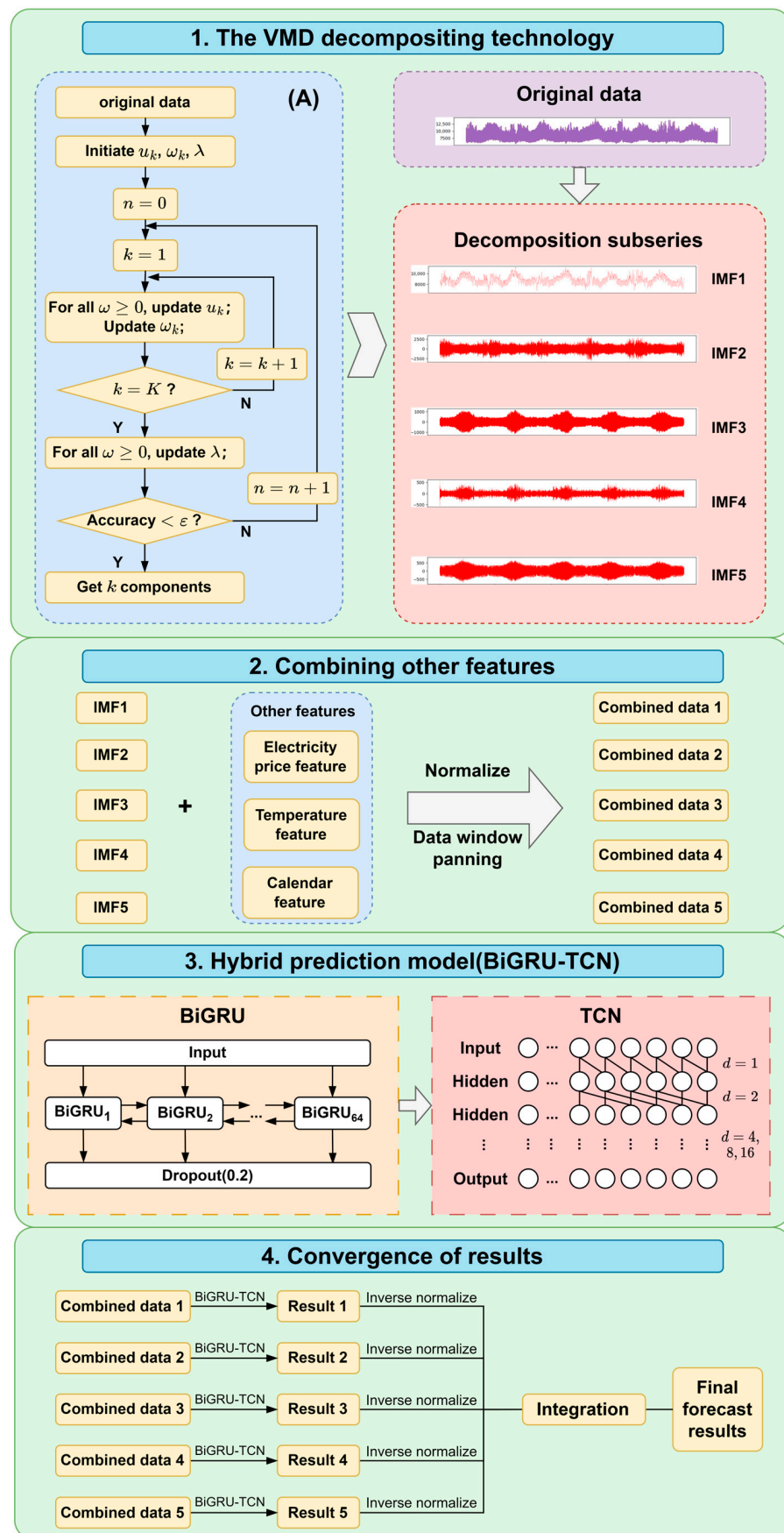


Figure 1. Flowchart of the VMD-TCN-BiGRU method. IMF represents termed intrinsic mode function. (A) VMD processing of load data.

This constrained problem can be solved through the alternating direction multiplier method, which involves fixing the other two variables and updating one of them, alternating updates for u_k^{n+1} , ω_k^{n+1} , and $\hat{\lambda}^{n+1}$ as follows:

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i > k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2}, \tag{4}$$

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega}, \tag{5}$$

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right), \tag{6}$$

where $\hat{f}(\omega)$, $\hat{u}_i^n(\omega)$, $\hat{\lambda}^n(\omega)$ and $u_k^{n+1}(\omega)$ are the Fourier transforms of $f(t)$, $u(t)$, $\lambda(t)$, and $\hat{u}_k^{n+1}(t)$, respectively; ω_k^{n+1} is the center frequency of the current mode; and $\hat{u}_k^{n+1}(\omega)$ represents the Wiener filtering of the current residuals.

The flowchart in Figure 1 illustrates the process of load sequence decomposition using the VMD algorithm. Initially, the load data are the input, and u_k , ω_k , λ are initialized. Next, u_k , ω_k , λ are updated alternately until the accuracy is lower than the set threshold, as described in the aforementioned method. Subsequently, K components are derived.

2.2. TCN

In the present study, TCN is utilized for short-term load data modeling. Its fundamental architecture comprises causal convolution, dilated convolution, and a residual module.

2.2.1. Causal Convolution

Each node’s data within the hidden layer of the causal convolution correlates only with the data at the same moment and those before it in the subsequent layer. This concept is illustrated in Figure 2. The primary goal of this approach is to mitigate the issue of information leakage that is prevalent in traditional convolutional structures, ensuring that no information from the future leaks into the past.

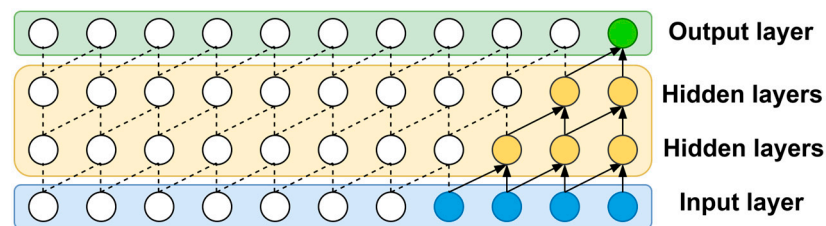


Figure 2. Structure of causal convolutions.

2.2.2. Dilated Convolution

To address the problem of information loss associated with historical data, the TCN model merges dilated convolution with causal convolution. This combination expands the field of view, capturing long-range dependencies within the input sequence. For a one-dimensional sequence $x_0(t)$, the filter f is f_0 to f_{k-1} . The convolution operation F for the sequence element s is as follows:

$$F(s) = (X *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i}, \tag{7}$$

where d is the dilation factor, k is the filter size, and $s-d \cdot i$ denotes the past direction.

Figure 3 presents the structure of the dilated convolution module. The inputs to the upper layer of the hidden layer neurons are discontinuous, allowing for an expanded field of view for the convolution kernel without the need for additional weights. As the dilation factor increases, the convolution kernel can capture increasingly distant dependencies.

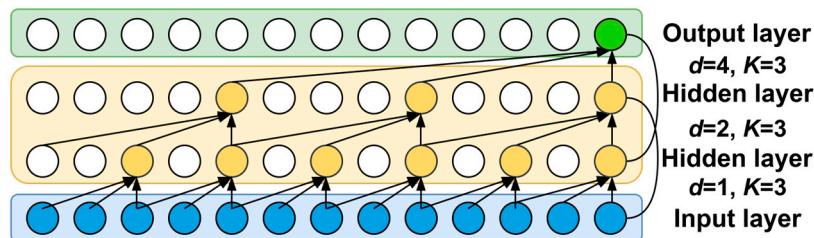


Figure 3. Structure of dilated convolutions.

2.2.3. Residual Module

The TCN model introduces the residual block to avoid the problems of information loss and instability caused by excessive network depth. The connection of the residual block is represented by the arc in Figure 3. As depicted in Figure 4, the residual module comprises two layers of dilated causal convolutional layers and their accompanying modules. To rectify the discrepancy of input and output widths, an additional one-dimensional convolution is implemented, ensuring that the two tensors involved in the summation operation (\oplus) maintain consistent shapes.

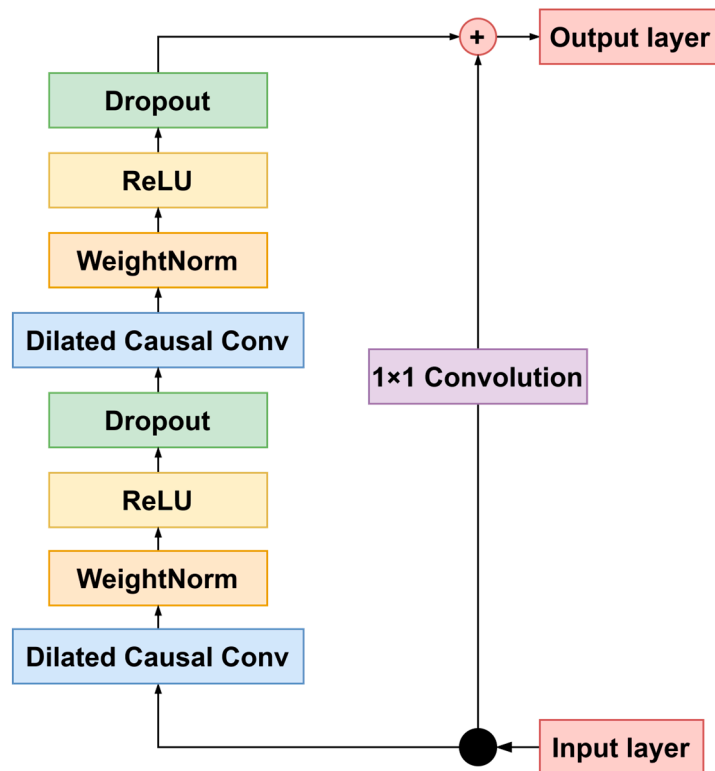


Figure 4. Structure of the residual block.

2.3. BiGRU

The BiGRU consists of two distinct GRU modules, namely, the forward and backward modules. This architecture improves upon the original LSTM by replacing the input gate, output gate, and forget gate with a reset gate and an update gate. This adjustment results in fewer parameters and accelerated training.

The calculations of the GRU update gate (z_t) and reset gate (r_t) are as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \tag{8}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \tag{9}$$

where h_{t-1} represents the hidden layer output at time $t-1$, while x_t denotes the current input. W_z and W_r are the weights in the update and reset gates, respectively, and σ is the Sigmoid function. The GRU output is the hidden layer at time t , i.e., h_t , and it is calculated as follows:

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t \times h_{t-1}, x_t]), \tag{10}$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t, \tag{11}$$

where \tilde{h}_t represents the candidate hidden layer; \tanh signifies the hyperbolic tangent function; and $W_{\tilde{h}}$ indicates the weight of the candidate hidden layer. The structure of the GRU is depicted in Figure 5.

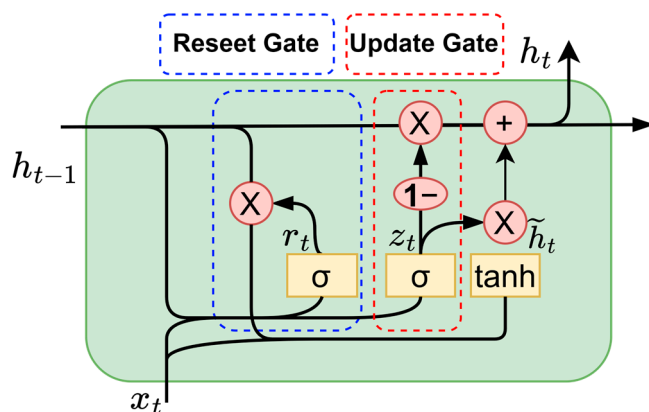


Figure 5. Structure of GRU.

The BiGRU utilizes two distinct GRUs for sequence modeling, capturing time series information features from two separate data transmission directions. This bidirectional modeling enhances prediction accuracy and robustness. The structure of the BiGRU is shown in Figure 6.

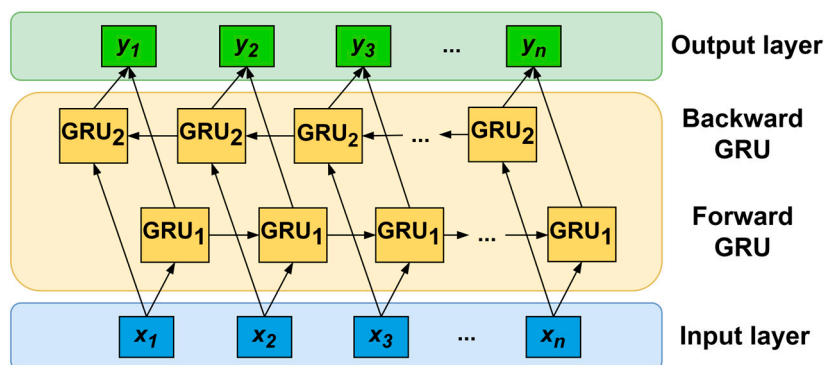


Figure 6. Structure of BiGRU.

2.4. Data Collection and Pre-Processing

The accuracy and generalizability of the proposed model were verified using two distinct datasets for training and model validation. The first data set (referred to hereafter as Dataset 1) was sourced from the Singapore National Electricity Market “<https://www.nems.emcsg.com/nems-prices> (accessed on 25 March 2023)”, encompassing data

from 8 January 2018 to 28 December 2019. The second dataset (henceforth referred to as Dataset 2) was derived from the Australian Energy Market Operator’s Australian database “<https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data> (accessed on 20 June 2023)”, including data from 1 January 2006 to 1 January 2011. Both datasets had a data collection interval of 0.5 h, with 48 data points collected throughout each day. Meteorological data were acquired from the National Oceanic and Atmospheric Administration “<https://www.noaa.gov/> (accessed on 25 March 2023)”.

As demonstrated in Figure 7, which represents the distribution of Dataset 1 and Dataset 2, a valley value occurs every 336 sampling points, i.e., roughly one week. This indicates that the load data exhibit weekly cyclical variation. And the electrical load is higher from Monday to Friday and lower on weekends. Thus, we labeled the work week as 0, and weekends as 1.

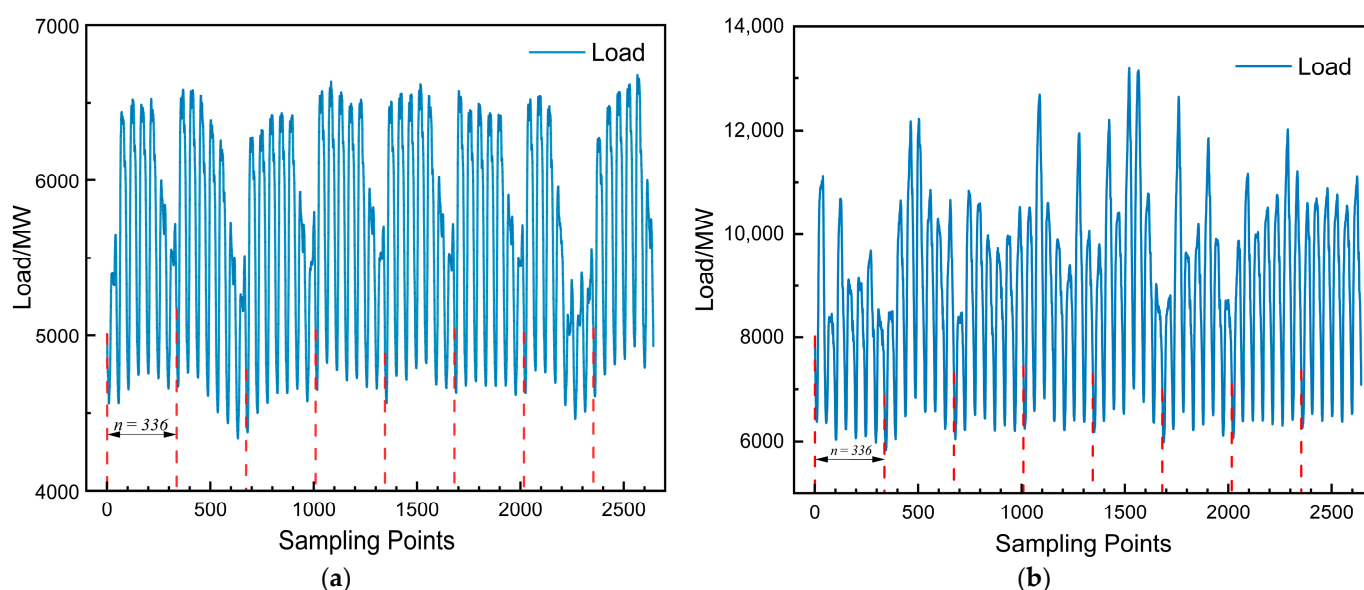


Figure 7. (a) Data distribution of Dataset 1. (b) Data distribution of Dataset 2.

The initial feature set was composed of 13 dimensions of data, including wind speed, wind direction, visibility, and precipitation. However, considering significant missing data for some features, we selected the features most relevant to the electric load data after a Pearson correlation analysis and a literature review [26].

The formula for calculating the Pearson correlation coefficient is as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{D_X}\sqrt{D_Y}}, \quad (12)$$

where $\rho_{X,Y}$ is the correlation coefficient between variables X , Y ; $\text{cov}(X, Y)$ is the covariance between variables X , Y ; and D_X and D_Y are the variance of X and Y .

According to Equation (12), the absolute values of correlation coefficients between electricity price, temperature, and load for the first five days of the two datasets were obtained as shown in Table 1. As shown in Table 1, the absolute values of the correlation coefficients between electricity price, temperature, and load were all greater than 0.5. Thus, electricity price and temperature can be considered as features to be introduced into the model.

For the calendar features, we labeled holidays as 0 and non-holidays as 1. The final feature set was narrowed down to the following dimensions: electricity price, temperature, and calendar features.

Table 1. Absolute values of correlation coefficients between electricity loads and each of the influencing factors for the first five days of both datasets.

Dataset	Date	Temperature	Price
Dataset 1	8 January 2018	0.87	0.73
	9 January 2018	0.14	0.96
	10 January 2018	0.93	0.99
	11 January 2018	0.51	0.91
	12 January 2018	0.58	0.89
Dataset 2	1 January 2006	0.66	0.75
	2 January 2006	0.60	0.92
	3 January 2006	0.85	0.62
	4 January 2006	0.86	0.52
	5 January 2006	0.80	0.61

The ratio of the training set, validation set, and test set was established as 8:1:1, and the time window was set to 7, applying super step load prediction [27,28]. Furthermore, the data were normalized to the maximum–minimum value and mapped to the [0, 1] interval.

3. Results and Discussion

3.1. Evaluation Indicators

This study employed the MAPE, RMSE, and R^2 as evaluation indicators. The smaller the MAPE value, the higher the accuracy of the prediction model; the smaller the RMSE, the smaller the prediction error of the model. The closer the R^2 is to 1, the better the fit of the model. These were calculated as follows:

$$\text{MAPE} = \frac{1}{m} \sum_{k=1}^m \frac{|\hat{y}_k - y_k|}{y_k} \times 100\%, \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{k=1}^m (\hat{y}_k - y_k)^2}, \quad (14)$$

$$R^2 = 1 - \frac{\sum_{k=1}^m (\hat{y}_k - y_k)^2}{\sum_{k=1}^m (\bar{y}_k - y_k)^2}, \quad (15)$$

where \hat{y}_k represents the forecasted electrical load data; y_k is the real electrical load data; \bar{y} is the average of all electrical load data; and m is the total number of electrical load data samples.

3.2. VMD Processing

Appropriate application of VMD can assist the combined model in mitigating the effects of noise and other interfering factors during model training. In this study, we applied VMD using a penalty factor α with a value of 2000 and a chosen threshold ϵ . We then conducted a comparative selection of K values. Table 2 presents the decomposition results of Dataset 1 for different K values. When the value of K equals or exceeds 6, the central frequencies of IMF3 and IMF4 are in close proximity, suggesting the occurrence of modal mixing in the system when K equals or exceeds 6. As a result, we selected $K = 5$ for the study of Dataset 1.

We subsequently subjected Dataset 2 to VMD. The decomposition results for Dataset 2 under varying K values are presented in Table 3. Consistent with the findings from Dataset 1, $K = 5$ was optimal for the study of Dataset 2.

Table 2. Process of VMD for Dataset 1.

K	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
3	0.09	725.36	1472.03				
4	0.09	725.14	1462.79	2874.54			
5	0.09	725.14	1462.69	2873.75	6502.09		
6	0.09	725.11	1461.81	1463.60	2893.02	6509.06	
7	0.09	725.15	1455.00	1450.39	2147.45	2919.46	6518.35

Table 3. Process of VMD for Dataset 2.

K	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
3	0.46	1822.48	3739.96				
4	0.45	1819.47	3651.89	5657.73			
5	0.44	1819.21	3647.96	5466.88	7414.18		
6	0.43	1819.02	3650.71	3592.43	7405.81	5470.70	
7	0.42	1819.10	3647.69	3626.22	5473.00	3631.91	7398.07

3.3. Analysis of Results

This section describes a comparative analysis conducted to highlight the strengths of the model proposed in this study. We compared our model with several others, including the artificial neural network (ANN), CNN, support vector regression (SVR), LSTM, and GRU models, as well as a few common combination models. This comparison verified the superior accuracy of our model. The models were trained with consistent parameters and then applied to predictions on Dataset 1 and Dataset 2 to demonstrate their generalization capabilities.

Table 4 details the specific parameters of our model and the comparison models. These parameters were obtained through several tuning experiments, and here, we show the tuning process of the model proposed in this paper under dataset 1, as shown in Table 5. The other models were similar and will not be listed in detail. With the current parameter configuration, optimal results were achieved for both our model and the comparison models.

Table 4. List of model parameters.

Model	Parameters *	Common Parameters
BiGRU	Units = 64	
TCN	Nb_filters = 64; Kernel_size = 3 Nb_stacks = 1; Dilations = (1, 2, 4, 8, 16) Activation = 'relu'; Padding = 'causal'	Dropout: 0.2 Loss: MSE
GRU	Units = 64	Optimizer: Adam
CNN	Filters = 128; Kernel_size = 1; Pool_size = 1	
ANN	Units = 64; Units = 32; Units = 1	
LSTM	Units = 64	
SVR	Kernel = 'rbf'; C = 100; Gamma = 0.001	

* Units represent the dimension of the output space; Nb_filters represents the number of filters to use in the convolutional layers; Kernel_size represents the size of the kernel to use in each convolutional layer; Dilations represents the list of dilations; Nb_stacks represents the number of stacks of residual blocks to use; Activation represents the activation used in the residual blocks; Padding represents the padding to use in the convolutional layers; Filters represents the dimensionality of the output space (i.e., the number of output filters in the convolution); Kernel specifies the kernel type to be used in the algorithm; C represents the penalty parameter C of the error term; and Gamma represents the kernel coefficient.

Table 5. The tuning process of the TCN-BiGRU model.

TCN	BiGRU	MAPE/%↓	RMSE/MW↓	R ² /%↑
Nb_filters = 32	Units = 32	1.73	131.40	85.38
Nb_filters = 64	Units = 32	1.64	128.95	85.92

Table 5. Cont.

TCN	BiGRU	MAPE/% \downarrow	RMSE/MW \downarrow	R ² /% \uparrow
Nb_filters = 128	Units = 32	2.15	141.02	83.16
Nb_filters = 32	Units = 64	1.28	96.18	92.16
Nb_filters = 64	Units = 64	0.42	29.35	98.27
Nb_filters = 128	Units = 64	1.17	80.34	94.53
Nb_filters = 32	Units = 128	1.57	110.04	89.74
Nb_filters = 64	Units = 128	1.55	108.03	90.12
Nb_filters = 128	Units = 128	1.91	149.17	81.16

Figure 8 presents the prediction results for each model under Datasets 1 and 2 as parts a and b, respectively, and Table 6 provides the corresponding evaluation metrics.

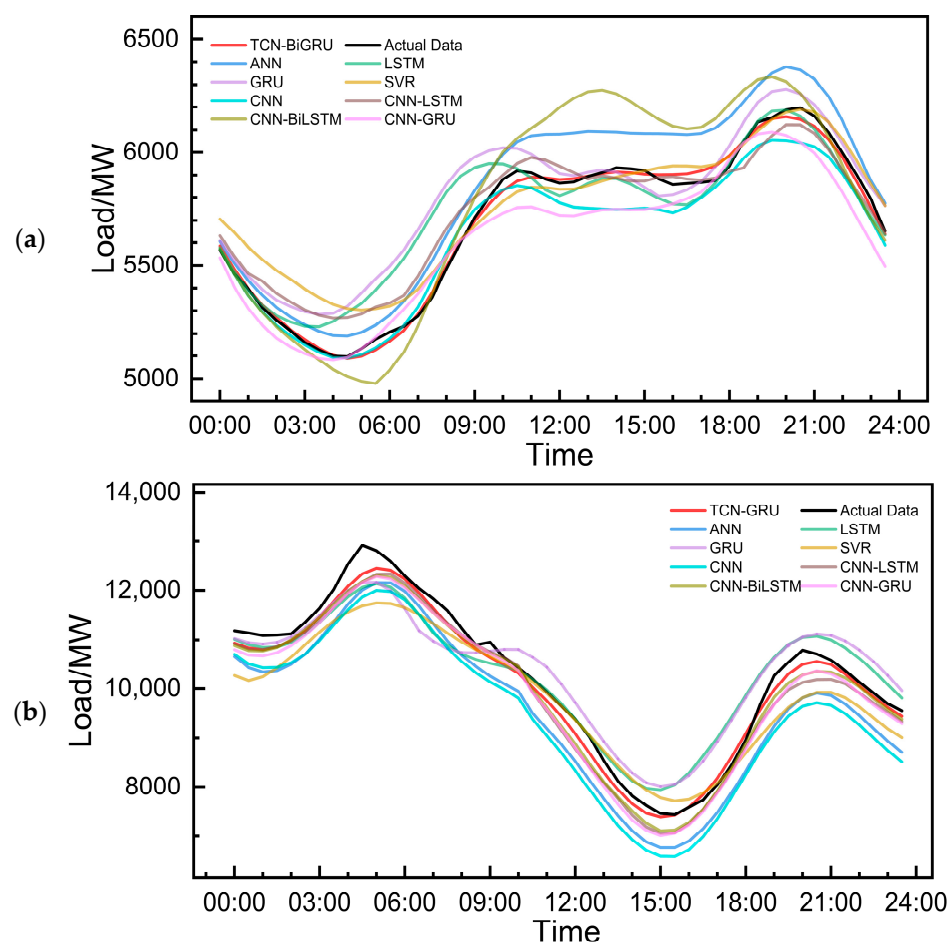


Figure 8. (a) Comparison of prediction results from the load model over 24 h for Dataset 1; (b) comparison of prediction results from the load model over 24 h for Dataset 2.

Figure 8 illustrates that the ANN model's overall prediction performance was unsatisfactory, with it only capturing the overall trend of the actual data and its error notably increasing over time. Furthermore, it struggled to handle large load fluctuations. The CNN model handled trough data better, but it encountered more significant errors when faced with frequent fluctuations. The SVR model roughly predicted the data direction, but initially exhibited a large error. Both the LSTM and GRU models failed to accurately predict long-term trough and peak fluctuations, but were sensitive to short- and medium-term changes. General combinations of CNN with the LSTM or GRU models did not yield appropriate results due to the CNN's inherent limitations. In contrast, the model proposed in this study fit the actual data appropriately; adeptly handled long-term fluctuations;

and provided accurate predictions for peaks, troughs, and short-term fluctuations. It can also precisely anticipate frequent short-term fluctuations, resulting in an adequate overall curve fit.

Table 6. Prediction results of different models.

Dataset	Model	MAPE/% \downarrow	RMSE/MW \downarrow	R ² /% \uparrow
Dataset 1	TCN–BiGRU	0.42	29.35	98.27
	ANN	2.40	149.52	81.08
	CNN	1.63	108.94	89.30
	SVR	1.67	115.01	88.80
	LSTM	1.76	139.76	83.46
	GRU	2.26	164.68	77.04
	CNN–LSTM	1.61	105.23	90.62
	CNN–BiLSTM	1.23	86.45	93.67
	CNN–GRU	1.51	101.62	91.25
Dataset 2	TCN–BiGRU	1.79	217.17	97.98
	ANN	6.89	701.17	78.94
	CNN	3.54	379.43	70.53
	SVR	4.42	573.92	85.89
	LSTM	4.17	485.24	89.91
	GRU	4.84	528.07	88.05
	CNN–LSTM	3.17	361.75	94.39
	CNN–BiLSTM	2.75	316.47	95.71
	CNN–GRU	3.47	380.71	93.79

Table 5 demonstrates that, regarding the experimental results from Dataset 1, the proposed model surpassed the comparative models in all evaluation metrics. Specifically, compared to the ANN, CNN, SVR, LSTM, and GRU models, the MAPE decreased by 1.98, 1.21, 1.25, 1.34, and 1.84 percentage points, respectively. This indicates a higher prediction accuracy of the model developed in this study. The RMSE also decreased, suggesting a lower prediction error for our model. Additionally, the R² of our model reached 98.27%, indicating well-fitted prediction results and excellent performance in completing the prediction task. Even when comparing the combined models, our model excelled over CNN–LSTM, CNN–BiLSTM, and CNN–GRU in terms of MAPE, RMSE, and R². The symbols of “ \downarrow ” or “ \uparrow ” beside these indicators represent the expected trend of each evaluation indicator.

The analysis of Dataset 2 also yielded similar results. By observing and comparing the results according to metrics of MAPE, RMSE and R², we found that our model, trained with the same parameters, outperformed both the individual and combined models.

3.4. Ablation Experiments

To further ascertain the viability and effectiveness of the method employed in this study, we conducted ablation experiments on both datasets. The ablation experiment results are presented in Table 7 and Figure 9. Groups A, B, C, D, and E in Table 7 represent the models after the deletion or alteration of modules, with consistent model parameters across all groups. By examining Figure 9 and comparing the experimental results from Groups A and B in Table 7, predictions with the BiGRU model are more accurate, and training performance markedly improved for both Datasets 1 and 2. Comparing Groups A and C, the results reveal that not decomposing data for Dataset 1 decreased the accuracy, and for Dataset 2, when VMD was not performed, the prediction results were adversely impacted, demonstrating the significance of VMD in the model. Moreover, omitting the BiGRU module also led to a decline in model prediction performance, thus emphasizing the critical role of BiGRU in the model. The ablation experiments allowed us to conclude that our combined model as well-constructed, with each structural element positively contributing to its prediction accuracy and efficiency.

Table 7. Results of ablation experiments.

Dataset	Group	VMD	TCN	GRU module	MAPE/% \downarrow	RMSE/MW \downarrow	R ² /% \uparrow	t/s \downarrow
Dataset 1	A	✓	✓	BiGRU	0.42	29.35	98.27	501.25
	B	✓	✓	GRU	1.71	123.18	87.16	1194.94
	C	/	✓	BiGRU	1.63	118.62	88.01	260.02
	D	✓	/	BiGRU	2.24	153.18	80.14	362.78
	E	✓	✓	/	1.33	90.93	93.00	444.26
Dataset 2	A	✓	✓	BiGRU	1.79	217.17	97.98	3666.43
	B	✓	✓	GRU	3.95	456.98	91.06	5501.26
	C	/	✓	BiGRU	/	/	/	/
	D	✓	/	BiGRU	3.37	392.79	93.39	1133.13
	E	✓	✓	/	2.28	304.89	96.01	5305.01

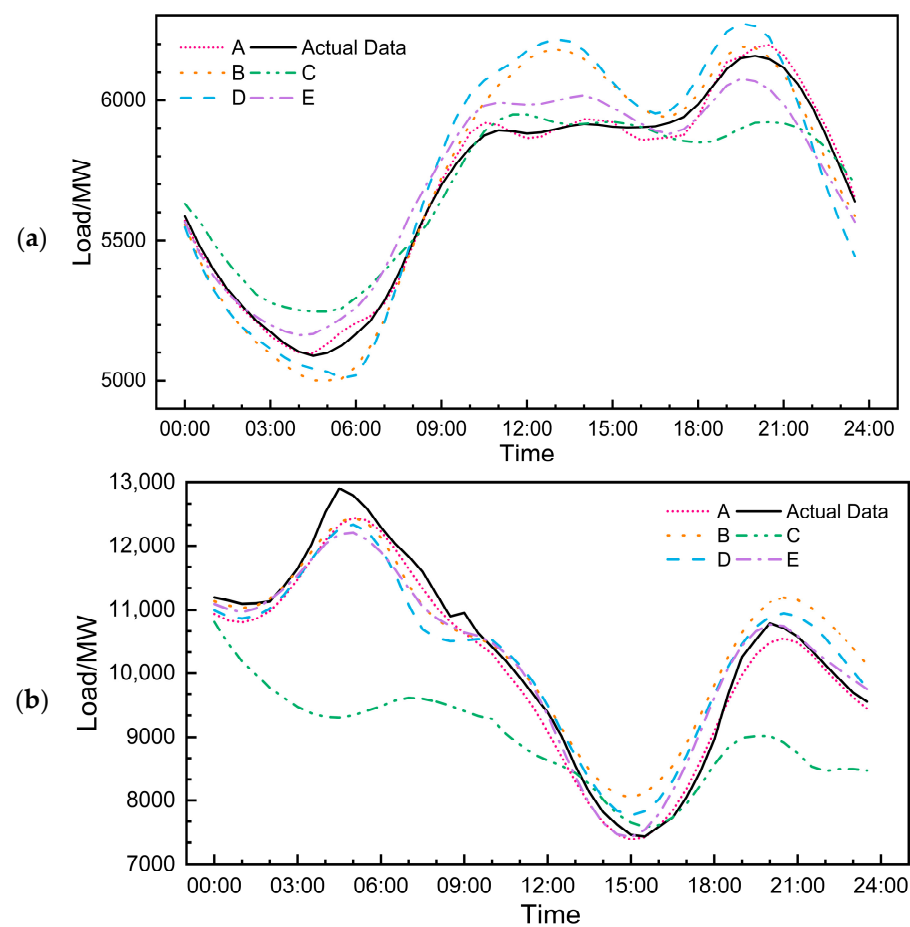


Figure 9. (a) Comparison of predicted results within 24 h of the ablation experiment for dataset 1. (b) Comparison of predicted results within 24 h of the ablation experiment for dataset 2.

4. Conclusions

This study addresses the challenges of uncertainty and non-linearity inherent in short-term electrical load series. Traditional prediction methods, often marked by low accuracy due to the inadequate consideration of influencing factors, are contrasted with a short-term electrical load forecasting approach based on the VMD–TCN–BiGRU model. This proposed method considers a broad range of factors, including electricity price, temperature, and calendar variables. Using Singapore load data for experimentation, the proposed method achieved a MAPE of 0.42% and an RMSE of 29.35 MW, suggesting high prediction accuracy with minimal errors. When the same set of parameters was applied to Australian load data, the model demonstrated 1.79% in MAPE and 217.17 in RMSE, thus confirming the

superior generalization of the proposed method. In summary, the following conclusions can be drawn:

- 1 By integrating a wide array of factors—natural, human, economic, and sequence characteristics—the predictive accuracy of the model can be significantly enhanced.
- 2 VMD can mitigate the impact of uncertainty and non-linearity in load series on prediction accuracy and stability.
- 3 The hybrid model employing TCN and BiGRU effectively captures the long- and short-distance dependencies in load data. This approach not only improves the model's performance and stability, but also exhibits robust adaptability to different datasets.

Future research will consider the incorporation of a broader range of meteorological factors to further improve the forecasting performance. Meanwhile, power load forecasting is a very sensitive application and is susceptible to many security issues. Therefore, in addition to pursuing high accuracy, we also need to consider the security of the model to ensure its ability to cope with various attacks. Therefore, it is also necessary to consider combining it with specific defense methods, such as [29], in future practical applications of the model.

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

Funding: This research was funded by the National Grid Limited Science and Technology Project, grant number 521750220003. College Student Innovation and Entrepreneurship Training Program, grant number X202210264160 and S202310264120.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://www.nems.emcsg.com/nems-prices> (accessed on 25 March 2023), <https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data> (accessed on 20 June 2023) and <https://www.noaa.gov/> (accessed on 25 March 2023).

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

Abbreviations

The following abbreviations are used in this manuscript:

Notation	Description
CNY	Chinese unit of currency
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
CNN	Convolutional neural network
CS	Cuckoo search
SVM	Support vector machine
XGBoost	Extreme gradient boosting
RF	Random forests
LSTM	Long short-term memory
MSE	Mean square error
MAE	Mean absolute error
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
SE	Sample entropy
RNN	Recurrent neural network
TCN	Temporal convolutional network
GRU	Gated recurrent unit
VMD	Variational modal decomposition
BiGRU	Bidirectional recurrent neural network
IMFs	Termed intrinsic mode functions

MAPE	Mean absolute percentage error
RMSE	Root mean square error
R ²	Goodness of fit
ANN	Artificial neural network
SVR	Support vector regression

Appendix A

Table A1. Summary of the literature review.

Reference	Method	Dataset (Period)	Location/County	Metrics	Pros	Cons
[10]	Threshold ARMA	Average daily residential electricity load data (from 1 May 2017 to 31 March 2020)	A prefecture-level city in the south-west of Zhejiang Province, China	MAPE: 4.167%	Considering the influence of temperature	Difficulty in adapting to the effects of non-linear factors on load data
[11]	ARIMA-CNN	Daily electricity consumption data (from 2016 to 2018)	Tai'an, Shandong Province, China	MAPE: 4.89%	Based on wavelet transform	
[12]	CS-SVM	PJM power market (from 1995 to 1998)	United States	MAPE: 13.43%	Considered demand price elasticity	It is difficult for traditional machine learning to extract its features deeply in non-linear time series data
[13]	RF	ENTSO-E repository (from 2012 to 2015)	Poland (PL), Great Britain (GB), France (FR) and Germany (DE).	MAPE: PL: 1.05% GB: 2.36% FR: 1.67% DE: 1.06%	RF has a low number of tuning hyperparameters; fast training and optimization	
[14]	CEEMDAN-SE-LSTM	Electric load data (from 13 May 2014, to 13 May 2017)	Changsha, China	MAPE: 1.649%	Decomposing the electric load data first	
[15]	LSTM-Informer	The power consumption data of the power grid (52,416 pieces of data in a 10 min window from 2017)	Tetouan, Morocco.	MSE: 0.2085% MAE: 0.3963%	Using the combined model	LSTM is not as fast as GRU
[16]	LSTM, GRU and RNN	Electricity load data (from 1 September 2021 to 31 August 2022)	Tubas Electricity Company, Palestine	GRU: MSE: 0.215% MAE: 3.266%	The GRU model obtained the best results.	Only a single factor was used and the considerations were not comprehensive enough
[17]	A novel CNN-GRU-Based hybrid approach	AEP and IHEPC datasets available (ten-minute resolution for about 4.5 months)	Public	MSE: 0.22 RMSE: 0.47 MAE: 0.33	The CNN-GRU model is better than base models such as XGBoost and RNN.	CNN is not as flexible as TCN
[18]	CEEMDAN-TCN-GRU-Attention	Power load data (f 5400 data points)	Quanzhou City, Fujian Province, China	MAE: 95.851 MW R ² : 98.2% RMSE: 125.23 MW MAPE: 1.099%	The data were first decomposed and a combined TCN and GRU model was applied.	Not combined with the holiday factor

References

1. Wang, N.; Fu, X.D.; Wang, S.B. Economic growth, electricity consumption, and urbanization in China: A tri-variate investigation using panel data modeling from a regional disparity perspective. *J. Clean. Prod.* **2021**, *318*, 128529. [\[CrossRef\]](#)
2. Kim, G.U.; Jin, B.Y.; Park, J.G. An Analysis on Causalities Among Economic Growth, Electricity Consumption, CO₂ Emission and Financial Development in Korea. *J. Ind. Econ. Bus.* **2020**, *33*, 2.
3. Vecchione, G. Economic Growth, Electricity Consumption and Foreign Dependence in Italy Between 1963–2007. *Energy Sources Part B Econ. Plan. Policy* **2011**, *6*, 3. [\[CrossRef\]](#)

4. Chen, Y.T.; Zhang, D.X. Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory. *Adv. Appl. Energy* **2021**, *1*, 100004. [[CrossRef](#)]
5. Lin, T.; Zhao, Y.; Feng, J.Y. Research on Short-Term Electric Load Combination Prediction Model Based on Feature Decomposition. *Comput. Simul.* **2022**, *39*, 91–95+251.
6. Dong, J.F.; Wan, X.; Wang, Y.; Ye, R.L.; Xiong, Z.J.; Fan, H.W.; Xue, Y.B. Short-term Power Load Forecasting Based on XGB-Transformer Model. *Electr. Power Inf. Commun. Technol.* **2023**, *21*, 9–18.
7. Zhu, Q.Z.; Dong, Z.; Ma, N. Forecasting of short-term power based on just-in-time learning. *Power Syst. Prot. Control* **2020**, *48*, 92–98.
8. Yao, G.F.; Li, T.J.; Liu, L.F.; Zheng, Y.N. Residential Electricity Load Forecasting Method Based on DAE and LSTM. *Control Eng. China* **2022**, *29*, 2048–2053.
9. Ouyang, F.; Wang, J.; Zhuo, H.X. Short-term power load forecasting method based on improved hierarchical transfer learning and multi-scale CNN-BiLSTM-Attention. *Power Syst. Prot. Control* **2023**, *51*, 132–140.
10. Sun, Y.Q.; Wang, Y.W.; Zhu, W.; Li, Y. Residential Daily Power Load Forecasting Based on Threshold ARMA Model Considering the Influence of Temperature. *Electr. Power Constr.* **2022**, *43*, 117–124.
11. Wang, A.D.; Zou, Y.; Jiang, T.Y.; Zhang, F. Short term load forecasting using ARIMA-CNN combination model based on wavelet transform. In Proceedings of the 21 National Simulation Technology Academic Conference, Virtual, 21–23 September 2021; pp. 167–171.
12. Su, J.; Fang, S.; Xing, G.J.; Du, S.H.; Shan, B.G. Short-term load forecasting method based on cuckoo search algorithm and support vector machine considering demand price elasticity. *J. Jiangsu Univ. Nat. Sci. Ed.* **2022**, *43*, 319–324.
13. Dudek, G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies* **2022**, *15*, 7547. [[CrossRef](#)]
14. Li, K.; Huang, W.; Hu, G.Y.; Li, J. Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. *Energy Build.* **2022**, *279*, 112666. [[CrossRef](#)]
15. Wang, K.; Zhang, J.; Li, X.; Zhang, Y. Long-Term Power Load Forecasting Using LSTM-Informer with Ensemble Learning. *Electronics* **2023**, *12*, 2175. [[CrossRef](#)]
16. Abumohsen, M.; Owda, A.Y.; Owda, M. Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms. *Energies* **2023**, *16*, 2283. [[CrossRef](#)]
17. 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](#)]
18. Hong, Y.; Wang, D.; Su, J.; Ren, M.; Xu, W.; Wei, Y.; Yang, Z. Short-Term Power Load Forecasting in Three Stages Based on CEEMDAN-TGA Model. *Sustainability* **2023**, *15*, 11123. [[CrossRef](#)]
19. Cheng, R.T.; Zhang, Y.J.; Li, L.C.; Ding, M.S.; Deng, W.Y.; Chen, H.Y.; Lin, J.C. Construction and Research Progress of Electricity Market for High-Proportion Renewable Energy Consumption. *Strateg. Study CAE* **2023**, *25*, 89–99.
20. Huang, S.; Zhang, J.; He, Y.; Fu, X.; Fan, L.; Yao, G.; Wen, Y. Short-Term Load Forecasting Based on the CEEMDAN-Sample Entropy-BPNN-Transformer. *Energies* **2022**, *15*, 3659. [[CrossRef](#)]
21. Wang, Z.C.; Wang, Q.Y.; Wu, T.H. A novel hybrid model for water quality prediction based on VMD and IGOA optimized for LSTM. *Front. Environ. Sci. Eng.* **2023**, *17*, 88. [[CrossRef](#)]
22. Cho, K.; Merriënboer, B.V.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv* **2014**, arXiv:1409.1259.
23. Bai, S.; Kolter, J.Z.; Koltun, V. Trellis Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
24. Zosso, D.; Dragomiretskiy, K. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544.
25. Yang, H.P.; Yu, Y.; Wang, C.; Li, X.J.; Hu, Y.T.; Rao, C.C. Short-Term Load Forecasting of Power System Based on VMD-CNN-BIGRU. *Electr. Power* **2022**, *55*, 71–76.
26. Xiong, M.M.; Li, M.C.; Ren, Y.; Xu, S.; Yang, Y.J. Characteristics of Electrical Load and Its Relationship to Meteorological Factors in Tianjin. *Meteorol. Sci. Technol.* **2013**, *41*, 577–582.
27. Zhao, Q.; Huang, J.T. On ultra-short-term wind power prediction based on EMD-SA-SVR. *Power Syst. Prot. Control* **2020**, *48*, 89–96.
28. Cui, J.H.; Bi, L. Research on photovoltaic power forecasting model based on hybrid neural network. *Power Syst. Prot. Control* **2021**, *49*, 142–149.
29. Abdalzaher, M.S.; Fouda, M.M.; Emran, A.; Fadlullah, Z.M.; Ibrahim, M.I. A Survey on Key Management and Authentication Approaches in Smart Metering Systems. *Energies* **2023**, *16*, 2355. [[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.