

Article

Multifeature-Based Variational Mode Decomposition–Temporal Convolutional Network–Long Short-Term Memory for Short-Term Forecasting of the Load of Port Power Systems

Guang Chen ^{1,2,3}, Xiaofeng Ma ^{2,3,*} and Lin Wei ^{2,3}

¹ Laboratory of Transport Pollution Control and Monitoring Technology, Beijing 100028, China; 333052@whut.edu.cn

² Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China; 346994@whut.edu.cn

³ National Engineering Research Center for Water Transport Safety, Wuhan 430063, China

* Correspondence: maxiaofeng@whut.edu.cn

Abstract: Accurate short-term forecasting of power load is essential for the reliable operation of the comprehensive energy systems of ports and for effectively reducing energy consumption. Owing to the complexity of port systems, traditional load forecasting methods often struggle to capture the non-linearity and multifactorial interactions within the factors creating power load. To address these challenges, this study combines variational mode decomposition (VMD), temporal convolutional network (TCN), and long short-term memory (LSTM) network to develop a multi-feature-based VMD-TCN-LSTM model for the short-term forecasting of the power load of ports. VMD is first used to decompose the power load series of ports into multiple, relatively stable components to mitigate volatility. Furthermore, meteorological and temporal features are introduced into the TCN-LSTM model, which combines the temporal feature extraction capability of the TCN and the long term-dependent learning capability of the LSTM. Comparative analyses with other common forecasting models using the observed power load data from a coastal port in China demonstrate that the proposed forecasting model achieves a higher prediction accuracy, with an R-squared value of 0.94, mean squared error of 3.59 MW, and a mean absolute percentage error of 2.36%.

Keywords: multifeatured short-term forecasting; port power load; variational mode decomposition; temporal convolutional network; long short-term memory network



Citation: Chen, G.; Ma, X.; Wei, L. Multifeature-Based Variational Mode Decomposition–Temporal Convolutional Network–Long Short-Term Memory for Short-Term Forecasting of the Load of Port Power Systems. *Sustainability* **2024**, *16*, 5321. <https://doi.org/10.3390/su16135321>

Academic Editor: Jack Barkenbus

Received: 26 April 2024

Revised: 18 June 2024

Accepted: 20 June 2024

Published: 22 June 2024



Copyright: © 2024 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 the economy develops rapidly, ports have become critical nodes for domestic logistics and international trade, leading to an increasing demand for energy. Traditional fossil fuels are causing severe environmental pollution. There is a pressing need to move toward more sustainable energy management techniques because ports must respond to ecological pressures by integrating sustainable considerations into port operation activities to promote a green transition [1]. Major ports worldwide have been electrified to varying degrees in terms of energy. A large number of studies have shown that electrification has a significant effect on improving energy efficiency in ports [2]. Electricity accounts for an increasing proportion of energy consumption in ports, making energy management and optimization in ports particularly important. In recent years, the growing usage of renewable energy sources such as wind and solar power in ports, the development of energy substitution technologies, virtual power plants, and energy storage technologies has necessitated the need for accurate short-term forecasting of port load [3,4]. Energy storage technologies have an important role to play during peak load hours in ports and by discharging during peak demand and charging during low demand, which can balance the energy supply and demand and maintain system stability [5]. Accurate port

load forecasting and fast response capabilities enable energy storage systems to optimize dispatch and improve the reliability and economics of port area power systems.

The traditional load forecasting methods include time series analysis [6] and regression analysis [7]. The Autoregressive Integrated Moving Average (ARIMA) time series forecasting model has gained popularity due to its exceptional ability to handle both smooth and non-smooth series. Nano et al. [8] used the cuckoo search algorithm (CS) to optimize the parameters of the ARIMA model to predict the actual power load data, and the results proved that ARIMA showed high accuracy in predicting the short-term power load. Jeong et al. [9] demonstrated good accuracy in multivariate time series forecasting while utilizing the Vector Autoregressive (VAR) model to predict building electrical loads. By taking data analysis and selecting logistic regression as the basic model, Feng et al. [10] proposed and developed a load forecasting method based on the combination of clustering and iterative logistic regression. Wu et al. [11] proposed an improved regression model based on mini-batch stochastic gradient descent to address the issues of slow prediction speed and low prediction accuracy in regression analysis models. The results demonstrated that the modified algorithm achieves significant improvement in prediction speed. The theoretical system of traditional load forecasting methods is relatively mature, and the calculation is simple. However, the prediction effect is unstable, and the accuracy is poor when dealing with high complexity and nonlinearity data, such as port load data [12,13]. As it is difficult to collect the reliable data required for predicting power loads in port areas, traditional models are difficult to adapt to rapidly changing environmental factors and complex port operations.

In recent years, deep learning models have become powerful tools for addressing such problems due to their excellent non-linear fitting ability and adaptability. Typical representatives are the long short-term memory (LSTM) network [14,15], the gated recurrent unit (GRU) [16], DeepAR [17], N-BEATS [18], transformers [19], etc. These methodologies exhibit proficiency in modeling complex nonlinear load dynamics, demonstrating an enhanced adaptability to nonlinear fluctuations and an improved ability to precisely capture the load data patterns. LSTM has gained widespread attention for its exceptional performance in predicting power load time series, and many researchers have improved the basic LSTM model. Buratto et al. [20] proposed a Seq2Seq-LSTM model based on an attention mechanism to predict Brazil's electricity load, achieving better capture of long-range dependencies in load sequences. Sheng et al. [21] proposed an improved residual LSTM-based framework for solving the short-term load forecasting problem, which avoids the problem of gradient vanishing when training deep neural networks. GRU is a simplified version of LSTM, reducing the model's complexity and computational overhead. Wang et al. [22] utilized GRU with a gorilla troop optimizer (GTO) to predict and optimize energy consumption in the HVAC systems of smart buildings. The GTO is employed to tune the parameters of the GRU model, enhancing the accuracy predictions. In addition, some researchers [23] combined LSTM with convolutional neural networks (CNNs) to develop a hybrid cross-channel CNN-LSTM model for smart grid load forecasting, which improved prediction efficiency and accuracy compared to a single model. Generally, these combined models, which integrate the advantages of multiple models, typically show higher accuracy. Temporal convolutional networks (TCNs) capture local features of sequence data through stacked convolutional layers, and dilated convolutions are used to effectively increase the receptive field, enabling the network to capture longer-range dependencies [24]. Zheng et al. [25] utilized TCNs and the Global Attention Mechanism (GAT) to model and process load time series data, improving model prediction accuracy by filtering input variables using Shapley Additive Explanation (SHAP) values.

To further address the strong volatility of load data and fully explore internal features, some researchers use modal decomposition algorithms such as empirical mode decomposition (EMD) [26], ensemble empirical mode decomposition (EEMD) [27], wavelet decomposition [28], and variational mode decomposition (VMD) [29] to decompose and smooth the load data before deep learning model training. However, EMD and EEMD

are prone to the mode-mixing problem, which can lead to significant errors in decomposition [30]. VMD can effectively avoid the mode-mixing problem through variational optimization. At the same time, it is more capable of adapting to the complex components of the sequence than wavelet decomposition, providing richer features for deep learning model prediction [31].

The energy demand at ports is significantly influenced by traffic demand, exhibiting notable differences across various periods and demonstrating strong temporal regularity. Additionally, it is affected by various environmental factors, especially meteorological elements such as wind speed and temperature. High wind speeds can restrict the operation of cranes and other loading and unloading equipment, thereby impacting the efficiency of cargo handling [32]. Temperature fluctuations, on the other hand, directly affect the energy consumption in the port area, such as cooling or heating demands. To address the nonlinear issues and complex multi-faceted influences in port load forecasting, deep learning algorithms have been proven effective. The integration of modal decomposition algorithms, called VMD, with deep learning offers a promising solution for efficient energy management in port operations. In pursuit of developing an accurate and efficient short-term port load forecasting model, this paper makes the following contributions:

- A VMD-TCN-LSTM model is proposed for port load forecasting. By leveraging VMD to mitigate data volatility and extract features of varying frequencies, along with the integration of TCN and LSTM, the model can effectively capture temporal patterns and long-term dependencies;
- By leveraging multi-feature modeling, we enhance the prediction accuracy of the port power load forecasting model by considering various feature variables such as the temperature, the 10 m wind speed, the quarter, and the hour as input to the model;
- Using real port load data, a case study was performed. The proposed model's superiority was confirmed through comparative experiments with other widely used load forecasting models.

The structure of the paper is as follows: Section 1 provides an introduction to the research background and related literature on short-term port load forecasting. Section 2 describes the research objective and the theoretical methods adopted for short-term port load forecasting. Section 3 focuses on data preprocessing, feature selection, and applying VMD decomposition to the original load data. Section 4 analyzes the data and compares various models with specific cases to verify the effectiveness of the proposed forecasting model. Section 5 discusses the limitations of this study and suggests future research directions. Finally, the conclusion section summarizes the main contributions of this study.

2. Methodology

2.1. The VMD-TCN-LSTM Load Forecasting Model

Given that port loads are susceptible to a variety of complex factors, exhibiting certain levels of non-stationarity and randomness, this study proposed the VMD-TCN-LSTM load forecasting model. This model combines decomposition algorithms with deep neural network strategies to address these problems, as illustrated in the basic process framework (Figure 1).

Firstly, the VMD is used to decompose the original port load data and compute the intrinsic mode functions (IMFs) to obtain a group of subsequence components with different center frequencies. This step aims to smooth the original sequence data and reduce the complexity of the sequence. Subsequently, a correlation analysis is conducted to identify the meteorological and temporal features that exhibit a strong correlation with port loads. These modal components, the selected meteorological and temporal feature variables, and the original historical load sequence data serve as inputs to the TCN-LSTM model. Before data input, the training and testing datasets are divided, standardized, and one-hot encoded. Then, using the sliding window method, the datasets are partitioned into multiple time windows, allowing for training and prediction within each window. After predictions are made within a particular window, the window proceeds to the next

for further processing, as shown in Figure 2. In this study, the model uses the previous L steps of the data to predict the subsequent port load value. The sliding time window retains the temporal sequence information within the data, ensuring the method's scientific validity and effectiveness. Finally, the model's output is inversely normalized to obtain the predicted load values.

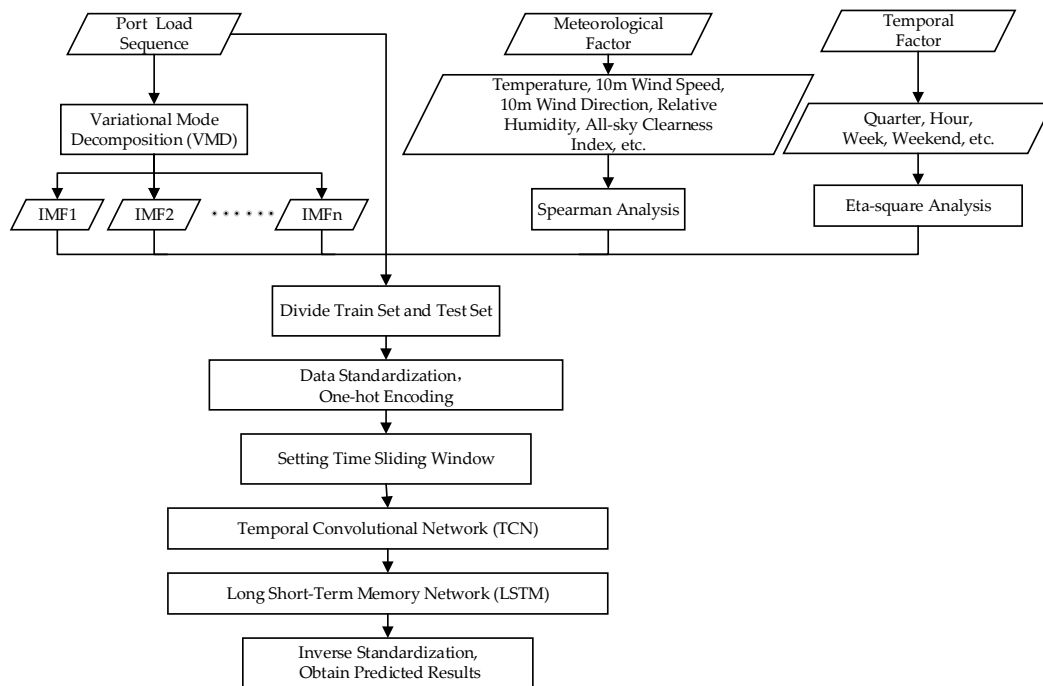


Figure 1. The basic process framework of the VMD-TCN-LSTM model.

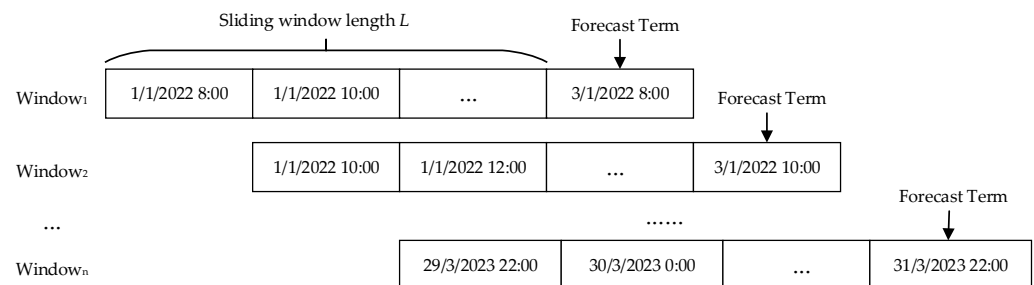


Figure 2. Schematic diagram of time-sliding window.

2.2. TCN-LSTM Model

2.2.1. Temporal Convolutional Network

The temporal convolutional network (TCN) combines the advantages of dilation convolution and causal convolution with a unique dilation causal convolution structure, supplemented by a residual connection structure to avoid gradient vanishing, and is a specialized convolutional network dedicated to time series forecasting [33]. If x_t is an element in the time series, then the output y_t is denoted by Equation (1).

$$y_t = F(x_{t-k}, \dots, x_{t-1}, x_t) \quad (1)$$

where k represents the size of the convolutional kernel; F represents the convolution operation; and y_t is the result of the convolution operation.

The dilation convolution of the TCN introduces a dilation factor to control the skipping of input segments, applying the convolutional kernel over a larger area [34]. Given a one-

dimensional sequence input $x \in R_n$ and convolutional kernels $f : \{0 \dots k - 1\} \rightarrow R$, the operation of dilated convolution F on sequence element s is defined in Equation (2).

$$F(s) = (x_d * f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-di} \quad (2)$$

where, d is the dilation factor; k is the size of the convolutional kernel; $s - di$ is the direction of the past time series.

The structure of the TCN's dilated convolution is shown in Figure 3. The dilation factor for different layers increases in the manner $d = 2^l$, enabling the TCN to achieve a larger receptive field with fewer layers, which is advantageous for handling time-series data with long-term historical dependencies.

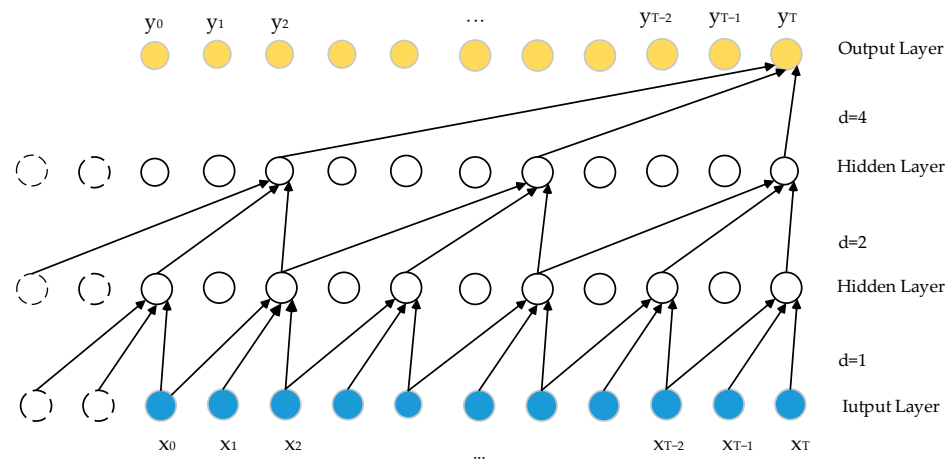


Figure 3. Schematic diagram of TCN's dilated convolution.

2.2.2. Long Short-Term Memory Network

The long short-term memory (LSTM) network is a variant of the recurrent neural network (RNN) capable of learning long-term dependencies. To address the issue of vanishing or exploding gradients that traditional RNNs face when training on a long sequence [16], LSTM introduces a gate control structure. The internal structure of LSTM is illustrated in Figure 4.

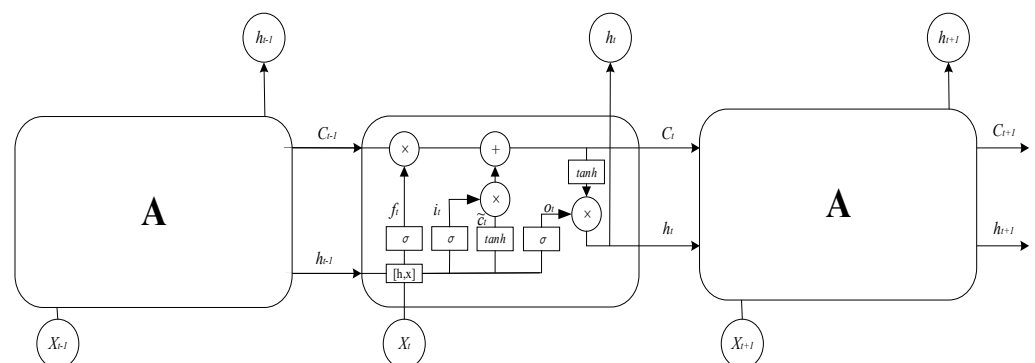


Figure 4. Schematic diagram of LSTM gate control structure.

The gating mechanisms of LSTM use σ function as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

The forget gate takes x_t and h_{t-1} as inputs and produces f_t as output, represented by Equation (4). Here, W_f denotes the weight matrix of the forget gate, and b_f denotes its bias term.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

The input gate determines which values are to be updated and how much of the current input x_t is to be saved to the cell state c_t . It also combines the current memory information $\tilde{c}_t$ and long-term memory information c_{t-1} to form a new cell state c_t . This process is depicted by Equations (5)–(7). The weight matrices for the input gate and cell state are denoted by W_i and W_c , and the corresponding bias terms of each weight matrix are denoted by b_i and b_c .

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (7)$$

The output gate determines how much information from the current cell state is to be outputted to h_t . The output of the output gate o_t is represented by Equation (8). The updated cell state c_t , after passing through a tanh unit, is multiplied by the output of the output gate o_t to yield the final output of the hidden layer h_t , as shown in Equation (9).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (9)$$

2.2.3. TCN-LSTM Model Framework

The TCN-LSTM model combines the temporal feature extraction capabilities of the TCN with the advantage of LSTM in learning long-term dependencies in time series data. This integration allows for a deeper exploration of data characteristics and facilitates the prediction of future trends. Consequently, it surpasses the performance of single models in uncovering potential relationships within the data [35]. The framework of the model is illustrated in Figure 5.

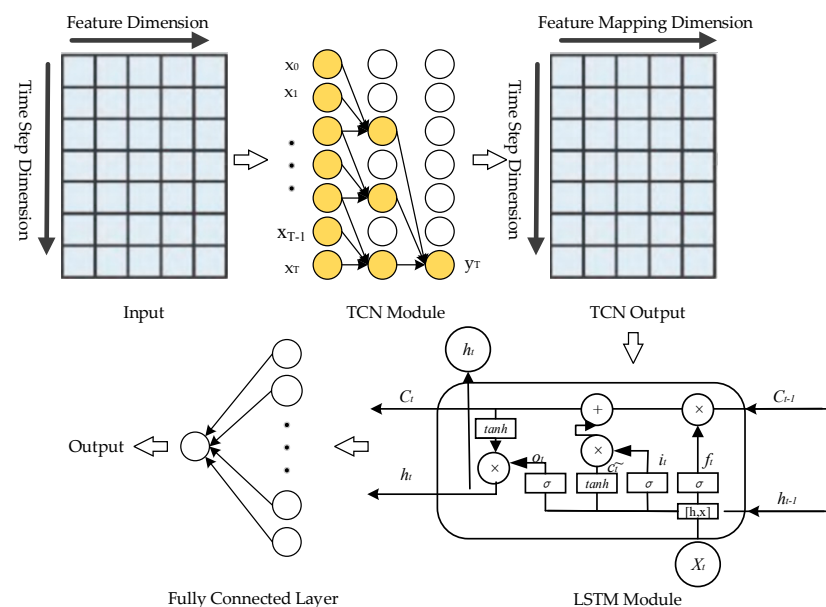


Figure 5. TCN-LSTM composite model framework.

In Figure 5, the model's input is a tensor formed by merging multiple input sequences, each representing different features or variables. The input data pass through the TCN

layer, where causal dilated convolutions output feature mappings with the same time step length, capturing the temporal features of the input data. The dimension size of the feature mapping equals the number of convolutional kernels in the TCN module. These temporal features extracted by the TCN module serve as inputs to the LSTM model, which learns the long-term dependencies of the sequence. In addition, a fully connected layer processes the high-dimensional output of LSTM, transforming it to match the dimensions of the prediction target. The number of output neurons is set to 1, indicating the prediction of the port load value at the next time step.

2.3. Variational Mode Decomposition

The variational mode decomposition (VMD) algorithm is frequently employed for processing non-stationary signals [36], decomposing complex signals into a series of intrinsic mode functions (IMFs). It finds extensive application across numerous domains, such as signal processing, fault diagnosis, and time series analysis. In the process of VMD, “intrinsic mode” refers to an independent frequency component separated from the original signal through an optimization process. Each mode has a specific center frequency and a finite bandwidth, where the center frequency represents the main oscillation frequency of that mode, and the finite bandwidth ensures that the frequency components are concentrated around the center frequency, thus reducing mode-mixing issues. VMD uses iterative optimization to achieve frequency separation and concentration for each mode, effectively decomposing the signal and analyzing its different frequency characteristics. This decomposition [37] is accomplished by solving a constrained variational problem, as shown in Formula (10).

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (10)$$

$$s.t. \sum_k u_k(t) = f(t)$$

where $u_k = \{u_1, u_2, \dots, u_k\}$ represents the set of all modes; ω_k are the center frequencies of each mode; ∂_t denotes the time derivative of the function; $\delta(t)$ signifies the unit impulse function; $f(t)$ is the main signal to be decomposed.

By introducing a quadratic penalty term a and the Lagrange multiplier λ , the problem is transformed into an unconstrained variational problem, as indicated in Formula (11).

$$L(\{u_k\}, \{\omega_k\}, \lambda) := a \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \quad (11)$$

$$+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle$$

The Alternating Direction Method of Multipliers (ADMMs) is employed to iteratively update w_k^{n+1} and u_k^{n+1} to obtain the optimal solution of the variational problem. The iteration process is described as follows:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2a(w - w_k)^2} \quad (12)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \quad (13)$$

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left(\hat{f}(w) - \sum_k \hat{u}_k^{n+1}(w) \right) \quad (14)$$

where τ represents the noise tolerance level. The iteration continues until either the convergence criteria are met or the maximum number of iterations is reached, with the convergence criteria defined as follows:

$$\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 < \varepsilon \quad (15)$$

3. Multifeature-Based VMD-TCN-LSTM Short-Term Port Load Forecasting

3.1. Correlation Analysis

The port load data used in this study were collected from a coastal port in China, covering the period from 1 January 2022 to 31 March 2023. Two hours were spaced between the sampling points, and the sample size was 5459, with a minimum value of 2.559 MW and a maximum value of 123.287 MW. The preliminary analysis indicates no missing or duplicate values in the temporal continuum of sampling. The load data curve and frequency histogram are depicted in Figures 6 and 7.

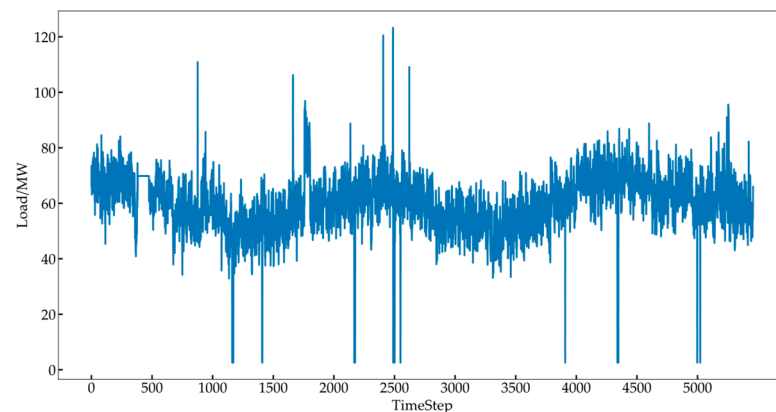


Figure 6. Port load curve.

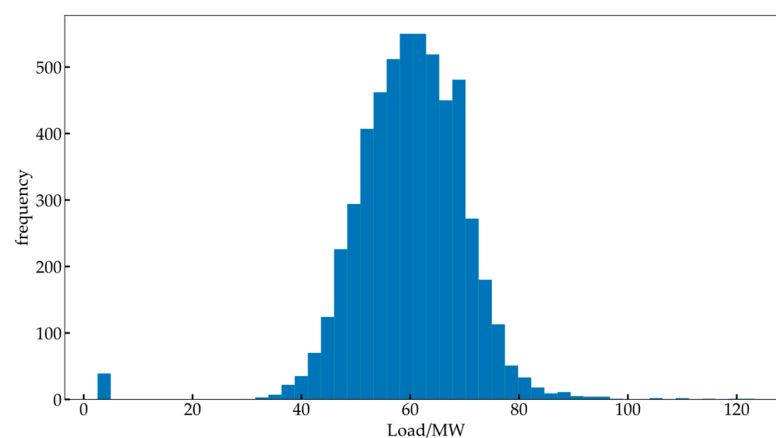


Figure 7. Port load frequency histogram.

As shown in Figure 6, the port power load data exhibit significant fluctuations, indicating a non-stationary series with extreme values. These extremes may result from instrument malfunction or anthropogenic factors. Upon examination, the frequency histogram of the port load data closely resembles a normal distribution. Therefore, this study employs the 3σ rule for outlier treatment, calculating the mean $\bar{X}$ and standard deviation σ of the data. The probability of data distribution within $(\bar{X} - 3\sigma, \bar{X} + 3\sigma)$ is 0.9973. Data points falling outside this range are considered outliers and replaced using the mean replacement method.

Various factors affect the port load, including the port operation environment, port throughput, and the amount of equipment used. This study mainly considered factors such as temperature, wind force, and wind direction in terms of the port operation environment. Since the port throughput has strong periodicity and a strong correlation with the amount of equipment in use, temporal feature factors can be used to indirectly reflect the amount of equipment in use and the traffic demand. The meteorological data were sourced from a meteorological data statistics website, encompassing temperature, wind speed (WS), wind direction (WD), relative humidity (RH), precipitation, the all-sky clearness index (ALLSKY_KT), and the surface ultraviolet index (SFC_UV_I), as shown in Table 1. To directly understand the correlation between meteorological factors and port loads, the graphs of the port temperature and 10 m WS are shown in Figure 8. By comparing Figures 6 and 8, it is clear that there are seasonal fluctuations in the port temperature. The port load values are higher when the port has high or low temperature periods. This may be due to the increase in cooling demand at high temperatures and heating demand at low temperatures, which leads to a rise in the electrical load of the port and also shows that temperature has an effect on port loads. Although port wind speed data are not significantly cyclical overall, their high volatility affects how port loads fluctuate. During high average wind speed periods, there is a slight decrease in loads, which may be due to high wind speeds causing some port operations to be interrupted or reduced for safety reasons.

Table 1. Meteorological influencing factors.

Variable Name	Unit	Variable Type
Temperature	°C	Continuous Variable
10 m WS	m/s	Continuous Variable
50 m WS	m/s	Continuous Variable
10 m WD	degree	Continuous Variable
50 m WD	degree	Continuous Variable
RH	%	Continuous Variable
Precipitation	mm/h	Continuous Variable
ALLSKY_KT	Dimensionless Units	Continuous Variable
SFC_UV_I	Dimensionless Units	Continuous Variable

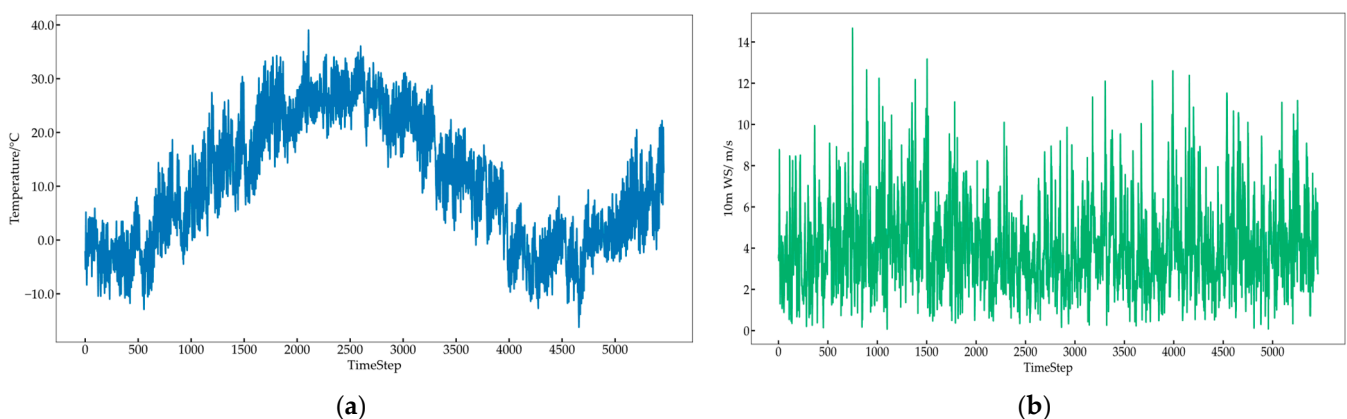


Figure 8. (a) Port temperature curve; (b) port 10 m wind speed curve.

The temporal feature factors primarily include quarters, workdays, and non-workdays, among others. The specific types of feature variables are detailed in Table 2.

Having numerous feature variables can increase the cost of model training and complicate the fitting process. Therefore, conducting correlation analyses is essential for the various factors, which enables the selection of variables with high correlation as features for model training. This study used the Spearman correlation coefficient, and the results are shown in Figure 9. For the analysis of correlations between discrete unordered and continuous variables, the Eta-squared (η^2) measure was used as an indicator of correlation.

Eta-squared is commonly utilized to measure the effect size in ANOVA to assess the degree of association between discrete unordered and continuous variables [38]. The analysis results are presented in Table 3.

Table 2. Temporal feature factors.

Variable Name	Variable Type
quarter	Discrete Unordered Variables
hour	Discrete Unordered Variables
week	Discrete Unordered Variables
non-workday or workday	Binary Variable
which day of the month	Discrete Unordered Variables

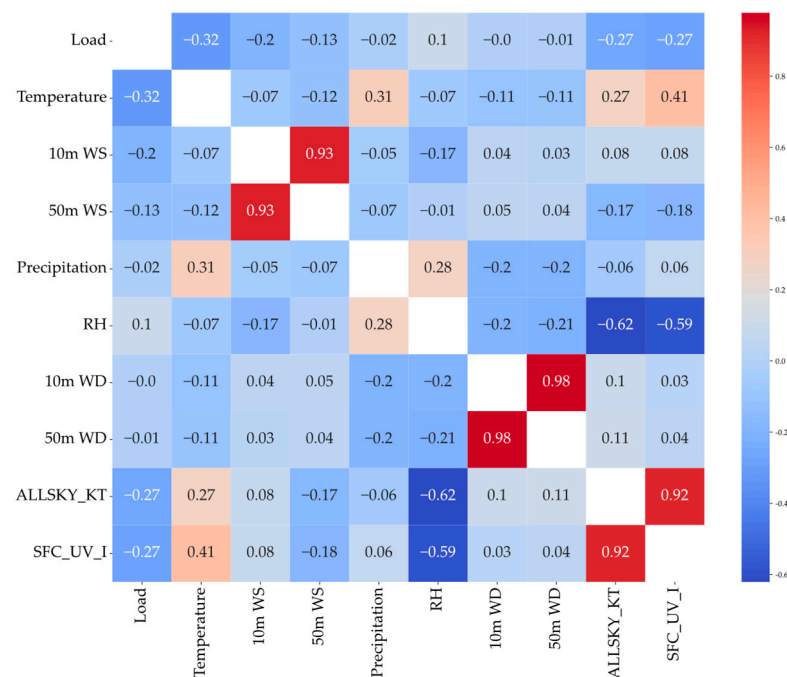


Figure 9. Heat map of meteorological factor correlation.

Table 3. Eta-squared of temporal feature factors calculation results.

Variable Name	Eta-Squared
quarter	0.0714
hour	0.0877
week	0.0056
non-workday or workday	0.0026
day of the month	0.0059

When the Eta-squared (η^2) is greater than 0.06 but less than 0.16, it indicates a moderate correlation between variables. An Eta-squared below 0.06 suggests a weak correlation. From Table 3, the Eta-squared for both “quarter” and “hour” exceed 0.06, being 0.0714 and 0.0877, respectively, indicating a moderate correlation with port load. A Spearman correlation coefficient with an absolute value greater than 0.2 indicates a certain level of correlation between variables, while a value below 0.2 suggests only a weak correlation or none. From Figure 9, variables with a correlation coefficient greater than 0.02 include the temperature, the 10 m wind speed, the all-sky clearness index, and the surface ultraviolet index, with respective values of -0.32 , -0.2 , -0.27 , and -0.27 . Due to the high degree of correlation between the all-sky clearness index and the surface ultraviolet index, selecting one of these variables suffices for model training. Thus, the all-sky clearness index was chosen

for subsequent model training. Combining the results of the above analysis, this study ultimately selected temperature, the 10 m wind speed, the all-sky clearness index, the quarter, and the hour as the feature variables for training the prediction model.

3.2. VMD Decomposition

To mitigate the impact of volatility in the original load data and get accurate prediction results, VMD decomposition was used to stabilize the original load data and reduce noise in the time series.

To avoid excessive decomposition of the original load data, the suitable number of modal decompositions is determined by gradually increasing the number of modal decompositions and observing the changes in the center frequencies of the components. As shown in Table 4, when the decomposition number K reaches 5, the center frequency stabilizes at a value of 4.192×10^{-1} . At a decomposition number of 6, the center frequency is 4.197×10^{-1} , very close to the center frequency when the decomposition number is 5, indicating over-decomposition. Therefore, set the VMD modal decomposition number to 5.

Table 4. Center frequency variation of VMD decomposition.

Number of Modal Decomposition	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
$K = 1$	1.574×10^{-5}	—	—	—	—	—
$K = 2$	1.563×10^{-5}	2.473×10^{-1}	—	—	—	—
$K = 3$	1.538×10^{-5}	1.621×10^{-1}	3.221×10^{-1}	—	—	—
$K = 4$	1.392×10^{-5}	8.196×10^{-2}	2.437×10^{-1}	3.337×10^{-1}	—	—
$K = 5$	1.379×10^{-5}	8.173×10^{-2}	1.663×10^{-1}	3.308×10^{-1}	4.192×10^{-1}	—
$K = 6$	1.373×10^{-5}	8.172×10^{-2}	1.652×10^{-1}	2.501×10^{-1}	3.326×10^{-1}	4.197×10^{-1}

After determining the number of modal decompositions, the noise tolerance parameter τ is set with the objective of minimizing distortion. The specific parameter determination method utilized the Residual Energy Index (REI) method [39] and the calculation formula shown in Equation (16). The noise tolerance parameter τ increases in steps of 0.01 from 0 to 1. When τ reaches 0.76, the REI achieves a minimum value of 5.171×10^{-7} . Thus, the noise tolerance is set to 0.76, and the penalty coefficient for VMD decomposition is empirically set to 2000.

$$REI = \frac{1}{N} \sum_{i=1}^N \left[\sum_{K=1}^K U_K - f \right] \quad (16)$$

where K represents the number of decomposed modal components; U represents the modal components; f represents the original load sequence; N represents the number of sampling points.

The overall results of VMD are shown in Figure 10, with the decomposition results for the last 1000 data sequences displayed in Figure 11.

Figure 11 demonstrates that the load sequence has been effectively decomposed into multiple subsequences with distinct frequency characteristics. The IMF1 low-frequency component reveals the overall trend of the data changes, while IMF2 and IMF3 represent mid-frequency components; IMF4 and IMF5 are high-frequency components, representing local fluctuations and noise. From the original signal, it is evident that the distortion level after decomposition is relatively low, with an average percentage error of 0.03%, indicating that a considerable amount of data information has been preserved.

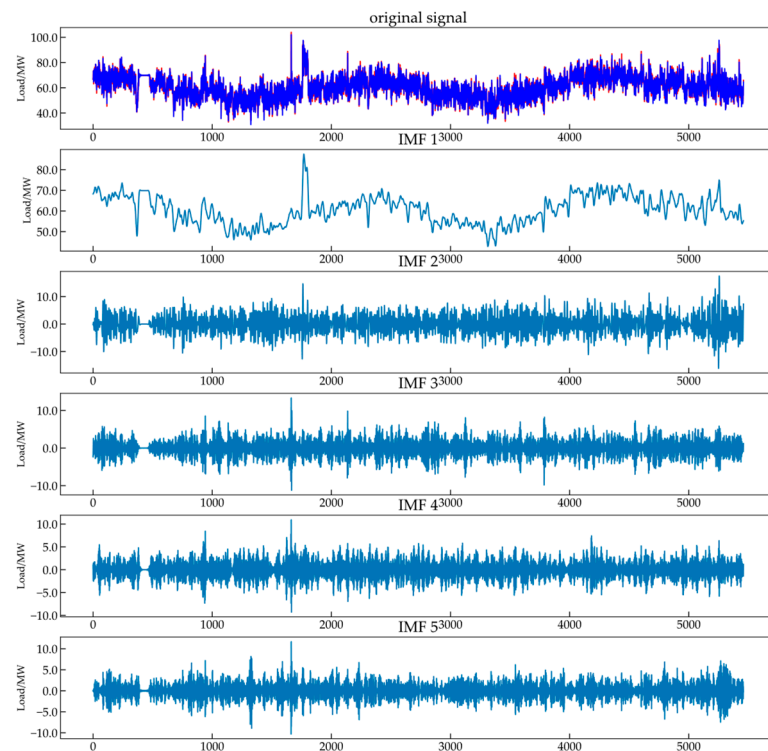


Figure 10. Overall VMD decomposition results for load data. The blue line represents the decomposed load curve, while the red line represents the original load curve.

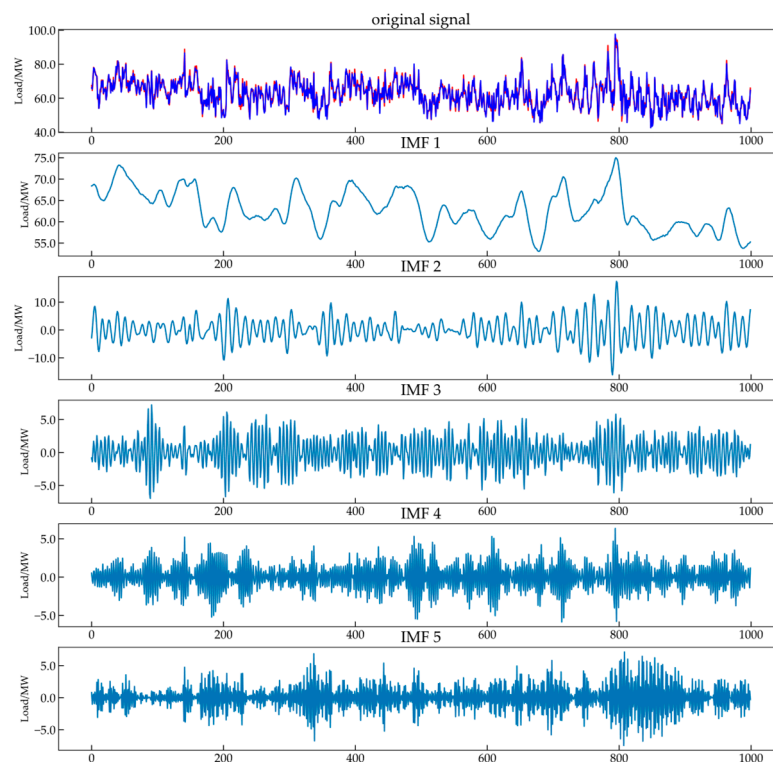


Figure 11. Decomposition results for the last 1000 data points of load data. The blue line represents the decomposed load curve, while the red line represents the original load curve.

3.3. Model Evaluation Metrics

To compare the performance and prediction effectiveness of different models, this study employs R-squared (R^2), the mean squared error (MSE), and the mean absolute per-

centage error (MAPE) for the quantitative assessment of each model's prediction accuracy. Smaller values of R^2 , MSE, and MAPE indicate better model prediction accuracy. The calculation methods for each metric are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (17)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (18)$$

$$MAPE = \frac{1}{n} \left(\sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \times 100\% \quad (19)$$

4. Case Study

4.1. Dataset Processing

The experimental dataset includes load data recorded by smart meters in a coastal port in China for the whole port area from 1 January 2021 to 31 March 2022. Additionally, port meteorological temperature data, 10 m wind speed data, sunshine clarity index data with the load data sampling times were collected from the NASA website.

First, the dataset was divided into a training set and a testing set, with 80% of the data allocated for training and 20% for testing. Next, the continuous variables in both the training and testing sets were standardized using the z-score method to eliminate differences in measurement units among feature variables, as shown in Equation (20). Discrete data, such as temporal feature factors, were processed using the one-hot encoding to avoid interference from the magnitude relationships among features in the model training. The encoding method is illustrated in Table 5. Finally, time-sliding windows were set for both the training and testing sets. The training set was then fed into the TCN-LSTM model for training. After the model weights were trained, the model was validated on the testing set, and the relevant evaluation metrics were calculated.

$$x'_t = \frac{x_t - \bar{X}}{\sigma} \quad (20)$$

Table 5. Example of one-hot encoding.

Quarter	One-Hot Encoding
First Quarter	[1,0,0,0]
Second Quarter	[0,1,0,0]
Third Quarter	[0,0,1,0]
Fourth Quarter	[0,0,0,1]

4.2. Hyperparameter Tuning of Forecasting Model

In this study, the random search (RS) algorithm was used to optimize the hyperparameters of the TCN-LSTM model. Unlike grid search, random search does not exhaustively explore the predefined parameter space. Instead, it randomly samples from the parameter space based on the performance metrics predicted on the validation set to find the optimal hyperparameter combination. The main advantage of this method is that it can efficiently explore a wide range of parameter spaces, reducing computational costs.

The TCN-LSTM model requires the tuning of both structural hyperparameters and training hyperparameters. This study utilized the advantages of RS to optimize some key hyperparameters to reduce the complexity of model tuning and improve training speed. The structural hyperparameters of the TCN include the time window size, the number of convolutional layers, and the kernel size. For LSTM, the hyperparameters

include the number of LSTM layers and the number of units per layer. The training hyperparameters encompass the initial learning rate and the dropout rate. Dropout is a regularization technique used during the training of deep neural networks, which randomly sets a portion of the neuron outputs to zero to reduce inter-neuron dependency, thereby preventing overfitting to some extent. The learning rate decay strategy adopted is the cosine annealing method, the activation function and optimizer are ReLU and Adam, with excellent performance, the loss function is the MSE, and the training process uses the early stopping method to prevent model overfitting, with the tolerance set to 10. The early stopping method prevents model overfitting by monitoring the performance on the validation set during model training and stopping training early when the validation performance stops improving or starts to get worse.

In summary, this study utilized the RS algorithm to optimize the above structural and training hyperparameters of the TCN-LSTM model. Finally, the model with the optimal combination of hyperparameters was used to forecast the load.

4.3. Comparison of Prediction Effects Based on Multifeature

To verify the enhancement of model prediction performance by incorporating multiple features, the TCN-LSTM model was used to conduct predictions under the following two scenarios: (1) a univariate prediction based solely on historical load data, without utilizing the VMD algorithm for decomposition; (2) a multivariate prediction based on multiple features (temperature, 10 m wind speed, the sunshine clarity index, quarters, hours), but without decomposition of the historical load data through the VMD algorithm.

The models of the two scenarios were tuned for hyperparameters using the random search algorithm and the early stopping method. The hyperparameter combination that performed best on the validation set was selected for prediction on the test set. The hyperparameter optimization results are shown in Table 6. By combining the learning loss curves of the models with the optimal hyperparameter combinations under the two scenarios in Figure 12, it is found that the learning loss curves of the validation set do not have a significant upward trend when the model terminates the training and do not exhibit overfitting.

Table 6. The hyperparameter optimization results of two scenarios.

Prediction Model	Number of Convolutional Layers	Kernel Size	Number of LSTM Layers	Number of Units	Initial Learning Rate	Dropout Rate	The Time Window Size
Univariate TCN-LSTM	3	2	1	64	0.0001	0.1	12
Multivariate TCN-LSTM	3	3	1	84	0.0005	0.15	12

Once training was completed, the models were used to predict the load on the test set. The predictions for the load of a specific port over two days, including 24 sampling points, are illustrated in Figure 13. The overall evaluation metrics for the test set predictions are presented in Table 7. After hyperparameter optimization through random search and preventing model overfitting through the early stopping method, the TCN-LSTM model with multi-feature input better fits the real load values and trends of the testing set efficiently. Furthermore, all performance indicators surpass those achieved in the prediction scenario (1), demonstrating that introducing multiple feature variables can effectively enhance the TCN-LSTM model's accuracy in predicting port electricity load, as shown in Table 7.

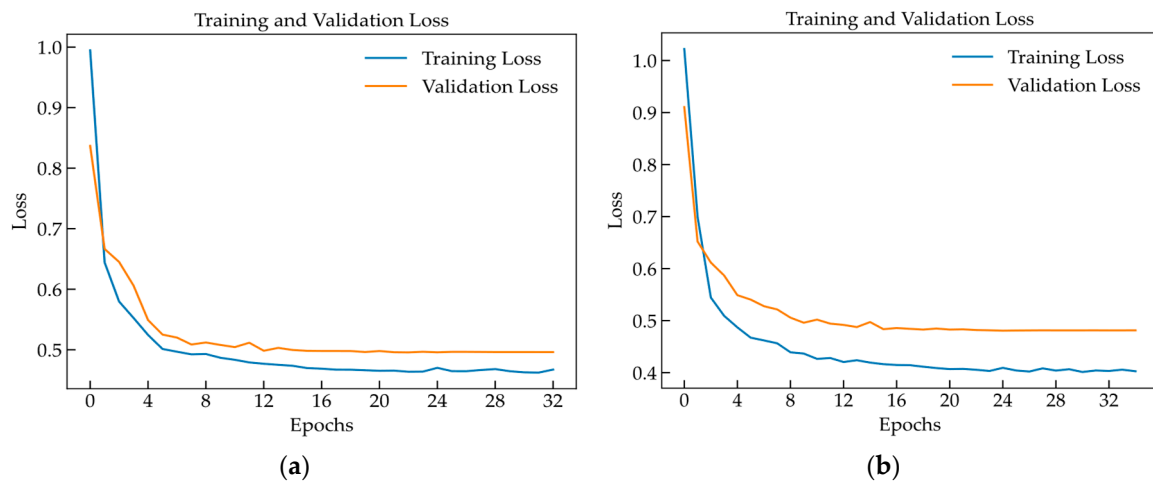


Figure 12. (a) The learning loss curves of the univariate prediction; (b) the learning loss curves of the multivariate prediction.

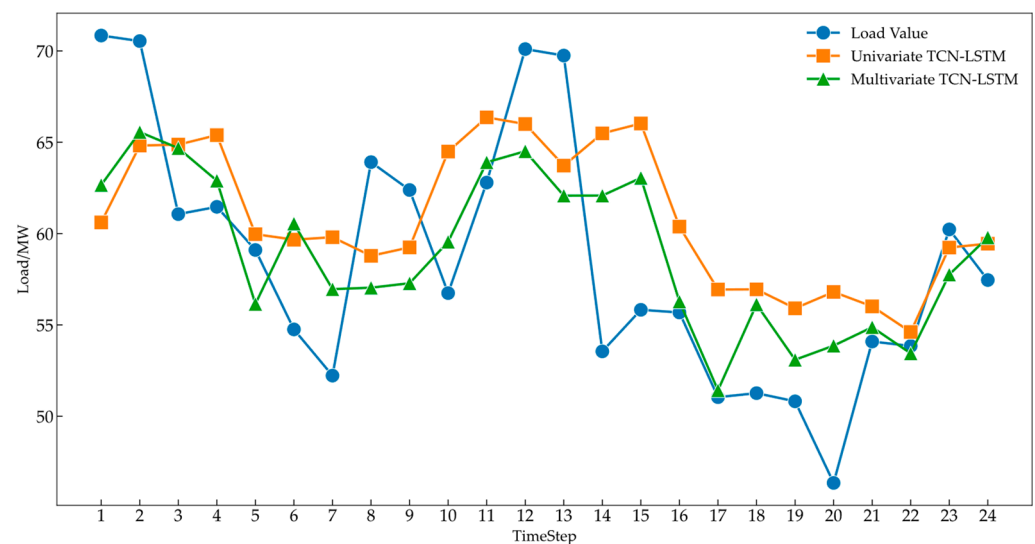


Figure 13. Comparison of model prediction results.

Table 7. Model evaluation results.

Prediction Model	R ²	MSE	MAPE
Univariate TCN-LSTM	0.48	39.77	7.78%
Multivariate TCN-LSTM	0.55	35.28	7.16%

4.4. Comparative Evaluation of Decomposition Algorithms

As observed from Figure 13, the inherent high volatility and unpredictability of the load sequence can lead to significant errors during peak and trough periods. This study employed decomposition algorithms to mine deep feature information within the load sequence to enhance prediction accuracy, thereby improving the model's adaptability to load fluctuations. To determine which decomposition algorithm is more suitable for uncovering the internal characteristics of port load sequences, this study conducted predictions using the TCN-LSTM model combined with the VMD, EEMD, and CEEMDAN decomposition algorithms. The prediction results of each model on the test set after training are shown in Figure 14, with the evaluation metrics provided in Table 8.

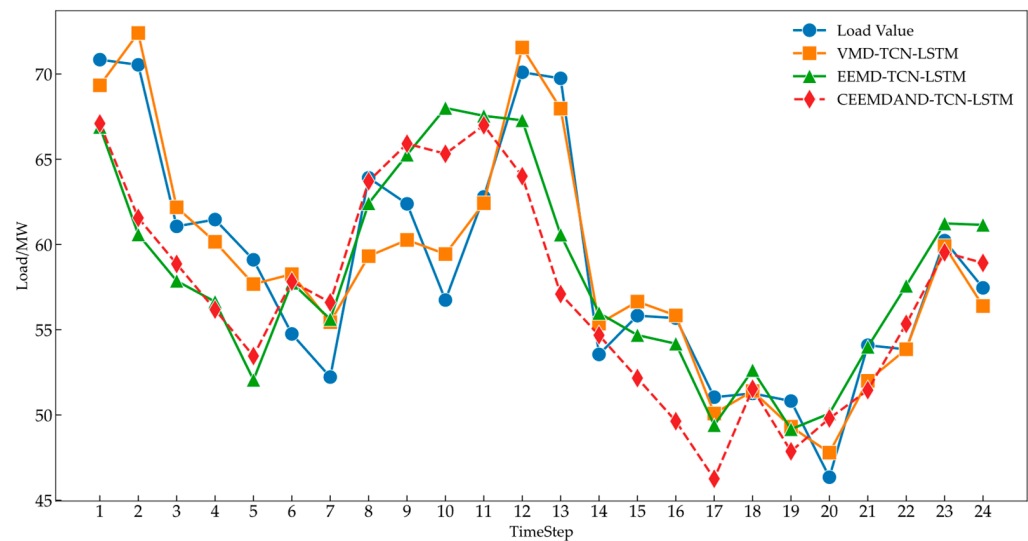


Figure 14. Prediction results with different decomposition algorithms.

Table 8. Evaluation results of predictions using different decomposition algorithms.

Prediction Model	R ²	MSE	MAPE
CEEMDAND-TCN-LSTM	0.72	19.77	5.71%
EEMD-TCN-LSTM	0.78	15.56	4.86%
VMD-TCN-LSTM	0.94	3.59	2.36%

Compared to the evaluation metrics of the TCN-LSTM model presented in Table 7, all three modal decomposition algorithms have improved the prediction accuracy of the TCN-LSTM model. However, due to the inability of EEMD to avoid the mode-mixing problem during decomposition, it falls short in extracting sufficient features from the sequence data, leading to subpar prediction accuracy. In contrast, the VMD-TCN-LSTM model significantly outperforms both the EEMD-TCN-LSTM and CEEMDAND-TCN-LSTM models in terms of R², MSE, and MAPE. Moreover, it better fits the actual load values during peak and trough periods, indicating that the VMD decomposition technology, relative to EEMD and CEEMDAND, can more effectively reduce the random volatility of load signals. This improves the model's prediction performance and enables the efficient mining of the load data's internal feature information.

4.5. Comparison of Different Prediction Models

The VMD-TCN-LSTM model proposed in this study is also compared with other commonly used time series prediction models, including the GRU, LSTM, XGBoost, and VMD-LSTM, to analyze their prediction performance. The multi-feature input method was used for all models. The results are illustrated in Figure 15, with the evaluation metrics detailed in Table 9.

The comparison reveals significant advantages of the model proposed in this study across key evaluation metrics, including R², the MSE, and the MAPE. When handling actual load data, the VMD-TCN-LSTM model, compared to the single LSTM model, shows a reduction in the MAPE by 4.99%, a decrease in the MSE by 33.38, and R² is improved by 0.43. This confirms the effectiveness of the TCN model in mining the potential temporal features of load sequences. Furthermore, the stabilization process of load sequences via VMD can further enhance the model's prediction performance, affirming the significant role of executing VMD decomposition in boosting the model's predictive capability. Thus, the short-term port load forecasting method proposed in this study demonstrates superior predictive performance.

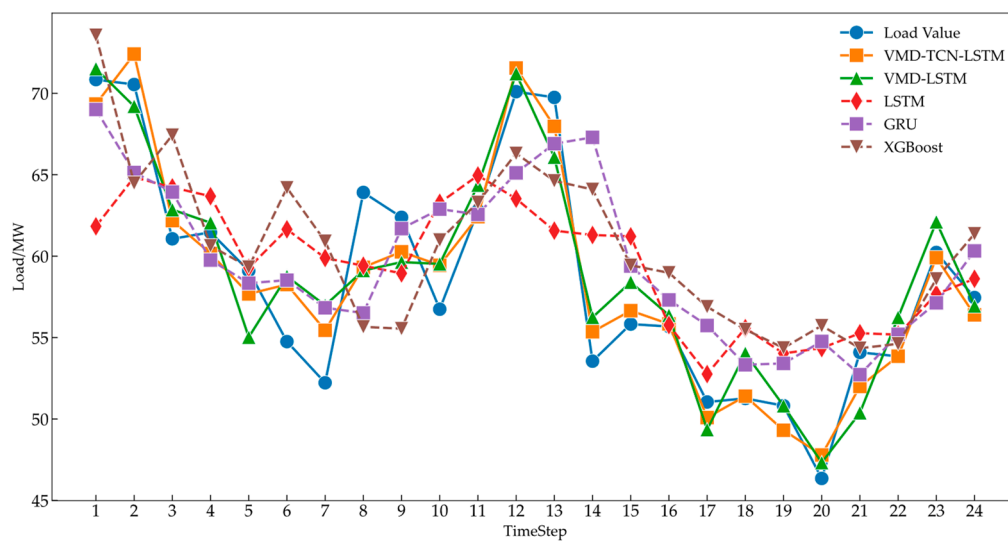


Figure 15. Comparison of prediction results among common forecasting models.

Table 9. Comparative evaluation results of common forecasting models.

Prediction Model	R ²	MSE	MAPE
LSTM	0.50	36.97	7.35%
XGBoost	0.52	35.49	7.41%
GRU	0.53	35.32	7.28%
VMD-LSTM	0.90	5.98	2.93%
VMD-TCN-LSTM	0.94	3.59	2.36%

5. Discussion

This study validates the proposed predictive model through a practical case analysis. The results demonstrate that the model can effectively predict short-term load trends in ports, which is crucial for sustainable energy management and planning in these areas.

By extracting key features across different frequencies from the original load data, the VMD method has proven to be effective in processing port load data. Moreover, while previous studies employing single predictive models have achieved acceptable results, combining the strengths of multiple models may reveal greater potential in future research. Exploring how to effectively integrate different predictive models will be an important direction for future load forecasting research. Additionally, analyzing key features that influence port load is also critical for port load forecasting. Choosing appropriate and highly relevant feature inputs can effectively enhance the model's accuracy.

Finally, this study also has some shortcomings. Usually, to ensure the generalization ability of the model, deep learning requires a relatively large training sample size. The port's actual load data used in this study may be insufficient. Therefore, it is essential to collect longer period load series data to further validate and optimize the model in future studies. However, the superiority of the forecasting method proposed in this study can also be demonstrated by comparing the experimental results with other common forecasting models in the case study. In addition, the load data of this study are from the coastal ports of China. Considering the differences in operation between coastal and inland ports, whether the model can be directly applied to inland ports still needs to be further verified. Future studies should be extended to different types of ports to enhance its applicability.

6. Conclusions

To enhance the accuracy of short-term port load forecasting and support port energy planning and management, this study proposes the VMD-TCN-LSTM model based on

multiple features. Through theoretical analysis and case studies, the following conclusions are drawn:

- By applying VMD decomposition to port load sequence data, the model excels in handling the high volatility and non-linearity of port load sequences. It significantly enhances the accuracy of short-term load forecasting.
- In addition to load data, meteorological factors and temporal characteristics are also included as prediction inputs, further improving the accuracy of the prediction model.
- The TCN-LSTM model effectively addresses the limitations of the single LSTM model in insufficiently extracting local features from sequence data, enabling the model to capture time dependencies across different ranges and further optimizing its understanding and predictive capabilities for time series data.

By improving the accuracy of short-term load forecasting through the method proposed in this study, ports can better manage their energy resources, thereby optimizing energy consumption, reducing operational costs, and minimizing environmental impact. This forecasting model provides port authorities and operators with the necessary predictive insights, enabling them to make informed decisions about energy procurement, infrastructure investments, and daily operational adjustments.

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

Funding: This research was funded by the National Key R&D Program of China, grant number 2021YFB2601300 and the Opening Foundation of the Laboratory of Transport Pollution Control and Monitoring Technology (2024) JH-F005.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to acknowledge the profound guidance and continuous support of Ming Zhong of Wuhan University of Technology in this research. Special thanks to Dung David Chuwang for his valuable assistance in polishing.

Conflicts of Interest: The authors declare that there are no conflicts of interests regarding the publication of this paper.

References

1. Platias, C.; Spyrou, D. EU-Funded energy-related projects for sustainable ports: Evidence from the port of Piraeus. *Sustainability* **2023**, *15*, 4363. [\[CrossRef\]](#)
2. Yu, P.; Wang, Z.Y.; Guo, Y.F.; Tai, N.L.; Wu, J. Application prospect and key technologies of digital twin technology in the integrated port energy system. *Front. Energy Res.* **2023**, *10*, 1044978. [\[CrossRef\]](#)
3. Pu, Y.; Chen, W.; Zhang, R.; Liu, H. Optimal operation strategy of port integrated energy system considering demand response. In Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October 2020.
4. Zhang, Q.; Qi, J.; Zhen, L. Optimization of integrated energy system considering multi-energy collaboration in carbon-free hydrogen port. *Transp. Res. E Logist Transp. Rev.* **2023**, *180*, 103351. [\[CrossRef\]](#)
5. Dimitriadis, C.N.; Tsimopoulos, E.G.; Georgiadis, M.C. Optimization-based economic analysis of energy storage technologies in a coupled electricity and natural gas market. *J. Energy Storage* **2023**, *58*, 106332. [\[CrossRef\]](#)
6. Anh, N.T.N.; Anh, N.N.; Thang, T.N.; Solanki, V.K.; Crespo, R.G.; Dat, N.Q. Online SARIMA applied for short-term electricity load forecasting. *Appl. Intell.* **2024**, *54*, 1003–1019. [\[CrossRef\]](#)
7. Dhaval, B.; Deshpande, A. Short-term load forecasting with using multiple linear regression. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 3911–3917. [\[CrossRef\]](#)
8. Nano, H.; New, S.; Cohen, A.; Ramachandran, B. Load forecasting using multiple linear regression with different calendars. In *Distributed Energy Resources in Microgrids*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 405–417. [\[CrossRef\]](#)
9. Jeong, D.; Park, C.; Ko, Y.M. Short-term electric load forecasting for buildings using logistic mixture vector autoregressive model with curve registration. *Appl. Energy* **2021**, *282*, 116249. [\[CrossRef\]](#)

10. Feng, R.; Xue, Y.; Wang, W.; Xiao, M. Saturated load forecasting based on clustering and logistic iterative regression. *Electr. Power Syst. Res.* **2022**, *202*, 107604. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)
12. Liu, Y.; Li, Y.; Li, G.; Lin, Y.; Wang, R.; Fan, Y. Review of multiple load forecasting method for integrated energy system. *Front. Energy Res.* **2023**, *11*, 1296800. [\[CrossRef\]](#)
13. Figlan, M.; Markus, E. Short term load demand forecasting for transnet port terminal (TPT) in east London using artificial neural network. *Int. J. Simul.* **2020**, *21*, 1–6. [\[CrossRef\]](#)
14. Caicedo-Vivas, J.S.; Alfonso-Morales, W. Short-term load forecasting using an LSTM neural network for a grid operator. *Energies* **2023**, *16*, 7878. [\[CrossRef\]](#)
15. Aduama, P.; Zhang, Z.; Al-Sumaiti, A.S. Multi-feature data fusion-based load forecasting of electric vehicle charging stations using a deep learning model. *Energies* **2023**, *16*, 1309. [\[CrossRef\]](#)
16. Abumohsen, M.; Owda, A.Y.; Owda, M. Electrical load forecasting using LSTM, GRU, and RNN algorithms. *Energies* **2023**, *16*, 2283. [\[CrossRef\]](#)
17. Qiu, X.; Ru, Y.; Tan, X.; Chen, J.; Chen, B.; Guo, Y. A k-nearest neighbor attentive deep autoregressive network for electricity consumption prediction. *Int. J. Mach. Learn. Cyber.* **2024**, *15*, 1201–1212. [\[CrossRef\]](#)
18. Pelekis, S.; Seisopoulos, I.K.; Spiliotis, E.; Pountridis, T.; Karakolis, E.; Mouzakitis, S.; Askounis, D. A comparative assessment of deep learning models for day-ahead load forecasting: Investigating key accuracy drivers. *Sustain. Energy Grids Netw.* **2023**, *36*, 101171. [\[CrossRef\]](#)
19. Upadhyay, A.; Garg, D.; Singh, M. Short term load forecasting for smart grids using apache spark and a modified transformer model. *Comput. Inform.* **2023**, *42*, 75–97. [\[CrossRef\]](#)
20. Buratto, W.G.; Muniz, R.N.; Nied, A.; González, G.V. Seq2Seq-LSTM with attention for electricity load forecasting in Brazil. *IEEE Access* **2024**, *12*, 30020–30029. [\[CrossRef\]](#)
21. Sheng, Z.; An, Z.; Wang, H.; Chen, G.; Tian, K. Residual LSTM based short-term load forecasting. *Appl. Soft Comput.* **2023**, *144*, 110461. [\[CrossRef\]](#)
22. Wang, B.; Wang, X.; Wang, N.; Javaheri, Z.; Moghadamnejad, N.; Abedi, M. Machine learning optimization model for reducing the electricity loads in residential energy forecasting. *Sustain. Comput.* **2023**, *38*, 100876. [\[CrossRef\]](#)
23. Saeed, F.; Paul, A.; Seo, H. A hybrid channel-communication-enabled CNN-LSTM model for electricity load forecasting. *Energies* **2022**, *15*, 2263. [\[CrossRef\]](#)
24. Jiang, L.; Wang, X.; Li, W.; Wang, L.; Yin, X.; Jia, L. Hybrid multitask multi-information fusion deep learning for household short-term load forecasting. *IEEE Trans. Smart Grid* **2021**, *12*, 5362–5372. [\[CrossRef\]](#)
25. Zheng, Q.; Zheng, J.; Mei, F.; Gao, A.; Zhang, X.; Xie, Y. TCN-GAT multivariate load forecasting model based on SHAP value selection strategy in integrated energy system. *Front. Energy Res.* **2023**, *11*, 1208502. [\[CrossRef\]](#)
26. Sulaiman, S.M.; Jeyanthi, P.A.; Devaraj, D.; Shihabudheen, K.V. A novel hybrid short-term electricity forecasting technique for residential loads using empirical mode decomposition and extreme learning machines. *Comput. Electr. Eng.* **2022**, *98*, 107663. [\[CrossRef\]](#)
27. Lin, G.; Lin, A.; Cao, J. Multidimensional KNN algorithm based on EEMD and complexity measures in financial time series forecasting. *Expert Syst. Appl.* **2021**, *168*, 114443. [\[CrossRef\]](#)
28. Alfieri, L.; Falco, D.P. Wavelet-based decompositions in probabilistic load forecasting. *IEEE Trans. Smart Grid* **2020**, *11*, 1367–1376. [\[CrossRef\]](#)
29. Zhang, Z.; Hong, W. Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads. *Knowl. Based Syst.* **2021**, *228*, 107297. [\[CrossRef\]](#)
30. Nhat, N.N.V.; Huu, D.N.; Hoai, T.N.T. Evaluating the EEMD-LSTM model for short-term forecasting of industrial power load: A case study in Vietnam. *Int. J. Renew. Energy Dev.* **2023**, *12*, 881–890. [\[CrossRef\]](#)
31. Wang, N.; Li, Z. Short term power load forecasting based on BES-VMD and CNN-Bi-LSTM method with error correction. *Front. Energy Res.* **2023**, *10*, 1076529. [\[CrossRef\]](#)
32. Yang, W.; Sun, Y.; Gao, Y.; Ren, J. Short-term forecasting of port electric power load based on wind condition similarity. *Electr. Power. Sci. Eng.* **2017**, *33*, 31–35.
33. Wang, Y.; Sun, S.; Chen, X.; Zeng, X.; Kong, Y.; Chen, J.; Guo, Y.; Wang, T. Short-term load forecasting of industrial customers based on SVM and XGBoost. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106830. [\[CrossRef\]](#)
34. Liu, M.; Qin, H.; Cao, R.; Deng, S. Short-term load forecasting based on improved TCN and DenseNet. *IEEE Access* **2022**, *10*, 115945–115957. [\[CrossRef\]](#)
35. Limouni, T.; Yaagoubi, R.; Bouziane, K.; Guissi, K.; Baali, E.H. Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. *Renew. Energy* **2023**, *205*, 1010–1024. [\[CrossRef\]](#)
36. Zhang, Z.; Hong, W.; Li, J. Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm. *IEEE Access* **2020**, *8*, 14642–14658. [\[CrossRef\]](#)
37. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)

38. Camacho, J.; Vitale, R.; Morales-Jiménez, D.; Gómez-Llorente, C. Variable-selection ANOVA simultaneous component analysis (VASCA). *Bioinformatics* **2023**, *39*, btac795. [[CrossRef](#)]
39. Li, J.; Cheng, X.; Li, Q.; Meng, Z. adaptive energy-constrained variational mode decomposition based on spectrum segmentation and its application in fault detection of rolling bearing. *Signal Process.* **2021**, *183*, 108025. [[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.