Electric Power Systems Research

Short-term power load forecasting based on multi-level decomposition and improved codec

--Manuscript Draft--

Manuscript Number:	EPSR-D-25-00723
Article Type:	Research Paper
Keywords:	Boruta algorithm; CEEMDAN; VMD; Depth-separable volume-spatial channel attention network; Adaptive Wavelet Convolutional Neural Network; Temporal Convolutional Network
Abstract:	Power load forecasting is crucial for reliable energy system operation and sustainable development, yet improving prediction accuracy and scalability remains challenging. To address these issues, this paper proposes a load forecasting method that leverages multi-level decomposition and an enhanced encoding-decoding architecture. Firstly, Boruta algorithm and SHAP algorithm are used to evaluate and select the feature information comprehensively. Subsequently, the original load data is decomposed using CEEMDAN. The resulting high and low-frequency signals are further segmented based on Sample Entropy, enabling more refined analysis and processing. For high-frequency signals, VMD is applied for secondary decomposition to effectively reduce data complexity. Subsequently, an improved Depth-separable volume-spatial channel attention network is employed as the encoder to enhance spatial feature representation and optimize channel utilization efficiency. Simultaneously, an Adaptive Wavelet Convolutional Neural Network is utilized to extract input data features, combined with an attention mechanism to strengthen the decoder's capacity for feature representation. Finally, a TCN serves as the decoder to achieve precise load forecasting. Case studies were conducted using datasets from four different regions, respectively. The experimental results demonstrate that the proposed method delivers excellent performance in the field of power load forecasting.

Short-term power load forecasting based on multi-level decomposition and improved codec

Yan Hong, Liang Xu*, Jingming Su, Ding Wang, Hantao Wang, Mushi Li

(School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: Power load forecasting is crucial for reliable energy system operation and sustainable development, yet improving prediction accuracy and scalability remains challenging. To address these issues, this paper proposes a load forecasting method that leverages multi-level decomposition and an enhanced encoding-decoding architecture. Firstly, Boruta algorithm and SHAP algorithm are used to evaluate and select the feature information comprehensively. Subsequently, the original load data is decomposed using CEEMDAN. The resulting high and low-frequency signals are further segmented based on Sample Entropy, enabling more refined analysis and processing. For high-frequency signals, VMD is applied for secondary decomposition to effectively reduce data complexity. Subsequently, an improved Depth-separable volume-spatial channel attention network is employed as the encoder to enhance spatial feature representation and optimize channel utilization efficiency. Simultaneously, an Adaptive Wavelet Convolutional Neural Network is utilized to extract input data features, combined with an attention mechanism to strengthen the decoder's capacity for feature representation. Finally, a TCN serves as the decoder to achieve precise load forecasting. Case studies were conducted using datasets from four different regions, respectively. The experimental results demonstrate that the proposed method delivers excellent performance in the field of power load forecasting.

Key words: Boruta algorithm; CEEMDAN; VMD; Depth-separable volume-spatial channel attention network; Adaptive Wavelet Convolutional Neural Network; TCN

1. Introduction

1.1. Motivation

With the ongoing advancement of power systems, research in power load forecasting has also evolved. However, traditional short-term load forecasting methods are increasingly inadequate in addressing the growing complexity of demand. The key to overcoming this challenge lies in optimizing neural network models to enhance forecasting accuracy[1]. Accurate power load forecasting is essential for the efficient operation of power systems. It not only improves the accuracy of power supply and demand planning but also ensures the stability and reliability of electricity distribution[2].

1.2. Literature review

Power load datasets are highly vulnerable to interference from various factors, making precise data preprocessing essential. Common feature selection methods include XGBoost[3], Random Forest[4] and Boruta algorithm[5]. While XGBoost and Random Forest are highly efficient and accurate for feature selection, XGBoost suffers from limited interpretability, and Random Forest is computationally intensive and prone to overfitting. To address these limitations, the Boruta algorithm has gained popularity. Its ability to automate feature selection minimizes random fluctuations in feature importance calculations and effectively mitigates overfitting. As a result, Boruta is widely employed in data preprocessing to reduce model complexity and enhance interpretability.

Short-term electricity load forecasting methods can be broadly categorized into three types: time series models, regression analysis models and artificial intelligence models. Time series models aim to analyze and forecast data over time, and commonly include

exponential smoothing[6], and autoregressive integral moving average (ARIMA)[7]. Regression analysis models mainly achieve prediction by establishing a mathematical relationship between the independent and dependent variables, covering random forest regression[8], polynomial regression[9], and support vector regression (SVR)[10], among others. Artificial intelligence modeling includes two major directions: deep learning and machine learning. Deep learning is suitable for dealing with tasks that require highly abstract feature learning, and representative methods include convolutional neural networks (CNN)[11], recurrent neural networks (RNN)[12], and long short-term memory networks (LSTM)[13]. Machine learning, on the other hand, is good at automatically extracting nonlinear features from data, and common methods include support vector machines (SVMs)[14], and extreme learning machines[15].

Power load data inherently contains fluctuation components at varying frequencies. To facilitate the analysis and prediction of these fluctuations, modal decomposition techniques are employed to break down the power load data into multiple intrinsic mode functions (IMFs). Yang et al. (2024) employed Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), incorporating adaptive noise and an enhanced algorithmic structure to effectively mitigate issues such as modal mixing, endpoint effects, reconstruction errors, and high computational costs associated with CEEMD, EEMD, and EMD in signal processing[16]. Ma et al. (2024) utilized Variational Modal Decomposition (VMD), which excels in enhancing decomposition accuracy, suppressing noise interference, and avoiding modal mixing, while effectively addressing variational constraint challenges[17].

Selecting an appropriate forecasting model is crucial for short-term electricity load

data. For instance, Xu et al. [18] introduced a time-series depthwise separable convolutional framework incorporating an attention mechanism to enhance forecasting accuracy and improve the scalability of existing power load prediction methods. Liao et al. [19] reconstructed the load sequence by wavelet transform in order to efficiently capture the periodic features of the load data, and on the basis of which, CNN was used to extract the potential features from the reconstructed data. Dai et al. [3] enhanced Seq2Seq model accuracy by employing TCN as both encoder and decoder, with CNN integrated for feature extraction. This approach significantly improved the encoder's representation capacity and the decoder's forecasting performance.

To enhance feature extraction and prediction accuracy, a novel method combining CEEMDAN-SE-VMD (CSV) and DSCAN-AWC-Attention-TCN (D_eAAT_d) is proposed. First, the Boruta algorithm is applied to eliminate redundant features, while SHAP analysis evaluates feature contributions. The original load data are then decomposed using CEEMDAN, splitting sample entropy into high-and low-frequency signals. High-frequency signals undergo secondary decomposition via VMD to further reduce data complexity. Next, a Depthwise Separable Convolutional Spatial Channel Attention Network (DSCAN) is introduced as the encoder, improving spatial feature representation and channel efficiency. An Adaptive Wavelet Convolutional Neural Network (AWC) extracts input data features, while the decoder leverages an attention mechanism to enhance feature representation, with a Temporal Convolutional Network (TCN) serving as the decoder. The proposed approach is validated using electricity data from two regions in New England and New York, USA. Additional analyses with datasets from Tetouan, Morocco and Singapore further demonstrate the methodology's robustness and

applicability.

The main contributions of this paper are as follows:

1. This paper introduces a novel feature engineering approach to enhance deep learning models' capacity to rapidly identify high-quality feature combinations across diverse data types.
2. This paper presents an innovative multi-level data decomposition method to reduce data complexity. The approach enhances decomposition precision, mitigates noise interference, and improves the efficiency and robustness of data processing.
3. This paper introduces an innovative power load forecasting model featuring an optimized encoder-decoder structure. The model significantly enhances feature representation, generalization capability, and robustness in power load prediction.

The paper is structured as follows: Section 2 outlines the methodologies. Section 3 introduces a novel power load forecasting framework. Section 4 presents a case study detailing the experimental data, evaluation metrics, model parameter settings, and compares the proposed model's performance with alternative approaches. Section 5 concludes with a summary of the study's findings.

2. Theoretical Basis

2.1. CEEMDAN

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)[20] represents a sophisticated advancement over Empirical Mode Decomposition (EMD). It integrates an adaptive noise mechanism, which dynamically modulates the amplitude of introduced white noise throughout the decomposition process, ensuring improved decomposition stability and accuracy. This adaptive noise mitigates

mode mixing and improves the separation of distinct oscillatory components. The steps for CEEMDAN decomposition are as follows:

(1) Add 1 set of Gaussian white noise to the original power load sequence $x(t)$ to construct the new signal $x_n(t)$.

$$x_n(t) = x(t) + \varepsilon_0 \delta_n(t) \tag{1}$$

(2) The nth decomposition (n=1,2,.....,N) of $x_n(t)$ is performed by the EMD method to obtain N eigenmode functions, and all of them are averaged to obtain the final first-order eigenmode component $C_{E1}(t)$, and the first-order residual $r_{E1}(t)$ is obtained at the same time.

$$C_{E1}(t) = \frac{1}{N}\sum_{n=1}^{N} C_{E1}^{n}(t) \tag{2}$$

$$r_{E1}(t) = x(t) - C_{E1}(t) \tag{3}$$

Where $C_{E1}^{n}(t)$ denotes the first IMF component decomposed by EMD.

(3) Gaussian white noise is added to $r_{E1}(t)$,EMD decomposition of the resulting new signal is performed to obtain the new first-order modal components, which are averaged to obtain the final second-order modal components $C_{E2}(t)$ and the second-order residuals $r_{E2}(t)$. The new signal is then analyzed by the EMD decomposition.

$$C_{E2}(t) = \frac{1}{N}\sum_{n=1}^{N} E_1\{r_{E1}(t) + \varepsilon_1 E_1[\delta_n(t)]\} \tag{4}$$

$$r_{E2}(t) = r_{E1}(t) - C_{E2}(t) \tag{5}$$

(4) Iterate the aforementioned steps until the resulting residual signal becomes a monotonic function that is no longer decomposable, at which point the algorithm terminates. Then the final K-order eigenmode component and the final residual r(t) are obtained.

$$C_{Ek}(t)=\frac{1}{N}\sum_{n=1}^{N}E_{k-1}\{r_{k-E1}(t)+\varepsilon_{k-1}E_{k-1}[\delta_n(t)]\} \tag{6}$$

$$r(t)=r_{k-E1}(t)-C_{Ek}(t) \tag{7}$$

(5) The original signal x(t) after CEEMDAN decomposition is denoted as:

$$x(t)=\sum_{k=1}^{k}C_{Ek}(t)+r(t) \tag{8}$$

Where $E_k(\bullet)$ is the k-order component obtained by CEEMDAN decomposition; ε_k is the weight coefficient of the Gaussian white noise; and $\delta_n(t)$ is the nth processed Gaussian white noise, where n=1,2,...,N.

2.2. VMD

Variational Modal Decomposition (VMD)[21] is an advanced, adaptive, and fully non-recursive variational signal processing approach. This technique presumes that the input signal $f(t)$ is decomposed into k distinct components, where each component represents a finite-bandwidth mode centered around a specific frequency. The algorithm seeks to minimize the total estimated bandwidth of all modes, subject to the constraint that their summation accurately reconstructs the original signal. The objective function calculation is shown by equation (9).

$$\begin{cases}\min\limits_{\{u_k\},\{\omega_k\}}\sum\limits_{k}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-j\omega_k t}\right\|_2^2\\ s.t.\sum\limits_{k}u_k(t)=f(t)\end{cases} \tag{9}$$

where k is the number of subsequences to be decomposed; ∂_t is the partial derivative with respect to t; $\delta(t)$ is the Dirac function; $u_k(t)$、ω_k are the decomposed modal components and center frequency, respectively; and $f(t)$ is the original signal.

$$\begin{aligned}L(\{u_k(t)\},\{\omega_k\},\lambda\})&=\alpha\sum_{k}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-j\omega_k t}\right\|_2^2\\&+\left\|f(t)-\sum_{k}u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_{k}u_k(t)\right\rangle\end{aligned} \tag{10}$$

The saddle points of the updated Lagrangian are obtained by solving equation (10) for alternating direction multiplier method iterations. The center frequency calculation is shown by equation (11).

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| U_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| U_k^{n+1}(\omega) \right|^2 d\omega} \tag{11}$$

Where ω_k^{n+1} is the center frequency of each modal component; ω is the frequency; $U_k(\omega)$ is the Fourier transform of $u_k(t)$; and n+1 is the number of iterations.

2.3. Attention

The attention mechanism[22] allocates resources to focus on and process key information while disregarding less relevant data. In computational applications, it enables selective focus on specific parts of large datasets. Softmax is used as the attention scoring function.

2.4. TCN

Temporal Convolutional Neural Network (TCN) is a transformed structure of Convolutional Neural Network[23], TCN is structurally composed of causal convolution, dilation convolution and residual module. By utilizing causal convolution, the leakage of future information is effectively prevented to obtain the correlation of time series.

3. Forecasting Methods

3.1. Improved DSCAN

In this paper, Depth-separable integrable spatial channel attention network (DSCAN) as the encoder of the model in this paper is the improved structure of the convolutional neural network module. DSCAN is composed of an Improved Spatial Reconstruction Unit (I'SRU) and Improved Channel Reconstruction Unit (I'CRU), a Depth-separable convolutional layer (DSC) is used in the spatial reconstruction unit to further improve the representation of spatial features, and a Convolutional Block Attention Module

(CBAM)[24] is introduced to enhance the network's attention to the input features and improve the network performance. Squeeze-Excitement Network (SENet) is introduced in the channel reconstruction unit to reduce redundant feature information, improve the utilization efficiency of channels, suppress unimportant channels, and automatically learn more important channels. The structure of DSCAN in this paper is shown in Fig. 1.

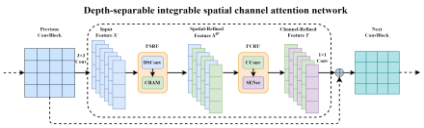

Fig. 1. Depth-separable integrable spatial channel attention network.

(1)The SRU suppresses spatial redundancy through separation-reconstruction operations, but it may increase the modeling and computational complexity for dealing with long time sequences. In this paper, the Improved Spatial Reconstruction Unit (I'SRU) introduces deeply separable convolution to effectively reduce the computational complexity, and introduces a spatial attention mechanism to further enhance the attention to the features.

(2)CRU reduces channel redundancy by segmentation-transformation-fusion method, but there is a potential risk of overfitting and increased model complexity. In this paper, the Improved Channel Reconstruction Unit (I'CRU) introduces a Squeezing-Excitation Network (SENet) that suppresses unimportant features to reduce the risk of overfitting, learns the importance of each feature channel and automatically adjusts the weights of each channel to reduce the model complexity.

3.2. Improved AWC

Wavelet Convolution (WC)[25] is a form of wavelet transform designed to multiply wavelet functions with signals at points. Wavelet convolution can analyze the properties of signals at different frequencies and time scales, but when dealing with noise-laden or

highly varying signals, fixing the convolution kernel may introduce unwanted noise, resulting in overfitting of the conventional convolutional layer or failure to capture useful signals. In this paper, the Adaptive Wavelet Convolution Neural Network (AWC) incorporates an adaptive strategy in the traditional wavelet convolution, generating the adaptive weight w and adaptive bias b of the convolution based on the input data, so that the convolution output x_{out} is multiplied element-by-element with the adaptive weights w and then added with the adaptive bias b. The model can dynamically reduce the influence of noise, improve the stability and prediction accuracy of the model, and at the same time enable the convolution operation to be dynamically adapted according to the different features of the input, thus enhancing the expressive power of the model. The AWC structure is shown in Fig. 2.

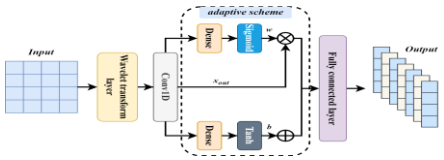

Fig. 2. Adaptive wavelet convolutional neural network.

3.3. Specific steps of the forecasting methodology

This paper presents a prediction method based on CSV and D_eAAT_d. The specific implementation process is shown in Fig. 3, and its steps are as follows.

(1) The collected data is preprocessed and divided into training set, verification set and test set. The collected data includes historical load and meteorological data, and the historical load is decomposed into multiple intrinsic modal functions (IMF). through CEEMDAN.

(2) SE is used to divide high-frequency and low-frequency signals, VMD is used to decompose high-frequency signals into a set of intrinsic modes, and the low-frequency signals and the signals after high-frequency decomposition are input into the model.

(3) The Boruta algorithm is used for feature screening, and the threshold value is set to 0.05 to automatically select important features. SHAP is used to analyze the contribution of important features to the model to reduce the redundancy of selected features to the model.

(4) In this paper, we adopt the improved DSCAN composed of spatial reconstruction unit and channel reconstruction unit as the encoder, and the feature information of each IMF and screening after CEEMDAN and VMD decomposition and reconstruction processing is used as the input of the model and transmitted to the DSCAN, which improves the ability to express spatial features and reduces the redundant feature information, and carries out the capturing of feature information. The Adaptive Wavelet Convolution and attention mechanism are integrated to avoid over-parameterization and enhance the decoder's attention to important features. Finally, TCN is used as the decoder to predict the load of the future time series, and Adam is used to optimize the model parameters.

(5) Conduct evaluation index analysis according to the prediction results to verify the robustness and practicability of the model in this paper.

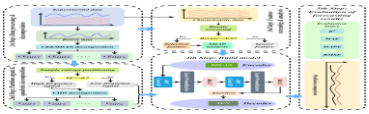

Fig. 3. Predictive model flow.

4. Case Study

To assess the efficacy of the proposed framework, this section is validated by a case study of four data sets. All experiments were conducted under Windows11, 2.11GHz Intel Core i5-10210U, 64-bit Python 3.8 and TensorFlow 2.13 with 8GB of RAM.

4.1. Data sources

This paper utilizes historical data spanning from January 1 to December 31, 2023, from New England, USA (Case 1, https://www.iso-ne.com/) and a state in New York, USA (Case 2, https://www.eia.gov/) for experimental analysis. The Case 1 dataset includes Load (MW), Dry_Bulb (°C), Dew_Point (°C), Electricity price ($/MWh), Distribution measurement ($/MWh), Congestion component ($/MWh), and Marginal loss component ($/MWh). The Case 2 dataset contains Load (MW), Dew point temperature (°C), Dry bulb temperature (°C), Relative humidity (%), Wet bulb temperature (°C), Wind speed (km/h), and Electricity price (USD/GJ). In both cases, the power load values are in MW, with a 1-hour sampling interval, totaling 8760 data points. Additionally, the dataset for Tetouan City, Morocco (Case 3, https://www.nems.emcsg.com/) from July 1 to August 31, 2017, with a 10-minute sampling interval, comprises 8928 data points. The data includes Load (KW), Temperature (°C), Humidity (%), Wind speed (m/s), General diffusion flow (g/(s·cm²)), and Diffusion flow (g/(s·cm²)). Finally, the dataset from Singapore, covering January 1 to December 31, 2023 (Case 4, https://www.nems.emcsg.com/), has a 30-minute sampling interval, yielding 17,520 data points. The data includes Load (MW), Electricity price ($/MWh), Generation cost ($/MWh), and Maximum capacity limit of the power system (MW). For experimental verification, the data were divided into training sets, verification sets and test sets according to the ratio of 6:2:2.

4.2. Assessment of indicators

In order to provide a comprehensive assessment of the model, the R-squared (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) are chosen as the assessment indexes of the model prediction in the

paper. The calculation formulas are shown by equation (12) to equation (15).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} | y_i - y_i^{'} |^2}{\sum_{i=1}^{n} | y_i - \overline{y_i} |^2} \tag{12}$$

$$MAE = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \left| y_i - y_i^{'} \right| \tag{13}$$

$$MAPE = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \left| \frac{\mathrm{y}_i - y_i^{'}}{y_i} \right| \times 100\% \tag{14}$$

$$RMSE = \sqrt{\frac{1}{\mathrm{n}} \sum_{i=1}^{n} (y_i - y_i^{'})^2} \tag{15}$$

Where y_i is the true value of the prediction set; $\overline{y_i}$ is the average value of the predicted data; $y_i^{'}$ is the predicted value of the test set; and n is the number of samples in the test set.

4.3. Parameterization

In configuring the model hyperparameters, this paper employs an experimental comparison approach, selecting the final parameters through iterative tuning, training, and optimization. The model parameter selection is presented in Table 1.

Table 1 Perparameter selection of model.

Model	Parameter
DSCAN	filters=64, kernel_size=3, dilation_rate=1, Dropout=0.2
AWC_1	filters=32, kernel_size=2, dilation_rate=2
AWC_2	filters=64, kernel_size=1, dilation_rate=3
Attention	filters=64, activation="Softmax"
TCN	filters=128, kernel_size=3, padding="causal", dilation_rate=2, activation="relu", Dropout=0.2
Dense	filters=32, activation=ReLU, epochs=50, batch_size=256, optimizer=Adam, loss function = MSE, learning rate(lr) = 0.001,
CNN-LSTM	CNN_1: filters=32, kernel_size=1 CNN_2: filters=64, kernel_size=1 LSTM_1: filters=64; LSTM_2: filters=32

4.4. Boruta-SHAP analysis

Given the significant impact of feature data on load forecasting, the Boruta algorithm is used to map nonlinear, high-dimensional data into a lower-dimensional space, selecting relevant features for power load forecasting. Based on Boruta analysis, the importance of historical load data and six feature variables in Case 1 and Case 2 is evaluated (as shown in Fig. 4 and 5). In this study, the Boruta threshold is set at 0.05, selecting features with an importance score above 0.05.

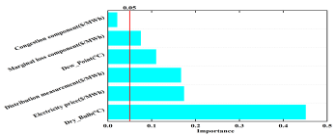

Fig. 4. Case 1 Boruta algorithm analysis.

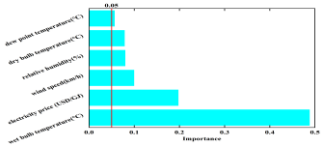

Fig. 5. Case 2 Boruta algorithm analysis.

The XGBoost model was interpreted and further analyzed using the SHAP algorithm to assess the contribution of each feature to the prediction outcomes (as shown in Fig. 6 and 7).

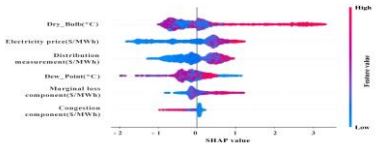

Fig. 6. Case 1 SHAP value analysis.

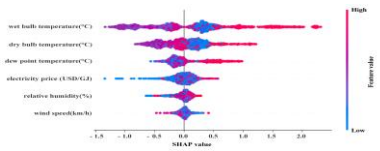

Fig. 7. Case 2 SHAP value analysis.

As illustrated in Fig. 7, higher values of wet bulb temperature (°C) and dry bulb temperature (°C) (red) are strongly associated with positive SHAP values, indicating a significant positive contribution to the target variables. In contrast, dew point temperature (°C), electricity price (USD/GJ), relative humidity (%), and wind speed (km/h) exhibit a narrower distribution of SHAP values, signifying a weaker influence. However, higher values (red) of these features occasionally show a positive impact, justifying their inclusion for consideration in the analysis.

4.5. CEEMDAN-SE-VMD analysis

Given the intricate nature of power load data, decomposition is essential. Initially,

the load data undergoes decomposition via the CEEMDAN method. Subsequently, the SE algorithm is applied to the decomposed subsequences to quantify their complexity, categorizing them into high-frequency and low-frequency components. The high-frequency signals are further processed using VMD, with the optimal value of K determined through experimental comparison. The results of the CEEMDAN decomposition are depicted in Fig. 8 and 9.

To differentiate between the high-frequency and low-frequency components of each CE_{IMF} derived from CEEMDAN, SE is utilized for classification. In this study, a threshold of 0.5 is applied, where SE values exceeding 0.5 are categorized as high-frequency, while those below 0.5 are designated as low-frequency. The SE values for each CE_{IMF} are detailed in Table 2.

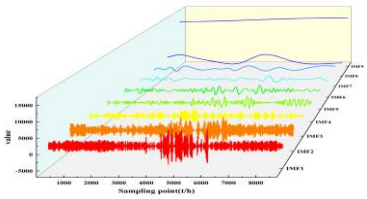

Fig. 8. Case 1 CEEMDAN decomposition.

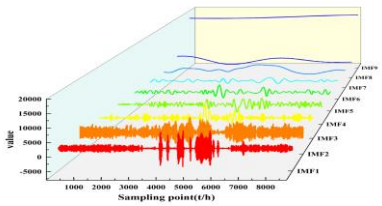

Fig. 9. Case 2 CEEMDAN decomposition.

Table 2 Sample entropy of each CE_{IMF}.

Case 1	CE_{IMF}	1	2	3	4	5	6	7	8	9
	$SE(H_Z)$	1.0551	1.0937	0.4789	0.3892	0.3836	0.3531	0.3647	0.3275	0.2870
Case 2	CE_{IMF}	1	2	3	4	5	6	7	8	9
	$SE(H_Z)$	0.6984	1.1015	0.4119	0.3896	0.3851	0.3674	0.3474	0.3040	0.2529

As observed in Table 2, both CE_{IMF1} and CE_{IMF2} for Case 1 and Case 2 exceed the threshold value of 0.5, classifying them as high-frequency signals. These high-frequency signals are further decomposed using VMD, and the reconstructed high-frequency signals from CEEMDAN are subsequently shown in Fig. 10 and 11.

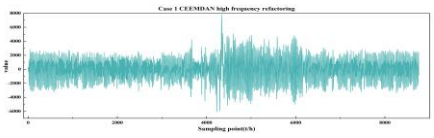

Fig. 10. Case 1 CEEMDAN high frequency refactoring.

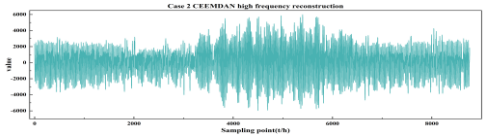

Fig. 11. Case 2 CEEMDAN high frequency refactoring.

The experimental comparison method was used to determine the value of K. The comparison of each parameter of VMD is shown in Table 3.

Table 3 Comparison of VMD parameters.

	alpha	K	R^2	MAE/MW	MAPE/%	RMSE/MW
Case 1	2000	2	0.995	99.425	0.773	122.851
		3	0.992	133.968	1.043	165.813
		4	0.989	147.906	1.131	188.879
		5	0.987	165.387	1.274	209.356
		6	0.986	170.508	1.325	216.449
	3000	2	**0.995**	**98.653**	**0.772**	**122.324**
		3	0.991	137.874	1.077	171.391
		4	0.988	155.085	1.196	196.935
		5	0.987	165.044	1.266	209.535
		6	0.986	166.635	1.284	212.748
Case 2	2000	2	0.990	152.385	0.934	191.581
		3	0.985	183.503	1.125	232.825
		4	0.981	196.273	1.190	262.224
		5	0.981	199.383	1.219	262.860
		6	0.984	190.097	1.166	246.036
	3000	2	**0.992**	**132.510**	**0.816**	**168.136**
		3	0.987	175.556	1.075	222.138
		4	0.982	195.323	1.190	259.289
		5	0.982	195.449	1.193	257.250
		6	0.980	205.247	1.248	271.664

As shown in Table 3, when alpha=3000 and K=2, the prediction accuracy surpasses that of other parameter settings for both Case 1 and Case 2. Therefore, these values are adopted. The high-frequency signals obtained from CEEMDAN are further decomposed using VMD to reduce data complexity further. The VMD decomposition results are illustrated in Fig. 12 and 13.

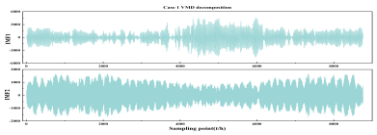

Fig. 12. Case 1 VMD decomposition.

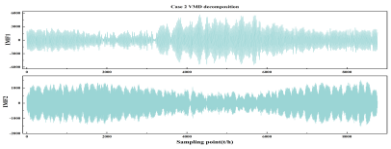

Fig. 13. Case 2 VMD decomposition.

Finally, all components are reconstructed to obtain the processed power load data after decomposition. A comparison between the reconstructed data and the original data is presented in Fig. 14 and 15.

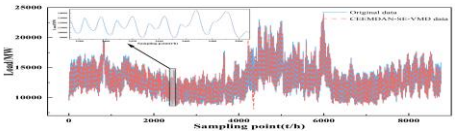

Fig. 14. Case 1 comparison of load data before and after decomposition.

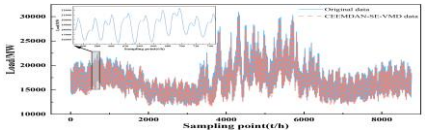

Fig. 15. Case 2 comparison of load data before and after decomposition.

The raw data and the decomposed processed data of Case 1 and Case 2 were analyzed for the assessment indicators, respectively, and the following data were obtained, as shown in Table 4.

Table 4 Decomposition before and after error comparison.

	R^2	MAE/MW	MAPE/%	RMSE/MW
Case 1	0.981	237.784	1.863	341.142
Case 2	0.989	259.730	1.515	332.050

As evidenced in Table 4, the data decomposed using the CSV method exhibits a high degree of alignment with the original data, achieving a fit of 98.1% and 98.9% for Case 1 and Case 2, respectively. Furthermore, this approach effectively mitigates the complexity of peak fluctuations. Consequently, the CSV methodology proposed in this study demonstrates its efficacy in reducing data complexity and enhancing the generalization capability of the model.

4.6. Experimental comparison and analysis

To confirm the efficacy of the CSV and D_eAAT_d models introduced in this study for short-term power load prediction, the proposed models were validated using datasets from two different regions. Four assessment metrics, R^2, MAE, MAPE, and RMSE, are used to determine the performance of each model.

4.6.1. Case 1

Prediction visualization of the test set of power data for Case 1 of this paper, where CSV-D_eCAT_d is the prediction method for the combination of CEEMDAN-SE-VMD and DSCAN-CNN-Attention-TCN, and the comparison visualization of the Case 1 model prediction is shown in Fig. 16.

In order to further analyze the fit of the test set and the prediction set, four assessment indicators are calculated for comparison, and the comparison of case one prediction of this paper's method is shown in Table 5, which shows that case one has a high fit in this paper's method.

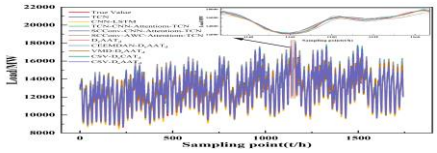

Fig. 16. Case 1 test set predicts results.

Table 5 Case 1 Test sets predict results.

Model	R^2	MAE/MW	MAPE/%	RMSE/MW
TCN	0.957	300.095	2.320	378.049
CNN-LSTM	0.969	284.094	2.238	319.248
TCN-CNN-Attention-TCN	0.974	230.698	1.791	292.969
SCConv-CNN-Attention-TCN	0.980	197.018	1.508	260.162
SCConv-AWC-Attention-TCN	0.984	174.919	1.326	233.253
D_eAAT_d	0.984	170.785	1.290	231.401
CEEMDAN-D_eAAT_d	0.984	179.551	1.393	228.835
VMD-D_eAAT_d	0.991	135.449	1.054	172.006
CSV-D_eCAT_d	0.995	101.684	0.804	125.914
CSV-D_eAAT_d	**0.995**	**98.653**	**0.772**	**122.324**

As presented in Table 5, the prediction accuracies of the methods proposed in this paper surpass those of other models. Specifically, the RMSE of our approach shows substantial improvements, outperforming TCN, CNN-LSTM, TCN-CNN-Attention-TCN, SCConv-CNN-Attention-TCN, SCConv-AWC-Attention-TCN, D_eAAT_d,

CEEMDAN-D_eAAT_d, VMD-D_eAAT_d, and CSV-D_eCAT_d by reductions of 67.64%, 61.68%, 58.25%, 52.98%, 47.56%, 47.14%, 46.54%, 28.88%, and 2.85%, respectively.

4.6.2. Case 2

A predictive visualization of the power data test set for Case 2 in this study has been conducted, with the results presented in Fig. 17.

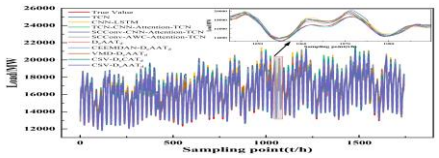

Fig. 17. Case 2 test set predicts results.

The four evaluation metrics were computed for comparative analysis, and the performance comparison of the Case 2 model predictions using the proposed approach is presented in Table 6.

Table 6 Case 2 Test sets predict results.

Model	R^2	MAE/MW	MAPE/%	RMSE/MW
TCN	0.938	374.825	2.286	483.898
CNN-LSTM	0.958	319.677	1.979	395.591
TCN-CNN-Attention-TCN	0.965	293.379	1.784	360.549
SCConv-CNN-Attention-TCN	0.969	279.342	1.725	343.985
SCConv-AWC-Attention-TCN	0.969	258.884	1.548	340.689
D_eAAT_d	0.976	224.987	1.359	299.775
CEEMDAN-D_eAAT_d	0.979	217.169	1.331	280.809
VMD-D_eAAT_d	0.981	196.010	1.194	262.447
CSV-D_eCAT_d	0.990	149.451	0.902	188.846
CSV-D_eAAT_d	**0.992**	**132.510**	**0.816**	**168.136**

As indicated in Table 6, the prediction accuracy of the methods proposed in this paper consistently outperforms other models. Specifically, the RMSE achieves reductions of 65.25%, 57.50%, 53.37%, 51.12%, 50.65%, 43.91%, 40.12%, 35.94%, and 10.97% compared to TCN, CNN-LSTM, TCN-CNN-Attention-TCN, SCConv-CNN-Attention-TCN, SCConv-AWC-Attention-TCN, D_eAAT_d, CEEMDAN- D_eAAT_d, VMD-D_eAAT_d,

and CSV-D_eCAT_d, respectively.

4.7. Experimental validation

To assess the applicability and performance of the proposed approach, the datasets from Case 3 and Case 4 were utilized for experimental validation.

4.7.1. Case 3

A comparison of the prediction curves for Case 3 of this paper is shown in Fig. 18.

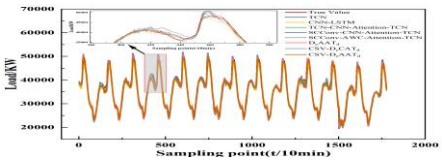

Fig. 18. Case 3 test set predicts results.

Four evaluation metrics were computed for comparative analysis, and the performance comparison of the Case 3 model predictions using the proposed approach is presented in Table 7.

Table 7 Case 3 Test sets predict results.

Model	R^2	MAE/MW	MAPE/%	RMSE/MW
TCN	0.965	946.781	2.716	1329.562
CNN-LSTM	0.972	850.015	2.229	1192.324
TCN-CNN-Attention-TCN	0.987	584.496	1.731	819.345
SCConv-CNN-Attention-TCN	0.987	510.139	1.478	800.950
SCConv-AWC-Attention-TCN	0.989	505.005	1.461	757.942
D_eAAT_d	0.990	467.301	1.380	714.684
CSV-D_eCAT_d	0.988	533.763	1.569	754.336
CSV-D_eAAT_d	**0.992**	**398.414**	**1.178**	**615.582**

As shown in Table 7, the RMSE of the proposed method is lower than that of TCN, CNN-LSTM, TCN-CNN-Attention-TCN, SCConv-CNN-Attention-TCN, SCConv-AWC-Attention-TCN, D_eAAT_d, and CSV-D_eCAT_d, with reductions of 53.70%, 48.37%, 24.87%, 23.14%, 18.78%, 13.87%, and 18.39%, respectively. Combined with the insights from Fig. 18, it is evident that the proposed method significantly enhances prediction

accuracy compared to other approaches, and the prediction curves demonstrate a strong fit.

4.7.2. Case 4

A comparison of the prediction curves for case 4 in this paper is shown in Fig. 19.

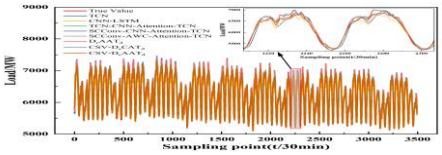

Fig. 19. Case 4 test set predicts results.

It can be deduced from Fig. 19 that the approach presented in this paper demonstrates superior fit compared to the other methods. Additionally, four evaluation metrics were computed for comparison purposes. The method proposed in this paper is used to compare the prediction results of the four evaluation indicators, and the results are shown in Table 8.

Table 8 Case 4 Test sets predict results.

Model	R^2	MAE/MW	MAPE/%	RMSE/MW
TCN	0.959	84.093	1.329	110.094
CNN-LSTM	0.960	84.747	1.288	109.561
TCN-CNN-Attention-TCN	0.983	58.877	0.930	71.309
SCConv-CNN-Attention-TCN	0.986	52.583	0.830	65.200
SCConv-AWC-Attention-TCN	0.988	48.989	0.775	60.320
D_eAAT_d	0.990	44.377	0.706	55.573
CSV-D_eCAT_d	0.992	38.897	0.617	49.103
CSV-D_eAAT_d	**0.994**	**32.348**	**0.513**	**42.112**

As evident from Table 8, the RMSE of the method proposed in this study is significantly lower than that of TCN, CNN-LSTM, TCN-CNN-Attention-TCN, SCConv-CNN-Attention-TCN, SCConv-AWC-Attention-TCN, D_eAAT_d, and CSV-D_eCAT_d, with reductions of 61.75%, 61.56%, 40.94%, 35.41%, 30.19%, 24.22%, and 14.24%, respectively.

5. Conclusion

This paper introduces a short-term forecasting method that incorporates signal multilevel decomposition processing and an enhanced encoder-decoder framework for power load sequences. The aim is to enhance the effectiveness of existing models in terms of forecasting accuracy and the utilization of feature information.

Acknowledgements

This work was accomplished by National Natural Science Foundation of China (Grant Nos. 62105004 and 52174141), Anhui Mining Machinery and Electrical Equipment Coordination Innovation Center, Anhui University of Science & Technology (Grant No. KSJD202304), and the Anhui Province Digital Agricultural Engineering Technology Research Center Open Project (Grant No. AHSZNYGC-ZXKF021), University Student Innovation and Entrepreneurship Fund Project (202210361053, 202310361037).

References

[1] Guo-Feng F ,Yan-Rong L ,Hui-Zhen W , et al.The new hybrid approaches to forecasting short-term electricity load[J].Electric Power Systems Research,2022,213:108759.

[2] Jianzhou W ,Linyue Z ,Zhiwu L .Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm[J].Applied Energy,2022, 305: 117911.

[3] Dai Y ,Yu W .Short-term power load forecasting based on Seq2Seq model integrating Bayesian optimization, temporal convolutional network and attention[J].Applied Soft Computing,2024,166:112248.

[4] Arshia A ,Mohsen G ,Burak K .Accuracy improvement of electrical load forecasting against new cyber-attack architectures[J].Sustainable Cities and Society,2022,77:103523.

[5] Li J ,Liu Y ,Gong H , et al.Stock price series forecasting using multi-scale modeling with boruta feature selection and adaptive denoising[J].Applied Soft Computing,2024,154:111365.

[6] Grzegorz D ,Pawel P ,Slawek S .A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting.[J].IEEE transactions on neural networks and learning systems,2021,33(7):2879-2891.

[7] Fuhua Y ,Qi Y ,Arda Y , et al.A novel hybrid deep correction approach for electrical load demand prediction[J].Sustainable Cities and Society,2021,74:103161.

[8] Magalhães B ,Bento P ,Pombo J , et al.Short-Term Load Forecasting Based on Optimized Random Forest and Optimal Feature Selection[J].Energies,2024,17(8):1926.

[9] Abu-Shikhah N ,Elkarmi F .Medium-term electric load forecasting using singular value decomposition[J].Energy,2011,36(7): 4259- 4271.

[10] Siwei L ,Xiangyu K ,Liang Y , et al.Short-term electrical load forecasting using hybrid model of manta ray foraging optimization and support vector regression[J].Journal of Cleaner Production,2023, 388:135856.

[11] Shiyun Z ,Runhuan C ,Jiacheng C , et al.A CNN and LSTM-based multi-task learning architecture for short and medium-term electricity load forecasting[J].Electric Power Systems Research,2023,222: 109507.

[12] Feifei Y ,Xueqian F ,Qiang Y , et al.Decomposition strategy and attention-based long short-term memory network for multi-step ultra-short-term agricultural power load forecasting[J].Expert Systems With Applications,2024,238(PF):122226.

[13] Zhining C ,Jianzhou W ,Li Y , et al.A hybrid electricity load prediction system based on weighted fuzzy time series and multi-objective differential evolution[J].Applied Soft Computing, 2023,149(PB):111007.

[14] Zhang X ,Wang J ,Zhang K .Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm[J].Electric Power Systems Research,2017,146:270-285.

[15] Yang Y ,Lou H ,Wang Z , et al.Pinball-Huber boosted extreme learning machine regression: a multiobjective approach to accurate power load forecasting[J].Applied Intelligence,2024,54(17-18): 8745-8760.

[16] Liang Y ,Zhang D ,Zhang J , et al.A state-of-the-art analysis on decomposition method for short-term wind speed forecasting using LSTM and a novel hybrid deep learning model[J].Energy,2024,313: 133826.

[17] Ma K ,Nie X ,Yang J , et al.A power load forecasting method in port based on VMD-ICSS-hybrid neural network[J].Applied Energy,2025, 377(PB):124246.

[18] Xu H ,Hu F ,Liang X , et al.A framework for electricity load forecasting based on attention mechanism time series depthwise separable convolutional neural network[J].Energy,2024,299(PP): 131258.

[19] Zhifang L ,Haihui P ,Xiaoping F , et al.Multiple Wavelet Convolutional Neural Network for Short-Term Load Forecasting[J].IEEE INTERNET OF THINGS JOURNAL,2021, 8(12):9730-9739.
[20] Ke L ,Wei H ,Gaoyuan H , et al.Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network[J]. Energy & Buildings,2023,279:112666.
[21] Fang Y ,Jinxing C .An ensemble multi-step M-RMLSSVR model based on VMD and two-group strategy for day-ahead short-term load forecasting[J].Knowledge-Based Systems,2022,252:109440.
[22] Jianguo W ,Lincheng H ,Xiuyu Z , et al.Electrical load forecasting based on variable T-distribution and dual attention mechanism[J]. Energy,2023,283:128569.
[23] Pang S ,Zou L ,Zhang L , et al.A hybrid TCN-BiLSTM short-term load forecasting model for ship electric propulsion systems combined with multi-step feature processing[J].Ocean Engineering,2025,316: 119808.
[24] Wang Y ,Pu J ,Miao D , et al.SCGRFuse: An infrared and visible image fusion network based on spatial/channel attention mechanism and gradient aggregation residual dense blocks[J].Engineering Applications of Artificial Intelligence,2024,132:107898.
[25] Stefenon F S ,Seman O L ,Silva D C E , et al.Hypertuned wavelet convolutional neural network with long short-term memory for time series forecasting in hydroelectric power plants[J].Energy,2024,313: 133918.