

Research article

Short-term multi-energy load forecasting of integrated energy systems based on temporal convolutional network and multi-task learning

He Jiang*, Tianhui Shi and Lixin Yuan

School of Renewable Energy, Shenyang Institute of Engineering, Shenyang 110136, China

* **Correspondence:** Email: jianghescholar@163.com.

Abstract: Accurate forecasting of electrical, heating, and cooling demands is crucial for the optimal design, operation, and coordinated planning of multi-energy coupling systems. In this study, we present a short-term multi-energy load forecasting method for integrated energy systems based on a temporal convolutional network and multi-task learning. First, the Pearson correlation coefficient was used to analyze the correlations among meteorological variables. Then, the maximum overlap discrete wavelet transform (MODWT) module decomposed the input time-series data into components at different frequency levels. These decomposed time series datasets were input into the multi-task learning model for sparse sharing. Subsequently, the learned time series datasets of different frequencies were input into the temporal convolutional network for temporal scale feature extraction. To improve prediction accuracy, the improved energy valley optimizer (IEVO) was introduced to optimize the model parameters. The feature vectors output by the optimized temporal convolutional network were then fed into the corresponding Transformer models for adaptive weight adjustment. Finally, the MODWT module reconstructed the predictions at different frequency levels to obtain the final multi-energy load forecasting results.

Keywords: integrated energy systems; multi-energy load forecasting; multi-task learning; temporal convolutional network; improved energy valley optimizer

1. Introduction

As the global intensification of the greenhouse effect, the continuous depletion of fossil fuels, and growing ecological issues [1], the establishment of an Integrated Energy System (IES) that is clean,

efficient, and safe, with the goal of reducing carbon emissions, has become imperative [2]. Traditional energy systems in their singular form lead to low energy efficiency and poor dispatch flexibility [3]. In contrast, an IES can meet energy demands while boosting the efficiency of energy utilization and reducing carbon footprints via the coordinated planning and adaptable management of multiple energy forms, including electrical, heating, and cooling [4]. Precise prediction of multi-energy loads proves essential to the efficient operation of IES, as it helps maintain the balance between electrical, heating, and cooling load supply and requirement. Enhancing the accuracy of these predictions is vital for optimizing the performance of IES [5].

Conventional methods for short-term multi-energy load forecasting fall into statistical modeling techniques and machine learning methods. Statistical modeling methods are simple to model and computationally efficient, performing well in handling linear data but showing limitations with nonlinear data [6]. Therefore, although statistical models provide a basic forecasting framework, their ability to describe the nonlinear and non-stationary characteristics of multi-energy load series is limited. In contrast, machine learning methods such as Extreme Learning Machine (ELM) [7], Least Squares Support Vector Machine (LSSVM) [8], and Random Forest (RF) [9] possess robustness and non-linear fitting capabilities, making them dominant in short-term load forecasting with satisfactory results. These methods improve the nonlinear fitting ability compared with statistical models, but their performance depends greatly on feature selection and they are not sufficiently effective in capturing deep temporal dependency information. However, the large-scale incorporation of renewable energy sources and the coupled connections between multi-energy loads within IES have led to increasingly complex and variable load fluctuations. As a subfield under the development of machine learning systems, deep learning frameworks, with their capability to accurately model nonlinear characteristics, has been widely applied in short-term load prediction [10]. Deep learning models mostly include Long Short-Term Memory network (LSTM) [11], Gated Recurrent Unit (GRU) [12], and Temporal Convolutional Network (TCN) [13]. Among them, LSTM and GRU are effective in learning sequential information and have been widely used in load forecasting. Zhang et al. [14] proposed a bi-directional GRU-enhanced twin support vector regression model with a seasonal mechanism for electric load forecasting, further confirming the effectiveness of GRU-based temporal modeling in this field. Studies on renewable-energy forecasting further support the effectiveness of deep sequential models and hybrid temporal architectures in strongly nonlinear and weather-sensitive energy prediction tasks. For example, Wang et al. [15] proposed a fusion model for enhanced ultra-short-term offshore wind power forecasting, showing that decomposition-based and BiGRU-based hybrid learning can significantly improve forecasting performance under complex offshore operating conditions. In addition, Xu et al. [16] developed a hybrid short-term and ultra-short-term wind power forecasting framework based on Weather Research and Forecasting (WRF), inverted Transformer (iTransformer), and particle swarm optimization (PSO), highlighting the importance of combining meteorological information, hybrid temporal modeling, and optimization strategies in energy forecasting.

Moreover, forecasting studies have increasingly incorporated graph neural networks, spatio-temporal dependency modeling, and heterogeneous data fusion. For example, a short-term regional electric vehicle (EV) charging load forecasting method based on graph attention network (GAT) and GRU with trip distribution estimation was developed to characterize regional travel behavior and charging-demand variations more accurately [17]. Moreover, a spatio-temporal data fusion framework based on a large language model was proposed to enhance the prediction of EV charging demand in smart grid management [18]. These studies illustrate the potential of graph-based learning, spatio-temporal fusion,

and emerging LLM-enhanced frameworks for complex demand forecasting. However, these methods mainly focus on EV charging demand rather than coordinated short-term forecasting of electricity, heating, and cooling loads in integrated energy systems. However, recurrent models such as LSTM and GRU may suffer from insufficient long-term information retention and relatively high computational burden when processing long load series. TCN provides another important solution using convolutional structures to extract temporal features more efficiently.

In reference [19], a predictive technique combining an improved TCN and a densely connected convolutional network was proposed. TCN effectively captures local dependencies in time series using dilated convolutions, thereby avoiding the gradient vanishing and gradient explosion problems commonly encountered in recurrent neural networks (RNNs). However, although TCN increases the receptive field through dilated convolution, it has limitations in handling long time series. This indicates that TCN is suitable for local temporal feature extraction, but it is necessary to strengthen its ability to capture global dependency relationships. The researchers in [20] proposed a predictive method integrating Patch and Transformer, where the Transformer module captures global dependency relationships within a sequence by leveraging its self-attention mechanism, thereby enhancing the effectiveness of processing long time series. However, the self-attention layer in the Transformer model requires substantial computational resources when dealing with long time series, which raises higher hardware requirements for practical applications. Thus, Transformer is advantageous in global feature modeling, but its computational cost limits its direct application in practical short-term multi-energy load forecasting. The integration of TCN and a Transformer enables the simultaneous capture of global and local time series patterns in practical applications. This approach reduces computational complexity to some extent while improving the model's prediction accuracy and stability.

Furthermore, due to factors such as differences in load physical characteristics, cross-energy coupling characteristics, and diverse user energy consumption patterns, multi-energy loads exhibit evident randomness and volatility [21]. To enhance the overall prediction accuracy, signal decomposition algorithms have gradually been applied in this field. For instance, the researchers in [22] utilized wavelet packet transform to carry out the decomposition of the load into parts with different frequency bands, calculated the cross-correlation coefficients of each part, and input the similar ones into the same neural network. This method verifies that frequency-domain decomposition can lower the complexity of the raw series, but the interaction among frequency components is not fully considered. In addition, Fan et al. [23] proposed a hybrid deep learning model for short-term power load forecasting based on feature extraction and statistical techniques, further demonstrating that extracting informative features from complex load sequences can effectively improve forecasting performance. In reference [24], a hybrid modal decomposition technique was proposed, which utilizes secondary modal decomposition for reducing the nonlinearity in the original load series, thus enhancing the forecasting precision of the model. Although modal decomposition can further weaken the non-stationarity of load data, it still mainly focuses on single-series preprocessing and does not sufficiently explore the coupling information among multiple energy loads. However, this signal decomposition method fails to deeply explore the intrinsic patterns of the original load data and reduce irrelevant features, hindering the achievement of required prediction accuracy in practical applications. Therefore, decomposition methods need to be combined with an effective multi-load learning framework so that multi-frequency characteristics and cross-energy coupling information can be fully utilized.

To improve the prediction accuracy of multi-energy load forecasting, it is necessary to examine the correlations among loads. In addition to the intrinsic characteristics of multi-energy loads, the

energy conversion and transmission processes among these loads involve complex dynamic temporal correlations. Conventional load forecasting methods typically have a single load type focus, such as electrical [25], heating [26], or cooling [27]. However, these methods often fail to fully utilize the coupling characteristics between loads, neglecting correlations, which can easily reduce forecasting accuracy [28]. This shows that single-task forecasting models are difficult to meet the demand of IES multi-energy cooperative operation. To address this issue, the researchers in [29] explored multi-energy coupling information by using the product of electrical, cooling, and heating matrices as the coupling feature matrix. This method explicitly introduces coupling features, but the coupling representation is relatively fixed and may not adaptively reflect the dynamic differences among load types. Moreover, Multi-Task Learning (MTL) has shown significant effectiveness in solving this problem. In reference [30], a joint prediction model of electrical, heating, and cooling loads was constructed from MTL and LSSVM using the weight-sharing mechanism in MTL. This work proves the feasibility of shared learning among multiple load forecasting tasks, but the traditional LSSVM structure limits deep temporal feature extraction. In reference [31], MTL was applied with the LSTM model serving as a shared layer to efficiently acquire and learn information from different tasks. This method improves the ability of MTL to learn sequential information, but the hard-sharing structure may ignore the individual differences among electrical, heating, and cooling loads. However, relevant studies have indicated that these methods suffer from low computational efficiency, learning information only from inputs at previous time points, and long spans that forget early-learned content. Additionally, when the strength of coupling among multi-energy loads is weak, the forecasting model struggles to consider differences in coupling strength, affecting prediction accuracy. Therefore, an improved MTL structure is required to balance shared information and task-specific information more effectively.

To mitigate the impact of model parameter changes on prediction accuracy, some researchers have employed random search algorithms to optimize model parameters [32]. However, random search algorithms are prone to local optima, affecting prediction accuracy. The Energy Valley Optimizer (EVO) algorithm simulates the movement of unstable particles in the universe and offers advantages such as fast convergence speed, not needing special parameters, and resistance to local optima. It has been applied in problems such as photovoltaic system maximum power point tracking [33] and thermoelectric generator system reconstruction [34], achieving remarkable success. These applications demonstrate the optimization potential of EVO in energy-related problems. However, research on its application in prediction model parameter optimization remains limited. Therefore, introducing an improved EVO into the parameter optimization of deep forecasting models may further improve forecasting accuracy and model stability.

According to the foregoing analytical results, in this article, we construct a short-period multi-energy demand forecasting model for an IES, denoted MODWT-MTL-IETCNT, which integrates MODWT, MTL, the improved energy valley optimizer (IEVO), a TCN, and a Transformer-based attention mechanism. The core innovation of this study resides not in any single standalone module, but in the coordinated integration of these components into a unified forecasting pipeline for electricity, heating, and cooling loads in an IES. Compared with other studies, the proposed framework is designed to address several remaining challenges: A single model often cannot simultaneously capture local details and global dependencies; decomposition-based methods may fail to sufficiently exploit multi-frequency coupling information; existing MTL methods often have difficulty balancing shared representations and task-specific characteristics; and inappropriate parameter settings may significantly affect forecasting accuracy. First, the Pearson correlation

coefficient is used to analyze the relationships between multi-energy load data and meteorological data, and the meteorological factors that significantly affect multi-energy loads are selected as model inputs. Next, the MODWT decomposes the input time-series data into components at different frequency levels, which are input into the MTL model for sparse sharing. In this way, the proposed framework can reduce the non-stationarity of the original load series while preserving useful multi-frequency information. Then, the frequency-divided time-series components output by MTL are input into the TCN for temporal feature extraction, where the TCN module is mainly used to efficiently capture local temporal patterns. To reduce the impact of deep learning model parameters on prediction accuracy, IEVO is employed to optimize the key model hyperparameters. The feature vectors extracted by the TCN are then fed into the corresponding Transformer modules to adaptively adjust feature weights. The Transformer module further captures global dependency relationships and enhances the representation of long-term temporal information. Finally, MODWT is used to reconstruct the forecast data at different frequency levels, yielding the final multi-energy load forecasting results. A simulation study is carried out using IES data from Arizona State University. Experimental findings indicate that the designed framework significantly boosts the precision of short-term multi-energy load forecasting and outperforms the benchmark models in terms of generalization capability. The core innovation of this work resides not in any single standalone module, but in the coordinated integration of these components into a unified forecasting pipeline for electricity, heating, and cooling loads in an IES.

2. Principles and methods

Using MODWT on multi-energy load time series makes it easier to identify and select features highly correlated with the multi-energy load sequences [35], thereby reducing irrelevant and redundant features. Additionally, the IEVO is utilized to regulate the internal parameters of the TCN-Transformer network, leading to an improvement in the stability of the model's output and its generalization capability [36]. Analyzing the coupling relationships among electrical, heating, and cooling types of loads enables more accurate forecasting related to future energy load demands, which supports operation and scheduling of IES. The principles of MODWT, IEVO, and the analysis of multi-energy load interactive relationships are introduced as follows.

2.1. MODWT algorithm

The MODWT algorithm represents an improved version of the Discrete Wavelet Transform (DWT). DWT obtains approximation coefficients and detail coefficients through a low-pass filter $g(l)$ and a high-pass filter $h(l)$, respectively. At each decomposition level, only the low-frequency component of the original sequence is further decomposed. The calculation of the decomposition coefficients is formulated as follows:

$$S_j(k) = \sum_{l=0}^{L-1} g(l)S_{j-1}(2k-l), k=0, \dots, N_l-1 \quad (1)$$

$$T_j(k) = \sum_{l=0}^{L-1} h(l)S_{j-1}(2k-l), k=0, \dots, N_l-1 \quad (2)$$

where $S_j(k)$ represents the coarse approximation coefficients at the j_{th} level; $T_j(k)$ denotes the

detail coefficients at the j_{th} level; $l=0, \dots, L-1$, L is the filter length; and N_l is the length of the original sequence.

DWT requires a high level of stationarity for multi-energy load data sequences, and its boundary effects can cause distortion near the edges of these sequences. The MODWT algorithm's approximation and detail coefficients possess translation invariance, are highly redundant, and non-orthogonal, with no specific restrictions on sequence length, enabling more effective capture of the local features associated with multi-energy load data sequences. The computation of MODWT decomposition coefficients is formulated as follows:

$$S'_j(k) = \sum_{l=0}^{L-1} g'(l) S'_j(k-l \bmod N), k=0, \dots, N_l-1 \quad (3)$$

$$T'_j(k) = \sum_{l=0}^{L-1} h'(l) S'_j(k-l \bmod N), k=0, \dots, N_l-1 \quad (4)$$

where $S'_j(k)$ stands for the coarse approximation coefficients at the j_{th} scale; $T'_j(k)$ indicates the detail components at the j_{th} decomposition scale; $g'(l) = g(l)/\sqrt{2}$ is the low-pass filter of the MODWT; $h'(l) = h(l)/\sqrt{2}$ is the high-pass filter of the MODWT; and $\bmod$ denotes the modulo operation.

2.2. IEVO algorithm

Since the internal parameters of deep learning models significantly impact prediction accuracy, we utilize the IEVO algorithm for optimizing the internal parameters of the TCNT network. This approach rapidly finds the optimal parameter values for predicting electricity, heating, and cooling load data. IEVO is an algorithm that improves upon the pinhole imaging theory, simulating the movement of unstable corpuscles in cosmic space as they move in the direction of the stable domain or an energy trough through α -decay, β -decay, γ -decay, electron capture event, or positron emission effect.

The sequence of steps involved in the algorithm's operation is as follows:

Step 1: Randomly initialize the particle swarm W . Use the pinhole imaging theory to generate the improved corresponding point positions W' . Compare the fitness $F(W)$ of the positional information of both particle individuals, and select n particles with better fitness to form the final initial population, and set the maximum iteration limit to $Z_{t_{max}}$.

$$W = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_i \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} W_1^1 & W_1^2 & \dots & W_1^j & \dots & W_1^d \\ W_2^1 & W_2^2 & \dots & W_2^j & \dots & W_2^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ W_i^1 & W_i^2 & \dots & W_i^j & \dots & W_i^d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ W_n^1 & W_n^2 & \dots & W_n^j & \dots & W_n^d \end{bmatrix} \quad (5)$$

$$W_i^{j'} = ((K+1)(U_{bi,j} + L_{bi,j}) - 2Wbest_i^{j'}) / 2K \quad (6)$$

The pinhole imaging initialization strategy is adopted to generate opposite candidate solutions, thereby enhancing the diversity of the initial population and improving the global exploration

capability of IEVO in the early optimization stage.

$$F(\mathbf{W}) = 1 - \max \sqrt{(\mathbf{Y}^* - \mathbf{Y})^2} \quad (7)$$

where W_i represents the i_{th} particle; W_i^j denotes the positional information of the i_{th} particle in the j_{th} dimensional axis, $i \in [1, n], j \in [1, d]$; K is the virtual candle length ratio coefficient; $W_{best_i}^{j'}$ is the position of the most stable particle currently; $W_i^{j'}$ is the corresponding point position in the pinhole imaging theory; and $U_{bi,j}, L_{bi,j}$ are the dynamic boundaries of the i_{th} particle in the j_{th} dimension. $\max \sqrt{(\mathbf{Y}^* - \mathbf{Y})^2}$ represents the maximum mean square error of the model's electricity, heating, and cooling load outputs; $\mathbf{Y}^*$ represents actual electrical, heating, and cooling loads data; and $\mathbf{Y}$ represents predicted electrical, heating, and cooling loads data.

Step 2: Evaluate the fitness function score $f(w)$ of the particle, serving as the corpuscle's neutron enrichment degree NEL . Then, compute the particle's enrichment boundary EB and determine the particle's stability level SL .

$$f(w) = 1 / \sqrt{\frac{\sum_{n=1}^N (\hat{y}_n - y_n)^2}{N}}, NEL_i = f(w)_i \quad (8)$$

$$EB = \frac{\sum_{i=1}^n NEL_i}{n} \quad (9)$$

$$SL_i = \frac{NEL_i - BS}{WS - BS} \quad (10)$$

where $\hat{y}_n$ represents the n_{th} predicted data; y_n represents the n_{th} actual data; N represents the number of samples; NEL_i serving as the i_{th} particle's neutron enrichment index; $f(w)_i$ represents the fitness scores for the i_{th} particle; SL_i indicates the i_{th} particle's stability level; and WS and BS represent the fitness function values of the least stable and most stable particles, respectively.

Step 3: When $NEL_i > EB$, a random number $[0, 1]$ is adopted as the stability boundary SB of the particle swarm. When $SL_i > SB$, the following equation is used to update the positions of the particles after α -decay and γ -decay.

$$W_i^{New1} = W_i(W_{BS}(W_i^j)), \begin{cases} i = 1, 2, \dots, g \\ j = \text{Alpha Index II} \end{cases} \quad (11)$$

$$W_i^{New2} = W_i(W_{NG}(W_i^j)), \begin{cases} i = 1, 2, \dots, g \\ j = \text{Gamma Index II} \end{cases} \quad (12)$$

$$D_i^a = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \begin{cases} i = 1, 2, \dots, n \\ a = 1, 2, \dots, n-1 \end{cases} \quad (13)$$

where W_i^{New1} denotes for the i_{th} corpuscle's new position after undergoing α -decay; W_{BS} denotes the most stable corpuscle; D_i^a stands for the separation distance between the i_{th} particulate and its

adjacent a_{th} particulate; (x_1, y_1) and (x_2, y_2) indicate the particle unit's positional coordinates within the search space; W_i^{New2} indicates the i_{th} particle unit's position after undergoing γ -decay; and W_{NG} denotes the positional vector corresponding to the adjacent corpuscle of the i_{th} corpuscle, determined by D_i^a .

Step 4: When $SL_i < SB$, the following equation is used to update the position after β -decay. In β -decay, more unstable particles are generated, which emit β -rays and jump significantly in space.

$$W_{CP} = \frac{\sum_{i=1}^n W_i}{n} \quad (14)$$

$$W_i^{New1} = W_i + \frac{(r_1 \times W_{BS} - r_2 \times W_{CP})}{SL_i} \quad (15)$$

$$W_i^{New2} = W_i + (r_3 \times W_{BS} - r_4 \times W_{NG}) \quad (16)$$

where W_{CP} represents the corpuscle center; and W_i^{New1} represents the updated position of the i_{th} corpuscle following β -decay, where the corpuscle moves towards W_{BS} and W_{CP} in a controlled manner. r_1 and r_2 are random scalars within the interval $[0,1]$. To better simulate the particle motion in the IEVO algorithm, an additional position update mechanism is deployed for particulates undergoing β -decay. At this point, the revised position of the i_{th} corpuscle following β -decay is W_i^{New2} . This corpuscle travels in a controlled manner toward W_{BS} and W_{NG} , and SL_i does not affect the movement process, with r_3 and r_4 being random scalars within the interval $[0,1]$.

Step 5: When $NEL_i \leq EB$, the following equation is utilized to revise the particle's new position after shifting toward the stability zone through electron capture or positron emission.

$$W_i^{New} = W_i + r \quad (17)$$

where W_i^{New} represents the new position of the i_{th} particle after electron capture or positron emission; $r \in [0,1]$.

Step 6: Compute the fitness value and refresh the current particle's position.

Step 7: Check whether the preset accuracy condition is satisfied. If it is satisfied, exit and output the result W_{BS} ; otherwise, repeat Steps 2 to 6 until the preset accuracy is attained or the maximum iteration count is reached.

The flowchart of the IEVO algorithm for optimizing the internal parameters of the TCNT network is shown in Figure 1.

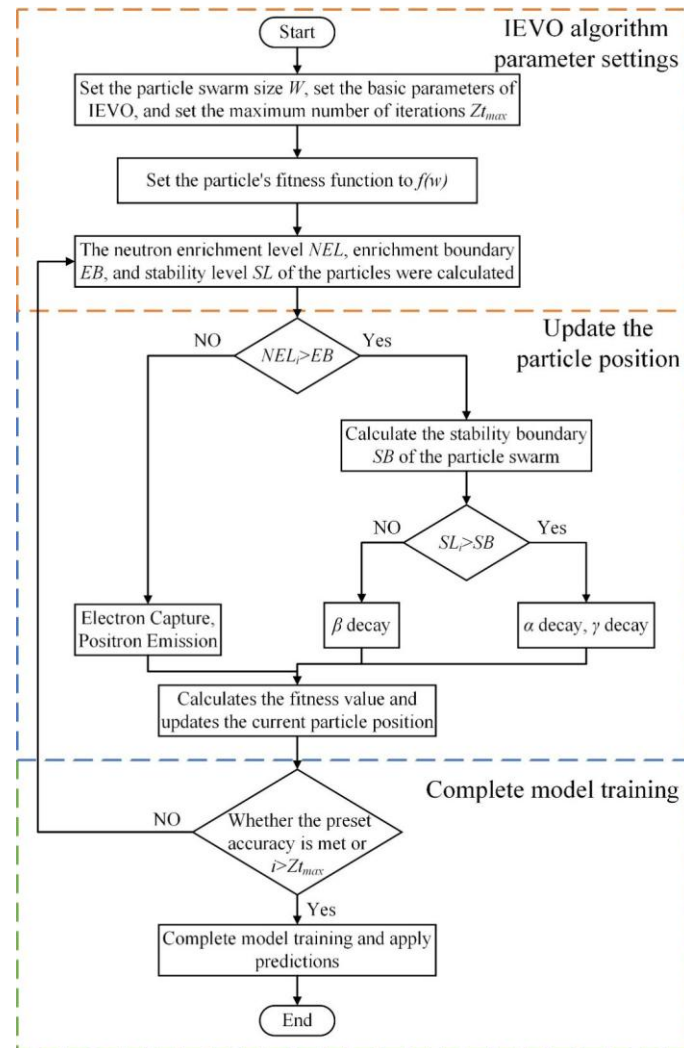


Figure 1. Flowchart of IEVO optimizing the TCNT network.

2.3. Analysis of multi-energy load coupling relationships

Traditional IES research typically focuses on a single form of load, often overlooking the coupling relationships among load types. However, modern IESs are gradually evolving into systems centered around the power subsystem. Utilizing various information technologies, these systems achieve horizontal “multi-source complementarity” of electrical, heating, and cooling, as well as vertical “source-grid-load-storage” coordination across all stages to maintain regional energy balance. They satisfy the diverse electrical, heating, and cooling load requirements on the terminal demand side, embodying interconnectedness and independence among each energy form. The specific energy flow relationships are presented in Figure 2.

Figure 2 explores the intrinsic coupling relationships among electrical, heating, and cooling loads. The electrical subsystem includes the external power grid, combined cooling, heating, and power (CCHP) units, photovoltaic (PV) systems, wind turbines (WTs), and energy storage. The heating subsystem consists of gas turbines, CCHP units, gas boilers, thermal storage devices, and electric heating devices. The cooling subsystem comprises electric cooling devices, CCHP units, absorption

chillers, and cool storage devices. Based on the interactive relationships among the electrical, heating, and cooling subsystems, the multi-energy coupling relationships are detailed and clarified, as shown in Table 1.

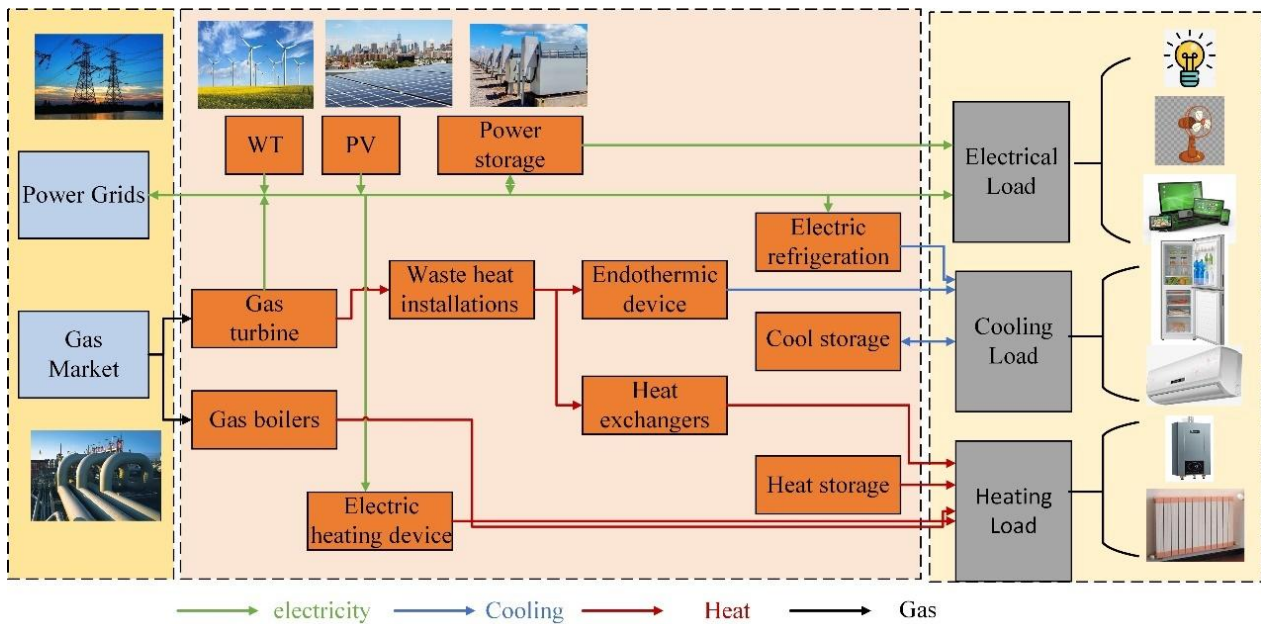


Figure 2. Energy flow relationship diagram of the integrated energy system.

Table 1. Coupling relationships of multiple energy types in the integrated energy system.

Energy subsystem	Energy input devices	Output to		
		Electrical subsystem	Thermal subsystem	Cooling subsystem
Electrical subsystem	External grid, WT, PV	Energy storage devices	Electric heating devices	Electric cooling devices
Thermal subsystem	External heating network	CCHP	Thermal storage Devices, Heat exchangers	Absorption chillers cooling subsystem
Cooling subsystem	CCHP	/	/	Energy storage devices

3. Multi-energy load forecasting model

The TCN model, optimized by IEVO, enhances its capability to process time series data and overcomes the gradient vanishing or exploding problems commonly encountered in traditional neural networks, thereby improving prediction accuracy. Integrating TCN with the Transformer attention mechanism enables the adaptive adjustment of feature weights, providing strong global learning capabilities and better extraction of globally relevant temporal features. Combining the MTL with the improved energy valley optimizer–temporal convolutional network–Transformer (IETCNT) model fully exploits the complex coupling information in the multi-energy conversion of IES. This integration

strengthens the impact of multi-energy load sequences with high correlation on the prediction results while reducing the influence of those with lower correlation. Consequently, it more effectively extracts features related to temporal information in multi-energy load sequences, thereby improving prediction accuracy.

3.1. MTL model

When using single-task learning for multi-energy load forecasting, multiple independent neural networks need to be established. Since the training processes are independent of each other, the model cannot fully account for the coupling characteristics among the loads [37]. In this paper, MTL is adopted to retrieve the coupling information among multi-energy loads that is difficult to capture using traditional methods through the sharing of underlying features [38]. It helps enhance the accuracy of each individual task.

Within the MTL, the correlation between tasks is leveraged through joint training of various tasks to improve the performance of each individual task. For this purpose, M target tasks and N data samples are constructed according to different load types, serving as the training dataset, denoted as $\{x_i, y_i^1, y_i^2, \dots, y_i^m, \dots, y_i^M\} (i=1, 2, \dots, N), m \in M$. Here, y_i^m represents the label of the i_{th} data in the m_{th} task. The MTL functional mapping is formulated as $f^m(x; \theta^{sh}, \theta^m): x \rightarrow y^m$, where θ^{sh} denotes the global shared parameters for load forecasting tasks, and θ^m denotes the task-specific parameters. The comprehensive optimization function of MTL is as follows:

$$\xi^m(\theta^{sh}, \theta^m) = \frac{\sum_{i=1}^N \xi(f^m(x_i; \theta^{sh}, \theta^m), y_i^m)}{N} \quad (18)$$

$$\min \sum_{m=1}^M \varpi^m \xi^m(\theta^{sh}, \theta^m) \quad (19)$$

where ϖ^m represents the task weight parameters, and $\xi^m(\theta^{sh}, \theta^m)$ denotes the loss function.

The information embedded in different tasks in a task learning process can be viewed as distinct tasks in MTL. To achieve a balance between the relevance and constraints between tasks, it is essential to simultaneously take into account the optimization goals of each task while aiming for improved learning results. However, tasks that are related do not imply similarity between the tasks; rather, they share common expressions that create interconnections among them.

The sharing mechanisms of MTL are divided into rigid and flexible sharing. Among them, the rigid sharing approach requires all subtasks use a unified feature-sharing layer and maintain consistency in feature parameters. This method is suitable for situations where there is a strong correlation between tasks, as it ensures consistent information transfer between tasks. Moreover, soft sharing provides greater flexibility, enabling each subtask to be equipped with different feature parameters based on its needs. However, this variability requires regularization techniques to coordinate effective information sharing among subtasks.

To better predict the close and complex coupling information among multi-energy loads, we apply a sparsity-sharing mechanism that integrates the advantages of both sharing methods. This mechanism can adapt to weak correlations between tasks and is suitable for situations with strong correlations. By employing the sparse-sharing MTL method, more accurately and efficiently multi-energy load forecasting can be achieved.

Sparsity sharing starts with an over-parameterized base network ψ . To handle multiple tasks, a subnetwork ψ^m is assigned to task m . A binary mask matrix θ_ψ is used to select a subnetwork for each task from the base network. Let $\ell_m \in \{0,1\}^{|\theta_\psi|}$ represent the mask for task m , then the parameters of the subnetwork connected to task m are $\ell_m \odot \theta_\psi$, where $\odot$ represent element-by-element multiplication. Thus, each task has a unique representation: $\psi^m(x) = \psi(x; \ell_m \odot \theta_\psi)$.

3.2. TCN model

The TCN serves as a variant of conventional Convolutional Neural Network (CNN) [39]. TCN primarily consists of multiple residual modules, each with residual connections. Each residual module is made up of several causal dilated convolution layers [40]. Causal dilated convolution empowers the network to tune the scope of historical contextual information, thereby fully analyzing the characteristics of load data. For the convolution kernel $f: (f_0, f_1, \dots, f_{k-1})$ and the input $X = \{x_1, x_2, \dots, x_n\}$, the causal expansion convolution formula is as follows:

$$F(s) = \sum_{i=0}^{h-1} f_i \cdot x_{s-(d \bullet i)} \quad (20)$$

where $F(s)$ denotes the expansion convolution of the s vector in the input X ; h the dimension of the convolutional kernel; f_i is the i_{th} convolutional kernel vector; d stands for the dilation rate; and $x_{s-(d \bullet i)}$ indicates the s_{th} element and the historical data preceding s .

The residual module effectively addresses the performance degradation problem in deep learning networks caused by an increasing number of layers. In the residual module, the Parametric ReLU (PReLU) activation function layer is used. This overcomes the limitation of the traditional ReLU activation function, which can solve the “vanishing gradient” problem but may cause the input to always be negative if the bias term changes too little. In such cases, the values in the negative interval become zero during backpropagation, potentially leading to non-updating neuron parameters. The formula for Prelu is as follows:

$$\text{Prelu}(x) = \begin{cases} x, & x \geq 0 \\ \partial x, & x < 0 \end{cases} \quad (21)$$

Parameter ∂ is designed for autonomous learning specific to different tasks and is not assigned a fixed value.

The structure of causal expansion convolution is shown in Figure 3. The white circles represent zero-padding, which ensures that the input and output sequences have the same length. As shown in the figure, the output value y_T at time step T is obtained by performing convolution operations using the historical values of x_T up to the T_{th} time step.

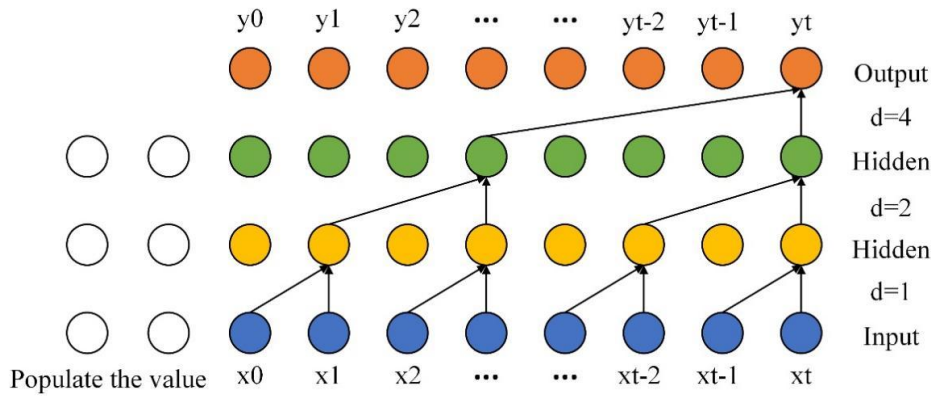


Figure 3. The structure of causally inflated convolutions.

3.3. Transformer model

The TCN-Transformer model adopts the multi-head self-attention module from the Transformer [41] to effectively capture the key information in multi-energy load sequences. By adjusting feature weights, this mechanism optimizes information extraction, reduces resource consumption, and improves prediction performance. It supports direct information exchange at any point in the sequence, enhancing the model's global learning capability, which is crucial for extracting global temporal features. Thus, we combine it with the TCN model to enhance the model prediction accuracy.

The Transformer's multi-head self-attention mechanism can analyze different input features from multiple perspectives, enabling the extraction and selection of important input data. The process is represented by the following formula:

$$[Q \ K \ V] = H[W_Q \ W_K \ W_V] \quad (22)$$

where Q denotes the query matrix; K denotes the key matrix; V denotes the value matrix; H is the temporal feature matrix output by the TCN model; and W_Q , W_K , W_V are the corresponding learnable weight matrices for Q , K , and V , used to calculate the correlation between the data.

Next, after performing z_{th} iterations of learning on Q , K , and V , the attention values are calculated. The softmax function is then used to normalize the attention values, completing the calculation of the temporal feature correlation attention values $Head_i$. The following formula represents the process:

$$\begin{aligned} \text{Attention}(Q, K, V) &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \\ Head_i &= \text{Attention}(QW_Q^i, KW_K^i, VW_V^i) \end{aligned} \quad (23)$$

where K^T is the transpose of K ; and d_K is the dimension of the K matrix. $Head_i$ represents the attention value of the i_{th} head; W_Q^i , W_K^i , W_V^i are the learnable weight matrices corresponding to Q , K , and V in the i_{th} attention head.

Finally, the Contact function is used to concatenate the attention values of all heads to obtain the end result produced by the Transformer. The calculation formula is given below:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Contact}(\text{Head}_1, \dots, \text{Head}_z) \mathbf{W}_O \quad (24)$$

where $\mathbf{W}_O$ is the weight matrix used to concatenate all single-head attention values Head_i .

It is well known that the standard self-attention mechanism in the Transformer exhibits quadratic computational complexity with respect to the input sequence length, i.e., $O(L^2)$, which may become computationally expensive when directly applied to long time-series forecasting tasks. However, in the proposed MODWT-MTL-IETCNT framework, the Transformer module is not applied to raw high-resolution input sequences. Instead, the TCN is first employed to extract local and multi-scale temporal representations from the original load series. This operation effectively transforms the input into a compact feature representation with reduced temporal redundancy. As a result, the subsequent Transformer module operates on a compressed feature space rather than the original long sequence, thereby significantly reducing the effective sequence length involved in the self-attention computation. In this way, the quadratic complexity of self-attention is mitigated in practice, since the computational cost becomes dependent on the reduced feature sequence length rather than the raw input length. Therefore, the overall computational efficiency of the proposed architecture is improved through hierarchical feature extraction rather than modification of the Transformer structure.

It is also worth noting that similar hybrid modeling strategies combining convolution-based temporal feature extraction with attention mechanisms have been widely adopted in recent time-series forecasting studies. These works generally demonstrate that convolutional or TCN-based front-end feature compression, followed by attention-based global dependency modeling, can achieve a favorable trade-off between forecasting accuracy and computational efficiency. Accordingly, the efficiency advantage claimed in this study is consistent with the literature and is primarily attributed to the reduction of effective sequence length before self-attention modeling, rather than a reduction in the intrinsic complexity of the Transformer. It should be emphasized that our focus of this study is not to redesign the Transformer complexity, but to improve its practical efficiency through architectural-level sequence compression and feature refinement. Therefore, the efficiency improvement is discussed from a structural and representation-learning perspective rather than through direct benchmarking of computational cost.

3.4. Combined forecasting model

The prediction flowchart of MODWT-MTL-IETCNT is shown in Figure 4.

The proposed model, MODWT-MTL-IETCNT, consists of two major stages.

In the data preprocessing stage, the Pearson correlation coefficient is first applied to optimize the interrelation metrics of multi-energy load data along with meteorological datasets, and meteorological parameters, which have a greater ramification on the multi-energy load, were segregated out as inputs to the model. Then, the original multi-energy load sequences and the meteorological factor sequences, filtered by Pearson correlation analysis, are decomposed into different frequency time series datasets using MODWT. The correlation among these frequency datasets is then analyzed.

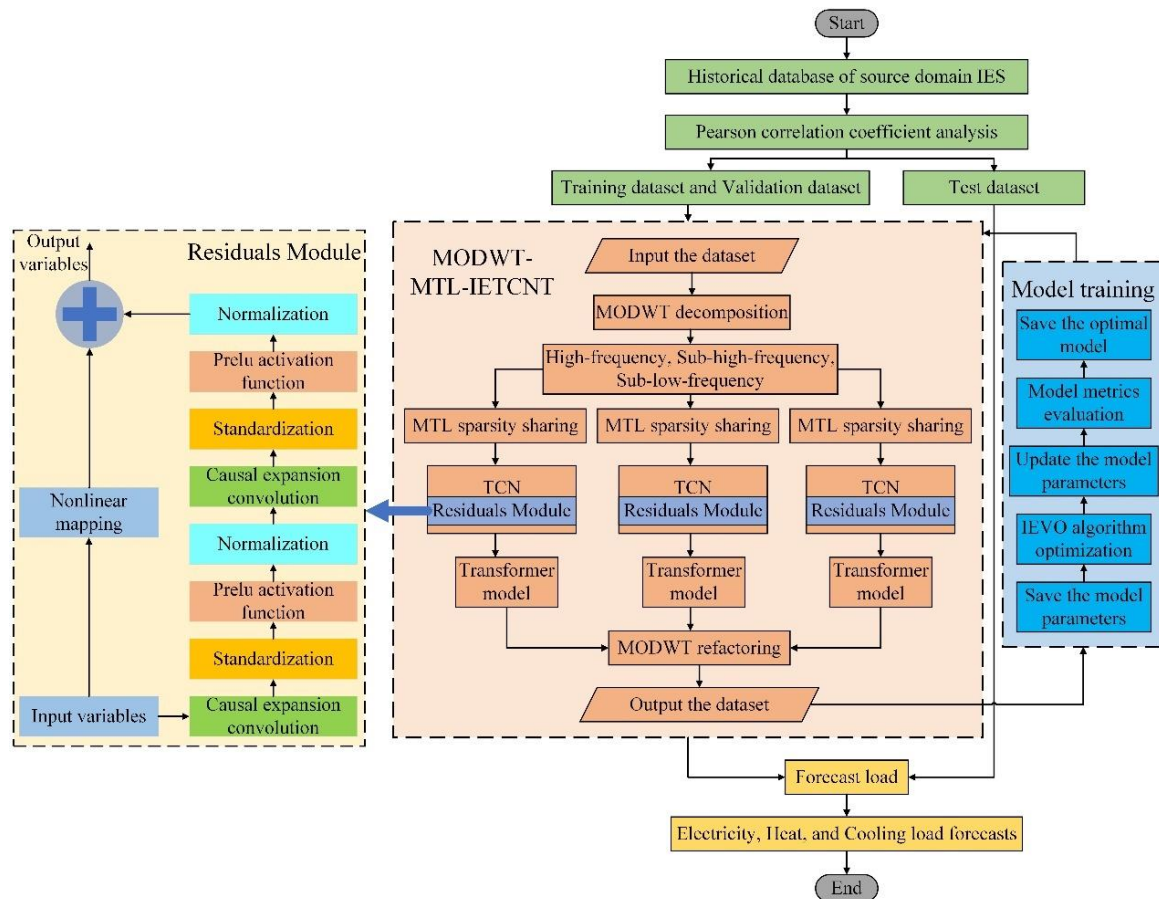


Figure 4. Flowchart of the MODWT-MTL-IETCNT forecasting model.

In the forecasting stage, the decomposed time series datasets are fed into the MTL for sparse sharing. The different frequency time series datasets output by the MTL are then input into the IETCNT for temporal scale feature extraction. Subsequently, the feature vectors output by the IETCNT are fed into their corresponding Transformer models to adaptively adjust the feature weights. Finally, the different frequency prediction data are reconstructed using MODWT to obtain the final predictions for electricity, heating, and cooling loads.

The overall framework of the proposed MODWT-MTL-IETCNT model is shown in Figure 5, which illustrates the complete forecasting process from input data to final multi-energy load prediction.

The MODWT-MTL-IETCNT model comprises an input module, an MODWT decomposition module, a multi-task sharing module, an IETCNT module, an MODWT reconstruction module, and an output module. Historical multi-energy load data for electricity, heating, and cooling, along with factors highly correlated with the load, are decomposed into different frequencies using the MODWT algorithm and then used as input data for the multi-task sharing module.

The multi-task sharing module uses the sparsity sharing mechanism to acquire the intercoupling information among loads and output to the IETCNT layer. The IETCNT layer extracts and adjusts the weights of the multi-energy load sequence timing features to achieve prediction.

Finally, the MODWT reconstruction layer combines the prediction data of different frequencies to produce the final forecast results. The output module outputs the predicted values of the target load.

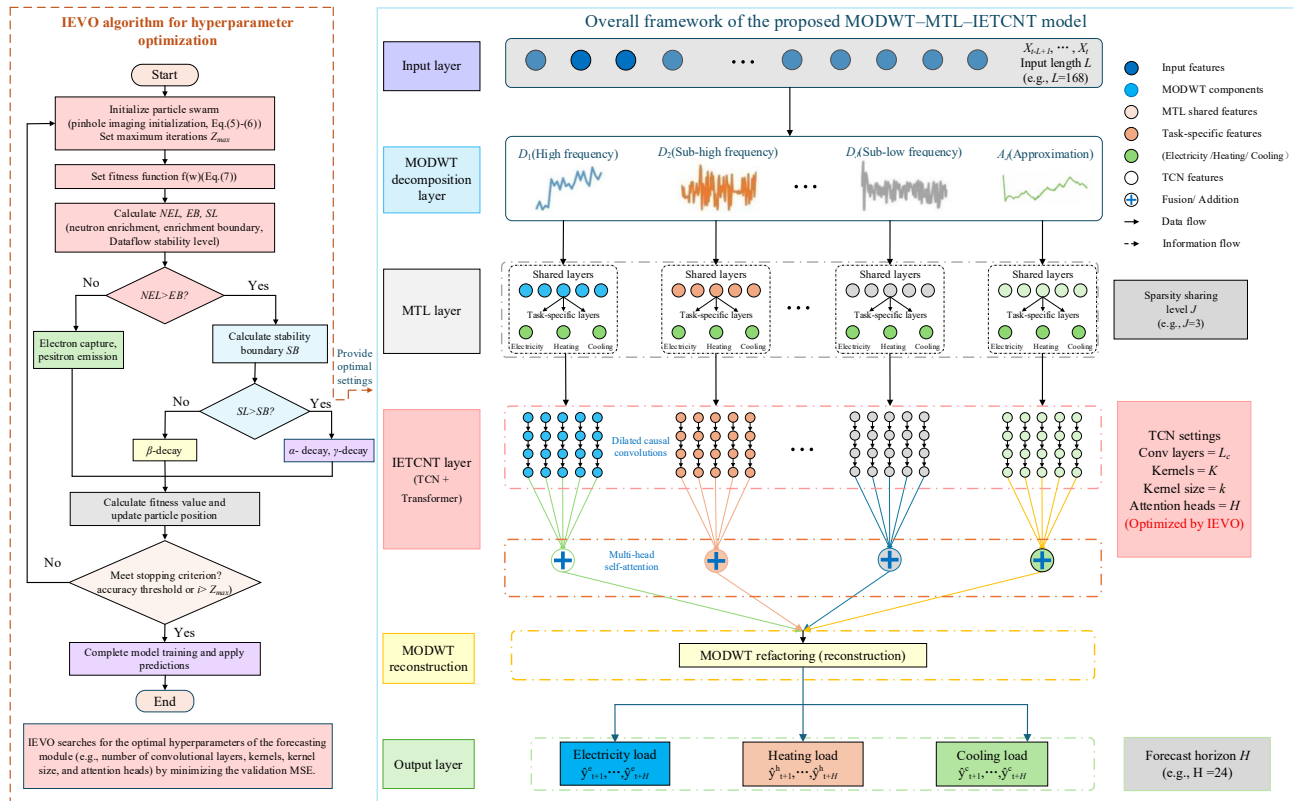


Figure 5. Overall framework of the proposed MODWT-MTL-IETCNT model with IEVO-based hyperparameter optimization.

3.5. Evaluation metrics

To better validate the predictive performance of forecasting models, we use E_{RMSE} , E_{MAE} , and E_{MAPE} to analyze the prediction errors of the model. Smaller values of these evaluation metrics signify lower prediction errors and better model performance. The computational formulas are as follows:

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (25)$$

$$E_{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (26)$$

$$E_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (27)$$

where N indicates the total number of predicted data; y_i indicates the real value of the i_{th} multi-energy load data; and $\hat{y}_i$ indicates the predicted result of the i_{th} multi-energy load data. It should be noted that the above metrics are designed for deterministic point forecasting. Therefore, we mainly evaluate the point-prediction accuracy of the proposed model, rather than the uncertainty of future load trajectories. Although deterministic forecasting is useful for assessing model precision, it cannot directly provide confidence information for risk-aware operation and scheduling in IES.

4. Case study

4.1. Experimental data and platform

In this study, the measured load data of the IES at Arizona State University's Tempe campus in the United States is used [42]. The system is a campus-type IES containing mostly electrical, heating, and cooling loads, with units of kilowatts (kW), tons of cooling (Ton), and million British thermal units per hour (MMBtu·h⁻¹), respectively. The integrated energy load data from January 1, 2022 to December 31, 2022 are selected with a temporal resolution of 1 h. Additionally, meteorological data corresponding to this period in the region is collected [43]. The collected meteorological variables include greenhouse gas emissions, dew point, wind speed, rainfall, surface reflectance, humidity, wind direction, temperature, and air pressure, with 8,760 data points for each variable.

To validate the generalization performance of the proposed forecasting model, the dataset is split into training, validation, and test sets in a 7:2:1 ratio. The experiments are completed based on an Intel (R) i5-1135G7 CPU @ 2.40GHz, GTX 1660Ti, and 16 GB RAM platform.

4.2. Data preprocessing

The input feature set is a crucial component of the multi-energy load forecasting model, as it directly affects the model's prediction performance. Therefore, the input feature set should include historical load data and cover sufficient influencing factors. In this study, calendar information is included to distinguish rest days and working days. Rest days mostly include national holidays and weekends, with weekends defined as Saturdays and Sundays. All remaining days are regarded as working days. Additionally, local meteorological factors are incorporated into the model's input. However, incorporating too many meteorological factors can lead to information redundancy and a decrease in model accuracy. Therefore, before model training, the Pearson correlation coefficient is adopted to select the meteorological factors and integrated energy load, selecting those meteorological factors that significantly impact multi-energy load, with the correlation levels shown in Figure 6.

Figure 6 shows a strong positive correlation of 0.92 between electrical and cooling loads, while the correlations between electrical and heating loads and between cooling and heating loads are -0.69 and -0.72, respectively, indicating strong negative correlations. Therefore, incorporating multi-energy load data simultaneously as inputs to the forecasting model enables it to fully learn the intercoupling characteristics among electrical, cooling load, and heating load.

Additionally, consider meteorological factors with correlation indices exceeding 0.3 in terms of absolute value as major influences and are also included as input features to assist in prediction. From Figure 6, the selected primary meteorological factors are greenhouse gas emissions, dew point, surface reflectance, rainfall, temperature, and air pressure.

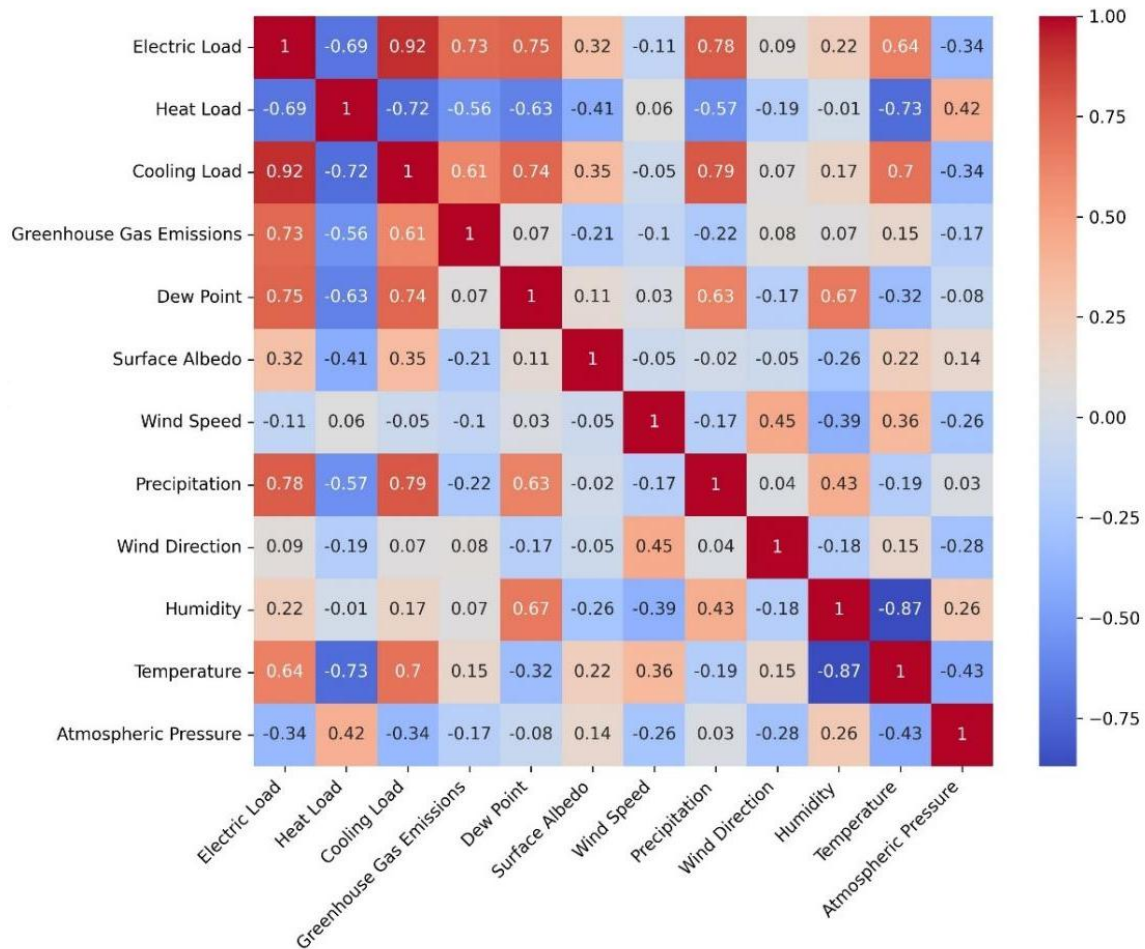


Figure 6. Correlation between multi-energy loads and meteorological factors.

Since the different magnitudes of each type of data can adversely affect the model's predictions, the multi-category multi-energy load data and meteorological factors are normalized utilizing the subsequent formula:

$$\tilde{x}_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (28)$$

where x_i is the sample value; $\min(x)$ is the minimum sample value; $\max(x)$ is the maximum sample value; and $\tilde{x}_i$ is the sample value after normalization.

Due to the potential noise generated by the equipment during the data gathering process, to further improve the completeness of the dataset, filtering is required. For multi-energy loads, due to their poor stationarity and nonlinear characteristics, we select the MODWT for further processing.

Using MODWT to process multi-energy load time series can address the randomness and volatility in the original data. It enables in-depth exploration of the intrinsic patterns of the load data while reducing irrelevant features, thus facilitating the extraction of the primary traits of the original data. Figure 7 shows the time series decomposition of electrical, heating, and cooling loads using MODWT, together with the results of their spectral analysis. The figure presents seven decomposition levels and the corresponding spectral analysis results for multi-energy loads.

It is evident that Levels 1 to 4 are relatively dense, exhibiting high frequency and randomness, which reveal the short-term undulations and abrupt variations in the multi-energy load profile. Then again, the randomness and volatility of Levels 5 to 7 decrease, reflecting the periodic patterns of the electrical, heating, and cooling loads. The residuals gradually stabilize, indicating the long-term variation trends of the multi-energy loads. MODWT decomposition provides time-frequency characteristics, aiding in the understanding of the dynamic features of load variations across time scales.

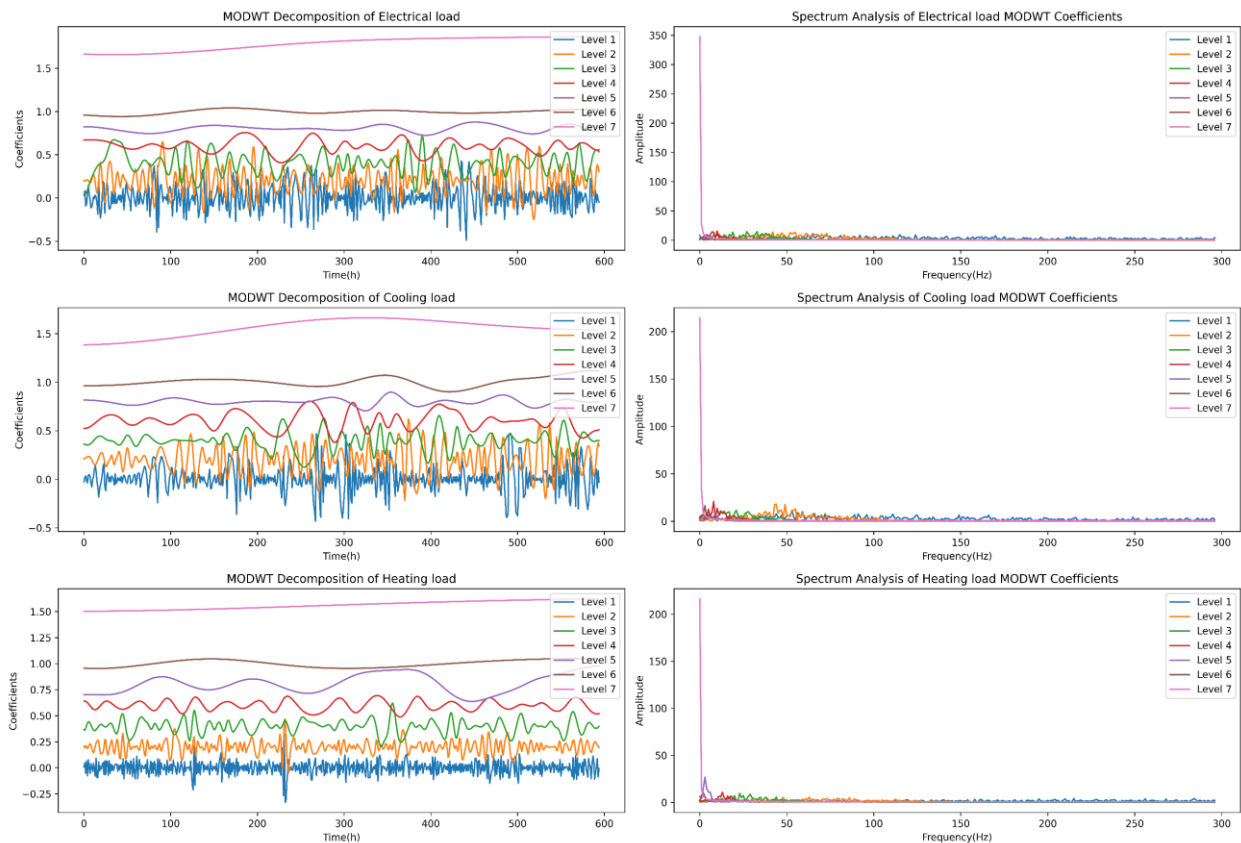


Figure 7. Time series decomposition and spectral analysis of electricity, cooling, and heating loads using MODWT.

In the spectral analysis, the primary frequency of noise is high, while the primary frequency of the signal is low. The spectral center of the dominant noise components is far from zero, while that of the dominant signal components is close to zero. The coefficient components are judged based on the deviation of the spectral center. The spectral information of different coefficient components varies significantly. As the decomposition level increases, the spectral peak gradually shifts to the left until it reaches the zero axis.

4.3. Model hyperparameter settings

Within this research, the IEVO algorithm serves to train the model. The training configurations include a batch size of 32, the Adam optimizer as the optimization method, a Mean Squared Error (MSE) loss function, and 50 training iterations. The objective of IEVO is to minimize the validation MSE by

automatically searching for the optimal hyperparameter combination. The decision variables optimized by IEVO include the number of decomposition layers, the number of sharing layers, the number of convolutional layers, the number of convolutional kernels, the convolution kernel size, and the number of attention heads. The starting step size parameter is initialized with a value of 0.001, and it is adjusted to 75% of the current learning rate whenever the MSE for the validation set does not decline for two successive epochs. The random seed is fixed at 42. The hyperparameters of MODWT-MTL-IETCNT are shown in Table 2.

Table 2. Hyperparameter settings of the model.

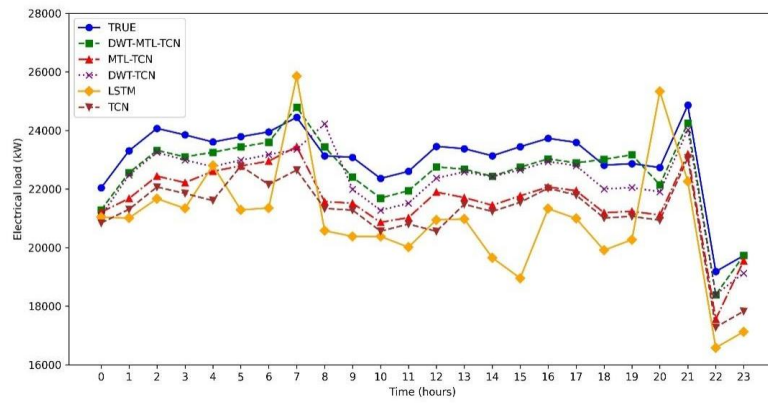
Model	Parameter	Value	Description
LSTM	Hidden layers	2	Controls network depth and nonlinear representation ability
	Hidden neurons	64	Determines feature learning capacity
	Convolution layers	2	Controls temporal feature extraction depth
TCN	Convolution kernels	64	Determines the number of extracted feature maps
	Kernel size	3	Controls local receptive field
MODWT-MTL-IETCNT	Decomposition layers	3	Controls multi-scale frequency decomposition
MODWT-MTL-IETCNT	Sharing layers	3	Balances shared and task-specific learning
	Attention heads	4	Captures multiple temporal dependency patterns

The hyperparameters listed in Table 2 are determined through preliminary experiments and validation-based tuning to balance forecasting accuracy, model complexity, and computational efficiency. Specifically, the number of hidden layers and neurons controls the representation capability of the LSTM model, while the number of convolutional layers, convolutional kernels, and kernel size determine the receptive field and local temporal feature extraction ability of the TCN. The number of decomposition layers in MODWT controls the granularity of multi-frequency information extraction, whereas the number of sharing layers in MTL balances shared feature learning and task-specific representation among electricity, heating, and cooling loads. The number of attention heads determines the ability of the Transformer module to capture multiple dependency patterns simultaneously. During optimization, IEVO searches these hyperparameters within predefined candidate ranges and selects the optimal combination according to the minimum validation loss, thereby improving forecasting accuracy and model stability.

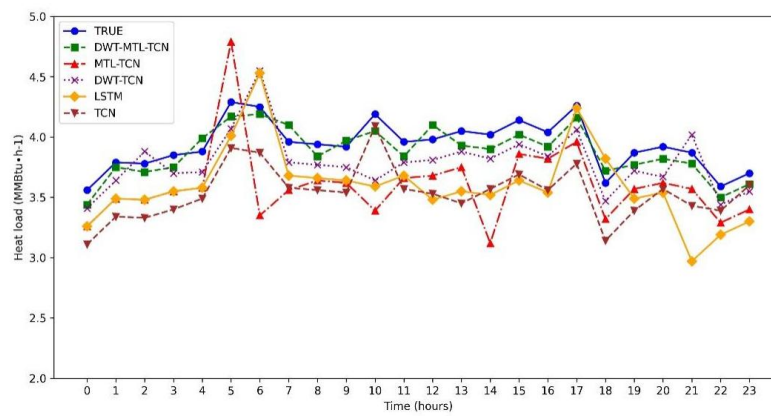
4.4. Comparison results and discussion

To confirm the superiority and generalization of our proposed model MODWT-MTL-IETCNT and to simultaneously consider the differences in load patterns between working days and rest days, November 18, 2022 is selected as a typical rest day, and November 23, 2022 as a typical working day for predictive analysis. The models MTL-IETCNT, MODWT-IETCNT, LSTM, TCN, and the proposed model are constructed for comparative analysis under the same experimental environment.

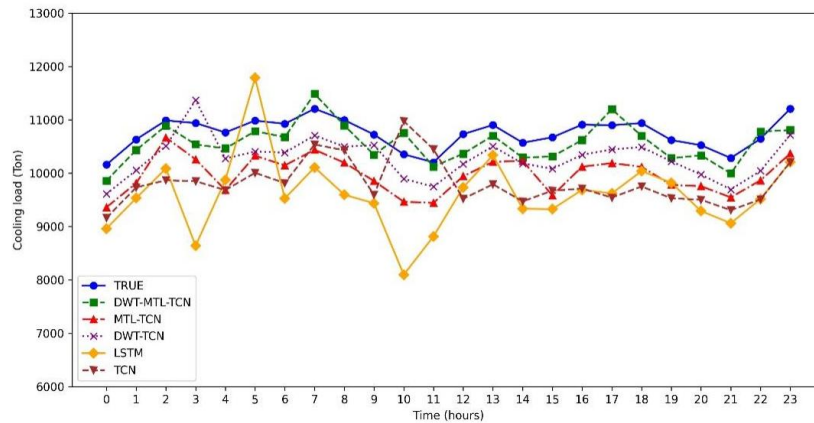
The evaluation metrics and prediction curves for each model on the typical rest day are shown in Table 3 and Figure 8, respectively. An analysis of the comparisons for the prediction errors among all models is presented in Figure 9. To facilitate the examination of the predictive performance of the presented model and the comparison models, each prediction curve graph displays the hourly predictions for the next day.



(a)

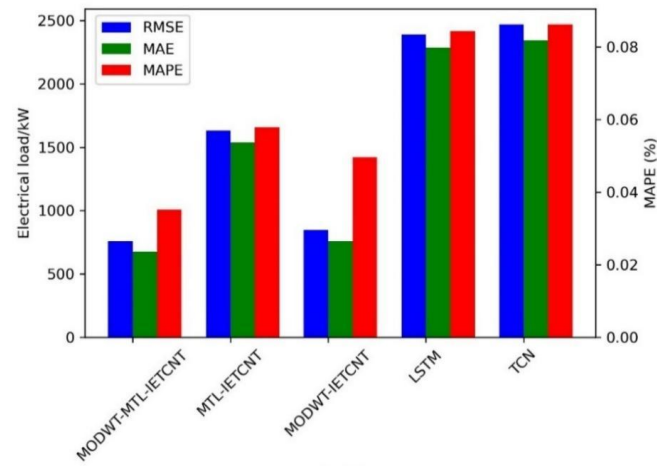


(b)

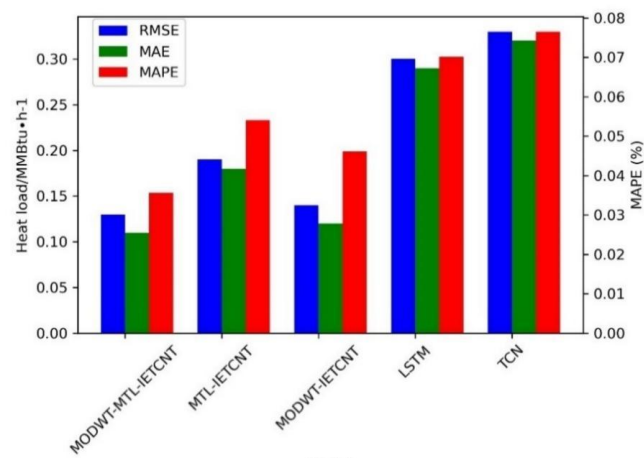


(c)

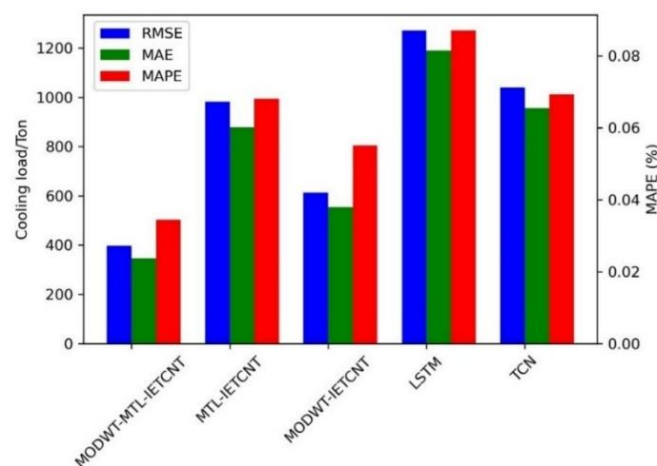
Figure 8. IES load forecasting results for typical rest days. (a) Electrical load, (b) Heat load, and (c) Cooling load.



(a)



(b)



(c)

Figure 9. Comparison of prediction error results of different models on typical rest days.
 (a) Electrical load, (b) Heat load, and (c) Cooling load.

Table 3. Evaluation metrics of each model on the rest day.

Model	Electric load/kW			Heating load/MMBtu·h ⁻¹			Cooling load/Ton		
	E_{RMSE}	E_{MAE}	E_{MAPE}	E_{RMSE}	E_{MAE}	E_{MAPE}	E_{RMSE}	E_{MAE}	E_{MAPE}
MODWT-MTL-IETCNT	757.37	676.00	3.51%	0.13	0.11	3.56%	397.73	345.7	3.44%
MTL-IETCNT	1631.89	1538.24	5.79%	0.19	0.18	5.40%	982.09	876.97	6.80%
MODWT-IETCNT	847.04	759.94	4.96%	0.14	0.12	4.61%	612.87	553.49	5.51%
LSTM	2390.07	2287.79	8.44%	0.30	0.29	7.01%	1271.12	1189.48	8.70%
TCN	2469.47	2344.84	8.62%	0.33	0.32	7.65%	1039.32	956.74	6.93%

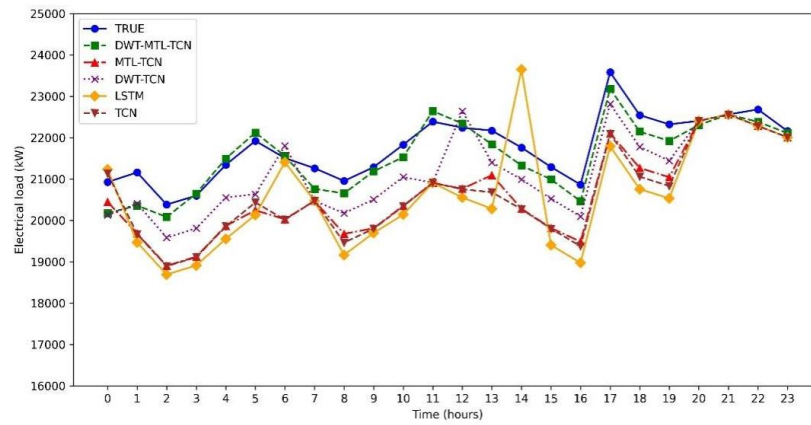
As illustrated in Figure 8, the prediction trajectory of MODWT-MTL-IETCNT is the closest to the true values among all the comparison models on the rest day, with the smallest fluctuation range around the actual values. This suggests that MODWT-MTL-IETCNT has excellent fitting ability and prediction effect. Table 3 and Figure 9 show the comparison of the final error results for different models. Among them, the prediction metrics of the MODWT-MTL-IETCNT model, such as E_{RMSE} , E_{MAE} , and E_{MAPE} , are the smallest for the electrical, heating, and cooling load predictions on the rest day compared to the other models, with values of 757.37 kW, 676 kW, 3.51%; 0.13 MMBtu·h⁻¹, 0.11 MMBtu·h⁻¹, 3.5%; and 397.73 Ton, 345.7 Ton, 3.44%, respectively.

Compared with single models like LSTM and TCN, the MODWT-MTL-IETCNT model's E_{RMSE} , E_{MAE} , and E_{MAPE} for electrical load forecasting decrease by 68.31%, 70.45%, and 58.41% and 69.33%, 71.17%, and 59.28%, respectively. For heating load forecasting, they decrease by 56.67%, 62.07%, and 49.22% and 60.61%, 65.63%, and 53.46%, respectively. For cooling load forecasting, they decrease by 68.71%, 70.94%, and 60.46% and 61.73%, 63.87%, and 50.36%, respectively. In consequence, the overall accuracy and generalization of MODWT-MTL-IETCNT is verified.

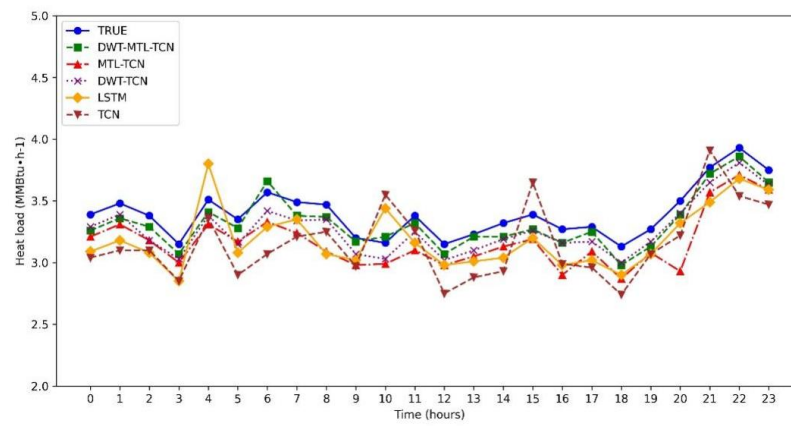
In contrast to the MTL-IETCNT model, the proposed model's E_{RMSE} , E_{MAE} , and E_{MAPE} for electrical load forecasting decreases by 53.58%, 56.05%, and 39.38%, respectively; for heating load forecasting, they decrease by 31.58%, 38.89%, and 34.07%; and for cooling load forecasting, they decrease by 59.5%, 60.58%, and 49.41%, respectively. This demonstrates that using MODWT for feature selection can effectively remove redundant features, improve the smoothness of the data, and reduce the learning difficulty of the model, thereby significantly improving prediction accuracy.

In comparison with the MODWT-IETCNT model, the proposed model's E_{RMSE} , E_{MAE} , and E_{MAPE} for electrical load forecasting decrease by 10.58%, 11.05%, and 29.23%, respectively; for heating load forecasting, they decrease by 7.14%, 8.33%, and 22.78%; and for cooling load forecasting, they decrease by 35.1%, 37.54%, and 37.57%. This indicates that MTL effectively addresses the strong coupling characteristics among multi-energy loads by deeply exploring shared features through the sparse sharing layer, thereby significantly enhancing prediction accuracy.

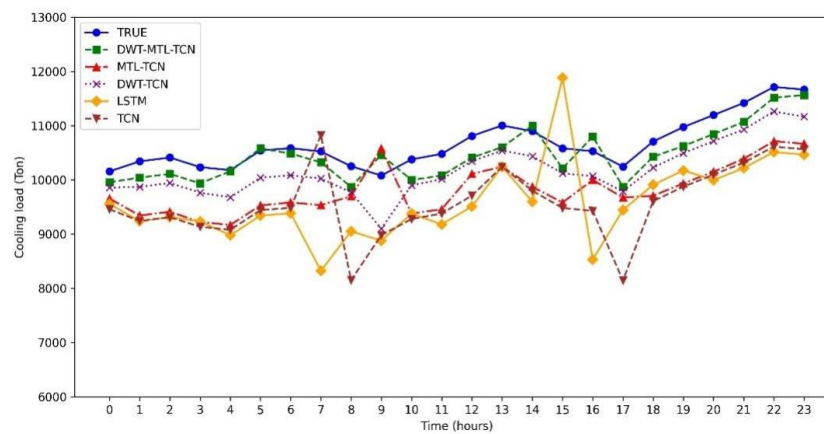
To confirm the performance validity of our proposed MODWT-MTL-IETCNT model on different dates, a comparison between a weekday and a weekend day is conducted. The evaluation metrics and prediction curves for each model on a representative working day are illustrated in Table 4 and Figure 10, respectively. A cross-model comparison of the predictive errors for each model is displayed in Figure 11.



(a)

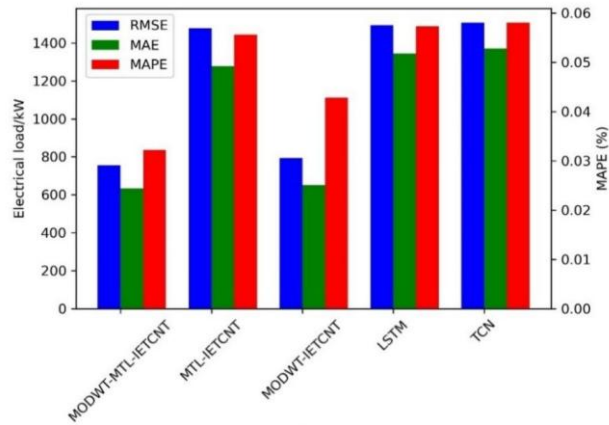


(b)

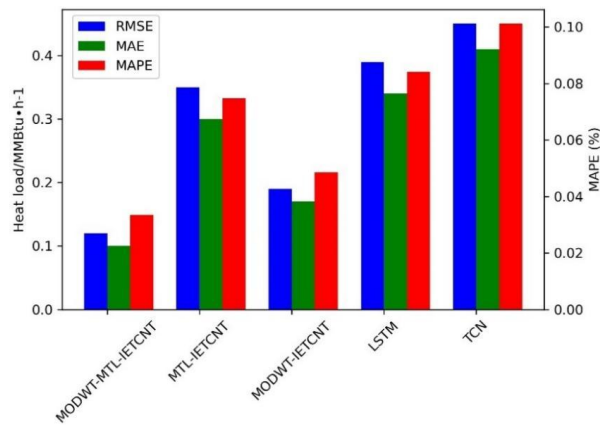


(c)

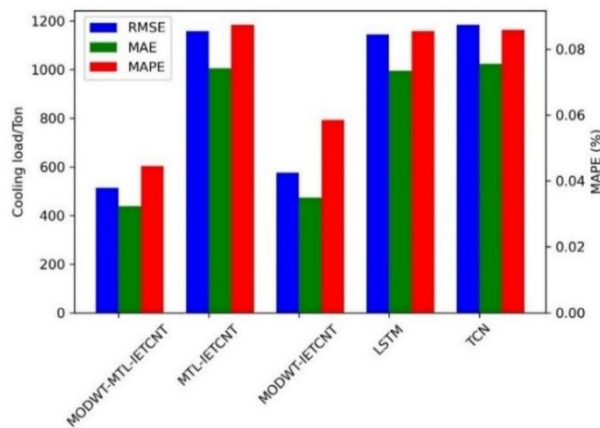
Figure 10. IES load forecasting results for typical working days. (a) Electrical load; (b) Heat load; and (c) Cooling load.



(a)



(b)



(c)

Figure 11. Comparison of prediction error results of different models on typical working days. (a) Electrical load, (b) Heat load, and (c) Cooling load.

Table 4. Evaluation metrics of each model on the working day.

Model	Electric load/kW			Heating load/MMBtu·h ⁻¹			Cooling load/Ton		
	E_{RMSE}	E_{MAE}	E_{MAPE}	E_{RMSE}	E_{MAE}	E_{MAPE}	E_{RMSE}	E_{MAE}	E_{MAPE}
MODWT-MTL-IETCNT	755.00	632.74	3.22%	0.12	0.10	3.35%	512.77	437.84	4.45%
MTL-IETCNT	1478.68	1277.76	5.56%	0.35	0.30	7.47%	1157.71	1005.3	8.73%
MODWT-IETCNT	794.06	651.91	4.28%	0.19	0.17	4.85%	575.56	472.57	5.85%
LSTM	1493.50	1343.27	5.73%	0.39	0.34	8.40%	1143.34	994.68	8.55%
TCN	1506.25	1370.73	5.80%	0.45	0.41	10.11%	1183.46	1024.17	8.58%

As shown in Figure 10, MODWT-MTL-IETCNT has the lowest prediction error of all the comparison model working days, with the best fitting degree. This indicates that MODWT-MTL-IETCNT has superior prediction accuracy compared to the reference models, demonstrating its effectiveness. Table 4 and Figure 11 present the error results for different models, showing that the prediction accuracy of the comparison models, including LSTM, TCN, MTL-IETCNT, and MODWT-IETCNT, declines to varying degrees compared to the proposed MODWT-MTL-IETCNT model.

A detailed analysis using the metric E_{MAPE} as an example reveals that, compared to the single models LSTM and TCN, the MODWT-MTL-IETCNT model's prediction indicators for electrical load forecasting decrease by 43.8% and 44.48%, respectively; for heating load forecasting, they decrease by 62.81% and 66.86%; and for cooling load forecasting, they decrease by 47.95% and 48.14%. This further verifies the generalization capability of the proposed model, indicating its applicability for short-term multi-energy load forecasting in IES across all dates.

When comparing the predictive performance of MTL-IETCNT and MODWT-MTL-IETCNT, the introduction of the MODWT algorithm leads to a decrease in E_{MAPE} for electrical, heating, and cooling load forecasting by 42.09%, 55.15%, and 49.03%, respectively. This indicates that using MODWT can deeply extract the temporal features of multi-energy load sequences, enabling effective knowledge acquisition and boosting the predictive precision of the model.

Furthermore, a comparison of the predictive results for MODWT-IETCNT and MODWT-MTL-IETCNT reveals that the inclusion of the MTL model results in a decrease in E_{MAPE} for electrical, heating, and cooling load forecasting by 24.77%, 30.93%, and 23.93%, respectively. This suggests that employing MTL fully considers the differences in coupling strength. Through the sparse sharing layer, it enables deep exploration of shared features across loads, significantly enhancing the prediction performance of the MODWT-MTL-IETCNT.

A comprehensive analysis of Table 3 and 4 and Figures 8–11 shows that the proposed MODWT-MTL-IETCNT model achieves superior prediction accuracy in short-term multi-energy load forecasting for IES. Additionally, the prediction performance on working days is slightly better than on rest days, which can be attributed to the larger amount of data on working days, reflecting the energy consumption patterns on the campus. It should also be emphasized that the forecasting results reported in this section are deterministic point forecasts. Although the proposed model achieves high predictive accuracy for electricity, heating, and cooling loads, the framework does not explicitly quantify

forecasting uncertainty caused by weather fluctuations, measurement noise, load randomness, or model misspecification. As a result, this method is more suitable for accuracy-oriented short-term forecasting than for fully risk-aware operational decision making. In practical IES applications, system operators may also require interval forecasts to evaluate the possible range of future loads and to support reserve scheduling and operational robustness. Therefore, incorporating uncertainty quantification into the proposed framework is an important direction for future work.

5. Conclusions

In this research, we present a multi-type energy load short-term forecasting model founded on MODWT-MTL-IETCNT. Through experimental examples, the validity and generalization capability of the presented model are validated. The experimental results reveal that, by applying MODWT breakdown, the input time series into varied frequency datasets are adopted as inputs to the forecast model, the temporal features can be deeply extracted, and redundant features effectively removed. This markedly boosts the model's prediction precision. The proposed MTL fully explores the complex dynamic coupling information in multi-energy conversions within the IES. Through the sparse sharing layer, it enables deep mining of shared features across loads, thereby improving the extraction of temporal information-related features in multi-energy load sequences. IEVO is used to optimize the hyperparameters of the TCNT, and it has advantages such as fast convergence, no requirement for special parameters, and the ability to avoid local optima. The integration of the Transformer mechanism within the TCN enables comprehensive consideration of the temporal feature information of IES multi-energy load sequences. It enables effective extraction and enhances the model's global learning capability, empowering the model to better extract the time-series dependencies in multi-energy load sequences and effectively improve its forecasting accuracy. In practical applications, the model demonstrates higher accuracy. Based on load forecasting results for different holidays, technicians can better formulate and adjust short-term regulation and distribution strategies for electrical, heating, and cooling energy in the system. This, in turn, improves the overall energy utilization efficiency, reduces carbon emissions, and minimizes unnecessary energy costs.

Nevertheless, it should be noted that this work centers on deterministic point forecasting and does not provide prediction intervals or other forms of probabilistic output. While the proposed MODWT-MTL-IETCNT model performs well in terms of point-forecast accuracy, uncertainty quantification is also important for risk-aware decision making in IES. In future work, we will extend the proposed framework to interval and probabilistic forecasting by combining this hybrid architecture with uncertainty modeling strategies such as quantile regression, residual-based interval estimation, or bootstrap-based uncertainty analysis. This extension is expected to further improve the practical value of the model for robust scheduling and secure operation of IESs.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by National Natural Science Foundation of China (62473269), Liaoning

Province Key Research and Development Project (2024JH2/102500093), Liaoning Revitalization Talents Program (XLYC2403160), and Basic Scientific Research Project of Liaoning Provincial Department of Education (LJ222411632036).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

He Jiang: Methodology, Formal analysis, Writing original draft; Tianhui Shi: Data curation, Validation; Lixin Yuan: Conceptualization.

References

1. Guo X, Xiao B, Song L (2020) Emission reduction and energy-intensity enhancement: The expected and unexpected consequences of China's coal consumption constraint policy. *J Clean Prod* 271: 122691. <https://doi.org/10.1016/j.jclepro.2020.122691>
2. Li J, Liu J, Yan P, et al. (2021) Operation optimization of integrated energy system under a renewable energy dominated future scene considering both independence and benefit: A review. *Energies* 14: 1103 <https://doi.org/10.3390/en14041103>
3. Guo Z, Zhang R, Wang L, et al. (2021) Optimal operation of regional integrated energy system considering demand response. *Appl Therm Eng* 191: 116860. <https://doi.org/10.1016/j.applthermaleng.2021.116860>
4. Fan H, Wang C, Liu L, et al. (2022) Review of uncertainty modeling for optimal operation of integrated energy system. *Front Energy Res* 9: 641337. <https://doi.org/10.3389/fenrg.2021.641337>
5. Liu X, Liu W, Zhou W, et al. (2024) Multi-energy load forecasting for IES considering meteorological causation and repetition cycles. *Electr Pow Syst Res* 237: 110977. <https://doi.org/10.1016/j.epsr.2024.110977>
6. Tan M, Liao C, Chen J, et al. (2023) A multi-task learning method for multi-energy load forecasting based on synthesis correlation analysis and load participation factor. *Appl Energy* 343: 121177. <https://doi.org/10.1016/j.apenergy.2023.121177>
7. Wang J, Lu S, Wang S, et al. (2021) A review on extreme learning machine. *Multimed Tools Appl* 81: 41611–41660. <https://doi.org/10.1007/s11042-021-11007-7>
8. Noori N, Kalin L, Isik S (2020) Water quality prediction using SWAT-ANN coupled approach. *J Hydrol* 590: 125220. <https://doi.org/10.1016/j.jhydrol.2020.125220>
9. Pasupuleti VR, Uhasri, Kalyan P, et al. (2020) Air quality prediction of data log by machine learning. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 1395–1399. <https://doi.org/10.1109/icaccs48705.2020.9074431>
10. Yazici I, Beyca OF, Delen D (2022) Deep-learning-based short-term electricity load forecasting: A real case application. *Eng Appl Artif Intel* 109: 104645. <https://doi.org/10.1016/j.engappai.2021.104645>

11. Wang S, Zhang Z (2021) Short-Term multiple load forecasting model of regional integrated energy system based on QWGRU-MTL. *Energies* 14: 6555. <https://doi.org/10.3390/en14206555>
12. Cen S, Lim CG (2024) Multi-Task learning of the PatchTCN-TST model for short-term multi-load energy forecasting considering indoor environments in a smart building. *IEEE Access* 12: 19553–19568. <https://doi.org/10.1109/access.2024.3355448>
13. Qiu Y, Wang S, Zhang S, et al. (2023) A novel gated dual convolutional neural network model with autoregressive method and attention mechanism for probabilistic load forecasting. *Appl Intell* 53: 20256–20271. <https://doi.org/10.1007/s10489-023-04589-2>
14. Zhang Z, Zhang C, Dong Y, et al. (2025) Bi-directional gated recurrent unit enhanced twin support vector regression with seasonal mechanism for electric load forecasting. *Knowledge-Based Syst* 310: 112943. <https://doi.org/10.1016/j.knosys.2024.112943>
15. Wang Q, Xu F, He J, et al. (2026) A new fusion model for enhanced ultra-short-term offshore wind power forecasting. *Renewable Energy* 256: 123876. <https://doi.org/10.1016/j.renene.2025.123876>
16. Xu F, Wang Q, Du X, et al. (2026) A novel hybrid short-term and ultra-short-term wind power forecasting method based on Weather Research and Forecasting: WRF-iTransformer-PSO. *Energy* 353: 140955. <https://doi.org/10.1016/j.energy.2026.140955>
17. He J, Liu QY, Hu YL, et al. (2026) Short-term regional EV charging load forecasting based on GAT and GRU with trip distribution estimation. *Appl Energy* 406: 127262. <https://doi.org/10.1016/j.apenergy.2025.127262>
18. Shang Y, Shang WL, Cui D, et al. (2026) Spatio-temporal data fusion framework based on large language model for enhanced prediction of electric vehicle charging demand in smart grid management. *Inf Fusion* 126: 103692. <https://doi.org/10.1016/j.inffus.2025.103692>
19. Liu M, Qin H, Cao R, et al. (2022) Short-Term load forecasting based on improved TCN and DenseNet. *IEEE Access* 10: 115945–115957. <https://doi.org/10.1109/access.2022.3218374>
20. Hong Q, Meng F, Maldonado F (2024) Advancing long-term multi-energy load forecasting with patchformer: A patch and transformer-based approach. *arXiv*. <https://doi.org/10.48550/arxiv.2404.10458>
21. Chen B, Wang Y (2021) Short-Term electric load forecasting of integrated energy system considering nonlinear synergy between different loads. *IEEE Access* 9: 43562–43573. <https://doi.org/10.1109/access.2021.3066915>
22. Imani M, Ghassemian H (2019) Residential load forecasting using wavelet and collaborative representation transforms. *Appl Energy* 253: 113505. <https://doi.org/10.1016/j.apenergy.2019.113505>
23. Fan GF, Han YY, Li JW, et al. (2024) A hybrid model for deep learning short-term power load forecasting based on feature extraction statistics techniques. *Expert Syst Appl* 238: 122012. <https://doi.org/10.1016/j.eswa.2023.122012>
24. Zhang T, Zhang J, Xue H, et al. (2021) Prediction of distribution network operation trend based on the secondary modal decomposition and LSTM-MFO algorithm. *2021 6th Asia Conference on Power and Electrical Engineering (ACPEE)*, Chongqing, China, 520–525. <https://doi.org/10.1109/acpee51499.2021.9436870>
25. Huang Q, Li J, Zhu M (2020) An improved convolutional neural network with load range discretization for probabilistic load forecasting. *Energy* 203: 117902. <https://doi.org/10.1016/j.energy.2020.117902>

26. Spoladore A, Borelli D, Devia F, et al. (2016) Model for forecasting residential heat demand based on natural gas consumption and energy performance indicators. *Appl Energy* 182: 488–499. <https://doi.org/10.1016/j.apenergy.2016.08.122>
27. Ding Y, Zhang Q, Yuan T (2017) Research on short-term and ultra-short-term cooling load prediction models for office buildings. *Energy Build* 154: 254–267. <https://doi.org/10.1016/j.enbuild.2017.08.077>
28. Li Y, Tian X, Liu T, et al. (2018) On better exploring and exploiting task relationships in multitask learning: Joint model and feature learning. *IEEE Trans Neural Netw Learn Syst* 29: 1975–1985. <https://doi.org/10.1109/tnnls.2017.2690683>
29. Hou J, Meng Y, Li Z, et al. (2024) Multi-energy load forecasting in integrated energy systems based on comprehensive correlation index and SA-BiGRU network. *Electr Power Constr* 45: 118–130. <https://doi.org/10.12204/j.issn.1000-7229.2024.05.012>
30. Tan Z, De G, Li M, et al. (2020) Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. *J Clean Prod* 248: 119252. <https://doi.org/10.1016/j.jclepro.2019.119252>
31. You W, Guo D, Wu Y, et al. (2023) Multiple load forecasting of integrated energy system based on sequential-parallel hybrid ensemble learning. *Energies* 16: 3268. <https://doi.org/10.3390/en16073268>
32. Yu X, Peng W, Xue D, et al. (2021) An improved spatial-temporal network based on residual correction and evolutionary algorithm for water quality prediction. In: Tan Y, Shi Y (eds) *Advances in Swarm Intelligence*. ICSI 2021. Lecture Notes in Computer Science, Springer, Cham, 12690. https://doi.org/10.1007/978-3-030-78811-7_46
33. Azad MA, Sajid I, Lu S, et al. (2023) Energy Valley Optimizer (EVO) for tracking the global maximum power point in a solar PV system under shading. *Processes* 11: 2986. <https://doi.org/10.3390/pr11102986>
34. Fathy A (2023) Efficient energy valley optimization approach for reconfiguring thermoelectric generator system under non-uniform heat distribution. *Renewable Energy* 217: 119177. <https://doi.org/10.1016/j.renene.2023.119177>
35. Wang J, Han L, Zhang X, et al. (2023) Electrical load forecasting based on variable T-distribution and dual attention mechanism. *Energy* 283: 128569. <https://doi.org/10.1016/j.energy.2023.128569>
36. Azizi M, Aickelin U, Khorshidi HA, et al. (2023) Energy valley optimizer: A novel metaheuristic algorithm for global and engineering optimization. *Sci Rep* 13: 226. <https://doi.org/10.1038/s41598-022-27344-y>
37. Jiang L, Wang X, Li W, et al. (2021) Hybrid multitask multi-information fusion deep learning for household short-term load forecasting. *IEEE Trans Smart Grid* 12: 5362–5372. <https://doi.org/10.1109/tsg.2021.3091469>
38. Li K, Duan P, Cao X, et al. (2024) A multi-energy load forecasting method based on complementary ensemble empirical model decomposition and composite evaluation factor reconstruction. *Appl Energy* 365: 123283. <https://doi.org/10.1016/j.apenergy.2024.123283>
39. Gopali S, Abri F, Siامي-Namini S, et al. (2021) A comparison of TCN and LSTM models in detecting anomalies in time series data. *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, 2415–2420. <https://doi.org/10.1109/BigData52589.2021.9671488>

40. Gu Y, Jazizadeh F, Wang X (2025) Toward large energy models: A comparative study of Transformers' efficacy for energy forecasting. *Appl Energy* 384: 125358. <https://doi.org/10.1016/j.apenergy.2025.125358>
41. Xu S, Yang D (2025) WLENet: An workload-aware physician recommendation system with patient collective learning. *J Supercomput* 81: 1589. <https://doi.org/10.1007/s11227-025-08091-z>
42. Geremicca F, Fascetti A, Brigham JC, et al. (2024) Urban metabolism and digital twin technologies for a sustainable built environment: Towards a framework for a campus application. *IOP Conference Series: Earth and Environment* 1363: 012081. <https://doi.org/10.1088/1755-1315/1363/1/012081>
43. Gebeyhu B (2024) Power data access viewer-based meteorological drought analysis and rainfall variability in the Nile River Basin. *Adv Meteorol*, 2024: 9985773. <https://doi.org/10.1155/2024/9985773>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)