EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

An MT-Transformer Framework for Coordinated Wind–Solar–Load Forecasting with Net-Load-Based Coal-Power Regulation Demand Identification

Meng Huang[1*], Lei Wang[1], Teng Luo[1], Chao Ju[1], Wenchao Guo[1] and Chenghao Ning[1]

[1]CHN Energy Technology & Economics Research Institute, Beijing 102211, China

Abstract

Data-driven forecasting has become increasingly important for describing the temporal interactions among heterogeneous variables in modern power systems. To capture the nonlinear coupling, heterogeneous fluctuations, and multi-scale temporal variations between renewable generation and load demand, this study develops an MT-Transformer framework for coordinated wind–solar–load forecasting. Meteorological variables, historical renewable output, historical load, and temporal labels are integrated as model inputs. A shared Transformer encoder is used to learn common temporal representations, and task-specific forecasting heads are designed to generate synchronized predictions for wind power, PV power, and load. The experimental results show that MT-Transformer achieves an MAE of 0.0848, an RMSE of 0.1254, and an R^2 of 0.9302. Compared with the Persistence Model, the MAE and RMSE decrease by 36.05% and 30.49%, respectively. The predicted outputs are further converted into a net-load sequence, from which fluctuation indicators are derived. The peak-valley difference reaches 0.462 p.u., and the maximum ramp rate reaches 0.087 p.u./h, indicating evident peak-shaving pressure and short-term regulation demand. These findings suggest that the proposed framework can improve coordinated forecasting performance and provides quantitative evidence for coal-power peak regulation, reserve capacity allocation, and ancillary service demand identification.

Keywords: MT-Transformer, joint wind–solar–load forecasting, multi-task learning, net-load fluctuation, coal-power regulation demand

Received on 15 May 2026, accepted on 29 June 2026, published on 11 August 2026

doi: 10.4108/ew.13035

*Corresponding author. Email: 20084536@ceic.com

1. Introduction

China's dual-carbon strategy is reshaping the power sector by increasing the share of renewable generation in the electricity mix. Wind and PV power have become important incremental sources in the restructuring of power supply [1]. However, renewable output is strongly driven by weather conditions and therefore shows randomness, intermittency, and volatility [2,3]. These characteristics make the net-load curve more nonlinear and more variable across multiple time scales. At the same time, coal power is shifting from a conventional baseload supplier to a resource that provides regulation, backup, and capacity support [4]. Its value is no longer evaluated only through electricity generation, but also through capacity adequacy, flexibility contribution, and low-carbon transition support[5]. Against this background, data-driven forecasting models are needed to describe the temporal interactions between renewable supply and load demand, and to identify peak-shaving, ramping, and reserve-capacity requirements caused by net-load fluctuations [6-8]. Therefore, a multi-source forecasting model for renewable output and electric load is essential for improving net-load analysis and coal-power regulation demand identification.

Renewable-output forecasting and load forecasting provide essential support for power-system dispatch and operation [9]. Forecasting research has gradually shifted from statistical time-series techniques to machine learning and deep neural-network models [10-12]. Statistical models are structurally clear and interpretable, but their adaptability is limited when renewable output contains random disturbances, non-stationary patterns, and nonlinear variations [13]. Machine learning methods improve nonlinear fitting ability, yet their performance often depends on manual feature design and relatively stable data distributions[14]. Deep learning models can jointly use weather information, historical generation, load records, and temporal labels, making them more suitable for extracting complex temporal features [15,16]. More recently, self-attention-based models have been used to capture global temporal dependencies in long-sequence, multivariate, and multi-scale forecasting tasks, providing a technical basis for coordinated modeling of wind, PV, and load [17,18]. Nevertheless, three limitations remain. First, existing studies often predict renewable output and load separately, while unified wind–solar–load forecasting is still insufficient. Second, many evaluations focus mainly on average errors, with less attention paid to multi-step stability, peak-valley identification, and ramping-related performance. Third, forecasting results are rarely transformed into net-load fluctuation indicators, which limits their direct use in coal-power peak regulation, reserve capacity planning, and ancillary service analysis [19]. Therefore, it is necessary to link artificial intelligence forecasting with net-load fluctuation identification.

Recent work has examined renewable-output prediction, demand forecasting, and long-horizon temporal modeling from different methodological perspectives. Pandžić et al. [20] emphasized that PV forecasting should consider data quality, forecasting horizon, and application scenario at the same time. Wang et al. [21] showed that deep learning can improve the extraction of nonlinear wind-power fluctuation features, although stable multi-step forecasting remains challenging. Manandhar et al. [22] compared Prophet [23], Random Forest, and LSTM for short-term load forecasting, suggesting that model performance should be assessed through multiple indicators rather than a single metric. Vaswani et al. [24] proposed the Transformer architecture, and Wu et al. [25] introduced graph information into Transformer-based wind-power forecasting for non-stationary multivariate spatio-temporal data. These studies indicate that self-attention is useful for modeling multivariate dependencies and complex periodic patterns. Overall, existing studies provide a foundation for renewable forecasting, load prediction, and Transformer-based temporal modeling. However, most of them still emphasize single-object prediction or accuracy improvement, while the collaborative modeling of wind, PV, and load and its connection with coal-power regulation demand remain insufficient.

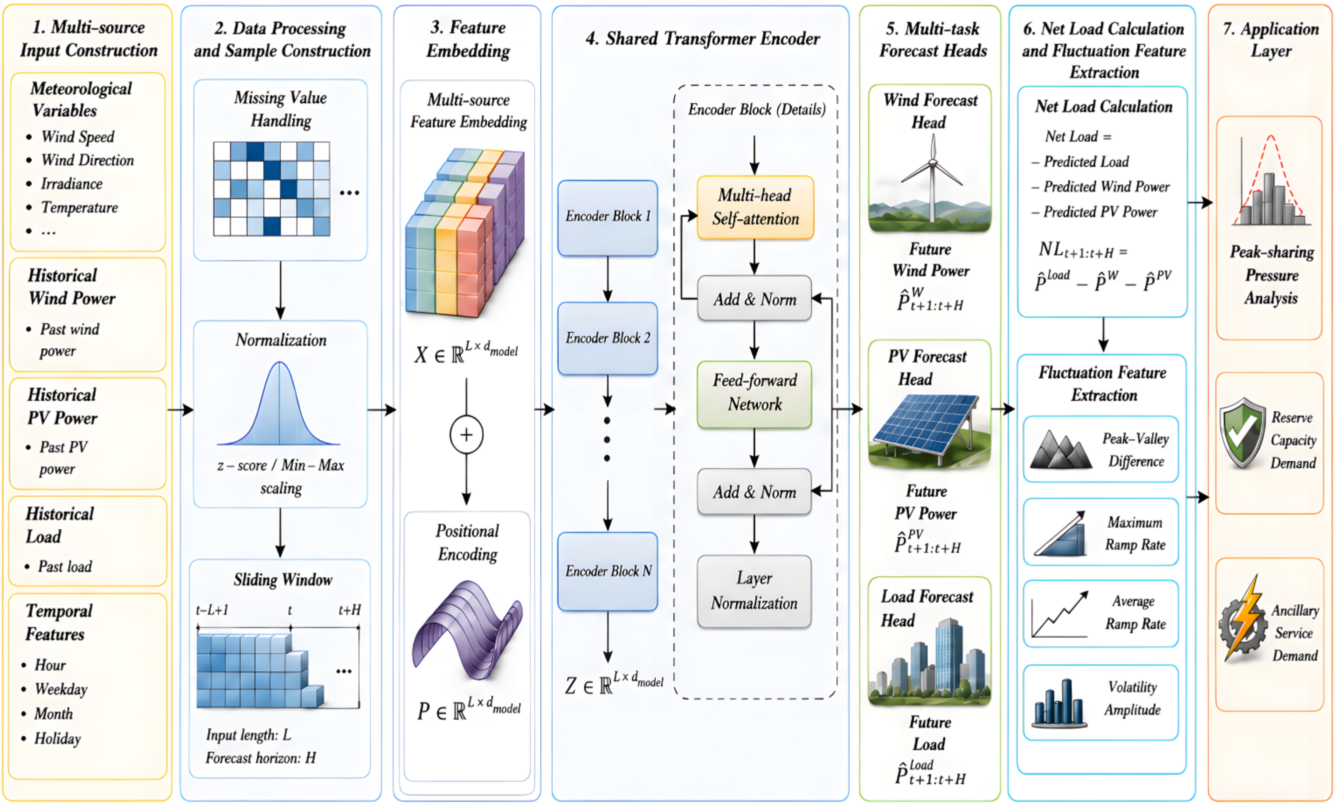


Figure 1. Overall framework of MT-Transformer-based joint forecasting of wind, PV, and load for coal-power policy adaptation

To address the above research gaps, this study constructs an MT-Transformer joint forecasting model for the three target variables. Based on the idea of multi-task learning, the model uses meteorological variables, historical wind power output, historical PV output, historical load, and temporal labels as multi-source inputs. A shared Transformer encoder is employed to learn common temporal features among wind, PV, and load, while three forecasting heads are designed to generate synchronized multi-step predictions for the three target variables [26]. This design enables the model to jointly describe the coupled evolution of renewable generation and electric load. To verify the effectiveness of the model, this study evaluates forecasting performance in terms of overall error, forecasting accuracy, multi-step forecasting stability, and ablation experiments. Furthermore, the forecasting results are used to calculate the net load and extract indicators including peak-valley difference, maximum ramp rate, average ramp rate, and fluctuation amplitude [27,28]. Through this design, the outputs of the artificial intelligence forecasting model are transformed into an analytical basis for identifying coal-power peak-shaving pressure, reserve capacity demand, and ancillary service demand, thereby improving the operational interpretability of joint forecasting results. The overall research framework is shown in Figure 1.

2. Construction of an MT-Transformer Joint Forecasting Model for Coal-Power Regulation Demand Identification

This section presents the design of the MT-Transformer model for joint forecasting of the three forecasting targets. It first defines the joint forecasting task and then introduces multi-source feature construction, sliding-window sample generation, the shared Transformer encoder, multi-task forecasting heads, and net-load fluctuation feature extraction. These components provide the methodological basis for subsequent experimental validation and coal-power regulation demand analysis.

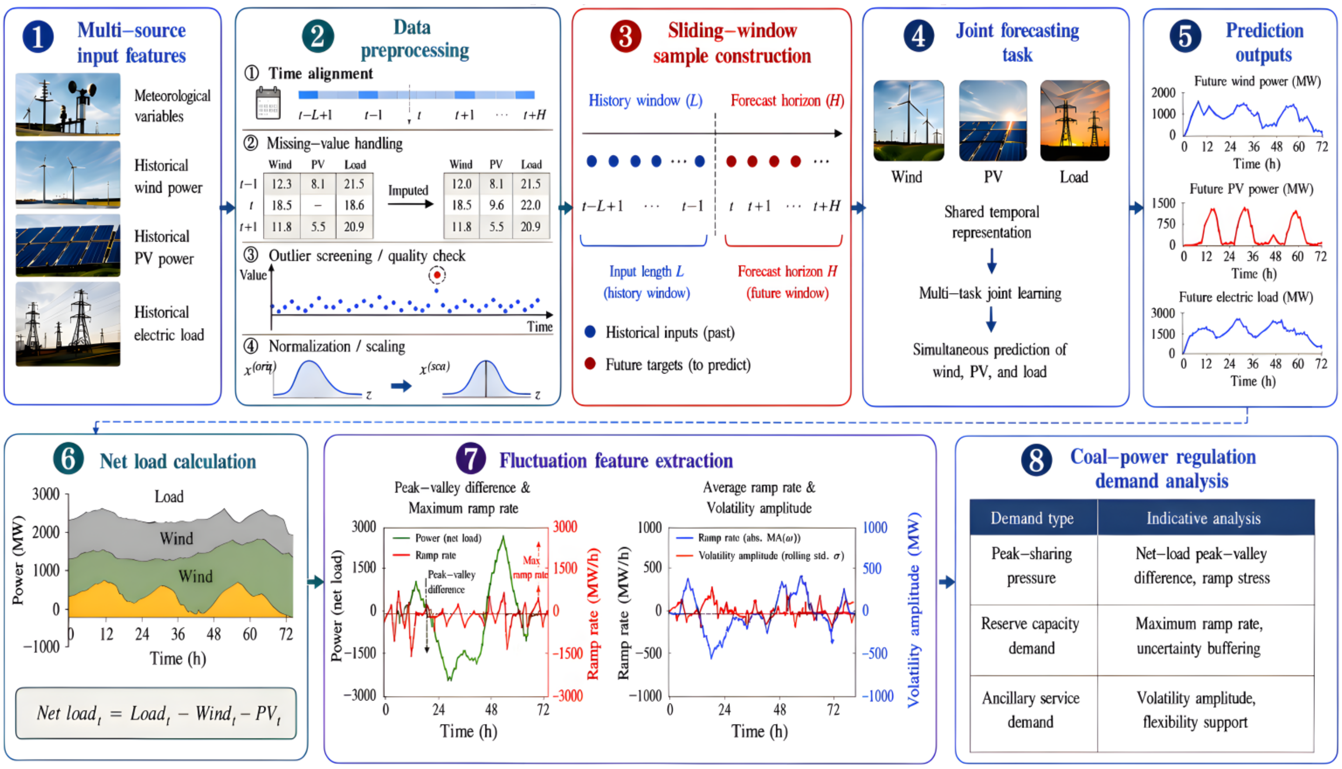


Figure 2. Relationship between the joint forecasting task and indicator transformation for wind, PV, and load

2.1. Joint Forecasting Task Formulation and Net Load Mapping Logic for wind, PV, and load

In this forecasting task, the three targets are treated as coupled temporal variables within the power-system supply－demand balance [29]. Wind and PV outputs describe renewable-side variations, while electric load reflects demand-side changes. Their joint evolution determines the system net-load level [30,31]. Therefore, the proposed model does not regard them as isolated prediction objects. Instead, it builds a unified wind－solar－load forecasting task, where multi-source historical information is used to generate synchronized predictions for the three targets. As shown in Figure 2, the predicted trajectories are then used for net-load calculation and fluctuation-feature extraction.

Let L denote the historical input length and H denote the forecasting horizon. At time t, the model uses the multi-source observations from the previous L time steps to predict the three targets over the following H time steps. The input sequence contains meteorological variables, historical wind output, historical PV output, historical load, and temporal labels, as shown in Eq. (1).

$$X_{t-L+1:t}=\{X^{met}_{t-L+1:t}, X^{w}_{t-L+1:t}, X^{pv}_{t-L+1:t}, X^{load}_{t-L+1:t}, X^{time}_{t-L+1:t}\} \quad (1)$$

$$\widehat{Y}_{t+1:t+H}=\{\widehat{P}^{w}_{t+1:t+H}, \widehat{P}^{pv}_{t+1:t+H}, \widehat{P}^{load}_{t+1:t+H}\} \quad (2)$$

In Eq. (1), X^{met}、X^{w}、X^{pv}、X^{load} and X^{time} denote meteorological features, historical wind output, historical PV output, historical electric load, and temporal features, respectively. The model output is given in Eq. (2), where the three components represent the future predictions of wind output, PV output, and load. This design enables the model to produce synchronized forecasts in one multi-task framework, while learning shared temporal information and preserving task-specific differences.

2.2. Multi-Source Feature Construction and Sliding Window Sample Generation

Joint wind–solar–load forecasting depends on the coordinated use of heterogeneous temporal information [32]. In this study, the input variables are organized into four groups: meteorological factors, historical renewable generation, historical load, and temporal descriptors. Meteorological factors provide external driving information for renewable output changes. Wind speed, wind direction, and air pressure are closely associated with wind-power variation, while solar irradiance, temperature, humidity, and cloud cover directly affect photovoltaic generation [33,34]. Historical renewable-generation records are used to describe inertia, periodicity, and short-term variations in wind and PV series. Historical load reflects daily demand changes, weekly patterns, and peak-valley characteristics. Temporal descriptors further help the model distinguish hour-of-day, weekday, month, season, holiday, and working-day effects.

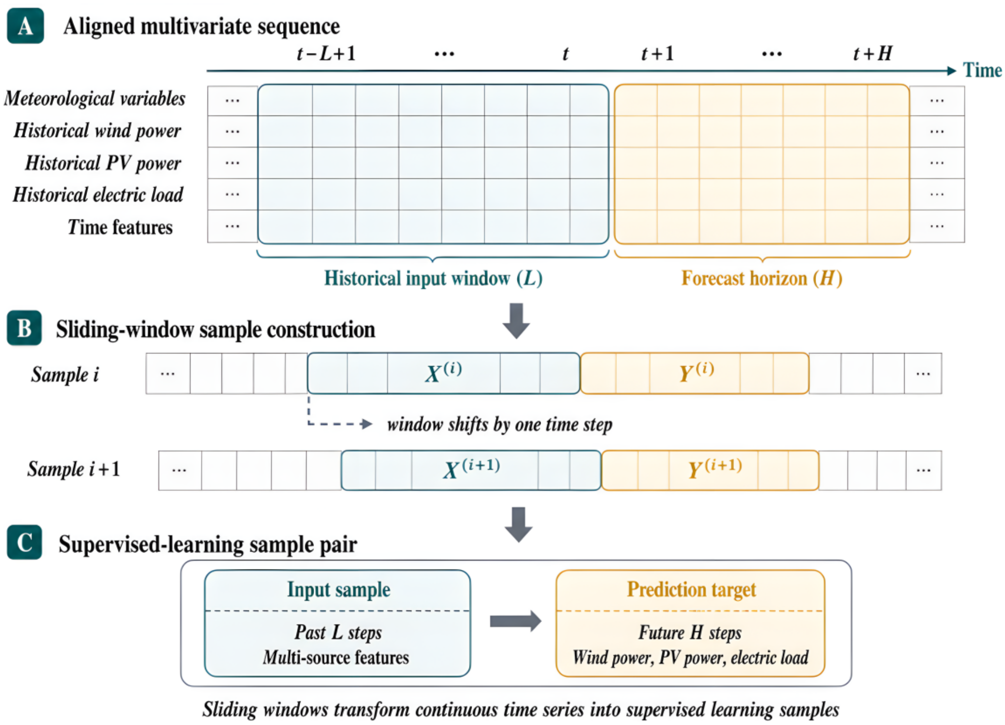


Figure 3. Sliding window sample construction

For continuous multivariate time series, the period from t−L+1 to t is taken as the historical observation window, and the period from t+1 to t+H is used as the forecasting horizon. As shown in Figure 3, each sample contains meteorological variables, historical wind power output, historical PV output, historical electric load, and temporal labels in the input window. The corresponding target contains the future values of the three forecasting variables. By rolling the window forward along the time axis, the original sequence is reorganized into supervised learning samples. This construction keeps the temporal order of the data and reduces the risk of future information leakage caused by random splitting [35].

After timestamp alignment, the above variables jointly form a multivariate temporal input matrix. For each time step t, all input features at that time can be represented as follows:

$$x_t=[x_t^{met}, x_t^{w}, x_t^{pv}, x_t^{load}, x_t^{time}] \tag{3}$$

where x_t denotes the multi-source feature vector at time t. The input features over consecutive time steps can be organized into a historical input sequence of length L:

$$X_{t-L+1:t}=[x_{t-L+1}, x_{t-L+2},\ldots, x_t] \tag{4}$$

To transform continuous time series into trainable samples, a sliding window method is adopted to construct supervised learning samples. Specifically, multi-source features from the past L time steps are used as the input, while the three forecasting targets over the future H time steps are used as the forecasting targets. When the temporal resolution is 1 hour, L=24 and H=24 can be set, meaning that the model uses the previous 24 hours of multi-source information to forecast the three target series over the next 24 hours.

The i-th sliding window sample can be expressed as:

$$X^{(i)}=X_{i:i+L-1} \tag{5}$$

$$Y^{(i)}=\{P^{w}_{i+L:i+L+H-1}, P^{pv}_{i+L:i+L+H-1}, P^{load}_{i+L:i+L+H-1}\} \tag{6}$$

where $X^{(i)}$ denotes the i-th input sample, and $Y^{(i)}$ denotes the corresponding multi-task forecasting target. The sliding window sample construction preserves the temporal continuity of the original series, avoids future information leakage caused by random shuffling, and provides a unified data structure for multi-step forecasting.

2.3. Shared Encoding and multi-task Forecasting Architecture of MT-Transformer

MT-Transformer is developed to predict renewable output and load demand in one multi-task framework. In this model, MT refers to multi-task learning. The architecture includes multi-source input processing, feature embedding, positional encoding, a shared Transformer encoder, and task-specific forecasting heads. Heterogeneous input variables are first projected into a common representation space. The shared encoder then extracts temporal features that are useful for all three forecasting tasks, and the forecasting heads separately generate the future trajectories of wind, PV, and load.

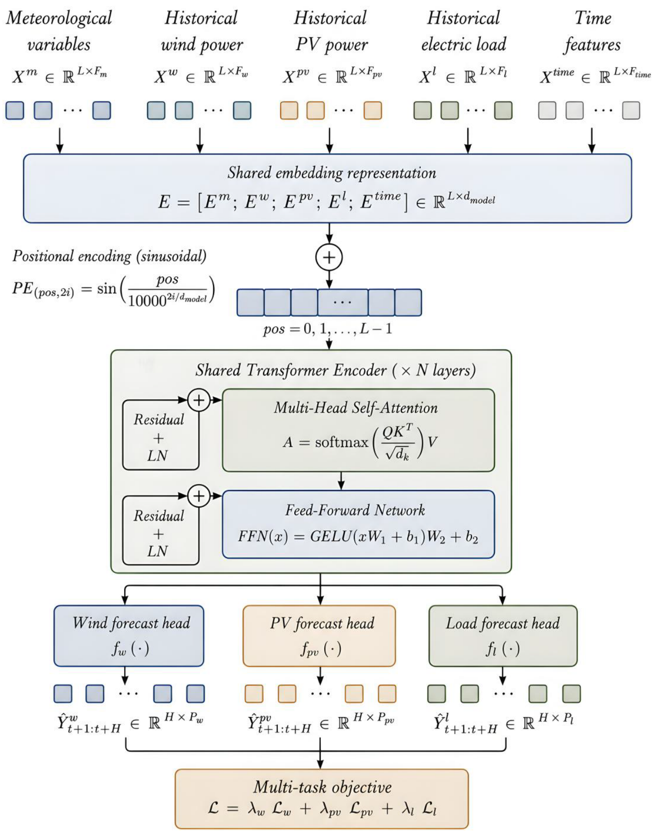


Figure 4. Architecture of the MT-Transformer model

As illustrated in Figure 4, the multi-source variables are embedded and combined with positional information before entering the shared Transformer encoder. The encoder output is a unified temporal representation, which is then delivered to three separate forecasting heads. This design allows the model to complete joint prediction in a unified structure while maintaining the individual variation characteristics of different targets.

Given the input sequence $X_{i:i+L-1}$, the original multi-source variables are first converted into a latent feature representation by a linear projection or embedding operation. Let d_{model} denote the hidden dimension of the model. The embedding process is formulated in Eq. (7).

$$E=\mathrm{Linear}(X) \tag{7}$$

where $E\in R^{L\times d_{model}}$ represents the embedded feature matrix, and Linear(•) is a learnable mapping function. Through this embedding process, meteorological variables, historical wind power output, historical PV output, historical load, and temporal labels are placed in the same feature space. This provides a consistent input representation for the subsequent shared Transformer encoder.

Because the Transformer structure itself does not explicitly record the chronological order of input observations, positional information is introduced to retain the temporal relationship among different time steps [36,37]. In this study, the positional encoding PE is added to the multi-source feature embedding E, and the initial representation entering the shared Transformer encoder is obtained as Eq. (8).

$$Z_0=E+PE \tag{8}$$

The positional encoding can be constructed by sinusoidal functions or by learnable positional embeddings. For the joint forecasting of wind, PV, and load, this design helps the model recognize time-related patterns, including the daily variation of photovoltaic output, the peak-valley pattern of load, and short-term fluctuations in wind power [38].

The shared encoder acts as the central representation-learning component of MT-Transformer. Several encoder blocks are stacked to update the temporal representation, with attention-based interaction, residual paths, normalization, and nonlinear transformation applied inside each block. Before the attention operation, the temporal representation is projected into Q, K, and V so that relevance among different time steps can be evaluated. The self-attention operation is defined in Eq. (9).

$$\mathrm{Attention}(Q, K, V)=\mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

where Q, K, and V denote the query matrix, key matrix, and value matrix, respectively, and dk is the dimension of the key vectors. Through the attention weights between different time steps, the model can identify which historical observations are more relevant to the current prediction. In this way, the model can identify cross-time dependencies within the coupled forecasting sequence.

The multi-head design allows several attention patterns to be learned in parallel rather than relying on a single temporal relation. In this way, the model can learn temporal relationships from different representation subspaces, as shown in Eq. (10) and Eq. (11).

$$\mathrm{MultiHead}(Q, K, V)=\mathrm{Concat}(head_1, head_2,\ldots, head_h)W^O \tag{10}$$
$$head_i=\mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{11}$$

where h is the number of attention heads, and W_i^Q、W_i^K、W_i^V and W^O are learnable parameter matrices. Different heads may emphasize different time scales or cross-variable interactions, allowing the model to represent the complex interactions among multi-source inputs.

In each encoder block, the multi-head attention output is first combined with the original representation through residual connection and layer normalization, and is then transformed by the feed-forward network. This process is expressed in Eq. (12) and Eq. (13).

$$\widetilde{Z}=\mathrm{LayerNorm}(Z+\mathrm{MultiHead}(Z)) \tag{12}$$
$$Z'=\mathrm{LayerNorm}(\widetilde{Z}+\mathrm{FFN}(\widetilde{Z})) \tag{13}$$

where FFN(·) denotes the feed-forward neural network, and Z' is the output of the encoder block. After several encoder blocks, the shared temporal representation is obtained as Eq. (14).

$$Z=\mathrm{Encoder}(Z_0) \tag{14}$$

The representation Z integrates common temporal information needed by the wind, PV, and load forecasting tasks.

After the shared temporal representation Z is generated, the model introduces three separate forecasting heads to produce the outputs of wind, PV, and load. These forecasting heads can be constructed by linear layers or multilayer perceptrons, and their output forms are shown in Eq. (15)-Eq. (17).

$$\hat{P}^{w}_{t+1:t+H}=\mathrm{Head}_w(Z) \tag{15}$$
$$\hat{P}^{pv}_{t+1:t+H}=\mathrm{Head}_{pv}(Z) \tag{16}$$
$$\hat{P}^{load}_{t+1:t+H}=\mathrm{Head}_{load}(Z) \tag{17}$$

where $\mathrm{Head}_w(\cdot)$、$\mathrm{Head}_{pv}(\cdot)$ and $\mathrm{Head}_{load}(\cdot)$ denote the wind power forecasting head, PV power forecasting head, and load forecasting head, respectively. Each head outputs H future values, corresponding to the forecasting horizon. The shared encoder provides common temporal representations for the three tasks, while the separate heads allow each target variable to retain its own changing pattern.

Compared with a standard Transformer used for single-target forecasting, MT-Transformer retains the shared Transformer encoder but replaces the single-output layer with three task-specific forecasting heads. In this study, MT-Transformer uses 3 encoder layers, 4 attention heads, and a hidden dimension of 64, enabling coordinated wind, PV, and load forecasting while preserving target-specific output patterns.

2.4. Multi-task Joint Optimization for Coordinated Forecasting of Wind, PV, and Load

Because the model generates wind, PV, and load predictions at the same time, the training objective needs to combine the errors of the three tasks. Let L_{wind}, L_{pv} and L_{load} represent the forecasting losses for wind, PV, and load, respectively. The total training loss is defined in Eq. (18).

$$Loss=\alpha L_{wind}+\beta L_{pv}+\gamma L_{load} \tag{18}$$

where α, β, and γ are the weighting coefficients of the three task losses. When the three targets are considered equally important, the weights can be set as α=β=γ=1. In preliminary validation experiments, different task-loss weight combinations were compared, and the equal-weight setting achieved more balanced overall performance across wind, PV, and load forecasting tasks. For each single task, the loss is calculated by the mean squared error between the predicted and actual values within the forecasting horizon, as shown in Eq. (19).

$$L_{task}=\frac{1}{H}\sum_{h=1}^{H}\left(P_{t+h}-\hat{P}_{t+h}\right)^2 \tag{19}$$

where P_{t+h} is the actual value at the t+h-th time step, Pt+h denotes the corresponding predicted value, and H is the forecasting horizon. The losses of wind, PV, and load are first calculated separately and then combined through Eq. (18) for joint optimization.

With this multi-task loss design, the three forecasting objectives are updated together during model training. As a result, the shared Transformer encoder can learn temporal representations that better reflect the relationships among system variables. Compared with independent single-task training, this optimization strategy reduces repeated feature extraction and improves the use of shared information among wind, PV, and load.

2.5 Construction of Net Load Fluctuation Indicators and Representation of Regulation Demand

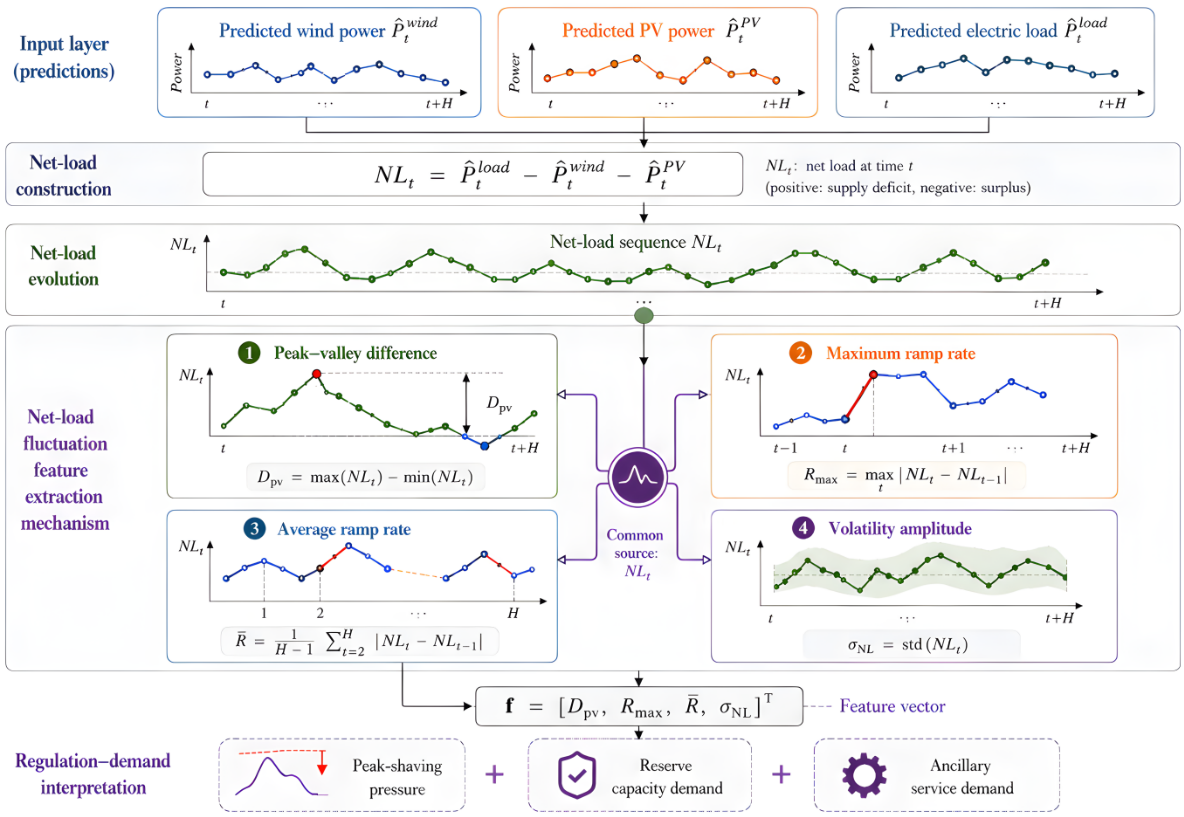


Figure 5. Net load calculation and fluctuation feature extraction

After the model outputs the forecasts of wind power, PV power, and electric load, this study further designs a module for net-load calculation and fluctuation-feature extraction. This module is arranged after the forecasting stage and functions as a post-processing component. It converts the three predicted trajectories into a residual-demand sequence of the power system and then extracts fluctuation indicators related to regulation pressure.

As illustrated in Figure 5, the predicted wind, PV, and load are first combined to calculate the net-load series over the future forecasting horizon. On this basis, four indicators are extracted, including peak-valley difference, maximum ramp rate, average ramp rate, and fluctuation amplitude. These indicators are used to describe system peak-shaving pressure, ramping demand, and operational uncertainty. Therefore, this step connects the forecasting results with the subsequent analysis of coal-power regulation demand.

Net load represents the remaining electricity demand after wind power and PV power are subtracted from the total load. This remaining demand needs to be supplied by coal-fired power units, energy storage, gas-fired units, or other dispatchable resources. For the future t+h-th time step, the predicted net load is calculated as Eq. (20).

$$\hat{P}^{net}_{t+h}=\hat{P}^{load}_{t+h}-\hat{P}^{w}_{t+h}-\hat{P}^{pv}_{t+h},\ h=1, 2,\dots, H \tag{20}$$

where $\hat{P}^{net}_{t+h}$ denotes the predicted net load, $\hat{P}^{load}_{t+h}$ denotes the predicted load, $\hat{P}^{w}_{t+h}$ denotes the predicted wind power output, and $\hat{P}^{pv}_{t+h}$ denotes the predicted PV output. Compared with examining wind power, PV power, or load curves individually, net load provides a more direct representation of the supply and regulation pressure that remains on the system after renewable generation has been absorbed.

To describe the fluctuation characteristics of net load, this study extracts four indicators: peak-valley difference, maximum ramp rate, average ramp rate, and fluctuation amplitude.

First, the peak-valley difference measures the gap between the maximum and minimum net-load values within the forecasting horizon, as shown in Eq. (21).

$$\Delta P=\max(\hat{P}^{net}_{t+1:t+H})-\min(\hat{P}^{net}_{t+1:t+H}) \tag{21}$$

A larger ΔP means that the net load varies more significantly between high-load and low-load periods, which corresponds to stronger peak-shaving pressure.

Second, the maximum ramp rate describes the largest net-load change between two adjacent time steps, as shown in Eq. (22).

$$R_{max}=\max_{h=2,\dots,H}\left|\hat{P}^{net}_{t+h}-\hat{P}^{net}_{t+h-1}\right| \tag{22}$$

A larger R_{max} indicates that the system faces a stronger requirement for short-term fast regulation.

Third, the average ramp rate reflects the overall changing intensity of the net-load sequence during the forecasting horizon, as expressed in Eq. (23).

$$R_{avg}=\frac{1}{H-1}\sum_{h=2}^{H}\left|\hat{P}^{net}_{t+h}-\hat{P}^{net}_{t+h-1}\right| \tag{23}$$

where R_{avg} represents the average variation level of net load over the forecasting horizon.

Fourth, the fluctuation amplitude measures the dispersion degree of the predicted net-load sequence, as shown in Eq. (24).

$$\sigma_{net}=\mathrm{std}(\hat{P}^{net}_{t+1:t+H}) \tag{24}$$

A larger σ_{net} indicates that the net-load series fluctuates more strongly and that the system faces a higher level of operational uncertainty.

Through these indicators, the forecasting outputs of wind, PV, and load generated by MT-Transformer are converted into interpretable net-load fluctuation features. This module retains the numerical information contained in the model outputs and further provides quantitative evidence for analyzing regulation pressure, ramping demand, reserve capacity demand, and ancillary service demand in the experimental section.

For engineering interpretation, the p.u. fluctuation indicators can be converted into MW or MW/h according to the reference system capacity. The peak-valley difference reflects MW-level peak-shaving demand, while the ramp-rate indicators describe MW/h-level regulation requirements. This conversion links normalized net-load indicators with practical coal-power regulation needs, including deep peak shaving, ramping capability, and reserve capacity.

3. Experimental Evaluation and Result Analysis

3.1. Data Description and Experimental Setup

The experimental dataset used in this study was constructed from China-related renewable generation and electricity load datasets. Wind and PV generation data were obtained from the Chinese State Grid Renewable Energy Generation Forecasting Competition dataset, and the electric load series was obtained from a provincial hourly electricity load dataset in China. The selected study period covers 2019-2020. The original renewable-generation data were aggregated to an hourly resolution and aligned with the load and meteorological records by timestamp. After preprocessing, approximately 17,500 valid hourly records were retained.

Wind power, PV power, and electric load were normalized into p.u. values based on their corresponding capacity or load bases. Missing records were filled by linear interpolation, and abnormal values were checked according to the physical range of each variable before normalization. The p.u. values reported in the experiments are normalized relative values, which are used to compare fluctuation intensity across models and time periods. Therefore, the net-load peak-valley difference and ramp rate discussed later represent relative fluctuation levels within the forecasting horizon, rather than the original unnormalized power values.

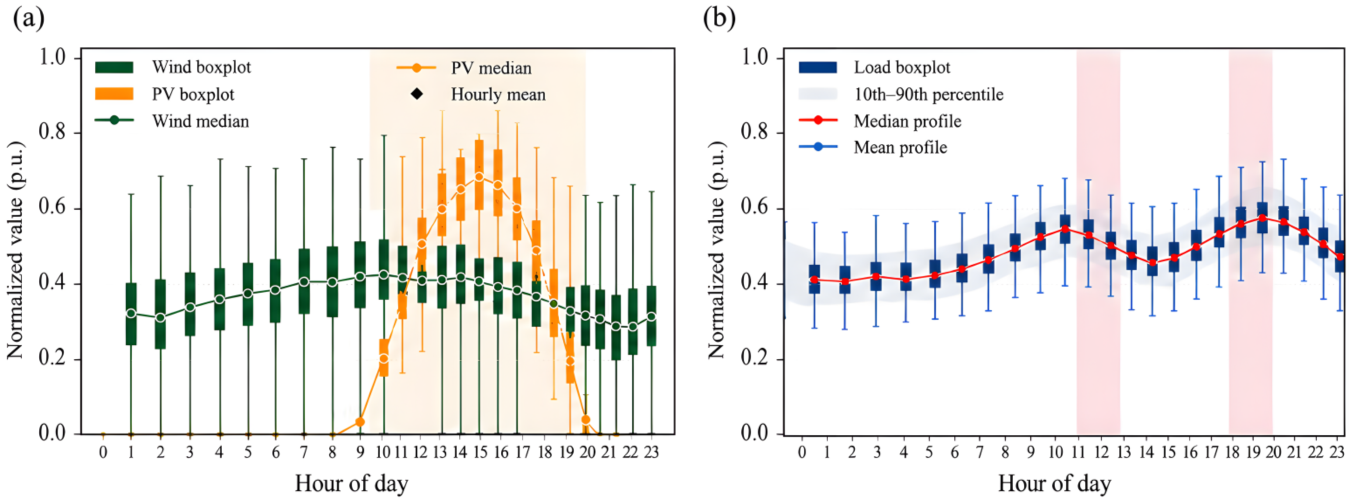


Figure 6. Intraday distribution characteristics of wind power output, PV output, and electric load
(a) Hourly distribution characteristics of wind power output and photovoltaic power output
(b) Hourly distribution characteristics of electric load.

The time resolution is one hour, and the forecasting horizon is the next 24 hours. The input set combines weather-related variables, historical wind and PV generation, previous load observations, and calendar-based labels. The outputs are the 24-hour-ahead forecasts of the three target series. The samples were divided chronologically, with the first 70% used for training, the following 15% for validation, and the final 15% for testing. Random partitioning was avoided because time-series samples have temporal dependence, and random shuffling may introduce future information into the training process. The training set was used to learn model parameters, the validation set was used for hyperparameter adjustment and early stopping, and the test set was reserved for evaluating generalization performance.

To present the intraday distribution patterns of the three targets, Figures 6(a) and 6(b) show their hourly boxplot distributions. Figures 6(a) and 6(b) indicate that the three targets have clearly different intraday distribution characteristics. Wind power output varies more widely across hours, PV output follows a distinct daytime generation pattern, and electric load shows a relatively stable peak-valley structure within the day. These differences in fluctuation range, periodicity, and distribution shape suggest that the three variables are governed by different temporal dynamics. This provides data-level support for adopting a multi-task joint forecasting framework.

For evaluating the effectiveness of MT-Transformer, this study selects baseline models from several categories, including persistence forecasting, traditional statistical modeling, machine learning, recurrent neural networks, and Transformer-based methods. The specific comparison models are Persistence Model, ARIMA, SVR, Random Forest, XGBoost, LSTM, GRU, Transformer, Informer, and Autoformer. This setting enables the proposed model to be compared under different forecasting paradigms.

For implementation, MT-Transformer was configured with 3 Transformer encoder layers, a hidden dimension of 64, 4 attention heads, and a dropout rate of 0.1. The model was trained using the Adam optimizer with a learning rate of 1e-3 and a batch size of 64. The maximum number of training epochs was set to 100, and early stopping was applied when the validation loss did not improve for 10 consecutive epochs. The random seed was fixed at 42 for reproducibility. For baseline models, the input window, forecasting horizon, chronological data split, and evaluation metrics were kept consistent with MT-Transformer. Their main hyperparameters were tuned on the validation set.

3.2. Forecasting Performance Comparison among Different Models

The comparative results in Table 1 show clear performance differences among the evaluated forecasting methods. The evaluation metrics include MAE, RMSE, MAPE, SMAPE, NRMSE, and R^2, where SMAPE denotes symmetric mean absolute percentage error. The Persistence Model and ARIMA produce larger errors, which suggests that simple continuation assumptions and linear time-series structures are insufficient for describing the nonlinear variations of wind power, PV power, and electric load. Compared with these traditional models, SVR, Random Forest, and XGBoost obtain better results to varying degrees. SVR performs well in MAE and MAPE, Random Forest shows relatively stable RMSE and R^2 values, and XGBoost achieves competitive results on several indicators. This result implies that well-designed input features can still make conventional machine learning models competitive.

Because photovoltaic output may approach zero during nighttime periods, MAPE can become sensitive and may overstate the relative error under low-output conditions. For this reason, SMAPE and NRMSE are introduced as supplementary metrics. MAE, RMSE, and R^2 are also considered together so that model performance can be assessed from multiple perspectives.

Table 1. Comparative forecasting results of different models

Model	MAE	RMSE	MAPE/%	SMAPE/%	NRMSE	R^2
Persistence Model	0.1326	0.1804	23.36	20.17	0.1879	0.8258
ARIMA	0.1214	0.1661	20.85	18.46	0.1718	0.8509
SVR	0.1068	0.1594	18.74	17.31	0.1652	0.8698
Random Forest	0.1142	0.1476	18.29	16.08	0.1537	0.8856
XGBoost	0.0976	0.1448	16.68	15.42	0.1491	0.8943
LSTM	0.1013	0.1365	15.52	14.16	0.1414	0.9072
GRU	0.0946	0.1402	16.21	13.71	0.1453	0.9041
Transformer	0.0924	0.1287	14.26	13.18	0.1346	0.9179
Informer	0.0881	0.1239	14.72	12.86	0.1305	0.9248
Autoformer	0.0869	0.1276	13.79	12.61	0.1324	0.9271
MT-Transformer	0.0848	0.1254	14.05	12.43	0.1269	0.9302

Deep temporal models generally improve prediction quality, although their advantages are distributed unevenly across different metrics. LSTM has slightly lower RMSE and higher R^2 than GRU, while GRU performs better in MAE and SMAPE. Among the Transformer-based models, Informer obtains the lowest RMSE, and Autoformer has an advantage in MAPE. MT-Transformer performs well in MAE, SMAPE, NRMSE, and R^2. Although it is not the optimal model for every single metric, it achieves a more balanced result across the main indicators. This suggests that the shared encoder and the multi-task forecasting heads help improve the joint prediction of wind, PV, and load.

To further evaluate the robustness of the forecasting results, repeated experiments were conducted using three random seeds, and Table 2 reports the mean and standard deviation of the main deep-learning models. The results show that MT-Transformer maintains stable and competitive performance across different random seeds. It achieves the lowest MAE, SMAPE, and NRMSE and the highest R^2, while remaining close to the best results in RMSE and MAPE. This indicates that its overall advantage is not caused by a single random initialization.

Table 2. Robustness comparison under different random seeds

Model	MAE	RMSE	MAPE/%	SMAPE/%	NRMSE	R^2
LSTM	0.0917 ± 0.0025	0.1331 ± 0.0032	15.42 ± 0.45	13.31 ± 0.36	0.1378 ± 0.0033	0.9181 ± 0.0063
Transformer	0.0895 ± 0.0022	0.1294 ± 0.0029	14.96 ± 0.39	13.06 ± 0.32	0.1340 ± 0.0030	0.9235 ± 0.0055
Informer	0.0881 ± 0.0019	0.1239 ± 0.0024	14.72 ± 0.35	12.86 ± 0.29	0.1305 ± 0.0026	0.9248 ± 0.0049
Autoformer	0.0869 ± 0.0018	0.1276 ± 0.0023	13.79 ± 0.32	12.61 ± 0.26	0.1324 ± 0.0024	0.9271 ± 0.0043
MT-Transformer	0.0848 ± 0.0013	0.1254 ± 0.0018	14.05 ± 0.28	12.43 ± 0.22	0.1269 ± 0.0019	0.9302 ± 0.0035

Table 3. Task-specific forecasting performance of MT-Transformer

Target	MAE	RMSE	SMAPE/%	R^2
Wind power	0.0896	0.1328	13.84	0.9187

PV power	0.0785	0.1169	11.76	0.9364
Electric load	0.0863	0.1265	11.69	0.9355

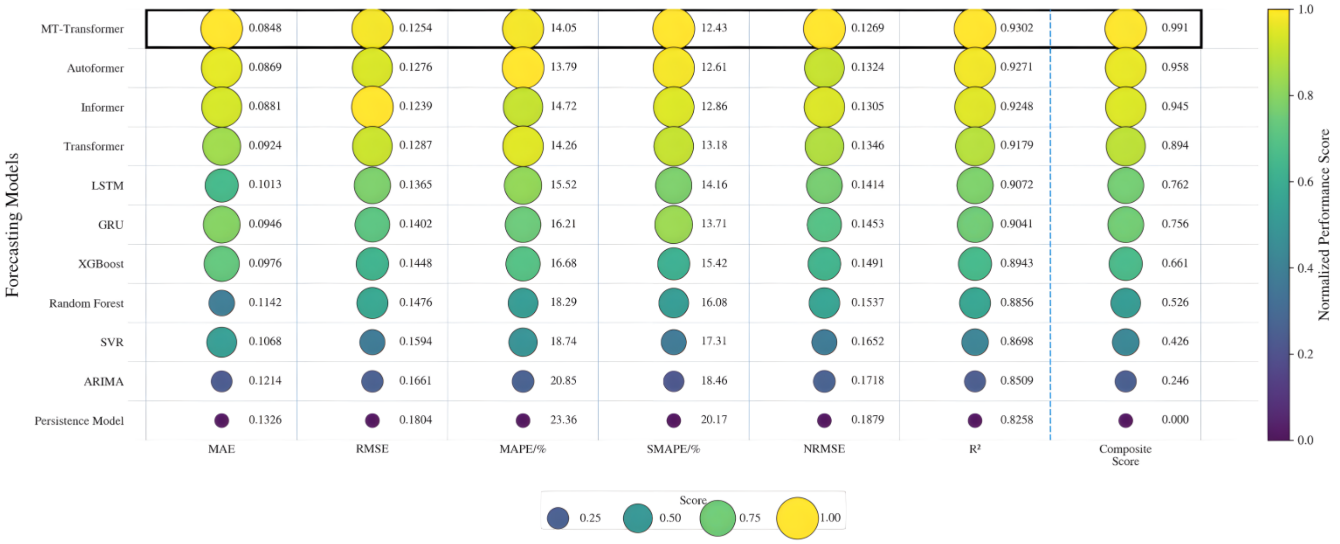


Figure 7. Bubble-based comparison of model performance metrics

Table 3 shows that MT-Transformer achieves stable forecasting performance across wind power, PV power, and electric load. The results suggest that the multi-task structure can support coordinated forecasting of all three targets, although the prediction difficulty varies among different targets.

Figure 7 further presents the multi-metric comparison among different models in the form of a bubble chart. The distribution of the indicators shows that MT-Transformer maintains a relatively balanced performance, while Informer and Autoformer retain advantages in some specific error metrics. This visual comparison is consistent with the numerical results reported in Table 1.

3.3. Forecasting Accuracy and Multi-Step Stability Analysis

For dispatch-oriented applications, mean error values are insufficient to describe the practical usefulness of a forecasting model. Therefore, this study further uses forecasting accuracy (FA), FA based on SMAPE (FA-SMAPE), and Acc10%, which denotes the percentage of predictions with relative error within 10%, to evaluate forecasting accuracy, and compares error variations under different forecasting horizons to assess multi-step stability [39]. The corresponding results are listed in Table 4.

Table 4. Accuracy and horizon-wise error comparison of forecasting models

Model	FA/%	FA-SMAPE/%	Acc10%	1h Error	6h Error	12h Error	24h Error
Persistence Model	76.94	80.06	61.20	0.0946	0.1198	0.1487	0.1873
ARIMA	79.11	82.03	64.85	0.0872	0.1104	0.1361	0.1712
SVR	81.38	82.74	70.12	0.0618	0.1126	0.1374	0.1698
Random Forest	80.52	84.61	67.84	0.0845	0.0954	0.1216	0.1539
XGBoost	84.06	84.36	72.91	0.0714	0.0918	0.1279	0.1467
LSTM	82.86	86.03	74.28	0.0756	0.0837	0.1089	0.1415
GRU	84.52	85.42	76.03	0.0667	0.0906	0.1158	0.1342
Transformer	86.18	86.49	76.62	0.0642	0.0788	0.1047	0.1328
Informer	85.16	87.52	78.35	0.0613	0.0824	0.0978	0.1296
Autoformer	85.74	86.91	78.92	0.0631	0.0756	0.1012	0.1264
MT-Transformer	86.08	87.71	79.64	0.0601	0.0774	0.0965	0.1236

Table 4 compares forecasting accuracy and multi-step errors under different horizons. Traditional models have lower accuracy, and their errors increase more obviously as the forecasting horizon becomes longer. This indicates that they have limited ability to adapt to multivariate nonlinear fluctuations. Machine learning models and recurrent neural network models improve the results to some extent, but error accumulation is still visible in longer-horizon forecasting [40]. Deep time-series models show better stability overall, although each structure has its own local advantage. LSTM is relatively stable in 6h and 12h Error, GRU performs better in Acc10% and 24h Error, Transformer has a slight advantage in FA, and Autoformer obtains a lower 6h Error. MT-Transformer performs well in FA-SMAPE, Acc10%, 1h Error, 12h Error, and 24h Error, showing balanced performance in overall accuracy, short-horizon error control, and long-horizon stability. Figures 8(a) and 8(b) show that the errors of all models increase from 1 hour to 24 hours, but the growth of MT-Transformer is relatively smaller. This pattern further suggests that shared representation learning and task-specific decoding help control error accumulation.

In addition to the horizon-based stability analysis, high-fluctuation periods in the test set were further reviewed to examine model behavior under abrupt-change conditions. These periods involve rapid variations in renewable output and load demand, including short-term changes in PV generation, wind power, and electric load. The results indicate that MT-Transformer can generally follow the main changing trend, although local prediction errors may increase when the fluctuation becomes more intense. This suggests that the proposed model has a certain level of robustness under non-stationary conditions, while forecasting under more extreme or strong-fluctuation scenarios remains a challenging issue for future work.

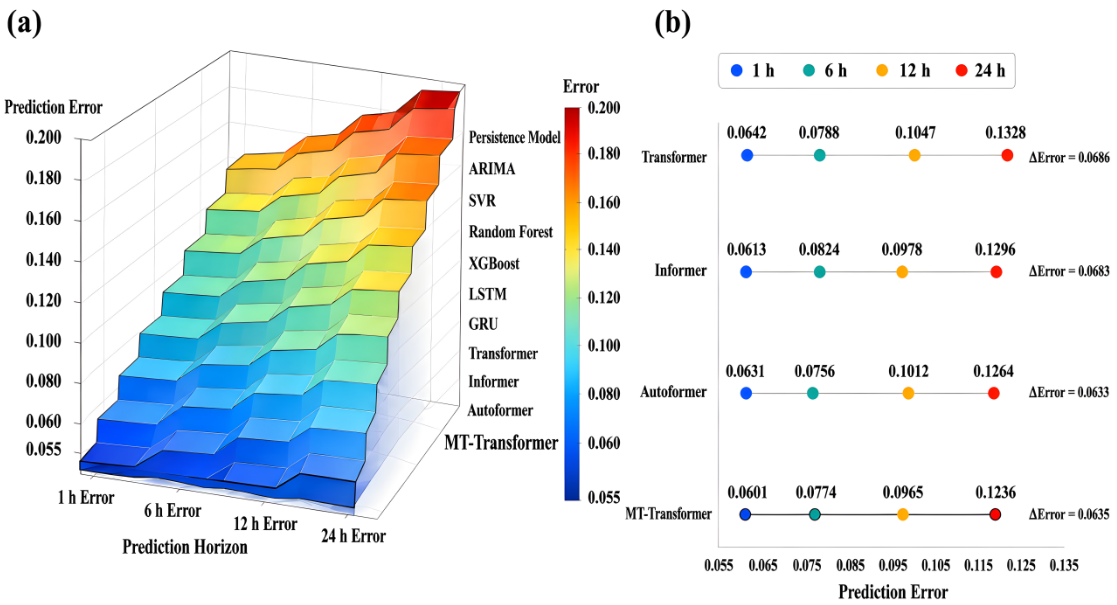


Figure 8. Horizon-wise error variation of forecasting models
(a) 3D error surface of all compared models
(b) Stability comparison of advanced models

3.4. Model Training Process and Variable Correlation Analysis

The training process is examined to evaluate the convergence behavior and stability of MT-Transformer. Figure 9 presents the comprehensive convergence diagnosis during model training. Both the training loss and validation loss decrease in the early stage and gradually stabilize in later epochs. This pattern indicates that the model can learn temporal features from wind, PV, and load series in a stable manner. The two curves do not show obvious separation, suggesting that serious overfitting does not occur during training.

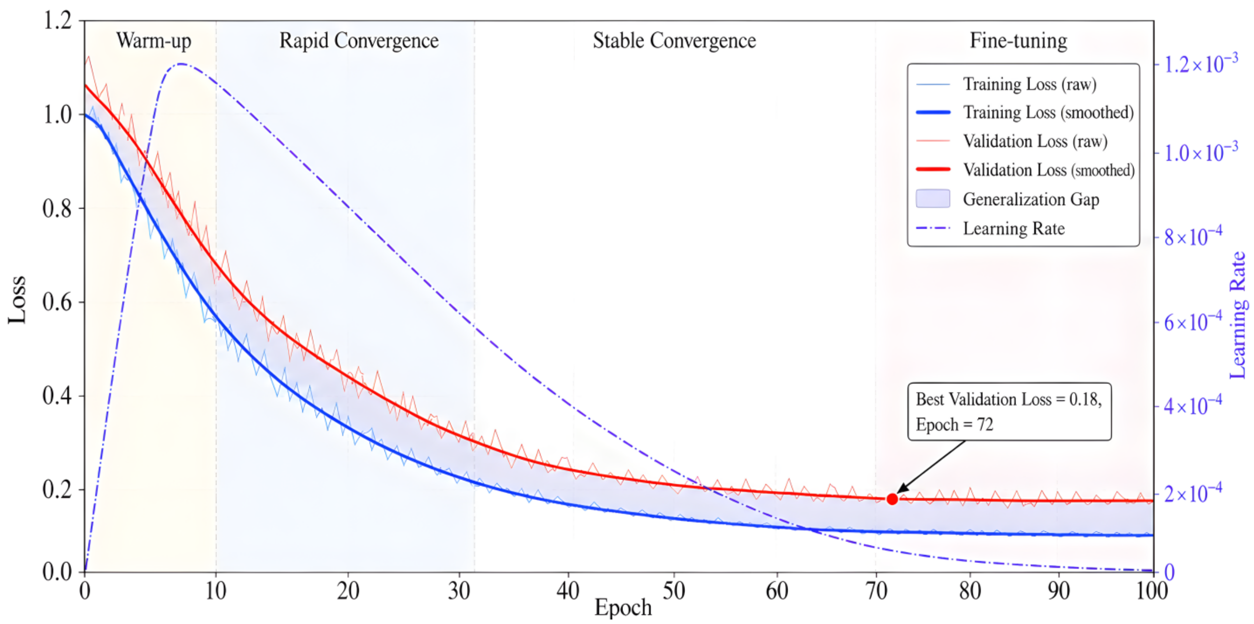
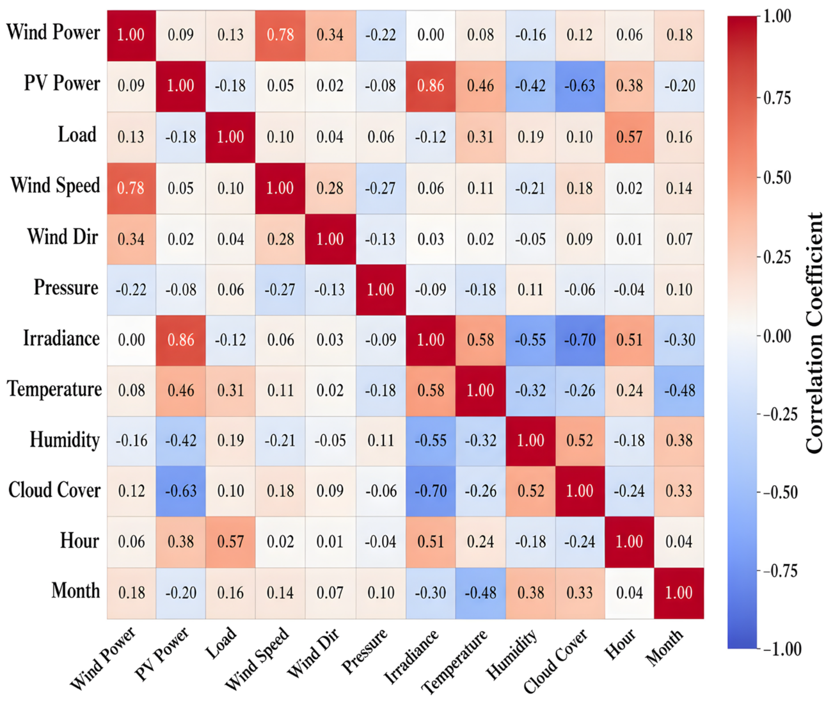


Figure 9. Training convergence diagnosis of MT-Transformer

Although wind, PV, and load series are driven by different mechanisms, they still have statistical relationships with meteorological factors, historical output, and temporal variables. Figure 10 shows the correlation distribution among the wind–solar–load related variables. Wind power output is closely associated with wind speed and wind direction, photovoltaic output is more strongly related to solar irradiance, temperature, and temporal labels, and electric load is influenced by historical load, time-related features, and meteorological conditions [41]. The heatmap therefore supports the necessity of multi-source feature integration and also provides data-level evidence for using a shared encoder in multi-task time-series modeling.



	Wind Power	PV Power	Load	Wind Speed	Wind Dir	Pressure	Irradiance	Temperature	Humidity	Cloud Cover	Hour	Month
Wind Power	1.00	0.09	0.13	0.78	0.34	-0.22	0.00	0.08	-0.16	0.12	0.06	0.18
PV Power	0.09	1.00	-0.18	0.05	0.02	-0.08	0.86	0.46	-0.42	-0.63	0.38	-0.20
Load	0.13	-0.18	1.00	0.10	0.04	0.06	-0.12	0.31	0.19	0.10	0.57	0.16
Wind Speed	0.78	0.05	0.10	1.00	0.28	-0.27	0.06	0.11	-0.21	0.18	0.02	0.14
Wind Dir	0.34	0.02	0.04	0.28	1.00	-0.13	0.03	0.02	-0.05	0.09	0.01	0.07
Pressure	-0.22	-0.08	0.06	-0.27	-0.13	1.00	-0.09	-0.18	0.11	-0.06	-0.04	0.10
Irradiance	0.00	0.86	-0.12	0.06	0.03	-0.09	1.00	0.58	-0.55	-0.70	0.51	-0.30
Temperature	0.08	0.46	0.31	0.11	0.02	-0.18	0.58	1.00	-0.32	-0.26	0.24	-0.48
Humidity	-0.16	-0.42	0.19	-0.21	-0.05	0.11	-0.55	-0.32	1.00	0.52	-0.18	0.38
Cloud Cover	0.12	-0.63	0.10	0.18	0.09	-0.06	-0.70	-0.26	0.52	1.00	-0.24	0.33
Hour	0.06	0.38	0.57	0.02	0.01	-0.04	0.51	0.24	-0.18	-0.24	1.00	0.04
Month	0.18	-0.20	0.16	0.14	0.07	0.10	-0.30	-0.48	0.38	0.33	0.04	1.00

Figure 10. Correlation heatmap of wind power, PV power, load, and related variables

3.5. Model Ablation and Module Contribution Analysis

Ablation experiments are carried out to examine how different components contribute to the forecasting performance of MT-Transformer. Three comparative settings are designed: removing the multi-task structure, removing temporal features, and removing meteorological features. In the w/o Multi-task setting, the shared encoder and multi-task forecasting heads are removed, and separate single-task models are built for wind, PV, and load. In the w/o Time Features setting, temporal variables such as hour, day of week, month, season, holiday, and working-day indicators are excluded. In the w/o Meteorological Features setting, meteorological variables including wind speed, wind direction, air pressure, solar irradiance, temperature, humidity, and cloud cover are removed. The results of the ablation experiments are reported in Table 5.

Table 5. Ablation experiment results of MT-Transformer

Model	MAE	RMSE	MAPE/%	SMAPE/%	FA/%	R^2
MT-Transformer	0.0848	0.1254	14.05	12.43	86.08	0.9302
w/o Multi-task	0.0924	0.1316	15.37	13.48	84.91	0.9176
w/o Time Features	0.0887	0.1334	14.69	13.31	85.38	0.9159
w/o Meteorological Features	0.0951	0.1357	15.82	13.86	84.27	0.9098

Table 5 shows that the full MT-Transformer obtains the best overall results among all ablation settings, which indicates that the multi-task structure, temporal features, and meteorological features all contribute to forecasting performance. When the multi-task structure is removed, MAE, MAPE, and SMAPE increase clearly. This suggests that separate modeling weakens the use of shared temporal information among wind, PV, and load. When temporal features are removed, MAE and MAPE increase slightly, and RMSE and R^2 also become worse to some extent. This result indicates that temporal labels help the model capture daily cycles, load peak-valley patterns, and photovoltaic output variation. When meteorological features are removed, the model shows the weakest performance, especially in MAE and MAPE. This is consistent with the strong dependence of renewable power output on exogenous weather factors such as wind speed, solar irradiance, and temperature. Figure 11 further presents the performance changes under different ablation settings and visually confirms the importance of meteorological inputs.

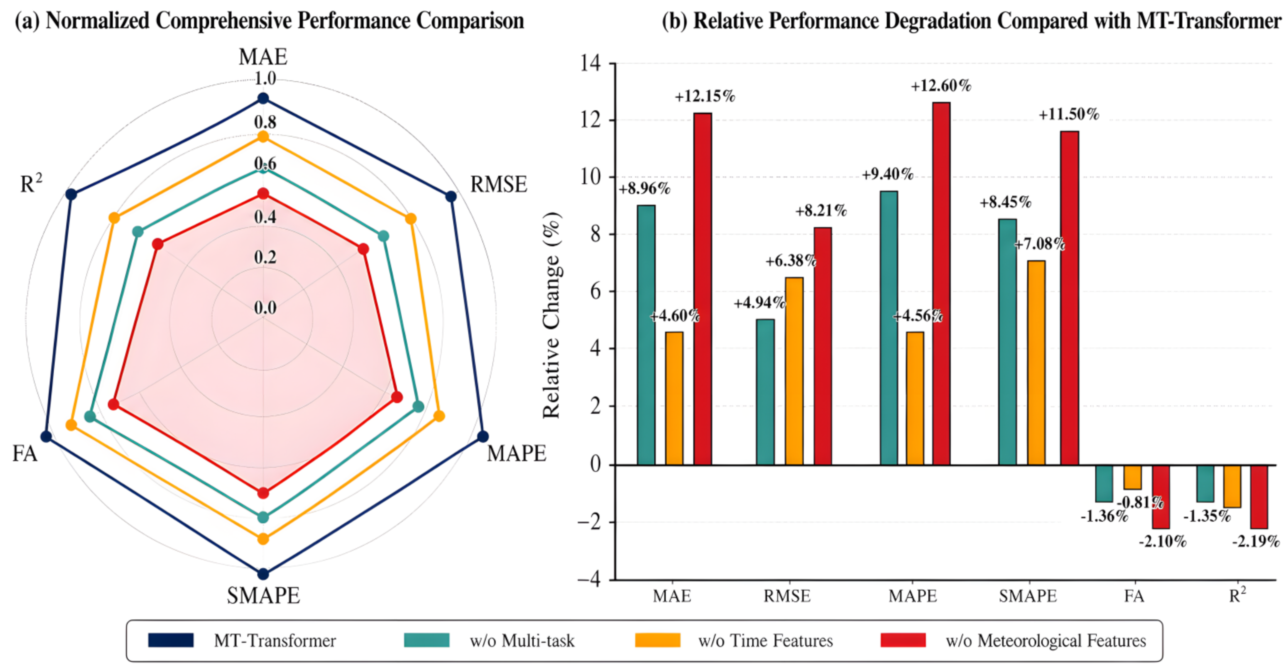


Figure 11. Ablation experiment comparison of MT-Transformer

3.6. Net Load Fluctuation Features and Coal-Power Regulation Demand Identification

After the forecasts of wind, PV, and load series are obtained, the predicted net load is further calculated, and four fluctuation indicators are extracted: peak-valley difference, maximum ramp rate, average ramp rate, and fluctuation amplitude. This section goes beyond forecasting-error comparison. It further examines whether the model outputs can be converted into operational indicators that describe system regulation pressure. Table 6 summarizes the net-load fluctuation indicators and the corresponding coal-power regulation demands.

Table 6. Net-load indicators and corresponding regulation requirements

Indicator	Numerical result	System operation implication	Corresponding coal-power regulation demand
Peak-valley difference	0.462 p.u.	Increased peak-shaving pressure	Deep peak-shaving demand
Maximum ramp rate	0.087 p.u./h	Rapid short-term variation pressure	Fast ramping capability demand
Average ramp rate	0.041 p.u./h	Continuous regulation pressure	Flexible regulation demand
Fluctuation amplitude	0.118 p.u.	Increased operational uncertainty	Reserve capacity and ancillary service demand

Table 7. Validation of predicted net-load fluctuation indicators

Indicator	Actual value	Predicted value	Absolute error	Relative error
Peak-valley difference	0.474 p.u.	0.462 p.u.	0.012 p.u.	2.53%
Maximum ramp rate	0.091 p.u./h	0.087 p.u./h	0.004 p.u./h	4.40%
Average ramp rate	0.043 p.u./h	0.041 p.u./h	0.002 p.u./h	4.65%
Fluctuation amplitude	0.122 p.u.	0.118 p.u.	0.004 p.u.	3.28%
Large ramping-event detection accuracy	91.67%	—	—	—

According to Table 6, the peak-valley difference of the predicted net load reaches 0.462 p.u., corresponding to 462 MW under the reference system capacity of 1000 MW. This value indicates that residual demand changes markedly between high-load and low-load periods within the forecasting horizon. Such variation implies evident peak-shaving pressure on the system. For coal-fired power units, a larger peak-valley difference means that they need to reduce output or participate in deep peak-shaving during low net-load periods, while providing capacity support when net load rises [42]. Therefore, the MW-level peak-shaving demand is closely related to the minimum stable output and deep peak-shaving capability of coal-fired units.

In terms of ramping behavior, the maximum ramp rate reaches 0.087 p.u./h, corresponding to 87 MW/h. This means that the net load changes sharply between some adjacent time steps and reflects the need for short-term response capability and fast load-following support. The average ramp rate is 0.041 p.u./h, corresponding to 41 MW/h, indicating that net-load variation is not only caused by local abrupt changes but also involves continuous regulation pressure over the forecasting horizon. The fluctuation amplitude is 0.118 p.u., corresponding to 118 MW, showing that the predicted net-load sequence still has dispersion and uncertainty. Therefore, reserve capacity and ancillary service resources are needed to support stable system operation.

The reliability of the net-load-based regulation demand identification was then validated by comparing predicted and actual net-load indicators. The actual net load was calculated using the observed wind power, PV power, and electric load in the test set, and the same fluctuation indicators were extracted from both the predicted and actual net-load sequences. As shown in Table 7, the predicted indicators are close to the actual indicators in terms of peak-valley difference, maximum ramp rate, average ramp rate, and fluctuation amplitude. The relative errors of the main indicators remain within 5%, and the large ramping-event detection accuracy reaches 91.67%. These results indicate that the predicted net-load indicators can effectively reflect the actual fluctuation characteristics and provide reliable support for coal-power regulation demand analysis.

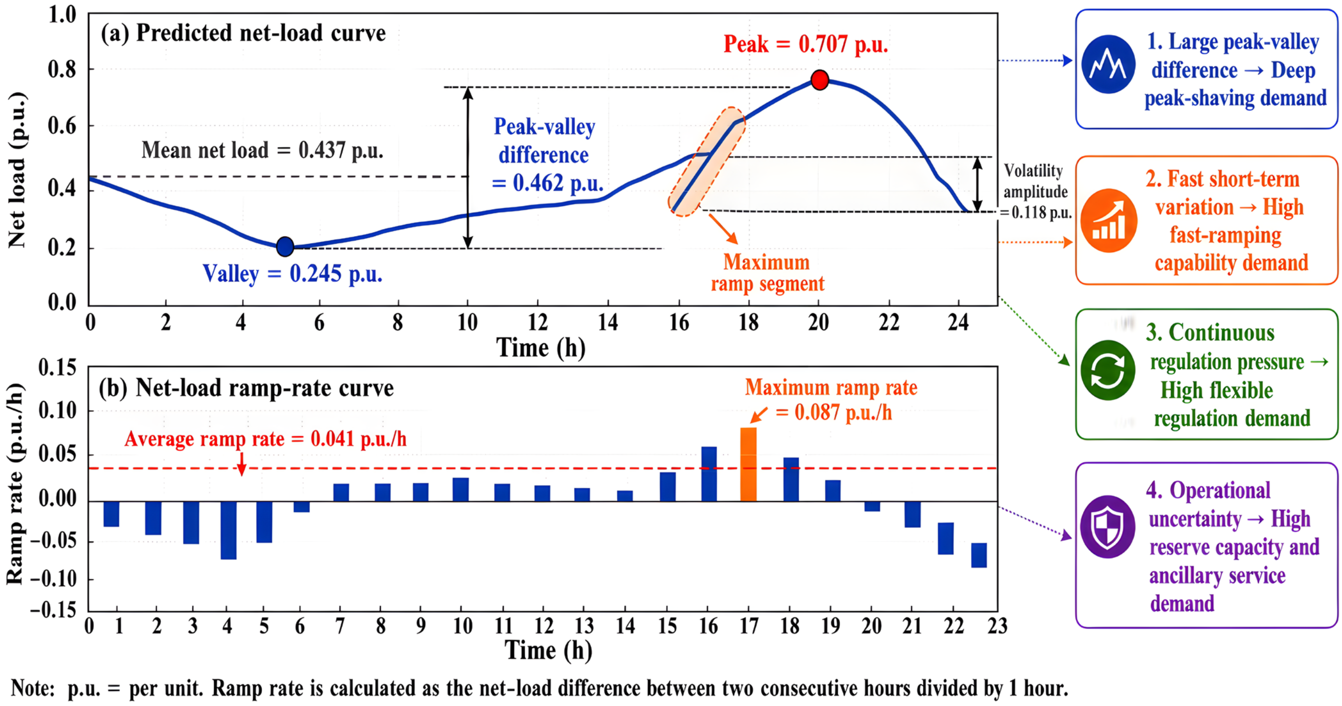


Figure 12. Predicted net-load curve and derived fluctuation features

Figure 12 shows the predicted net-load curve and the corresponding fluctuation features. The curve presents the high and low variations of net load within the forecasting horizon. The distance between the peak and valley corresponds to the peak-valley difference, while the changes between adjacent time periods reflect ramping demand [43]. The curve rises and falls noticeably in several periods, which indicates that the system requires both deep peak-shaving capability and rapid response capability.

These results show that net-load fluctuation indicators can convert the forecasting outputs of MT-Transformer into operationally interpretable indicators. Therefore, joint forecasting of wind, PV, and load can support not only forecasting error evaluation, but also the identification of coal-power peak regulation pressure, fast ramping demand, reserve capacity demand, and ancillary service demand.

4. Conclusions and Policy Implications

This study examines coordinated forecasting of renewable output and electric load under high renewable-energy penetration. An MT-Transformer-based multi-task framework is developed, and its prediction outputs are further used to derive net-load and fluctuation indicators. The results show that the proposed framework can describe both renewable-side variability and demand-side changes. Compared with single-target forecasting, this approach provides a more direct basis for analyzing net-load variation and identifying the peak regulation, ramping, and reserve-support requirements of coal-fired power units.

In terms of forecasting performance, MT-Transformer achieves strong results in overall error, prediction accuracy, and multi-step stability. The model obtains an MAE of 0.0848, an RMSE of 0.1254, and an R^2 of 0.9302. Compared with the Persistence Model, its MAE and RMSE decrease by 36.05% and 30.49%, respectively. Compared with other deep learning and Transformer-based models, MT-Transformer remains competitive on most key indicators. This suggests that the shared Transformer encoder and multi-task forecasting heads can effectively learn common temporal features among wind, PV, and load. The ablation results further confirm the contribution of the multi-task structure, temporal features, and meteorological features. Among them, removing meteorological features leads to the most obvious performance decline.

From the net-load perspective, the peak-valley difference within the forecasting horizon reaches 0.462 p.u., and the maximum ramp rate reaches 0.087 p.u./h. These values indicate that the system faces both peak-shaving pressure and short-term fast regulation demand under the joint influence of renewable output fluctuation and load variation. The average ramp rate and fluctuation amplitude further suggest that net-load variation includes not only local peak-valley differences but also continuous fluctuation characteristics. Therefore, the MT-Transformer forecasting outputs can be used not only for joint forecasting accuracy evaluation, but also for constructing net-load fluctuation indicators. These indicators provide quantitative support for coal-power peak regulation, rapid ramping, reserve capacity allocation, and ancillary service demand identification.

Based on these findings, coal-power policy adaptation should place greater emphasis on capacity value and regulation value. Indicators such as peak-valley difference, ramp rate, and fluctuation amplitude can be introduced into the evaluation of capacity compensation, ancillary services, and flexibility retrofitting. This would provide quantitative evidence for coal-power peak regulation, reserve capacity allocation, and fast ramping services. In future policy implementation, regional renewable-energy structures and net-load fluctuation levels should also be considered. Differentiated flexibility retrofitting and low-carbon transformation of coal-fired power units should be promoted, so that coal power can provide stronger regulation support in power systems with high renewable energy penetration.

Overall, the study links artificial intelligence forecasting with net-load fluctuation analysis and provides a quantitative route for identifying coal-power regulation demand. Future work can extend the model to longer time spans and multi-region datasets. Probabilistic forecasting, energy storage participation, and electricity market mechanisms may also be incorporated to enhance model generalization and practical interpretability in complex power-system scenarios.

Acknowledgments

This work was supported by the Research on Regional Industrial Structure Optimization Underpinning the Formulation of China Energy Investment Corporation's Strategy Map.

References

[1] Zhang S, Chen W. Assessing the energy transition in China towards carbon neutrality with a probabilistic framework[J]. Nature Communications, 2022, 13: 87.
[2] An K, Zheng X, Shen J, et al. Repositioning coal power to accelerate net-zero transition of China's power system[J]. Nature Communications, 2025, 16(1): 2400.
[3] Yin G, Li B, Fedorova N, et al. Orderly retire China's coal-fired power capacity via capacity payments to support renewable energy expansion[J]. iScience, 2021, 24(11): 103287.
[4] Lin B, Liu Z. Assessment of flexible coal power and battery energy storage system in supporting renewable energy[J]. Energy, 2024, 313: 133805.
[5] Hirth L. The market value of variable renewables: The effect of solar wind and solar power variability on their relative price[J]. Energy Economics, 2013, 38: 218-236.
[6] Tziolis G, Livera A, Theocharides S, et al. Net load forecasting: A comprehensive literature review[J]. Sustainable Energy Technologies and Assessments, 2025, 82: 104450.
[7] Kerkau S, Sepasi S, Howlader H O R, et al. Day-ahead net load forecasting for renewable integrated buildings using XGBoost[J]. Energies, 2025, 18(6): 1518.
[8] Telle J S, Upadhaya A, Schönfeldt P, et al. Probabilistic net load forecasting framework for application in distributed integrated renewable energy systems[J]. Energy Reports, 2024, 11: 2535-2553.
[9] Song Z, Gu Y, Liu H, et al. Application of deep learning in wind, solar, and ocean energy: An analysis of prediction, optimization, and operation & maintenance[J]. Renewable and Sustainable Energy Reviews, 2026, 230: 116663.
[10] Hasan M, Mifta Z, Papiya S J, et al. A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications[J]. Energy Conversion and Management: X, 2025, 26: 100922.
[11] Malakouti M. Machine learning for renewable energy forecasting: Advances, challenges, and future directions[J]. Sustainable Energy Research, 2026, 13: 14.
[12] Gupta M, Arya A, Varshney U, et al. A review of PV power forecasting using machine learning techniques[J]. Progress in Engineering Science, 2025, 2(1): 100058.
[13] Box G E P, Jenkins G M, Reinsel G C, et al. Time Series Analysis: Forecasting and Control[M]. 5th ed. Hoboken: Wiley, 2015.
[14] Ding J W, Chuang M J, Tseng J S, et al. Reanalysis and ground station data: Advanced data preprocessing in deep learning for wind power prediction[J]. Applied Energy, 2024, 375: 124129.
[15] Cheng J, Luo X, Jin Z. Integrating domain knowledge into Transformer for short-term wind power forecasting[J]. Energy, 2024, 312: 133511.
[16] Piantadosi G, Dutto S, Galli A, et al. PV power forecasting: A Transformer based framework[J]. Energy and AI, 2024, 18: 100444.
[17] Hussan U, Wang H, Peng J, et al. Transformer-based renewable energy forecasting: A comprehensive review[J]. Renewable and Sustainable Energy Reviews, 2026, 226: 116356.
[18] Hu J, Hu W, Cao D, et al. Probabilistic net load forecasting based on transformer network and Gaussian process-enabled residual modeling learning method[J]. Renewable Energy, 2024, 225: 120253.
[19] Ahmad A, Xiao X, Mo H, et al. TFTformer: A novel transformer based model for short-term load forecasting[J]. International Journal of Electrical Power & Energy Systems, 2025, 166: 110549.
[20] Pandžić F, Capuder T. Advances in short-term solar forecasting: A review and benchmark of machine learning methods and relevant data sources[J]. Energies, 2023, 17(1): 97.
[21] Wang Y, Zou R, Liu F, et al. A review of wind speed and wind power forecasting with deep neural networks[J]. Applied Energy, 2021, 304: 117766.
[22] Manandhar P, Rafiq H, Rodriguez-Ubinas E, et al. New forecasting metrics evaluated in Prophet, Random Forest, and Long Short-Term Memory models for load forecasting[J]. Energies, 2024, 17(23): 6131.
[23] Taylor S J, Letham B. Forecasting at scale[J]. The American Statistician, 2018, 72(1): 37-45.
[24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017, 30: 5998-6008.
[25] Wu X, Zhan H, Hu J, et al. Non-stationary GNNCrossformer: Transformer with graph information for non-stationary multivariate spatio-temporal wind power data forecasting[J]. Applied Energy, 2025, 377: 124492.
[26] Vandenhende S, Georgoulis S, Van Gansbeke W, et al. Multi-task learning for dense prediction tasks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3614-3633.
[27] Dou X, Wang Y, Zhang H, et al. An adaptive forecasting method for net load ramping demand in power systems[J]. IET Generation, Transmission & Distribution, 2025, 19(1): e70176.
[28] Verdone A, Galletti A, Mazzoni S, et al. A review of solar and wind energy forecasting: From single-site to spatio-temporal

deep learning approaches[J]. Applied Energy, 2025, 392: 125843.
[29] Lim B, Zohren S, Roberts S. Time-series forecasting with deep learning: A survey[J]. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200209.
[30] Tao K, Zhang X, Wang Y, et al. Operational day-ahead photovoltaic power forecasting using transformer-based deep learning models[J]. Applied Energy, 2024, 373: 123986.
[31] Liu H, Zhang Z, Wang Y, et al. Probabilistic forecasting of multiple plant day-ahead renewable power generation sequences with data privacy preserving[J]. Energy and AI, 2025, 19: 100463.
[32] Yang T, Yang Z, Li F, et al. A short-term wind power forecasting method based on multivariate signal decomposition and variable selection[J]. Applied Energy, 2024, 360: 122759.
[33] Caruana R. Multi-task learning[J]. Machine Learning, 1997, 28(1): 41-75.
[34] Hyndman R J, Athanasopoulos G. Forecasting: Principles and Practice[M]. 3rd ed. Melbourne: OTexts, 2021.
[35] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient Transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(12): 11106-11115.
[36] Bashir T, Wang H, Tahir M, et al. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models[J]. Renewable Energy, 2025, 239: 122055.
[37] Lim B, Arık S Ö, Loeff N, et al. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4): 1748-1764.
[38] Liu L, Zhang J, Xue S. Photovoltaic power forecasting: Using wavelet threshold denoising combined with VMD[J]. Renewable Energy, 2025, 249: 123152.
[39] Zhai C, Zhang Y, Liu X, et al. Photovoltaic power forecasting based on VMD-SSA-Transformer: Multidimensional analysis of dataset length, weather mutation and forecast accuracy[J]. Energy, 2025, 324: 135971.
[40] Wang B, Chen J, Zhu Y, et al. SP-Transformer: A medium- and long-term photovoltaic power forecasting model integrating multi-source spatiotemporal features[J]. Applied Sciences, 2025, 15(21): 11846.
[41] Wu M, Feng W, Li X, et al. Short-term power load forecasting using an improved model integrating GCN and Transformer[J]. Applied Sciences, 2025, 15(13): 7003.
[42] Lafuente-Cacho M, Martínez S, López G, et al. State of the art for solar and wind energy-forecasting methods for sustainable grid integration[J]. Journal of the Knowledge Economy, 2025, 16: 16432-16467.
[43] Feng W, Deng B, Chen T, et al. Probabilistic net load forecasting based on sparse variational Gaussian process regression[J]. Frontiers in Energy Research, 2024, 12: 1429241.