

Article

An Ensemble Deep Learning Model for Provincial Load Forecasting Based on Reduced Dimensional Clustering and Decomposition Strategies

Kaiyan Wang ^{1,2}, Haodong Du ^{1,*}, Jiao Wang ², Rong Jia ^{1,2} and Zhenyu Zong ¹¹ School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China² Key Laboratory of Smart Energy in Xi'an, Xi'an University of Technology, Xi'an 710048, China

* Correspondence: 3170421159@stu.xaut.edu.cn

Abstract: The accurate prediction of short-term load is crucial for the grid dispatching department in developing power generation plans, regulating unit output, and minimizing economic losses. However, due to the variability in customers' electricity consumption behaviour and the randomness of load fluctuations, it is challenging to achieve high prediction accuracy. To address this issue, we propose an ensemble deep learning model that utilizes reduced dimensional clustering and decomposition strategies to mitigate large prediction errors caused by non-linearity and unsteadiness of load sequences. The proposed model consists of three steps: Firstly, the selected load features are dimensionally reduced using singular value decomposition (SVD), and the principal features are used for clustering different loads. Secondly, variable mode decomposition (VMD) is applied to decompose the total load of each class into intrinsic mode functions of different frequencies. Finally, an ensemble deep learning model is developed by combining the strengths of LSTM and CNN-GRU deep learning algorithms to achieve accurate load forecasting. To validate the effectiveness of our proposed model, we employ actual residential electricity load data from a province in northwest China. The results demonstrate that the proposed algorithm performs better than existing methods in terms of predictive accuracy.

Keywords: short-term load forecasting; clustering; decomposition; forecasting strategies; ensemble deep learning

MSC: 60G25



Citation: Wang, K.; Du, H.; Wang, J.; Jia, R.; Zong, Z. An Ensemble Deep Learning Model for Provincial Load Forecasting Based on Reduced Dimensional Clustering and Decomposition Strategies.

Mathematics **2023**, *11*, 2786. <https://doi.org/10.3390/math11122786>

Academic Editor: Faheim Sufi

Received: 6 April 2023

Revised: 12 June 2023

Accepted: 16 June 2023

Published: 20 June 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

As China's electrification level rises, so does demand for electrical energy [1]. Residential, commercial, and industrial electricity consumption all show a year-on-year growth trend. However, the electricity load exhibits obvious randomness and volatility due to objective and social factors such as weather changes, holidays, and unexpected situations [2,3]. This complicates load forecasting and has an impact on the reliability and efficiency of power system operation [4–7]. Highly accurate load forecasting enables power dispatching authorities to develop more scientific and cost-effective power generation plans, reducing fossil fuel consumption and slowing environmental degradation [8]. Improved load forecasting accuracy can lead to better peak regulation in power systems with energy storage. As a result, finding high-accuracy load forecasting methods has been a hot and difficult research topic in this field [9–11].

In terms of time scales, there are currently four main types of load forecasts: ultra-short-term load forecasts [12], short-term load forecasts [13,14], medium-term load forecasts [15–17], and long-term load forecasts [18–21]. Ultra-short-term forecasts (one hour or 10 min in the future) are mainly used for real-time security analysis, real-time economic dispatch, and automatic generation control. Short-term forecasts (one day or one week ahead) are used for

scheduling daily start-up and shutdown plans and generation plans. Medium-term forecasts (one year in the future) are used for monthly maintenance plans, operation methods, and reservoir scheduling plans. Long-term forecasts (10 years in the future) provide key basic data for grid planning or for determining annual maintenance plans, operation methods, etc.

Load forecasting models can be mainly classified as statistical [22,23], artificial intelligence [24,25], and combined models [26–28].

Statistical models mainly include autoregressive integrated moving average, seasonal autoregressive integrated moving average, multiple linear regression, exponential smoothing method, etc. Wang Bo et al. constructed an autoregressive and moving average model with exogenous variables to achieve the short-term forecasting of load [29]. Luiz Felipe Amaral et al. developed the smooth transition periodic autoregressive model and evaluated the load forecasting performance of the model [30]. Traditional statistical methods will no longer be applicable when forecasting complex non-linear trends.

In recent years, artificial intelligence models have been widely applied in the field of load forecasting with excellent results. Common artificial intelligence models can be broadly classified into machine learning models [31–33] and deep learning models [34–36]. Support vector machines and neural network models are the most representative machine learning models. Jian Luo et al. [37] constructed a weighted quadratic surface support vector regression model to achieve efficient load prediction. The results show that the support vector function can handle non-linear time series better, but the parameter setting of the method is too cumbersome. Haoming Liu et al. [38] constructed a combinatorial model based on support vector regression for short-term load forecasting of integrated energy systems. Pham et al. [39] used BPNN as a core forecasting algorithm for load forecasting. Yusha Hu et al. [40] established a parameter-optimised BPNN network to avoid the problem of prediction results falling into local optima. Deep learning models are a field of algorithms that has grown from neural network models in machine learning models. Deep learning models contain more complex structures and are suitable for processing large amounts of non-linear data. A stacked autoencoder structure based on a deep LSTM was innovatively proposed by Zahra Fazlipour et al. [41]. Notably, the study shows that the deep structure is useful for improving prediction accuracy. A stack LSTM model was developed by Hongbo Ren et al. [42] for implementing load forecasting. Temporal convolutional networks incorporating an attention mechanism were constructed by XianlunTang et al. [43].

Various single prediction models have their limitations, and prediction accuracy is difficult to meet production needs. Combinatorial prediction methods have started to emerge in recent years. Combinatorial models can be divided into two main categories: combinations of optimisation algorithms and forecasting models and the integration of multiple forecasting models. A combined forecasting model, i.e., the Elman neural network (ENN) model optimised using the particle swarm optimization algorithm for load power forecasting was proposed by Kun Xie et al. [44]. Considerable work has also been performed by groups of multiple researchers in the area of integration of multiple predictive models. Xifeng Guo et al. [45] used a convolutional neural network to cascade four different scales of features to fully exploit the potential relationships between continuous and discontinuous data in the feature maps. The feature vectors at different scales are fused as input to the LSTM network, and the LSTM neural network is used for short-term load prediction. Umar Javed et al. [46] combined an expanded causal convolutional network with short sensory fields and a bi-directional LSTM to build a new load forecasting architecture, achieving higher accuracy in load forecasting. Sana Arastehfar et al. [47] integrated a graph convolutional neural network and a long- and short-term memory network into a unified network. The network can extract both temporal and spatial information. In addition to the combined models mentioned above, the combination of decomposition methods and deep learning models is a new trend in the field of load forecasting. Weimin Yue et al. [48] combined ensemble empirical modal decomposition with long- and short-term memory neural networks to address the problem of poor load prediction accuracy.

Qian Zhang et al. [49] combined a variational modal decomposition model with a stacked integrated model for load prediction. These studies show the outstanding advantages of combined models in the field of load forecasting.

In everyday life and production processes, there is differentiation in the electricity consumption behaviour of customers. Load characteristics therefore vary, making it difficult for generalised forecasts to meet the requirements of forecasting accuracy. When forecasting the load in an area, if each subarea is forecast individually and then integrated, the forecast granularity is too fine and prone to over-fitting. At the same time, it would be more time-consuming to forecast the load using this method. Conversely, if all customer loads are aggregated and then predicted, the prediction granularity is too large, and the differentiated characteristics of customer electricity consumption cannot be obtained. Therefore, the extraction of customer electricity characteristics is also a key technique that affects the accuracy of load forecasting. KOIVISTO M et al. [50] used principal component analysis for dimensionality reduction and clustering using K-means. The method can effectively achieve the clustering of loads, but the stability of the principal component analysis method is poor, and the accuracy is not high. ZHONG S et al. [51] used the Fourier transform as a dimensionality reduction method to extract the main load features and achieve classification of loads, but the method did not specify the weights of the dimensionality reduction indicators.

Based on previous research, an ensemble load forecasting algorithm based on load clustering and load decomposition strategies is proposed in this study. The improved prediction strategy includes three aspects. The first is a load clustering method based on principal characteristic extraction and dimensionality reduction. The extracted primary features reflect the fluctuation characteristics of various loads. The dimensionality reduction strategy can preserve important features while reducing the complexity of the clustering model, improving clustering efficiency and reducing memory requirements. The clustering algorithm groups loads with similar trends in variation. By predicting a class of loads with similar trends in a provincial region, the forecasting accuracy can be improved while reducing time costs. The second aspect is the load decomposition strategy. Multiple classes of loads with different characteristics are obtained after clustering. The sum of each type of load is decomposed into separate components with a single and uncoupled frequency, and then separate prediction models are constructed for the different frequency components. By decomposing each class of load, not only can the information contained in the data be fully explored, but also the interaction between different components at the characteristic scale can be reduced. Thirdly, an integrated deep learning model based on the LSTM and CNN-GRU model is constructed. In this model, the LSTM and CNN-GRU models are used for the prediction of low-frequency and high-frequency components, respectively. The LSTM can fully reflect the overall trend of the load and has high accuracy in predicting low-frequency time series. The CNN-GRU neural network, on the other hand, has a strong non-linear fitting capability and can achieve accurate prediction of high-frequency components with high randomness. The advantages of these two prediction methods complement each other. The prediction methods were validated by simulation, yielding desirable prediction results.

This paper is organised as follows. Section 2 presents the novel load forecasting model and describes the relevant theoretical background of the algorithms involved in this paper. Section 3 presents a case study of the proposed model. Section 4 concludes the paper.

2. Methodology

2.1. An Ensemble Forecasting Model Based on an Improved Load Clustering and Decomposition Strategy

In this paper, a forecasting strategy based on load clustering and decomposition is designed to achieve accurate load forecasting for provincial areas containing multiple cities, and the following detailed steps are given.

Step 1: A load clustering method based on principal feature extraction was used to cluster loads from multiple cities. The load characteristics are first calculated for all cities. Secondly, the SVD method is applied to reduce the load characteristics in order to

extract the main ones. Finally, a K-means algorithm is used to cluster the loads of several cities based on their main characteristics. In this study, the load data of one province (containing 10 cities) are used as the study data. The load data of 10 cities are processed by dimensionality reduction–clustering.

Step 2: The total load of each category is obtained from the clustering results. The VMD algorithm is used to decompose the various types of loadings obtained by clustering. Several different frequency components are obtained.

Step 3: Based on the improved prediction strategy, the ensemble prediction algorithm is proposed. The LSTM and CNN-GRU models are used to predict the low-frequency and high-frequency IMF components obtained using the VMD algorithm, respectively, and then the prediction results of each component are superimposed to obtain the final prediction results of each type of load. Finally, the forecast results of each type of load are superimposed to obtain the load forecast results of the province.

The flowchart of the ensemble model considering clustering and decomposition strategies to achieve provincial short-term load forecasting is given in Figure 1.

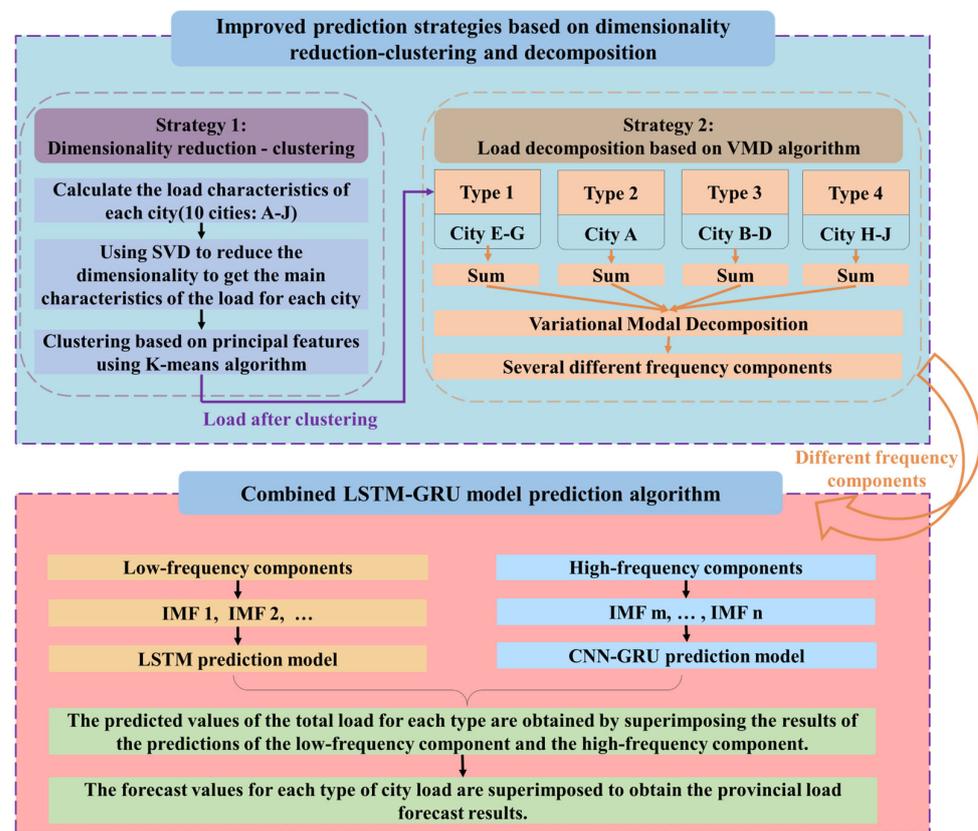


Figure 1. An ensemble model based on clustering and decomposition strategies for the provincial short-term load forecasting.

2.2. Load Characteristic Dimensionality Reduction Based on Singular Value Decomposition (SVD)

As the number of dimensions of load data increases, the efficiency of load clustering decreases significantly. By reducing the dimensionality of the load characteristics, the efficiency of clustering can be improved and the memory requirements for data storage can be reduced. SVD, as a matrix decomposition method, enables the dimensionality reduction of matrices.

Assume that the n load characteristics of m users form a real matrix $A = [a_1, a_2, \dots, a_m]^T$ of order $m \times n$. The n load characteristics of each user are denoted as $a_k = [a_{k,1}, a_{k,2}, \dots, a_{k,n}]^T$. For matrix A , there exist orthogonal matrices $U \in R^{m \times m}$ and $V \in R^{n \times n}$, such that the following equation holds [52].

$$A = U\Lambda V^T = [u_1, u_2, \dots, u_m] \begin{bmatrix} \Lambda_1 \\ 0 \end{bmatrix} [v_1, v_2, \dots, v_n]^T = [\lambda_1 u_1, \lambda_2 u_2, \dots, \lambda_n u_n] \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} = \sum_{i=1}^n \lambda_i u_i v_i^T \quad (1)$$

where Λ is the diagonal matrix.

The magnitude of the singular value λ_i indicates the importance of the load characteristics. A larger singular value indicates that the feature is more important, while a smaller value indicates that the feature is unimportant and can be ignored. In Formula (1), only the first q ($q < n$) dominant singular values are retained. Then, the matrix A and a_k are reduced to

$$A \approx \sum_{i=1}^q \lambda_i u_i v_i^T = [\lambda_1 u_1 \ \lambda_2 u_2 \ \dots \ \lambda_q u_q] \cdot [v_1^T \ v_2^T \ \dots \ v_q^T]^T \quad (2)$$

$$a_k = [\lambda_1 u_{1,k} \ \lambda_2 u_{2,k} \ \dots \ \lambda_q u_{q,k}] \cdot [v_1^T \ v_2^T \ \dots \ v_q^T]^T \quad (3)$$

From Equations (2) and (3), it can be seen that the coordinate system v_1, v_2, \dots, v_n can be reduced to the low-dimensional coordinate system v_1, v_2, \dots, v_q after neglecting the direction of the small variance of the data variation. Accordingly, the coordinate values of the load characteristic a_k in the low-dimensional coordinate system $\lambda_i u_{i,k}$ can be used to reflect the main characteristics of the load characteristic. In addition, the singular values corresponding to each axis describe the importance of the load characteristic. When clustering coordinates, the higher the singularity value, the more important the corresponding load characteristic is. Therefore, the singular values of the axes are chosen as the weights of the dimensionality reduction indicators and are then normalised.

2.3. K-Means Clustering Algorithm

In this study, the K-means algorithm is used to cluster the loads of several cities in the province based on the main features after dimensionality reduction. The method needs to determine the number of clusters k in advance, and this paper adopts the sum of squared error (SSE) as the criterion for evaluating the effectiveness of clustering and determines the number of clusters accordingly. The SSE metric is defined in the following equation [53].

$$I_{SSE} = \sum_{i=1}^k \sum_{x \in C_i} d^2(c_i, x) \quad (4)$$

where C_i is the class C_i sample. c_i is the cluster centre of the class i sample, and $d^2(c_i, x)$ is the squared Euclidean distance between c_i and sample x . Smaller I_{SSE} means better quality of clustering.

Suppose X is a collection of n metadata with s dimensions, denoted $X = \{x_1, x_2, \dots, x_n\} \in R^s$. The steps for clustering load data are as follows:

- (1) Determine the number of clusters k according to the clustering validity index SSE.
- (2) Randomly select the initial k clustering centres $u_1, u_2, \dots, u_k \in R^s$. Calculate for each data sample the class $lable_s$ it belongs to.

$$lable_s = \operatorname{argmin} \|x^i - u_j\|^2 \quad (5)$$

- (3) For each class j , recalculate the cluster centre of that class:

$$x_j^{t+1} = \frac{1}{|N_i^t|} \sum_{x_j \in lable_i^t} x_j \quad (6)$$

where N_i^t is the number of clustering centres recalculated for the t -th time for class i .

- (4) Update the class centre with the class mean.
- (5) Repeat (3) and (4) until the class centres are unchanged.
- (6) Output the clustering results.

2.4. Load Decomposition Based on VMD Algorithm

Considering the non-linear and non-smooth nature of the load series, the decomposition method is used to decompose the total load for each category. Each type of load is decomposed into multiple IMFs of different frequencies, and then each load sequence component is predicted separately.

Variational empirical modal decomposition (VMD) [54] is a new type of adaptive decomposition algorithm. The algorithm decomposes a complex time series into a number of single-frequency components based on a pre-determined number of decompositions M . The optimal solution of the model is obtained by alternating directional multiplication and iterative updating.

Assume a signal to be decomposed:

$$y(t) = \sum_{m=1}^M v_m(t) = \sum_{m=1}^M A_m(t) \cos[\varphi_m(t)] \tag{7}$$

where $y(t)$ is the original load signal to be decomposed. $v_m(t)$ ($m = 1 \sim M$) is the single frequency signal after load decomposition. M is the number of decompositions. $A_m(t)$ is the amplitude of the signal $v_m(t)$. $\varphi_m(t)$ is the phase angle of $v_m(t)$.

The VMD extracts M modal components when the original signal is non-smooth, such that the sum of the frequency bandwidths of each component is minimised and the sum of each modal component is equal to the original signal. The constraint model is

$$\left\{ \begin{array}{l} \min_{\{\mu_m\}, \{\omega_m\}} \left\{ \sum_m \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \cdot \mu_m(t) \right] e^{-j\omega_m t} \right\|_2^2 \right\} \\ s.t. \sum_m \mu_m = f \end{array} \right. \tag{8}$$

where $\{\mu_m\} = \{\mu_1, \mu_2, \dots, \mu_m\}$ is the decomposed modal component, $\{\omega_m\} = \{\omega_1, \omega_2, \dots, \omega_m\}$ is the decomposed centre frequency, $\delta(t)$ is the shock function, and f is the original load signal.

The Lagrangian multiplier λ and the second-order parametric penalty factor α are used to construct the extended Lagrangian function, and then the alternating direction multiplier method is used to iteratively find the global optimal solution of the objective function. The mathematical model of the augmented Lagrangian function is as follows.

$$L(\{\mu_m\}, \{\omega_m\}, \lambda) = \alpha \sum_m \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \cdot \mu_m(t) \right] e^{-j\omega_m t} \right\|_2^2 + \left\| f(t) - \sum_m \mu_m(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_m \mu_m(t) \right\rangle \tag{9}$$

The optimal solution to the above equation is found by alternately solving the multiplicative updates: μ_m^{n+1} , ω_m^{n+1} , and λ_m^{n+1} . The value of μ_m^{n+1} is expressed as follows:

$$\mu_m^{\Lambda^{n+1}} = \operatorname{argmin}_\alpha \cdot \left\| j\omega \left[(1 + \operatorname{sgn}(\omega + \omega_m)) \cdot \mu_m^\Lambda(\omega + \omega_m) \right] \right\|_2^2 + \left\| f^\Lambda(\omega) - \sum_i \mu_i^\Lambda(\omega) + \frac{\lambda^\Lambda(\omega)}{2} \right\|_2^2 \tag{10}$$

The minimal value of each IMF component from Fourier transform is as follows:

$$\mu_m^{\Lambda^{n+1}}(\omega) = \frac{f^\Lambda(\omega) + \sum_i \mu_i^\Lambda(\omega) + \frac{\lambda^\Lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_m)^2} \tag{11}$$

Similarly, find the central frequency minima as

$$\omega_m^{n+1} = \frac{\int_0^\infty \omega \left| \hat{\mu}_m^\Lambda(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{\mu}_m^\Lambda(\omega) \right|^2 d\omega} \quad (12)$$

2.5. LSTM

Long Short-Term Memory (LSTM) [55] is a type of recurrent neural network model used for processing sequence data such as text, speech, and time series data. The advantage of LSTM models is that they can capture long-term dependencies, which traditional recurrent neural network models cannot achieve. The memory unit in the LSTM model can remember information in the input sequence and pass it on to the next time step for better prediction of future values. In addition, the LSTM model can control the flow of information through gate mechanisms to reduce the problem of vanishing gradients. These advantages make LSTM models widely applicable in areas such as speech recognition and time prediction. The low-frequency component of the load fluctuates smoothly, and accurate prediction can be achieved using only the LSTM model. Therefore, the LSTM model is used to implement the low-frequency component prediction in this study.

The structure of the LSTM [54] consists of three gates: the forgetting gate, the input gate, and the output gate. According to the internal structure of the three gates, the value of the LSTM hidden layer at this moment depends on the joint action of the current moment and the previous moment. The three gates act as three control switches to extract and process the information: delete the historical information that is not useful, retain the information related to the output characteristics, update the state of the hidden layer, and improve the convergence of the model.

2.6. CNN-GRU

The CNN-GRU model is a novel neural network architecture that combines the convolutional neural network (CNN) and gated recurrent unit (GRU) models. The CNN-GRU model is designed to process sequential data, such as time series data, and is particularly well-suited for high-frequency component prediction. The CNN-GRU model consists of two main components: a CNN and a GRU. The CNN is responsible for extracting local features from the input data, while the GRU is used to capture the temporal dependencies in the data. The output of the CNN is fed into the GRU, which then produces time series prediction results. This allows the model to accurately predict high-frequency components in the data, which are often difficult to predict using traditional time series models. Overall, the CNN-GRU model represents a significant advancement in the field of time series prediction and can greatly improve the accuracy and reliability of high-frequency component predictions.

2.6.1. CNN

The convolutional neural network (CNN) is a deep learning model used for image classification, feature extraction, target detection, speech recognition, etc. The CNN mainly consists of multiple convolutional layers, where the convolutional layers are used to extract potential features of the data and pooling layers are used to reduce the size of the feature map. Convolutional layers play a key role in feature extraction. They are responsible for detecting local features in the input data by sliding a set of learnable filters over the input volume and computing the dot product between the filter and the corresponding patch of the input. The result of this operation is a feature map that captures potential fluctuating features in the input data. Overall, the convolutional layer enables the neural network to learn hierarchical representations of the input data that are increasingly abstract and discriminative, thereby improving the accuracy and robustness of the model.

2.6.2. GRU

GRU is a variant of LSTM that uses its specific memory and forgetting structure to model time series dynamically in time. It addresses the phenomenon of gradient disappearance and gradient explosion during the training of loaded time series. Compared with LSTM, GRU reduces the number of gates, which ensures both the accuracy of load prediction and the training time.

The GRU has two gates. The GRU integrates the forgetting gate and the input gate from the LSTM to form a new update gate. The output gate in the LSTM is replaced by a reset gate, which picks the state at the previous moment and writes it to the candidate set at this moment [54].

2.7. Performance Evaluation

In this paper, the error evaluation index root mean square error (RMSE) [56] and mean absolute percentage error (MAPE) [57] are used to evaluate the accuracy of the prediction model. The mathematical expressions are shown in Equations (13) and (14).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2} \tag{13}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - p_i}{y_i} \right| \times 100\% \tag{14}$$

where N is the number of samples. y_i is the true value of the load data. p_i is the load forecast value.

3. Case Analysis

To verify the application effectiveness of the proposed ensemble forecasting method based on an improved forecasting strategy, simulations were carried out using actual load data. Actual load data from 15 July to 15 August 2017 were collected from 10 cities in a province in northwest China to forecast the load on August 16. The data sampling period was 15 min, and a total of 96 points were sampled in 1 day. The typical daily load curves of 10 cities are shown in Figure 2.

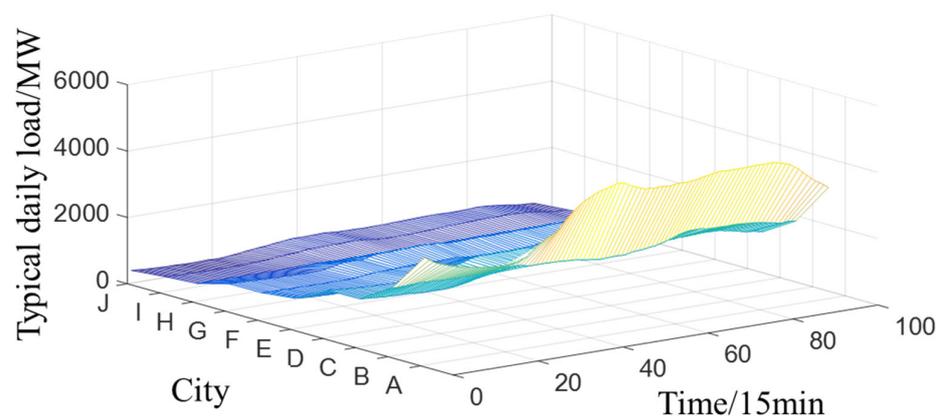


Figure 2. Typical daily load curve of 10 cities in a province.

3.1. Load Characteristic Extraction

In 2005, the National Grid Limited issued the “In-depth Requirements for Load Characteristics Study Content and Interpretation of Indicators”. In this document, the definitions of various types of load characteristic indicators are specified. In this paper, seven types of daily load characteristic indicators are selected and defined, as shown in Table 1.

Table 1. Load characteristic index definition.

Indicators	Calculation Formula
Maximum daily load	$\alpha_1 = P_{\max}$
Minimum daily load	$\alpha_2 = P_{\min}$
Average daily load	$\alpha_3 = P_{\text{sum}}/n$
Daily load factor	$\alpha_4 = (P_{\text{sum}}/n)/P_{\max}$
Minimum daily load factor	$\alpha_5 = P_{\min}/P_{\max}$
Daily peak-to-valley difference	$\alpha_6 = P_{\max} - P_{\min}$
Daily peak-to-valley ratio	$\alpha_7 = (P_{\max} - P_{\min})/P_{\max}$

In the table, P_{\max} , P_{\min} , and P_{sum} are the maximum daily load, minimum daily load, and total daily load, respectively.

Based on the above-defined equations of load characteristics, the load characteristics of each municipality are calculated, and the singular values of each load characteristic are calculated using the SVD. Seven types of load characteristic singular values and their Pareto diagrams are shown in Figure 3.

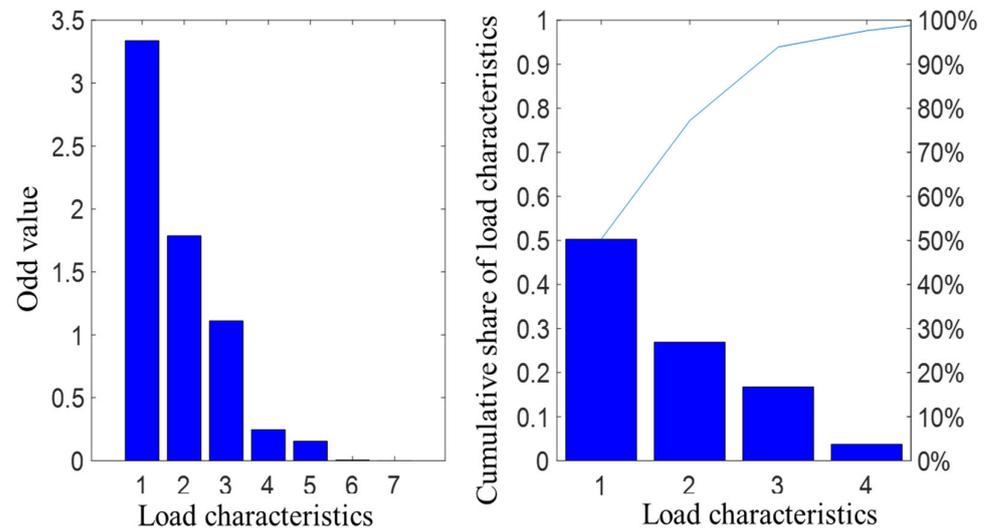


Figure 3. The singular values of load characteristics and their Pareto graphs.

As can be seen from Figure 3, the first three load characteristics have larger singularity values and the last four have smaller ones. According to the SVD principle, the larger the singularity value, the more information the load characteristic reflects and the more important the load characteristic is. The Pareto diagram shows that the cumulative contribution of the first three load characteristics is high, reaching 93.92%. It can be observed that the first three load characteristics contain most of the load information, so the first three load characteristics are extracted as the main characteristics.

3.2. Analysis of Clustering Results for Ten City Loads

Before clustering, the number of clusters needs to be determined based on the SSE index to ensure the clustering effect, as shown in Figure 4.

It can be seen that when the number of clusters is four, the SSE index is relatively small. The number of clusters is chosen to be four, as each category contains a certain number of cities after clustering. Using the load characteristic matrix obtained after SVD dimensionality reduction, the K-means algorithm is applied to achieve clustering of different urban loads. The clustering results after the SVD dimensionality reduction are shown in Figure 5.

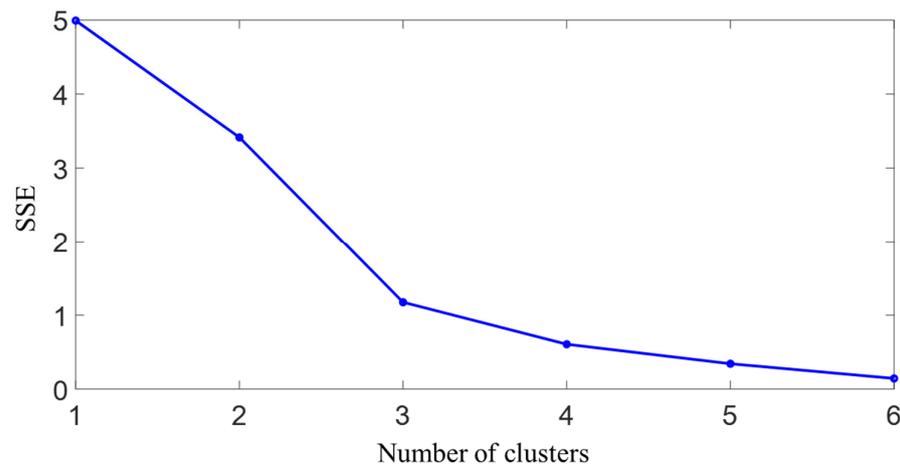


Figure 4. Clustering effectiveness index after dimensionality reduction.

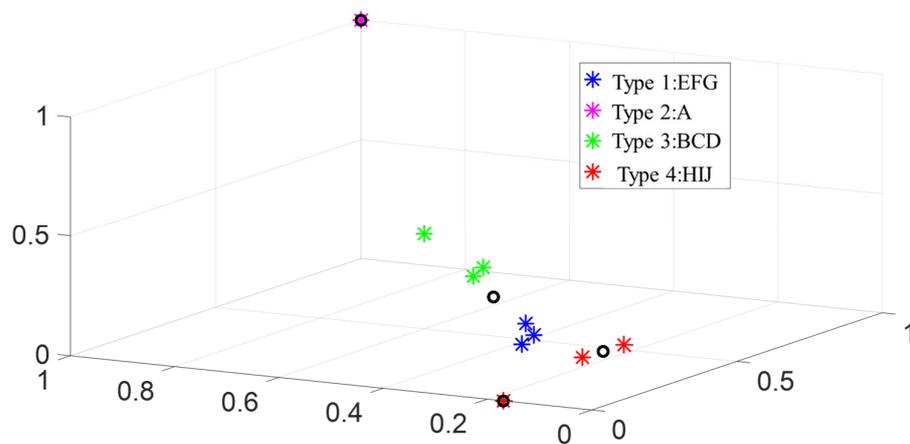


Figure 5. Load clustering results of 10 cities in a province.

The load characteristics of each class of cities after clustering are shown in Figure 6. In this figure, for each city load, each colour bar represents a characteristic of that city load. A total of seven load characteristics are plotted for each city load. The seven colour bars represent, from left to right, the load characteristics: maximum daily load, minimum daily load, average daily load, daily load factor, minimum daily load factor, daily peak-to-valley variance, and daily peak-to-valley ratio. The seven load characteristics are defined in Table 1 above.

As can be seen from Figure 6, the load characteristics of cities clustered into the same category are similar, and the load characteristics of cities in different categories vary considerably. The first category contains three cities, E–G. The second category contains only city A, which is the capital city of the province and has the largest electricity load. The third category contains cities B–D. The fourth category contains three cities, H–J.

Figure 7 shows the gross regional product (GDP) of the 10 municipalities in the province.

Electricity load is used as a barometer to measure economic and social development, and the results of load clustering are correlated with the economic development of each region.

The comparative analysis revealed that, with the exception of cities C and G, the clustering results corresponded to the regional GDP for all cities. Geographically, cities C and G are adjacent to the provincial capital, located to the east and west of the provincial capital, respectively. The two cities’ GDP, climate, and population statistics are extremely comparable. This demonstrates the effectiveness of the proposed clustering method based on the extraction of load principal characteristics. Clustering the loads of cities with

similar characteristics into the same class simplifies and reduces the randomness of load forecasting.

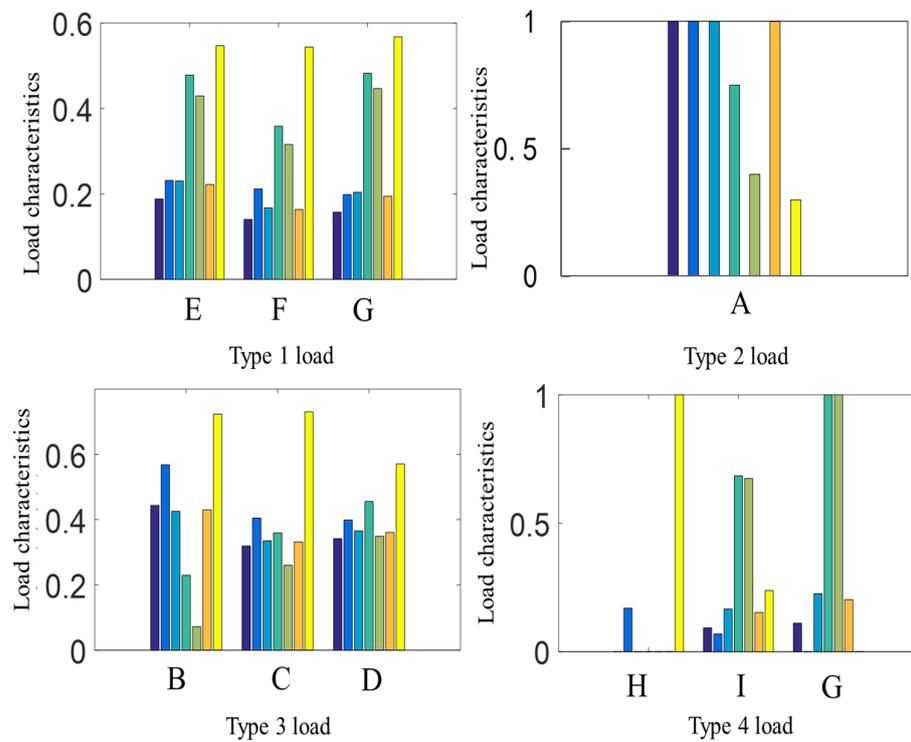


Figure 6. Load characteristics of various cities after clustering.

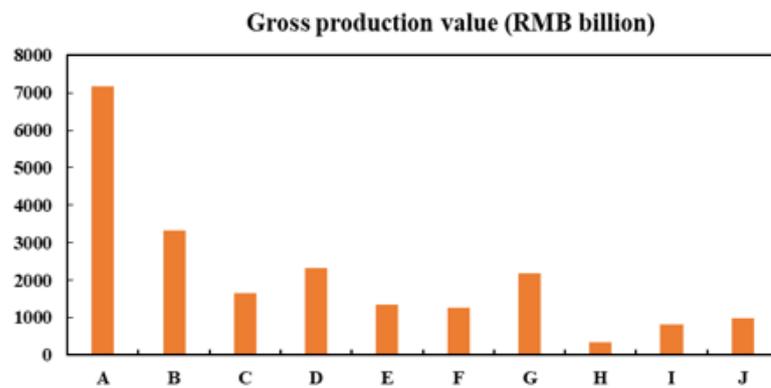


Figure 7. Regional GDP of 10 cities in a province.

3.3. Analysis of the Results of the Frequency Domain Decomposition of the Load

VMD was carried out for each type of load after clustering, and the number of decompositions is four. Due to space limitations, only the results of the fourth type of load decomposition are shown in Figure 8.

As can be seen in Figure 8, each type of load is decomposed into components with a smooth sequence and a single frequency. The intrinsic variation pattern of the load is extracted, reducing the complexity of the original load sequence. The effect between different components at the characteristic scale is greatly reduced, providing a simpler sequence for the fine prediction of load.

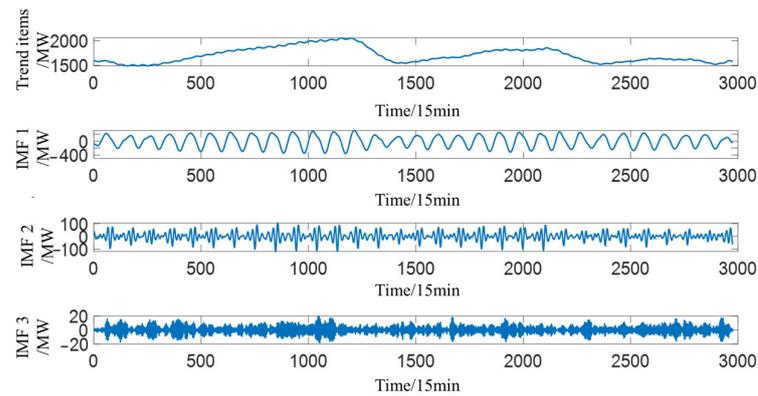


Figure 8. VMD of the second cluster load.

3.4. Analysis of Prediction Results Based on a Proposed Ensemble Model

The proposed ensemble forecast model is used to forecast four types of loads. Firstly, the low-frequency component (trend term, IMF1, IMF2) and the high-frequency component (IMF3) are predicted separately, and then the predictions for each component are superimposed.

For each type of load, the prediction results of high-frequency components and low-frequency components are superimposed to obtain the final prediction results, as shown in Figure 9. The forecast errors RMSE and MAPE for each type of load are shown in Table 2.

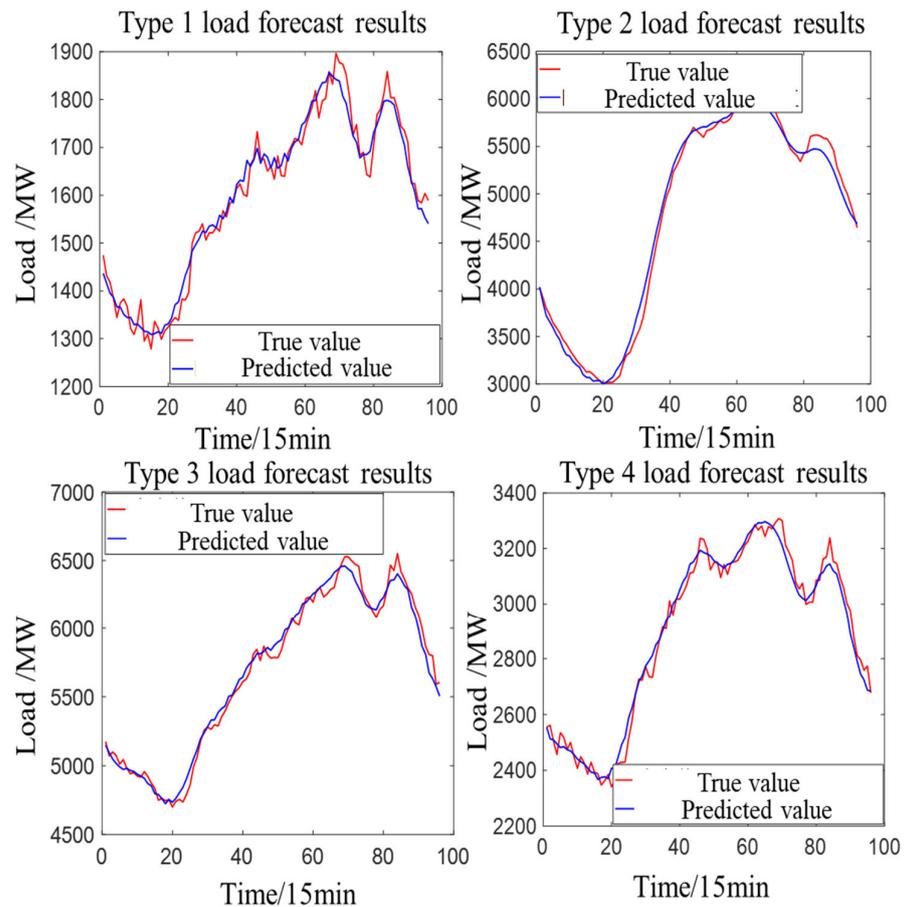


Figure 9. The prediction results of the proposed ensemble model.

Table 2. Error of the proposed ensemble prediction algorithm.

Error	Type 1	Type 2	Type 3	Type 4
RMSE	30.50	94.12	66.69	29.41
MAPE	1.12%	1.76%	0.95%	1.47%

As seen in Table 2, the greater the peak-to-valley difference, the greater the root mean square error (RMSE) of the load forecast. The average absolute percentage error MAPE is 1.12% for the first category of load, 1.76% for the second category, 0.95% for the third category, and 1.47% for the fourth category. The mean absolute percentage error MAPE for each category of load did not exceed 2%, and the accuracy of the predictions is high.

3.5. Comparison of the Proposed Forecast and Baseline Schemes

3.5.1. Five Comparative Baseline Schemes

In this section, five forecasting schemes are proposed. All five schemes are used to compare and validate the effectiveness of the proposed ensemble forecasting method with improved forecasting strategies.

Scheme 1: Based on the prediction strategy proposed in this paper, LSTM is used to predict the decomposed components. That is, “reduced dimensional clustering—decomposition—LSTM”.

Scheme 2: The support vector regression (SVR) is applied directly to each class of load after clustering for prediction, i.e., “reduced dimensional clustering—SVR”.

Scheme 3: The clustered loads of each class are predicted directly using BP neural networks, i.e., “reduced dimensional clustering—BPNN”.

Scheme 4: The LSTM is used directly to predict the load for each class after clustering, i.e., “reduced dimensional clustering—LSTM”.

Scheme 5: Direct forecasting of the total load of the 10 municipalities in the province using LSTM. This is known as the “LSTM”.

A clearer presentation of the six prediction schemes is given in Table 3. The values of each model parameter are shown in Table 4.

Table 3. Comparison of six prediction schemes.(√ indicates that the policy is adopted, × indicates that the policy is not adopted).

Scheme	Whether Reduced Dimensional Clustering Is Adopted	Decomposition Is Adopted or Not	Prediction Model for Each Component	Prediction Models for Various Types of Loads When Decomposition Is Not Used	Prediction Model for the Overall Load of 10 Cities
Proposed	√	√	LSTM, CNN-GRU	-	-
Scheme 1	√	√	LSTM	-	-
Scheme 2	√	×	-	SVR	-
Scheme 3	√	×	-	BPNN	-
Scheme 4	√	×	-	LSTM	-
Scheme 5	×	×	-	-	LSTM

Table 4. The value of each model parameter.

Scheme	Algorithm	Parameter Value
Proposed	LSTM	num_layers = 3, units = 128, activation = 'relu', epochs = 150, optimizer = 'adam'
	CNN-GRU	CNN: num_layers = 1, filters = 96, kernel_size = 2, padding = valid
		GRU: num_layers = 3, units = 128, dropout = 0.02 epochs = 150, optimizer = 'adam'
Scheme 1	LSTM	num_layers = 3, units = 128, activation = 'relu', epochs = 150
Scheme 2	SVR	Kernel = rbf, C = 1
Scheme 3	BPNN	num_layers = 3, units = 128, epochs = 150, optimizer = 'adam'
Scheme 4	LSTM	num_layers = 3, units = 128, activation = 'relu', epochs = 150, optimizer = 'adam'
Scheme 5	LSTM	num_layers = 3, units = 128, activation = 'relu', epochs = 150, optimizer = 'adam'

In the following study, the prediction performance of the schemes using the reduced dimensional clustering approach will be compared first (i.e., proposed and Schemes 1–4). In detail, the prediction performance of these five prediction schemes for each type of load is compared separately. The forecast results for each type of load under the five schemes are then integrated to obtain the total load forecast results for the province for the coming day under the five prediction schemes. Finally, the forecast results of these five schemes are compared with Scheme 5.

3.5.2. Comparison of Predictive Results for Various Types of Loads (Proposed Scheme and Schemes 1–4)

Figures 10–13 show a comparison of the results of the four types of load prediction using the first four schemes and the proposed scheme.

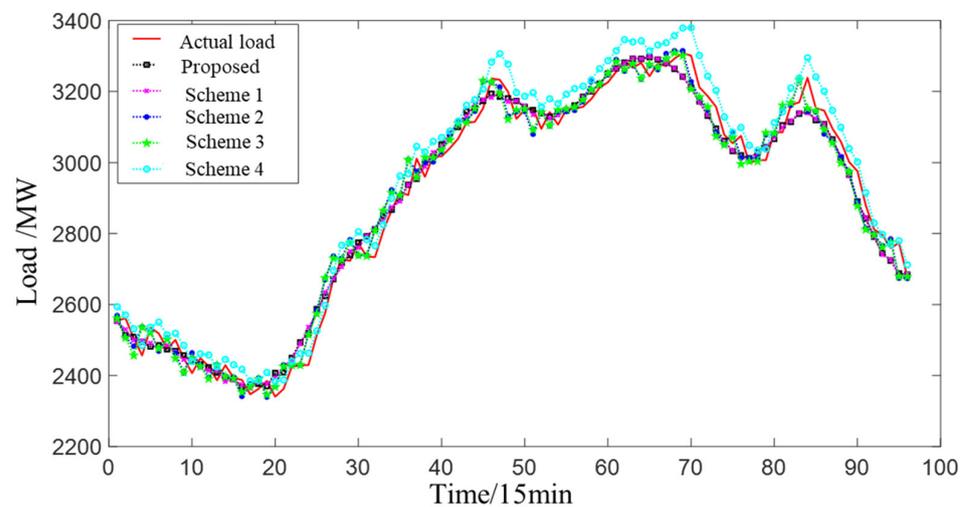


Figure 10. Forecast results for the first type of load using different forecast schemes.

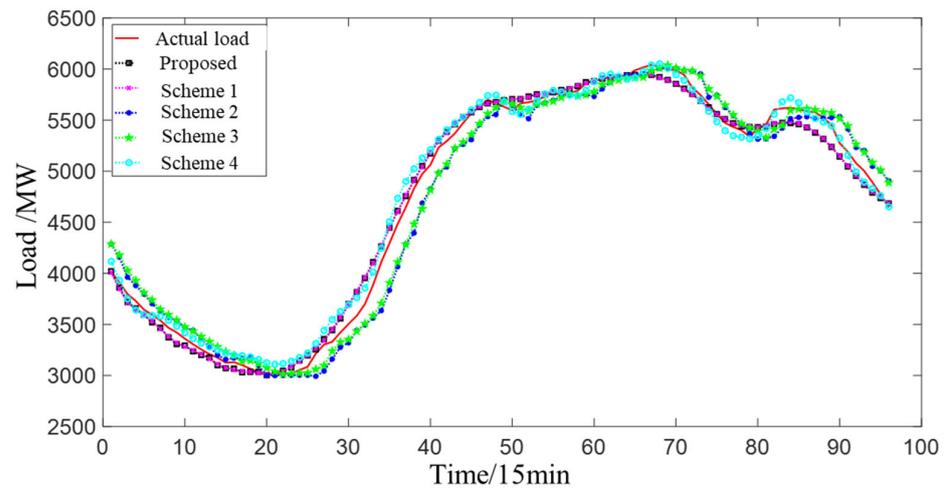


Figure 11. Forecast results for the second type of load using different forecast schemes.

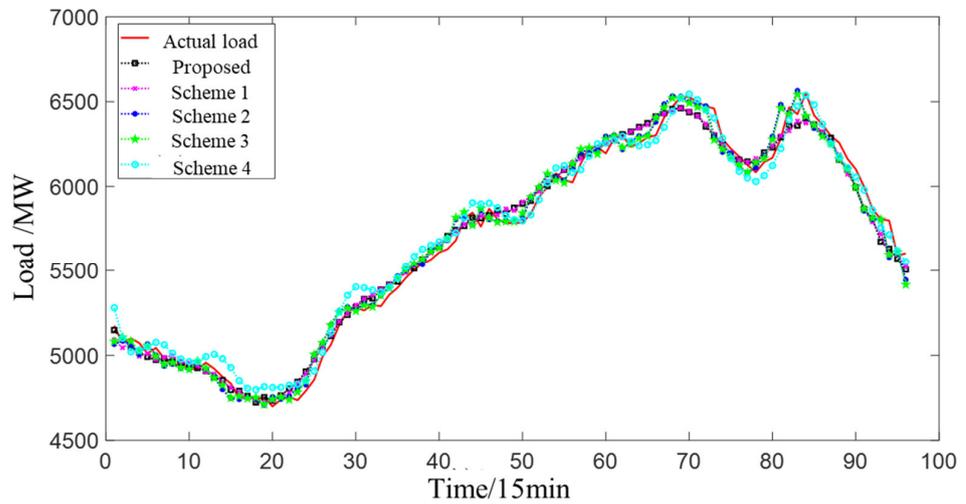


Figure 12. Forecast results for the third type of load using different forecast schemes.

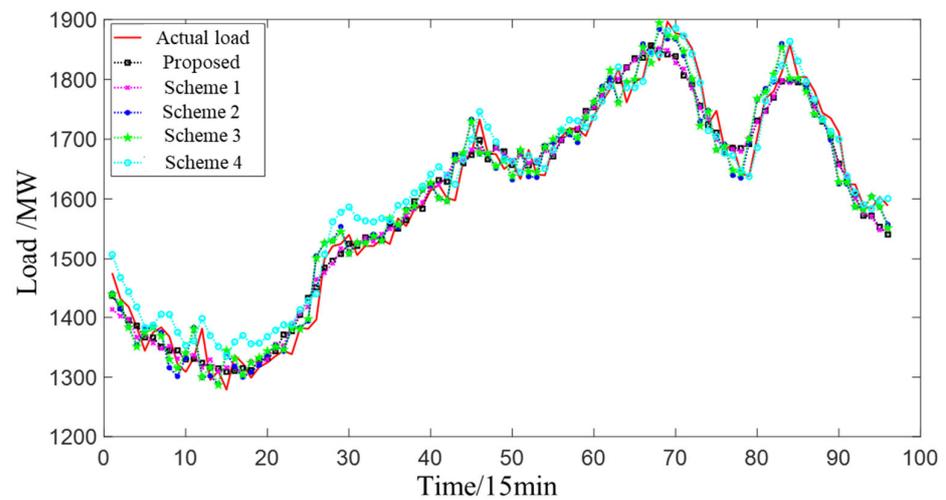


Figure 13. Forecast results for the fourth type of load using different forecast schemes.

A comparison of the error results for the four types of load forecast is shown in Table 5. Figure 14 shows a histogram of the forecast error for each type of load under various schemes.

Table 5. Comparison of prediction errors of different prediction schemes.

Model	Error Indicators	Type 1	Type 2	Type 3	Type 4
Proposed	MAPE	1.12%	1.76%	0.95%	1.47%
Scheme 1	MAPE	1.19%	1.84%	0.97%	1.56%
Scheme 2	MAPE	1.32%	3.04%	0.98%	1.79%
Scheme 3	MAPE	1.30%	2.96%	0.99%	1.77%
Scheme 4	MAPE	1.29%	1.96%	0.98%	1.59%
Proposed	RMSE	30.50	94.12	66.69	29.41
Scheme 1	RMSE	31.04	98.84	70.37	31.12
Scheme 2	RMSE	45.69	172.80	71.99	35.98
Scheme 3	RMSE	45.16	163.10	71.79	35.52
Scheme 4	RMSE	44.99	107.81	71.47	32.02

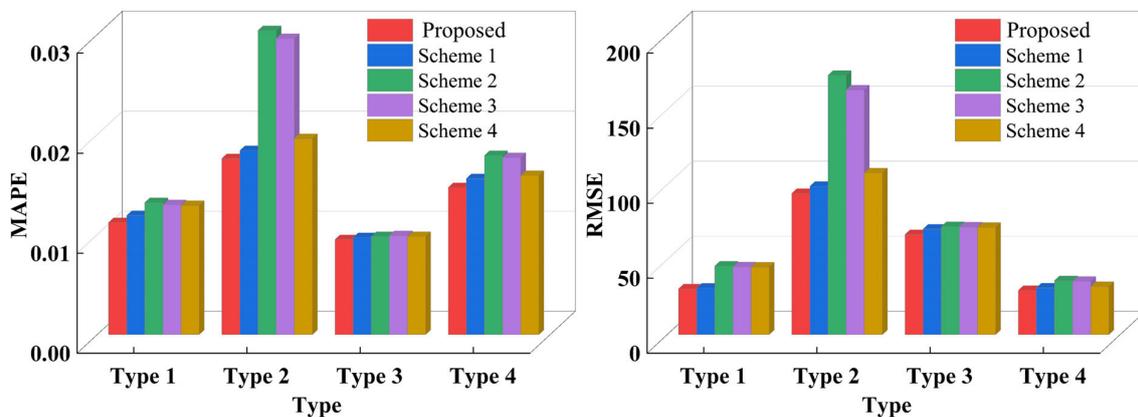


Figure 14. Histogram of forecast errors for each type of load under various schemes.

It can be seen that both the MAPE and RMSE values of the proposed method are lower. The dimensionality reduction–clustering and decomposition strategy is used in Scheme 1, but only the LSTM algorithm is used for the prediction of each component. The MAPE prediction errors for all types of loads using Scheme 1 are 2.11–6.25% higher than the proposed method. The RMSE prediction errors for all types of loads using Scheme 1 are 1.77~5.81% higher than the proposed method. This is due to the use of CNN-GRU to predict the more non-linear stochastic component separately, which improves the prediction accuracy of the high-frequency component. The results verify the effectiveness and accuracy of CNN-GRU for predicting the high-frequency components, and the ensemble prediction can achieve accurate prediction for the characteristics of various load components.

Schemes 2–4 all use a method of forecasting each class of load after clustering the loads. Compared to the proposed forecasting scheme and Scheme 1, the forecasting error for each class of load increases to varying degrees. This shows that the decomposition of the loads can improve the prediction accuracy. Comparing Schemes 2–4, the error index MAPE for Scheme 4 is smaller than that of Schemes 2 and 3. The results show that LSTM has a higher prediction accuracy than the SVR and BPNN algorithms, which reflects the effectiveness of the LSTM algorithm.

3.5.3. Comparison of the Predicted Results of the Total Provincial Load (Proposed Scheme and Schemes 1–5)

For the proposed scheme and Schemes 1–4, the predicted results of each type of load are superimposed to obtain the predicted load for the entire province on August 16th under

each scheme. The direct forecast of the province’s load on August 16 using Scheme 5 yields the forecast of the province’s load on August 16 under Scheme 5. The forecast results of these six schemes are compared with the actual load of the province on August 16, and the forecast result curves are shown in Figure 15. A comparison of the forecast errors is shown in Table 6.

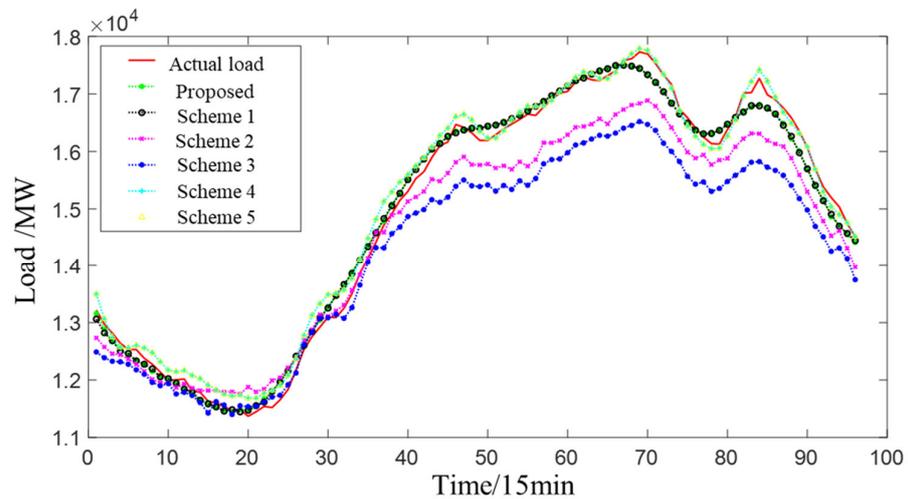


Figure 15. Comparison of forecast results for the total provincial load.

Table 6. Comparison of total load forecast errors in the province.

Model	MAPE	RMSE
Proposed	1.09%	195.40
Scheme 1	1.11%	195.87
Scheme 2	2.97%	535.20
Scheme 3	4.25%	808.90
Scheme 4	1.16%	198.56
Scheme 5	1.30%	214.97

The comparison shows that the proposed prediction scheme is able to achieve higher prediction accuracy. The table also shows that the prediction error of Scheme 5 is higher than that of Scheme 4. The MAPE error indicator value of Scheme 5 is 12.07% higher than that of Scheme 4. The RMSE error indicator of Scheme 5 is 8.26% higher than that of Scheme 4. It is inferred that the use of the clustering method can effectively improve the prediction accuracy of provincial load. The effectiveness of frequency domain decomposition of load sequences in improving prediction accuracy can be readily seen by comparing the error metrics of Schemes 1 and 4. The proposed scheme has a 1.8% reduction in the MAPE error metric and a 0.23% reduction in the RMSE error metric compared to Scheme 1. Thus, the validity of the ensemble model is verified. After a comprehensive comparison and analysis, the proposed forecasting scheme in this paper has significant advantages, and the scheme can achieve accurate forecasting of provincial loads.

4. Conclusions

Considering the high demand for short-term load forecasting accuracy in power dispatch, an ensemble forecasting algorithm based on an improved forecasting strategy is proposed in this paper. The proposed model is validated with actual load data from a province in northwest China, leading to the following conclusions:

- (1) The proposed load clustering based on principal characteristic extraction improves the efficiency and quality of clustering. At the same time, the randomness of the load sequence for each class of users is reduced, as users with similar load characteristics

- are clustered into the same class. This reduces the complexity of load prediction and facilitates the improvement of prediction accuracy.
- (2) After VMD, the load sequence has a single frequency and is not prone to modal confusion. The VMD method can fully exploit the implicit features of the data and avoid the mutual interference between different local features. It lays the foundation for the accurate prediction of load sequences.
 - (3) The proposed prediction model fully considers the fluctuation characteristics of high-frequency components and low-frequency components and fully utilizes the respective advantages of the LSTM and CNN-GRU models. At the same time, the total prediction error after the superposition of each class of load can offset the prediction error of each class to a certain extent, so that the prediction error after superposition can be further reduced. The results show that the proposed model can achieve better prediction results, and the proposed model can not only predict the change trend of electric load but also predict the local details. The forecast error (RMSE) of the proposed scheme is 0.23%, 63.49%, 75.84%, 1.59%, and 9.10% lower than that of the benchmark scheme.
 - (4) The CNN-GRU component in the proposed model can better extract the local features of the high-frequency component, which ensures that the proposed model can track the load trend more accurately. Compared with the scheme that only uses LSTM to predict each frequency component, the proposed model increases the complexity of the model, but the prediction accuracy is improved.

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

Funding: This work was funded by the Shaanxi Province Science and Technology Department, grant number 2022JM-208.

Data Availability Statement: Not applicable.

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

References

1. Wu, Z.; Mu, Y.; Deng, S.; Li, Y. Spatial–temporal short-term load forecasting framework via K-shape time series clustering method and graph convolutional networks. *Energy Rep.* **2022**, *8*, 8752–8766. [[CrossRef](#)]
2. Yang, Y.; Zhou, H.; Wu, J.; Liu, C.-J.; Wang, Y.-G. A novel decompose-cluster-feedback algorithm for load forecasting with hierarchical structure. *Int. J. Electr. Power Energy Syst.* **2022**, *142*, 108249. [[CrossRef](#)]
3. Yang, Y.; Wang, Z.; Gao, Y.; Wu, J.; Zhao, S.; Ding, Z. An effective dimensionality reduction approach for short-term load forecasting. *Electr. Power Syst. Res.* **2022**, *210*, 108150. [[CrossRef](#)]
4. Haque, A.; Rahman, S. Short-term electrical load forecasting through heuristic configuration of regularized deep neural network. *Appl. Soft Comput.* **2022**, *122*, 108877. [[CrossRef](#)]
5. Deng, X.; Ye, A.; Zhong, J.; Xu, D.; Yang, W.; Song, Z.; Zhang, Z.; Guo, J.; Wang, T.; Tian, Y.; et al. Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Rep.* **2022**, *8*, 8661–8674. [[CrossRef](#)]
6. Han, X.; Su, J.; Hong, Y.; Gong, P.; Zhu, D. Mid- to Long-Term Electric Load Forecasting Based on the EMD–Isomap–Adaboost Model. *Sustainability* **2022**, *14*, 7608. [[CrossRef](#)]
7. Lee, G.-C. Regression-Based Methods for Daily Peak Load Forecasting in South Korea. *Sustainability* **2022**, *14*, 3984. [[CrossRef](#)]
8. Zhao, X.; Shen, B.; Lin, L.; Liu, D.; Yan, M.; Li, G. Residential Electricity Load Forecasting Based on Fuzzy Cluster Analysis and LSSVM with Optimization by the Fireworks Algorithm. *Sustainability* **2022**, *14*, 1312. [[CrossRef](#)]
9. Li, M.; Xie, X.; Zhang, D. Improved Deep Learning Model Based on Self-Paced Learning for Multiscale Short-Term Electricity Load Forecasting. *Sustainability* **2022**, *14*, 188. [[CrossRef](#)]
10. Son, N. Comparison of the Deep Learning Performance for Short-Term Power Load Forecasting. *Sustainability* **2022**, *13*, 12493. [[CrossRef](#)]
11. Aslam, S.; Ayub, N.; Farooq, U.; Alvi, M.J.; Albogamy, F.R.; Rukh, G.; Haider, S.I.; Azar, A.T.; Bukhsh, R. Towards Electric Price and Load Forecasting Using CNN-Based Ensembler in Smart Grid. *Sustainability* **2021**, *13*, 12653. [[CrossRef](#)]

12. Wang, H.; Zhang, N.; Du, E.; Yan, J.; Han, S.; Liu, Y. A comprehensive review for wind, solar, and electrical load forecasting methods. *Glob. Energy Interconnect.* **2022**, *5*, 9–30. [[CrossRef](#)]
13. Yang, D.; Guo, J.-E.; Li, Y.; Sun, S.; Wang, S. Short-term load forecasting with an improved dynamic decomposition-reconstruction-ensemble approach. *Energy* **2023**, *263*, 125609. [[CrossRef](#)]
14. Geng, G.; He, Y.; Zhang, J.; Qin, T.; Yang, B. Short-Term Power Load Forecasting Based on PSO-Optimized VMD-TCN-Attention Mechanism. *Energies* **2023**, *16*, 4616. [[CrossRef](#)]
15. Matrenin, P.; Safaraliev, M.; Dmitriev, S.; Kokin, S.; Ghulomzoda, A.; Mitrofanov, S. Medium-term load forecasting in isolated power systems based on ensemble machine learning models. *Energy Rep.* **2022**, *8*, 612–618. [[CrossRef](#)]
16. Wang, S.; Wang, S.; Wang, D. Combined probability density model for medium term load forecasting based on quantile regression and kernel density estimation. *Energy Procedia* **2019**, *158*, 6446–6451. [[CrossRef](#)]
17. Ahmad, T.; Zhang, H. Novel deep supervised ML models with feature selection approach for large-scale utilities and buildings short and medium-term load requirement forecasts. *Energy* **2020**, *209*, 118477. [[CrossRef](#)]
18. Kalhori, M.R.N.; Emami, I.T.; Fallahi, F.; Tabarzadi, M. A data-driven knowledge-based system with reasoning under uncertain evidence for regional long-term hourly load forecasting. *Appl. Energy* **2022**, *314*, 118975. [[CrossRef](#)]
19. Xiang, W.; Xu, P.; Fang, J.; Zhao, Q.; Gu, Z.; Zhang, Q. Multi-dimensional data-based medium- and long-term power-load forecasting using double-layer CatBoost. *Energy Rep.* **2022**, *8*, 8511–8522. [[CrossRef](#)]
20. Guan, Y.; Li, D.; Xue, S.; Xi, Y. Feature-fusion-kernel-based Gaussian process model for probabilistic long-term load forecasting. *Neurocomputing* **2021**, *426*, 174–184. [[CrossRef](#)]
21. Wen, Z.; Xie, L.; Fan, Q.; Feng, H. Long term electric load forecasting based on TS-type recurrent fuzzy neural network model. *Electr. Power Syst. Res.* **2020**, *179*, 106106. [[CrossRef](#)]
22. Xie, J.; Zhong, Y.; Xiao, T.; Wang, Z.; Zhang, J.; Wang, T.; Schuller, B.W. A multi-information fusion model for short term load forecasting of an architectural complex considering spatio-temporal characteristics. *Energy Build.* **2022**, *277*, 112566. [[CrossRef](#)]
23. Deng, S.; Chen, F.; Wu, D.; He, Y.; Ge, H.; Ge, Y. Quantitative combination load forecasting model based on forecasting error optimization. *Comput. Electr. Eng.* **2022**, *101*, 108125. [[CrossRef](#)]
24. Ning, Y.; Zhao, R.; Wang, S.; Yuan, B.; Wang, Y.; Zheng, D. Probabilistic short-term power load forecasting based on B-SCN. *Energy Rep.* **2022**, *8*, 646–655. [[CrossRef](#)]
25. Kim, H.; Jeong, J.; Kim, C. Daily Peak-Electricity-Demand Forecasting Based on Residual Long Short-Term Network. *Mathematics* **2022**, *10*, 4486. [[CrossRef](#)]
26. Yang, W.; Shi, J.; Li, S.; Song, Z.; Zhang, Z.; Chen, Z. A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior. *Appl. Energy* **2022**, *307*, 118197. [[CrossRef](#)]
27. Nie, Y.; Jiang, P.; Zhang, H. A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting. *Appl. Soft Comput.* **2020**, *97*, 106809. [[CrossRef](#)]
28. Xiao, L.; Shao, W.; Yu, M.; Ma, J.; Jin, C. Research and application of a combined model based on multi-objective optimization for electrical load forecasting. *Energy* **2017**, *119*, 1057–1074. [[CrossRef](#)]
29. Wang, B.; Tai, N.-L.; Zhai, H.-Q.; Ye, J.; Zhu, J.-D.; Qi, L.-B. A new ARMAX model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting. *Electr. Power Syst. Res.* **2008**, *78*, 1679–1685. [[CrossRef](#)]
30. Amaral, L.F.; Souza, R.C.; Stevenson, M. A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *Int. J. Forecast.* **2008**, *24*, 603–615. [[CrossRef](#)]
31. Ibrahim, B.; Rabelo, L.; Gutierrez-Franco, E.; Clavijo-Buritica, N. Machine Learning for Short-Term Load Forecasting in Smart Grids. *Energies* **2022**, *15*, 8079. [[CrossRef](#)]
32. Ma, D.; Wu, R.; Li, Z.; Cen, K.; Gao, J.; Zhang, Z. A new method to forecast multi-time scale load of natural gas based on augmentation data-machine learning model. *Chin. J. Chem. Eng.* **2022**, *48*, 166–175. [[CrossRef](#)]
33. Nguyen, V.H.; Besanger, Y.; Tran, Q.T. Self-updating machine learning system for building load forecasting—Method, implementation and case-study on COVID-19 impact. *Sustain. Energy Grids Netw.* **2022**, *32*, 100873. [[CrossRef](#)]
34. 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](#)]
35. Bian, H.; Wang, Q.; Xu, G.; Zhao, X. Load forecasting of hybrid deep learning model considering accumulated temperature effect. *Energy Rep.* **2022**, *8*, 205–215. [[CrossRef](#)]
36. Deepanraj, B.; Senthilkumar, N.; Jarin, T.; Gurel, A.E.; Sundar, L.S.; Anand, A.V. Intelligent wild geese algorithm with deep learning driven short term load forecasting for sustainable energy management in microgrids. *Sustain. Comput. Informatics Syst.* **2022**, *36*, 100813. [[CrossRef](#)]
37. Luo, J.; Hong, T.; Gao, Z.; Fang, S.-C. A robust support vector regression model for electric load forecasting. *Int. J. Forecast.* **2022**, *39*, 1005–1020. [[CrossRef](#)]
38. Liu, H.; Tang, Y.; Pu, Y.; Mei, F.; Sidorov, D. Short-term Load Forecasting of Multi-Energy in Integrated Energy System Based on Multivariate Phase Space Reconstruction and Support Vector Regression Mode. *Electr. Power Syst. Res.* **2022**, *210*, 108066. [[CrossRef](#)]
39. Pham, M.-H.; Nguyen, M.-N.; Wu, Y.-K. A Novel Short-Term Load Forecasting Method by Combining the Deep Learning with Singular Spectrum Analysis. *IEEE Access* **2021**, *9*, 73736–73746. [[CrossRef](#)]

40. Hu, Y.; Li, J.; Hong, M.; Ren, J.; Lin, R.; Liu, Y.; Liu, M.; Man, Y. Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—A case study of papermaking process. *Energy* **2019**, *170*, 1215–1227. [[CrossRef](#)]
41. Fazlipour, Z.; Mashhour, E.; Joorabian, M. A deep model for short-term load forecasting applying a stacked autoencoder based on LSTM supported by a multi-stage attention mechanism. *Appl. Energy* **2022**, *327*, 120063. [[CrossRef](#)]
42. Ren, H.; Li, Q.; Wu, Q.; Zhang, C.; Dou, Z.; Chen, J. Joint forecasting of multi-energy loads for a university based on copula theory and improved LSTM network. *Energy Rep.* **2022**, *8*, 605–612. [[CrossRef](#)]
43. 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](#)]
44. Xie, K.; Yi, H.; Hu, G.; Li, L.; Fan, Z. Short-term power load forecasting based on Elman neural network with particle swarm optimization. *Neurocomputing* **2020**, *416*, 136–142. [[CrossRef](#)]
45. Guo, X.; Zhao, Q.; Zheng, D.; Ning, Y.; Gao, Y. A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. *Energy Rep.* **2020**, *6*, 1046–1053. [[CrossRef](#)]
46. Javed, U.; Ijaz, K.; Jawad, M.; Khosa, I.; Ansari, E.A.; Zaidi, K.S.; Rafiq, M.N.; Shabbir, N. A novel short receptive field based dilated causal convolutional network integrated with Bidirectional LSTM for short-term load forecasting. *Expert Syst. Appl.* **2022**, *205*, 117689. [[CrossRef](#)]
47. Arastehfar, S.; Matinkia, M.; Jabbarpour, M.R. Short-term residential load forecasting using Graph Convolutional Recurrent Neural Networks. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105358. [[CrossRef](#)]
48. Yue, W.; Liu, Q.; Ruan, Y.; Qian, F.; Meng, H. A prediction approach with mode decomposition-recombination technique for short-term load forecasting. *Sustain. Cities Soc.* **2022**, *85*, 104034. [[CrossRef](#)]
49. Zhang, Q.; Wu, J.; Ma, Y.; Li, G.; Ma, J.; Wang, C. Short-term load forecasting method with variational mode decomposition and stacking model fusion. *Sustain. Energy Grids Netw.* **2022**, *30*, 100622. [[CrossRef](#)]
50. Koivisto, M.; Heine, P.; Mellin, I.; Lehtonen, M. Clustering of Connection Points and Load Modeling in Distribution Systems. *IEEE Trans. Power Syst.* **2013**, *28*, 1255–1265. [[CrossRef](#)]
51. Zhong, S.; Tam, K.-S. Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain. *IEEE Trans. Power Syst.* **2015**, *30*, 2434–2441. [[CrossRef](#)]
52. Falini, A. A review on the selection criteria for the truncated SVD in Data Science applications. *J. Comput. Math. Data Sci.* **2022**, *5*, 100064. [[CrossRef](#)]
53. Jenssen, R.; Eltoft, T. A new information theoretic analysis of sum-of-squared-error kernel clustering. *Neurocomputing* **2008**, *72*, 23–31. [[CrossRef](#)]
54. Zhao, L.; Li, Z.; Qu, L.; Zhang, J.; Teng, B. A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China. *Ocean Eng.* **2023**, *276*, 114136. [[CrossRef](#)]
55. Wang, K.; Du, H.; Jia, R.; Jia, H. Performance Comparison of Bayesian Deep Learning Model and Traditional Bayesian Neural Network in Short-Term PV Interval Prediction. *Sustainability* **2022**, *14*, 12683. [[CrossRef](#)]
56. Hu, X.; Li, K.; Li, J.; Zhong, T.; Wu, W.; Zhang, X.; Feng, W. Load forecasting model consisting of data mining based orthogonal greedy algorithm and long short-term memory network. *Energy Rep.* **2022**, *8*, 235–242. [[CrossRef](#)]
57. Hu, H.; Xia, X.; Luo, Y.; Zhang, C.; Nazir, M.S.; Peng, T. Development and application of an evolutionary deep learning framework of LSTM based on improved grasshopper optimization algorithm for short-term load forecasting. *J. Build. Eng.* **2022**, *57*, 104975. [[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.