New Energy Power Load Forecasting and multi-time Scale Scheduling Optimization Based on Informer-Transformer

Li Zhang[1], Wei Chen[1], Weibo Yuan[2]
[1]State Grid Anhui Electric Power Co. Ltd, Hefei 230022, Anhui, China
[2]Electric Power Research Institute of State Grid Anhui Electric Power Co. Ltd, Hefei 230022, Anhui, China
Corresponding author: Li Zhang (e-mail: 13023009868@163.com)

ABSTRACT To address the load uncertainty and limited dispatch flexibility inherent in power systems with high renewable energy penetration, this paper proposes an integrated architecture that combines the Informer prediction model with a multi-time-scale scheduling optimization strategy. Accurate forecasting and coordinated scheduling are increasingly important for maintaining the stability and efficiency of modern electromagnetic energy systems and smart grid infrastructures. The study first constructs a high-quality time series dataset from multi-source information, including load, renewable generation, and meteorological data, through systematic preprocessing. Leveraging the ProbSparse sparse attention mechanism, the Informer model performs high-precision rolling forecasting at minute-, hour-, and day-level time scales. A hierarchical multi-time-scale dispatch framework is then established to tightly couple prediction with control, covering day-ahead economic dispatch, hourly source–load balancing, and minute-level energy storage regulation. By integrating energy storage systems and enforcing operational constraints, the proposed framework jointly optimizes economic efficiency and system stability. Experimental results demonstrate that the proposed method achieves RMSE and MAE values of 0.152 and 0.118 for one-hour forecasting, significantly outperforming baseline approaches. In dispatch optimization, the operating cost is reduced to 4,281 yuan while the renewable energy curtailment rate decreases to 2.1%, indicating superior economy and robustness. The proposed framework provides an effective solution for intelligent operation of high-renewable power systems and offers valuable support for reliable electromagnetic energy management and sustainable grid operation.

INDEX TERMS informer, multi-time-scale scheduling, renewable power forecasting, sparse attention mechanism, electromagnetic energy systems

I. INTRODUCTION

WITH large-scale access to renewable energy, such as wind and photovoltaic power, the power system faces profound structural changes—a transformation whose stability directly impacts energy-intensive industries [1,2]. Although promoting new energy is conducive to promoting green and low-carbon transformation, its inherent volatility and intermittent characteristics pose severe challenges to the dispatch and operation stability of the power system [3]. In this context, high-precision prediction of load and new energy output has become an important prerequisite for achieving safe and economical operation of the power system and the basis for supporting the implementation of intelligent dispatch optimization strategies [4,5]. Among them, the Informer model effectively reduces the computational complexity of ultra-long sequence prediction by applying a sparse attention mechanism while enhancing its adaptability to long-term time series patterns, providing a new method path for new energy output and load forecasting tasks [6].

Based on the above requirements, this paper proposes a comprehensive system architecture that integrates the Informer prediction model with a multi-time-scale optimization strategy, aiming to achieve close coupling of prediction and dispatch and improve the intelligent operation capability of the power system under the condition of high proportion of new energy penetration. The constructed system mainly consists of two parts: first, based on the Informer model, historical load, meteorological information and time characteristics are deeply modeled to achieve high-precision prediction of future output trends; second, based on the prediction results, a multi-level dispatch optimization model is constructed, covering different time granularities such as day, hour and minute, comprehensively considering the volatility of new energy, energy storage control strategy and system operation constraints, to achieve multi-scale adaptive adjustment and system-level optimization of dispatch decisions.

II. RELATED WORK

As one of the important directions of the smart grid [7], the forecasting of renewable energy power load has accelerated its evolution towards deep learning methods in recent years. Traditional methods such as ARIMA, SVM, and BP neural networks were mainly used [8-10]. Still, they gradually showed limitations in dealing with the strong volatility and multi-scale characteristics of wind and solar power output. Subsequently, Recurrent Neural Network (RNN) with time-dependent modeling capabilities and their extended structures became mainstream [11], among which LSTM and GRU were widely used in short-term forecasting tasks [12,13]. Although the above methods have achieved good results in short-term prediction, there are still certain bottlenecks. The Transformer architecture has been applied to the power load forecasting task to break through the performance limitations of deep time series modeling. It has received widespread attention due to its long-distance dependency modeling capabilities and highly parallel computing advantages [14,15]. Among them, the Informer model effectively reduces the computational complexity of ultra-long sequence modeling through a sparse self-attention mechanism and probabilistic sampling strategy, showing good application prospects in new energy scenarios. In short-term scheduling from minutes to hours, Model Predictive Control (MPC) and dynamic programming are widely used for rapid system response and stability control [16,17]; in medium- and long-term scheduling, intelligent optimization methods such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) have shown strong solving capabilities in multi-objective and multi-constraint scenarios [18,19]. However, most existing methods still design the prediction and scheduling modules independently, lacking a closed-loop optimization mechanism that drives the scheduling strategy update based on prediction output, which limits the depth of application of prediction results in actual scheduling decisions.

III. CONSTRUCTION OF NEW ENERGY FORECASTING AND SCHEDULING SYSTEM BASED ON INFORMER

System Overall Architecture

This system is designed with modular architecture based on the closed-loop control concept of "prediction-scheduling-feedback". The overall structure consists of four parts: data processing module, prediction modeling module, scheduling optimization module and feedback correction module, as shown in Figure 1:

The data processing module is responsible for accessing multi-source heterogeneous data such as original load, new energy output, and meteorology, performing preprocessing operations such as normalization, missing value filling, and time alignment, and generating high-quality feature sequences in a unified format for subsequent modeling. The prediction modeling module uses a deep prediction network based on the Informer architecture to model the long-term dependencies between multiple variables through a sparse attention mechanism, supporting the simultaneous prediction of load and output sequences at multiple time scales, including minute, hour, and day levels. Specifically, the initial state of charge (SOC) and total charge-discharge volume for minute-level energy storage are constrained by pre-set values, while the upper and lower limits of unit output and total energy budget for hourly scheduling are determined by the day-ahead economic dispatch output, forming a top-down rigid boundary.

Data Collection and Preprocessing

To ensure the effectiveness and reliability of system model training and evaluation, this study constructs a comprehensive data set covering load, renewable energy output, and meteorological elements based on actual power grid operation data. It carries out systematic data cleaning and feature engineering processing, providing a solid data foundation for the construction of subsequent prediction and scheduling models.

(1) Original data collection

The data used in this study mainly come from the power dispatching control center in a certain area of Shaanxi Province, the SCADA (Supervisory Control and Data Acquisition) system connected to the dispatching platform, and the National Meteorological Information Center, covering three core data types: historical load, new energy output, and meteorological information. The historical load data is provided by the power grid dispatching system in the region, with a time range of January 2020 to December 2023, a time granularity of 15 minutes, and a total of more than 100,000 records, which can fully reflect the changing characteristics of power load in different periods. The new energy output data comes from the SCADA system connected to the dispatching platform, which automatically collects the actual power output of photovoltaic and wind farms connected to the power grid in the region. The time granularity is between 15 minutes and 1 hour, and the cumulative data volume exceeds 60,000, reflecting the high volatility and non-stationary characteristics of new energy output. Meteorological data comes from the national meteorological data sharing platform, including key environmental variables such as temperature, humidity, wind speed, and sunshine intensity, with a time granularity of hours. It is accurately aligned and integrated with load and output data through a timestamp mechanism to ensure the consistency and synchronization of multi-source data in the time series dimension. The entire data collection process uses automated scripts to periodically call API interfaces and schedule database extraction mechanisms to ensure real-time and continuous data updates. At the same time,

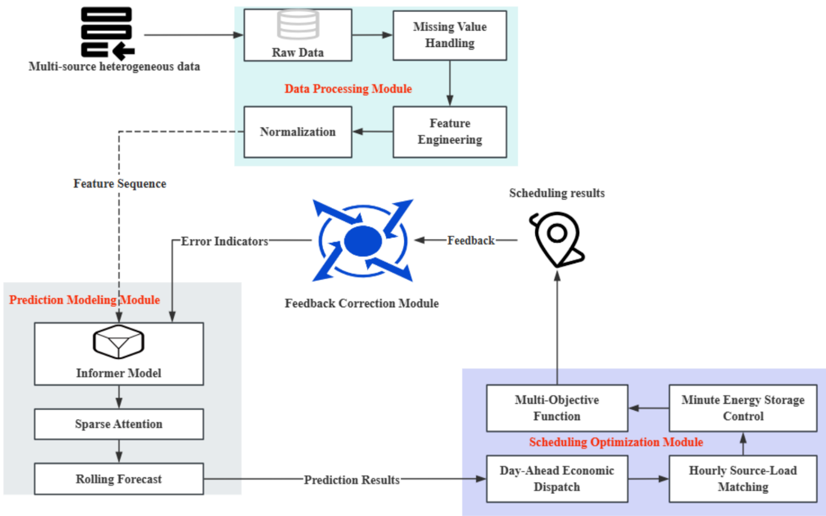

FIGURE 1. System overall architecture

a supplementary collection and local backup mechanism has been established to prevent the loss of key data due to communication interruptions or equipment abnormalities and to ensure data integrity and stability.

(2) Data preprocessing

To improve the model's learning ability and generalization performance for input data, this study systematically and carefully preprocessed the collected raw data. First, in the data cleaning stage, the sliding window statistics combined with the interquartile range (IQR) method are used to detect abnormal values during the collection process, such as sudden, instantaneous load changes and negative output of new energy sources. The outliers are defined as data points outside the range of the first quartile minus 1.5 times the IQR or the third quartile plus 1.5 times the IQR. The detected abnormal data are corrected by using mean interpolation, median interpolation, or smooth interpolation based on the previous and next time points to ensure the continuity and smoothness of the data. For data with a missing value ratio of less than 1%, the K nearest neighbor interpolation method (K=5) is used to fill in the missing data. This method achieves a reasonable estimate of the missing data by referring to the values of the 5 samples adjacent to the missing point.

In the feature engineering phase, the periodicity and trend characteristics of the time series are fully explored, and auxiliary information such as hours, weeks, and holidays in the timestamps are extracted as supplementary features of the model input to enhance the model's perception of the time context. Z-score standardization is used for numerical variables to convert the data into a standard normal distribution with a mean of 0 and a standard deviation of 1. The standardization formula is:

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

x is the original data; μ is the mean; σ is the standard deviation. For variables with physical upper and lower limits, Min-Max normalization is used to scale the values to the [0, 1] interval. The formula is:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

Finally, in terms of data division, the principle is strictly followed of time series continuity and divides the data set into a training set, validation set, and test set in a ratio of 7:2:1. All divisions are completed in chronological order to avoid information leakage and ensure the authenticity and scientificity of model evaluation.

Informer Prediction Model Design

This paper applies the Informer model in the system prediction module to solve the long-term dependency modeling and multi-step prediction problems in the process of renewable energy output and load changes.

It has a sparse attention mechanism and efficient decoding capabilities, which can consider both prediction accuracy and computational efficiency. It is suitable for large-scale power data-driven multi-time scale prediction tasks.

(1) Input feature construction and rolling prediction mechanism

This study aims to construct a multi-dimensional input feature system for the non-stationary and strong coupling characteristics of new energy load and wind power and photovoltaic output data in time series to fully capture their periodicity, trend, and external driving influence. The input variables include historical load, wind power and photovoltaic output power in the power grid, temperature, wind speed, solar radiation intensity, and other meteorological elements, as well as time labels such as hours, weeks, and holidays. After normalization, they are spliced into an input feature matrix in the time dimension $X \in \mathbb{R}^{L \times d_{in}}$. Among them, L represents the time length of the sliding window, and d_{in} is the feature dimension of the input. To adapt to the Encoder-Decoder architecture of the Informer model [20,21], this study constructs a sliding window-style time series input and prediction output construction mechanism, which is suitable for rolling prediction tasks. Assuming the length of the complete time series is T, the historical feature sequence constructed at any time t is:

$$X(t) = \{x_{t-L+1}, x_{t-L+2}, \ldots, x_t\}, \quad X(t) \in \mathbb{R}^{L \times d_{in}} \tag{3}$$

The corresponding future target prediction sequence is:

$$Y(t) = \{y_{t+1}, y_{t+2}, \ldots, y_{t+H}\}, \quad Y(t) \in \mathbb{R}^{H \times 1} \tag{4}$$

The model aims to learn the function f_θ that generates the corresponding predicted output based on $X^{(t)}$, that is:

$$\hat{Y}^{(t)} = f_\theta(X^{(t)}), \quad \hat{Y}^{(t)} \in \mathbb{R}^{H \times 1} \tag{5}$$

θ is the model parameter, and the training goal is to minimize the error between the predicted value and the actual value. The sliding window generates samples by sliding on the time series with a fixed step size S, thereby achieving high-frequency and continuous multi-step prediction. The prediction step size H can be set according to application requirements to support rolling scheduling tasks at multiple time scales.

In addition, to improve the stability and generalization ability of multi-step prediction, Decoder also applies historically known target value fragments $\hat{Y}_{input} \in \mathbb{R}^{H \times 1}$ as auxiliary input when generating the prediction sequence to improve the time-dependent modeling effect and error control ability.

(2) Model architecture and core mechanism

This study's new energy prediction model is based on the Informer architecture. An Encoder-Decoder structure with strong adaptability and scalability is designed to aim at the nonlinear periodicity, trend changes, and high noise characteristics of wind power, photovoltaic, and other output sequences on a long time scale. In each component of the model, the attention mechanism, feature embedding method, and decoding strategy are customized and optimized to meet the needs of high-precision rolling prediction in actual power dispatching. The overall structure of the model is shown in Figure 2:

The Informer model uses a typical Encoder-Decoder architecture. The input feature matrix X is first increased in dimension by a linear mapping layer, and a fixed position encoding PE is added to obtain a time-aware embedding representation:

$$X_{emb} = \text{Embedding}(X) + PE \tag{6}$$

The embedded sequence is fed into the Encoder, which consists of three layers of stacked attention submodules. Each layer includes the ProbSparse self-attention mechanism, feedforward network, residual connection and layer normalization structure. The temporal encoding of the Encoder output is represented as:

$$X_{enc} = \text{Encoder}(X_{emb}) \tag{7}$$

The core attention mechanism uses ProbSparse Attention to effectively reduce redundant calculations in long sequences. It selects the top u positions T_i with the largest amount of information corresponding to each Query vector Q_i and calculates its sparse attention output:

$$\text{Attention}(Q, K, V) = \sum_{j \in T_i} \alpha_{ij} V_j \tag{8}$$

α_{ij} is the normalized attention weight, and V_j is the corresponding Value vector. Compared with the $O(L^2)$ complexity of Full Attention, this mechanism can reduce the amount of calculation to $O(L \log L)$, greatly improving the efficiency of long-term processing. The Decoder receives two types of input: one is the historically known target value fragment $\hat{Y}_{input}$, which is used to guide the generation of future sequences; the other is the global encoding output X_{enc} of the Encoder. This helps capture historical context features. The Decoder is also composed of a three-layer stack, including self-attention, cross-attention, and Feedforward Neural Network (FNN) modules, which ultimately generates the predicted value for the next H steps:

$$\hat{Y} = \text{Decoder}(\hat{Y}_{input}, X_{enc}) \tag{9}$$

The output end uses linear mapping and Dropout regularization to enhance generalization ability and suppress overfitting. In addition, the position encoding is constructed in a fixed sine and cosine form, which is expressed as follows:

$$\begin{aligned} PE(pos, 2i) &= \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right), \\ PE(pos, 2i+1) &= \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \end{aligned} \tag{10}$$

pos represents the time step position and i is the embedding dimension subscript.

(3) Model training and optimization

After completing data preprocessing and sliding window sequence construction, this study systematically trains and optimizes the constructed Informer model based on the divided training set and validation set to achieve high-precision rolling prediction of new energy load, wind power and photovoltaic output data. During the training process, the mean squared error (MSE) is selected as the loss function, and its expression is:

$$\mathcal{L}_{MSE} = \frac{1}{H} \sum_{i=1}^{H} (\hat{y}_{t+i} - y_{t+i})^2 \tag{11}$$

H is the number of prediction steps, and $\hat{y}_{t+i}$ and y_{t+i} represent the predicted value and true value of the $t+i$-th step, respectively. The optimizer uses the Adam algorithm,

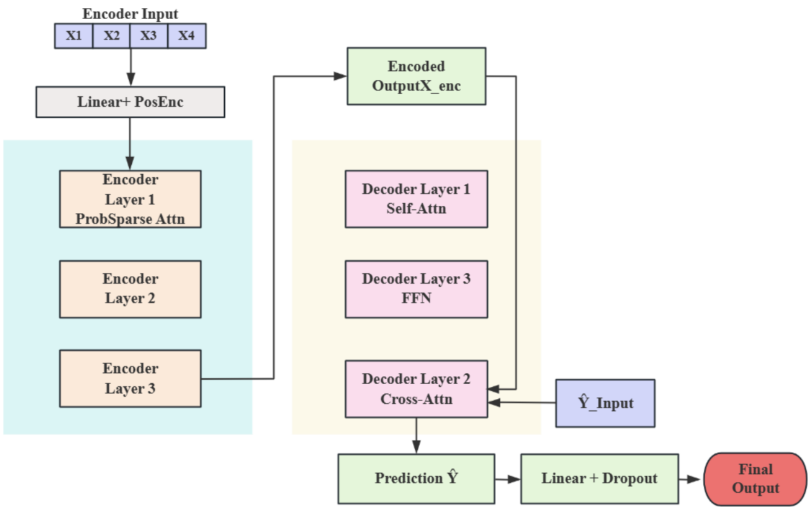

FIGURE 2. Overall structure of the model

the momentum factors are $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the L2 weight decay term is applied to enhance the stability of convergence and generalization ability. To improve the model's performance at multiple time scales, this study performed hyperparameter tuning on the model. A combination of grid search and random sampling is used to construct a hyperparameter combination space around key dimensions such as learning rate, sliding window length, number of prediction steps, hidden layer dimension, number of attention heads, number of Encoder layers, and batch size, as shown in Table 1:

TABLE 1. Parameter combination space table

Hyperparameters	Candidate value
Learning Rate	{1e-3, 5e-4, 1e-4, 5e-5}
Sliding window length (L)	{48, 72, 96, 120}
Hidden layer dimensions	{256, 512, 768}
Number of attention heads	{4, 8}
Number of Encoder layers	{2, 3, 4}
Batch size	{16, 32, 64}
Prediction steps (H)	{24, 48, 72, 96}

The complete training process is completed for each candidate configuration, and the MSE and accuracy on the validation set are used as the evaluation basis. The optimal hyperparameter combination is finally determined: sliding window length $L = 96$, prediction step number $H = 96$, hidden layer dimension of 512, number of attention heads of 4, number of Encoder layers of 3, batch size of 32, and learning rate of 1×10^{-4}. The model is iteratively trained for 100 rounds, and the model accuracy and training loss are recorded in each round. To avoid overfitting, the model uses an early stopping mechanism. If the model loss on the validation set does not decrease significantly after 10 rounds, the training ends automatically. Figure 3 shows the changes in model training loss and accuracy under the optimal hyperparameter settings:

Figure 3 shows the trend of training loss and accuracy over the training rounds during the training process of the Informer model under the optimal hyperparameter configuration. The blue dotted line represents the training loss, and the green solid line represents the model accuracy. In the early stage of training, the model shows rapid learning ability, the training loss drops rapidly, and the accuracy increases significantly, indicating that the model can efficiently capture the data's key features and potential laws. When the training rounds reach about 50, the model performance is optimal. Then, the training loss and accuracy gradually stabilize, indicating that the model enters the convergence stage.

Construction of Multi-time Scale Scheduling Optimization Model

(1) Time scale division and scheduling architecture

This study divides power dispatching into three time scales: minute, hour and day, corresponding to the three task levels of rapid response, intraday adjustment and economic planning. Minute-level dispatching has a cycle of 1 to 5 minutes, mainly dealing with sudden changes in the output of new energy sources such as wind power and photovoltaic power and short-term load fluctuations, and adopts instant charging and discharging of energy storage systems and flexible load regulation to respond quickly. Hour-level dispatching is carried out every 15 to 60 minutes. Based on the rolling forecast results, the unit output plan, energy storage arrangement and load adjustment strategy are dynamically optimized to correct the deviation between the day-ahead plan and the current operation. Daily dispatching has a cycle of 24 hours. Based on the forecast information, the next day's unit start-up and shutdown, energy procurement and carbon emission control strategies are formulated, taking into account both operating costs and environmental constraints. In practical deployment, minute-level control relies on existing Automatic Generation Control (AGC)

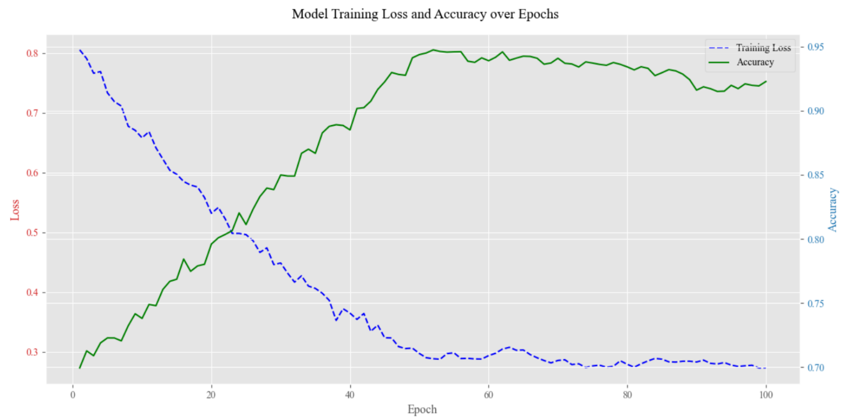

FIGURE 3. Training loss and accuracy curve of the model under the optimal hyperparameter configuration

or rapid communication infrastructure, whose 1–5-minute cycle settings have been engineered to accommodate the constraints of prediction updates, optimization calculations, and execution delays.

(2) Optimization objective function design

To achieve efficient access to new energy, economical operation of the power system, and carbon emission control, this study constructs a multi-objective joint optimization function, which comprehensively considers factors such as power curtailment minimization, operating cost control, carbon emission constraints and energy storage charging and discharging efficiency. The overall objective function of scheduling optimization is defined as follows:

$$\min J = \omega_1 J_{\text{spillage}} + \omega_2 J_{\text{cost}} + \omega_3 J_{\text{emission}} + \omega_4 J_{\text{storage}} \tag{12}$$

ω_1, ω_2, ω_3 and ω_4 are weight coefficients of each sub-objective function, which can be adjusted dynamically according to different operation strategy requirements. The power loss part is measured by the positive difference between the predicted value of wind power and photovoltaic power and the actual access amount, which is defined as:

$$J_{\text{spillage}} = \sum_{t=1}^{T} \Big[\max(0, P_{\text{wind}}^{pred}(t) - P_{\text{wind}}^{actual}(t)) + \max(0, P_{\text{solar}}^{pred}(t) - P_{\text{solar}}^{actual}(t)) \Big] \tag{13}$$

The economic cost part is composed of the unit output cost of thermal power, start-up and shutdown costs, energy storage operation costs and the price difference between purchase and sale of electricity, and the expression is:

$$J_{\text{cost}} = \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \big(C_i^{\text{fuel}} P_i(t) + C_i^{\text{start}} u_i(t)\big) + C^{\text{ess}} P_{\text{ess}}(t) + C^{\text{price}}(t) \cdot (P_{\text{buy}}(t) - P_{\text{sell}}(t)) \right) \tag{14}$$

The carbon emission target is weighted and accumulated based on the unit emission factor of each power generation unit to constrain the overall emission level, as follows:

$$J_{\text{emission}} = \sum_{t=1}^{T} \sum_{i=1}^{N} EF_i \cdot P_i(t) \tag{15}$$

The energy storage part focuses on optimizing the energy loss and charge state stability during the charging and discharging process. The objective function is as follows:

$$J_{\text{storage}} = \sum_{t=1}^{T} \big(C_{ch} \cdot P_{ch}(t) + C_{dis} \cdot P_{dis}(t) + \lambda \cdot |SOC(t) - SOC_{ref}| \big) \tag{16}$$

$P_{ch}(t)$ and $P_{dis}(t)$ represent the charging and discharging power of the energy storage system at time t; $SOC(t)$ is the state of charge at the corresponding time; SOC_{ref} is the target reference level; λ is the deviation penalty factor. By constructing this multi-objective function, the model can consider the efficiency of new energy utilization, system operation economy, carbon emission controllability, and energy storage regulation performance, and has strong flexibility and scalability and can adjust the optimization focus according to the operation scenario.

(3) Constraints and physical limitations

To ensure the feasibility of the dispatch results in engineering implementation and the safety of system operation, this study embeds multiple types of constraints in the multi-objective optimization model, covering the grid operation boundary, energy storage equipment performance, power supply and demand balance, and unit operation conditions. In terms of grid operation, the stable range of bus voltage amplitude, branch current and system frequency is set, and the constraints are expressed as follows:

$$\begin{aligned} &V_{\min} \le V_i(t) \le V_{\max}, \\ &I_{ij}(t) \le I_{ij}^{\max}, \\ &f_{\min} \le f(t) \le f_{\max} \end{aligned} \tag{17}$$

$V_i(t)$ represents the voltage amplitude of node i at time t; $I_{ij}(t)$ is the current of branch $i-j$; $f(t)$ is the system frequency. All values must be within the allowable range of the operating standard.

In terms of energy storage system constraints, its capacity boundary, charge and discharge power limit, charge and discharge mutual exclusion logic, and state of charge (SOC) update rules are considered. Its constraint model is as follows:

$$\begin{aligned} &0 \leq SOC(t) \leq SOC_{\max}, \\ &0 \leq P_{ch}(t) \leq P_{ch}^{\max}, \\ &0 \leq P_{dis}(t) \leq P_{dis}^{\max} \end{aligned} \tag{18}$$

$$\begin{aligned} &P_{ch}(t) \cdot P_{dis}(t) = 0, \\ &SOC(t+1) = SOC(t) + \eta_{ch} \cdot P_{ch}(t) \cdot \Delta t \\ &\qquad - \eta_{dis}^{-1} \cdot P_{dis}(t) \cdot \Delta t \end{aligned} \tag{19}$$

η_{ch} and η_{dis} are energy conversion efficiencies, and mutually exclusive conditions are used to prevent simultaneous charging and discharging at the same time.

At every dispatching time, the system's total production must equal the total load due to supply and demand balancing limitations. A certain safety margin R(t) is added to account for the uncertainty of new energy predictions:

$$\sum_{i=1}^{N} P_i(t) + P_{dis}(t) + P_{RES}^{actual}(t) = P_{load}(t) + P_{ch}(t) + R(t) \tag{20}$$

P_{RES}^{actual} is the power of the new energy source that has been accepted. Notably, the safety margin in the scheduling model is not a fixed value but dynamically adjusted based on the historical error statistics of the Informer's predictions, functioning as an implicit reserve capacity driven by prediction uncertainty. This design ensures the system's robustness against minute-level prediction deviations, preventing control failures caused by point prediction errors. In terms of the operation of the generator set, its start-stop logic and climbing ability are constrained to prevent frequent start-stop or power mutation. The upper and lower limits of the unit output and the climbing constraints are as follows:

$$P_i^{\min} \cdot u_i(t) \leq P_i(t) \leq P_i^{\max} \cdot u_i(t) \tag{21}$$

$$|P_i(t+1) - P_i(t)| \leq R_i^{up/down} \tag{22}$$

$u_i(t) \in \{0,1\}$ is the start/stop state variable, and $R_i^{up/down}$ represents the upper/lower limit rate of ramp rate of unit i. The above constraints are all uniformly modeled as mixed integer programming problems, and are embedded in the scheduling solver together with the optimization objective function for joint optimization, ensuring that the scheduling strategy output by the model is not only optimal in theory but also feasible and safe in engineering implementation.

Module Coupling and System Construction

(1) Module coupling logic

This study uses a data flow-driven approach to achieve close linkage between the prediction module and the scheduling module. The Informer model receives new energy output, load changes, and meteorological data from the historical database, performs rolling predictions at a fixed step size, and generates a multivariate prediction sequence for several future time steps. The prediction results are standardized, including timestamps, variable codes, and corresponding values, and are encapsulated in structured JSON (JavaScript Object Notation) format and transmitted to the scheduling module through a RESTful API. The scheduling module is divided into three scheduling tasks at the minute level, hour level and day level according to the received forecast time window, and generates control plans at the corresponding time scales respectively. The minute-level scheduling modifies the energy storage system's power and the flexible load's operational status in real time using the forecast's short-term fluctuation data. The hour-level scheduling redistributes the unit output and energy storage timing within the day based on the forecast trend distribution. The daily scheduling formulates the start-stop and power purchase strategies of the units for the next day according to the overall forecast curve, and outputs the baseline load plan to be executed. The nested structure of "upper-level output as lower-level boundary" is used to achieve coupling between different time scales.

(2) System construction

The system as a whole adopts a container-based microservice architecture and uses Kubernetes to implement module deployment and scheduling. The Informer prediction module and the scheduling optimization module are encapsulated as independent services and deployed on different nodes. The modules interact synchronously through the RESTful API, and the key mission data is streamed through Kafka to ensure that the system has high throughput and low latency processing capabilities. The prediction module supports configuration changes of input feature dimensions, prediction length and sliding window size, while the scheduling module supports flexible setting of objective function weights, time scale division and constraint parameters. Multi-time scale optimization solutions are executed in parallel by configuring the task scheduler. The solver can switch between Mixed Integer Linear Programming (MILP), genetic algorithm and deep reinforcement learning, supporting adaptation of multiple scenarios.

IV. SYSTEM PERFORMANCE EVALUATION AND EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL SETUP

(1) Experimental environment configuration

All experiments in this study are completed on a server platform with high-performance computing capabilities. In terms of hardware, the experimental platform is equipped with an NVIDIA RTX 3090 GPU, an Intel Xeon Gold 6248 processor, 64GB DDR4 memory, and a 512GB NVMe solid-state drive, which can fully meet the high concurrency and high precision requirements of deep learning model training and multi-time scale scheduling optimization calculations. In terms of software environment, the operating system uses Ubuntu 20.04 LTS to ensure system stability and compatibility; the deep learning part is developed based on the PyTorch 1.12 framework, supports GPU parallel acceleration, and significantly improves model training efficiency; data processing and visualization mainly rely on mainstream scientific computing and graphics libraries such as NumPy, Pandas and Matplotlib; the scheduling optimization model is solved using the Gurobi 9.5 mixed integer programming solver to efficiently support joint optimization tasks under complex constraints.

(2) Baseline system selection

To verify the effectiveness of the Informer-based new energy forecasting and multi-time scale dispatching system in this study, three groups of baseline systems are designed. Baseline group 1 is a traditional rule-based system that uses historical mean and smoothing methods to forecast load and new energy output, and dispatches based on fixed start-stop rules and empirical energy storage management. Baseline group 2 is a deep learning system that uses a Bidirectional Long Short-Term Memory Network (BiLSTM) and an attention mechanism for forecasting, and uses multi-time scale mixed integer programming optimization for dispatching. Baseline group 3 is a fusion system of machine learning and physical modeling, combining the Gradient Boosting Decision Tree (GBDT) with the physical mechanism model for prediction, and the scheduling strategy combines the hierarchical and robust optimization methods.

Experimental Process and Result Analysis

(1) Prediction performance experimental analysis

To comprehensively evaluate the prediction performance of the constructed system, this study designs a comparative experiment with three groups of baseline systems based on the historical load, renewable energy output and related meteorological data collected in the previous paper. The experiment carries out multi-step rolling prediction for different prediction window lengths (1 hour, 6 hours, 24 hours), and records the RMSE and MAE of each system in each prediction window. The experimental results are shown in Figure 4:

As can be seen from Figure 4, under different prediction window lengths, the prediction system based on the Informer architecture constructed in this paper always outperforms the baseline systems in terms of MAE and RMSE, showing stronger prediction accuracy and robustness. To further verify the actual prediction effect of the model, the study selected a representative full day (24 hours) of time series data in the test set. This time period contains the typical fluctuation characteristics of renewable energy output, including peak-to-valley changes and sudden fluctuations. Based on the trained Informer model, rolling multi-step predictions are performed step by step for this time period, and the prediction window covers continuous output from 1 hour to 24 hours in the future. Subsequently, the prediction results are compared point by point with the actual observation data at the corresponding time. The comparison results are shown in Figure 5:

From the prediction curve in Figure 5, it can be seen that the Informer model performs well in the 24-hour rolling prediction task. Its predicted values are highly consistent with the changing trends of the actual observed values. Especially in the areas of stable load changes and typical fluctuations, the model shows strong fitting ability. In the new energy power load sequence simulated by virtual data, Informer not only accurately captures periodic characteristics such as morning and evening peaks and night troughs but also responds well to sudden fluctuations, reflecting its advantages in modeling long time series dependencies and complex nonlinear relationships. The prediction curve is close to the actual value in most time periods, with small errors and stable distribution, indicating that the model has good robustness and generalization ability. Even at some mutation points, the prediction value can maintain a high tracking accuracy, indicating that the model has a certain dynamic adaptability

(2) Analysis of dispatch optimization effect

Based on the prediction results of renewable energy output obtained in the previous paper, the corresponding prediction data are input for each designed system, and a multi-time scale dispatch optimization experiment is carried out. The dispatch process is executed within a typical 24-hour time window. During the experiment, key dispatch parameters such as wind power and photovoltaic output power, energy storage system discharge power, etc. are recorded in real time. The experimental results are shown in Figure 6:

As shown in Figure 6, the scheduling of wind power output, photovoltaic output and energy storage discharge power all show higher stability and adaptability, especially in dealing with fluctuations in new energy output and energy storage management. The multi-level scheduling optimization strategy adopted by the model in this paper covers different time granularities such as day, hour, and minute, comprehensively considers the volatility of new energy, energy storage control strategy, and system operation constraints, and realizes multi-scale adaptive adjustment and system-level optimization of scheduling decisions. In contrast, due to the lack of multi-time scale collaborative control mechanisms, Baseline 1, Baseline 2, and Baseline 3 have difficulty achieving efficient scheduling of energy storage at different time scales, resulting in large deviations and fluctuations in scheduling results.

In addition, combined with the scheduling results of each system, this study further statistically calculated the operating cost and power abandonment rate of each system scheduling scheme. The statistical results are shown in Table 2:

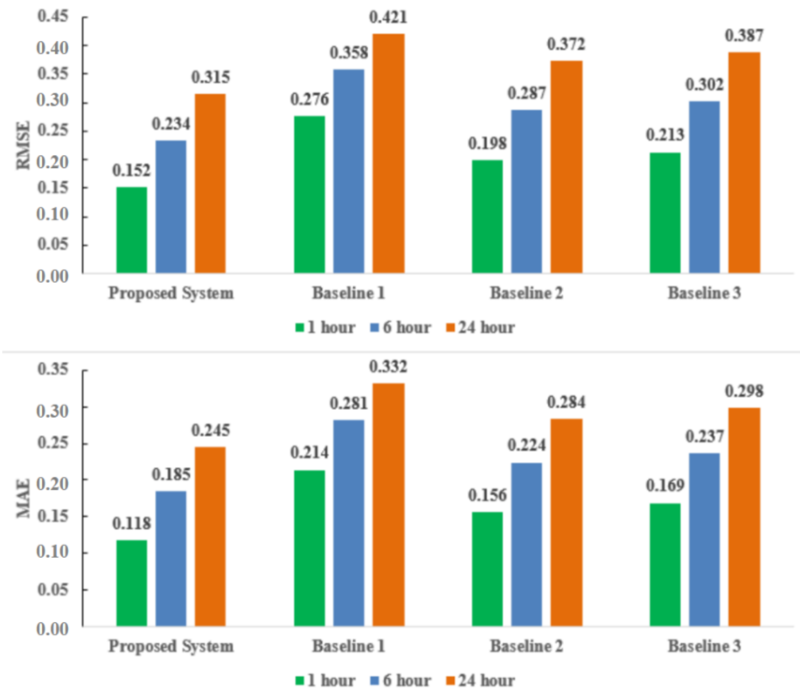

FIGURE 4. Comparison of RMSE and MAE of various systems under different prediction windows

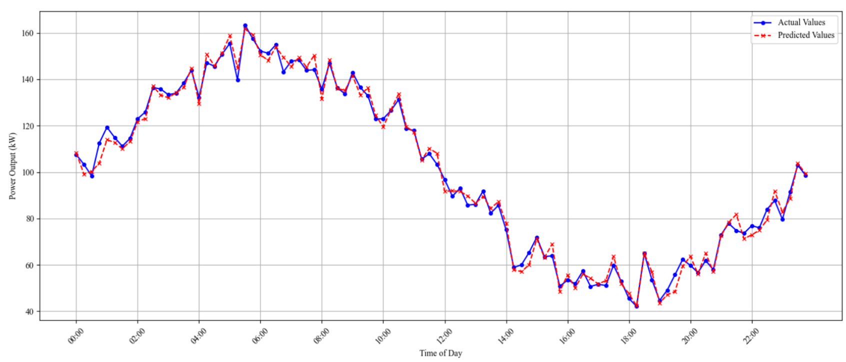

FIGURE 5. Comparison of 24-hour prediction values of the Informer model and actual observation values

TABLE 2. Statistics of operating costs and power abandonment rates of various system dispatching schemes

System Name	Operating cost (yuan)	Power abandonment rate (%)
Proposed System	4,281	2.1
Baseline 1	5,763	6.8
Baseline 2	4,957	4.5
Baseline 3	5,328	5.7

From the data in Table 2, it can be seen that the Informer-based new energy power load forecasting and multi-time scale scheduling optimization system proposed in this paper is significantly better than the three control systems in terms of operating cost and power abandonment rate. The operating cost is 4,281 yuan, which is significantly lower than the 5,763 yuan for Baseline 1, 4,957 yuan for Baseline 2, and 5,328 yuan for Baseline 3, indicating that the system has stronger economic advantages in resource coordination, energy storage control, and load matching. At the same time, its power abandonment rate is only 2.1%, which is much lower than 6.8% of Baseline 1 and 5.7% of Baseline 3, and better than 4.5% of Baseline 2, reflecting a higher capacity for absorbing new energy and flexibility in dispatching.

(3) Multi-time scale scheduling performance and robustness experiments

To systematically evaluate the impact of multi-time scale scheduling mechanism on scheduling performance and system robustness, this paper conducts comparative experiments based on the prediction-scheduling integrated system constructed in the previous paper. During the experiment, the input data set is unified, the prediction model structure and parameter configuration are kept consistent, and variables are set only in the scheduling time granularity and

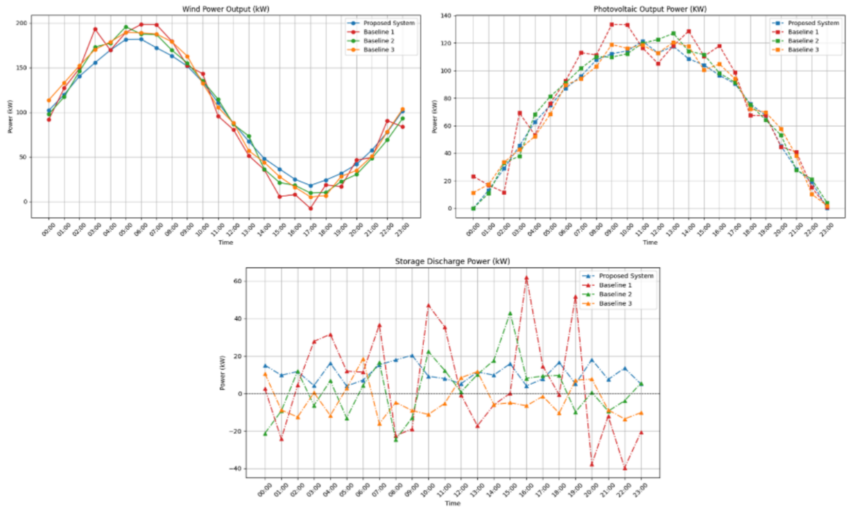

FIGURE 6. The dispatching and allocation of wind power, photovoltaic output and energy storage system discharge power of each system

operating conditions to ensure the fairness and repeatability of the experiment. In the scheduling granularity comparison experiment, three scheduling scales are constructed: minute level (15 minutes), hour level (1 hour) and day level (24 hours). The operating economic cost and power abandonment rate corresponding to the scheduling schemes generated by the system at different time granularities are statistically analyzed to evaluate the impact of scheduling resolution on system performance. The experimental results are shown in Figure 7:

As shown in Figure 7, the scheduling granularity has a significant impact on system performance. From the overall trend, as the scheduling time scale is refined from the daily level to the hourly level and minutely level, the system's operating cost gradually decreases, and the rate of renewable energy abandonment also decreases, indicating that finer time resolution helps improve the economic efficiency of scheduling and the utilization rate of renewable energy. Combined with the specific data for analysis, under minute-level scheduling, its operating cost is 4,382 yuan, and the abandonment rate is 2.2%; in the hourly scheduling, the operating cost rose to 4,505 yuan, and the power abandonment rate is 2.4%; under the daily scheduling, the operating cost further increases to 4,821 yuan, and the power abandonment rate also rises to 2.8%. To further verify the stability and adaptability of the multi-time scale scheduling mechanism in various operating environments, this paper selects three typical complex working conditions in the test set: continuous rainy weather, high load fluctuation scenario, and intermittent wind superimposed on cloudy weather. Repeating the experiment on the three scheduling scales in the above scenarios, and record the operating cost and power abandonment rate of the scheduling scheme under each scheduling scale in different scenarios. The statistical results are shown in Table 3:

TABLE 3. Statistical results of economic cost and curtailment rate of dispatching schemes of various dispatching scales under different scenarios

Scene type	Scheduling standards	Operating cost (yuan)	Power abandonment rate (%)
Continuous rainy weather	Daily-level	5,870	4.9
	Hourly-level	5,260	4.4
	Minute-level	4,710	3.1
High load fluctuation scenario	Daily-level	6,040	5.7
	Hourly-level	5,420	5.3
	Minute-level	4,830	4.8
Intermittent wind + cloudy weather	Daily-level	5,920	6
	Hourly-level	5,390	5.5
	Minute-level	4,800	4.8

From the data in Table 3 and the previous data, it can seen that the multi-time scale scheduling performance of the system in this paper shows good stability and adaptability in different complex operating scenarios. Whether it is in the high-demand period when continuous rain leads to limited photovoltaic output and drastic load fluctuations, or in the uncertain environment of intermittent changes in wind power and cloudy weather, the system can significantly reduce operating costs and reduce the power abandonment rate through minute-level scheduling. In the most challenging intermittent wind + cloudy weather scenario, the daily dispatch cost is as high as 5,920 yuan and the power abandonment rate reaches 6%. After adopting minute-level dispatch, the operating cost drops to 4,800 yuan and the power abandonment rate is also controlled by 4.8%.

(4) Strategy ablation experiment

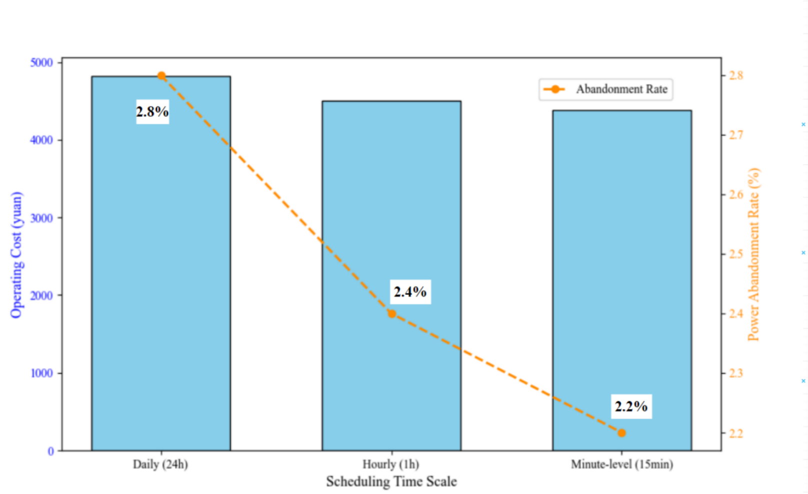

FIGURE 7. Comparison of dispatch economic cost and power abandonment rate under different dispatch time scales

To quantitatively analyze the actual contribution of the Informer prediction module and the multi-time scale scheduling mechanism to the system performance, this study designs three groups of ablation experiments. Group 1 replaces the Informer module with LSTM and keeps other structures unchanged. Group 2 removes the time-layered scheduling mechanism and changes it to single-scale hourly scheduling. Group 3 removes the above two modules at the same time as a weakened baseline system. Under the unified test set and scheduling environment, the performance of each version of the system in terms of RMSE, MAPE, operating cost, and power abandonment rate is evaluated and compared with the complete system. The results are shown in Figure 8:

Figure 8 shows the performance comparison of the proposed system under different ablation configurations.

By comparing the performance of the complete system with the three ablation experimental groups in terms of RMSE, MAPE, operating cost, and power abandonment rate, it can be seen that the Informer prediction module and the multi-time scale scheduling mechanism have a significant contribution to the overall performance of the system.

V. CONCLUSION

This study proposes a new energy power load forecasting and multi-time-scale scheduling optimization system based on the Informer model, which effectively addresses the limitations of traditional approaches and directly supports energy-intensive industries. Traditional forecasting methods are not accurate enough when dealing with multi-variable coupling, strong nonlinearity and time series characteristics, and existing scheduling strategies are also difficult to cope with the regulation needs of new energy volatility and multi-time scale dynamic changes. This study applies the Inforemr model and combines it with the ProbSparse sparse attention mechanism to significantly improve the accuracy of new energy load forecasting and perform well in long-time series forecasting. At the same time, the multi-time scale dispatch optimization model constructed realizes the coordinated optimization at the daily, hourly and minute levels, comprehensively considers the volatility of new energy, energy storage control strategy and system operation constraints, effectively reduces the operating cost and reduces the power abandonment rate, and significantly improves the flexibility and economy of the system. According to the experimental results, this system outperforms the baseline system in terms of dispatch optimization and forecast accuracy, offering solid support for the power system's intelligent operation under the high percentage of new energy access.

AUTHOR CONTRIBUTIONS

Conceptualization –Li Zhang, Wei Chen and Weibo Yuan; methodology – Li Zhang, Wei Chen and Weibo Yuan; investigation – Li Zhang and Weibo Yuan; writing-original draft preparation – Li Zhang, Wei Chen and Weibo Yuan. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] P. N. Papadopoulos, S. Chatzivasileiadis, and A. Marot, "Can Machine Learning Help Keep the System Secure?: Power Systems and Change Addressing the Increasing Complexity and Uncertainty During the Energy Transition," IEEE Power and Energy Magazine, vol. 22, no. 6, pp. 100-111, 2024, doi: 10.1109/MPE.2024.3421388.
[2] S. Shahzad and E. B. Jasińska, "Renewable Energy Revolution: A Review of Strategic Flexibility in Future Power Systems," Sustainability, vol. 16, no. 13, Art. no. 5454, 2024, doi: 10.3390/su16135454.

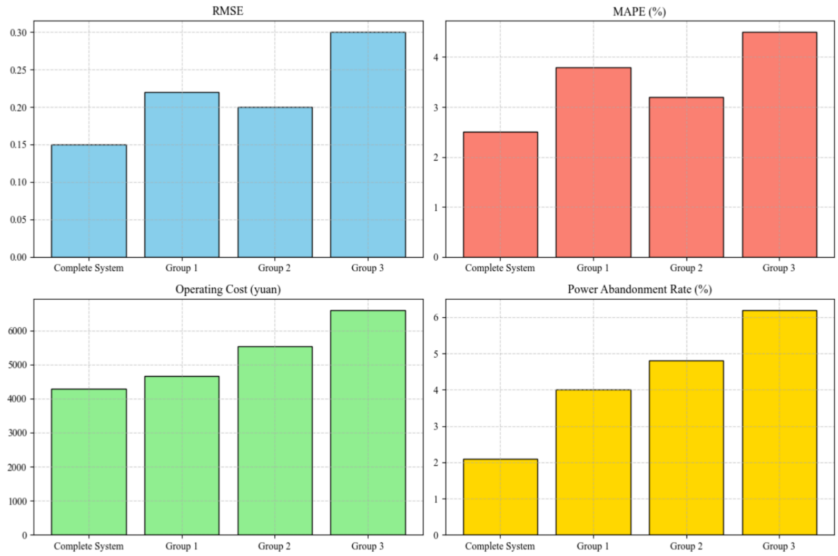

FIGURE 8. Comparison of ablation experiment results

[3] C. Medina, C. R. M. Ana, and G. González, "Transmission grids to foster high penetration of large-scale variable renewable energy sources–A review of challenges, problems, and solutions," International Journal of Renewable Energy Research (IJRER), vol. 12, no. 1, pp. 146-169, 2022, doi: 10.20508/ijrer.v12i1.12738.g8400.
[4] N. Ahmad, Y. Ghadi, M. Adnan, and M. Ali, "Load forecasting techniques for power system: Research challenges and survey," IEEE Access, vol. 10, pp. 71054-71090, 2022, doi: 10.1109/ACCESS.2022.3187839.
[5] Y. Yuan, Q. Yang, J. Ren, X. Mu, Z. Wang, Q. Shen, et al., "Short-term power load forecasting based on SKDR hybrid model," Electrical Engineering, vol. 107, no. 5, pp. 5769-5785, 2025, doi: 10.1007/s00202-024-02821-x.
[6] Y. Tan, Y. Huang, J. Liu, and R. Yang, "Long and short-term multivariate load forecasting based on MMoE-CNN-Informer model for power systems," Journal of Electrical Engineering, vol. 20, no. 2, pp. 253-263, 2025, doi: 10.11985/2025.02.025.
[7] F. Huang, H. Zhao, P. Yi, P. Li, and J. Peng, "An improved power load forecasting method based on transformer," Modern Electric Power, vol. 40, no. 1, pp. 50-58, 2023, doi: 10.19725/j.cnki.1007-2322.2021.0209.
[8] N. Q. Dat, N. T. Ngoc Anh, N. Nhat Anh, and V. K. Solanki, "Hybrid online model based multi seasonal decompose for short-term electricity load forecasting using ARIMA and online RNN," Journal of Intelligent & Fuzzy Systems, vol. 41, no. 5, pp. 5639-5652, 2021, doi: 10.3233/jifs-189884.
[9] Y. Lu and G. Wang, "A load forecasting model based on support vector regression with whale optimization algorithm," Multimedia Tools and Applications, vol. 82, no. 7, pp. 9939-9959, 2023, doi: 10.1007/s11042-022-13462-2.
[10] T. Bashir, H. Chen, M. F. Tahir, and L. Zhu, "Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN," Energy Reports, vol. 8, pp. 1678-1686, 2022, doi: 10.1016/j.egyr.2021.12.067.
[11] L. Lv, Z. Wu, J. Zhang, L. Zhang, Z. Tan, and Z. Tian, "A VMD and LSTM based hybrid model of load forecasting for power grid security," IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 6474-6482, 2021, doi: 10.1109/TII.2021.3130237.
[12] M. Abumohsen, A. Y. Owda, and M. Owda, "Electrical load forecasting using LSTM, GRU, and RNN algorithms," Energies, vol. 16, no. 5, pp. 2283-2314, 2023, doi: 10.3390/en16052283.
[13] X. Fang, W. Zhang, Y. Guo, J. Wang, M. Wang, and S. Li, "A novel reinforced deep rnn–lstm algorithm: Energy management forecasting case study," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5698-5704, 2021, doi: 10.1109/TII.2021.3136562.
[14] A. L'Heureux, K. Grolinger, and M. A. M. Capretz, "Transformer-based model for electrical load forecasting," Energies, vol. 15, no. 14, pp. 4993-5016, 2022, doi: 10.3390/en15144993.
[15] Z. Zhao, C. Xia, L. Chi, X. Chang, W. Li, T. Yang, et al., "Short-term load forecasting based on the transformer model," Information, vol. 12, no. 12, pp. 516-538, 2021, doi: 10.3390/info12120516.
[16] J. Moreno-Castro, V. S. Ocaña Guevara, L. T. León Viltre, Y. Gallego Landera, O. Cuaresma Zevallos, and M. Aybar-Mejía, "Microgrid management strategies for economic dispatch of electricity using model predictive control techniques: A review," Energies, vol. 16, no. 16, pp. 5935-5959, 2023, doi: 10.3390/en16165935.
[17] Y. Lan, Q. Zhai, X. Liu, and X. Guan, "Fast stochastic dual dynamic programming for economic dispatch in distribution systems," IEEE Transactions on Power Systems, vol. 38, no. 4, pp. 3828-3840, 2022, doi: 10.1109/TPWRS.2022.3204065.
[18] A. Kalakova, H. K. Nunna, P. K. Jamwal, and S. Doolla, "A novel genetic algorithm based dynamic economic dispatch with short-term load forecasting," IEEE Transactions on Industry Applications, vol. 57, no. 3, pp. 2972-2982, 2021, doi: 10.1109/TIA.2021.3065895.
[19] D. Wang, D. Peng, D. Huang, L. Ren, M. Yang, and H. Zhao, "Research on short-term and mid-long term optimal dispatch of multi-energy complementary power generation system," IET Renewable Power Generation, vol. 16, no. 7, pp. 1354-1367, 2022, doi: 10.1049/rpg2.12366.
[20] K. Yang and F. Shi, "Medium-and long-term load forecasting for power plants based on causal inference and Informer," Applied Sciences, vol. 13, no. 13, pp. 7696-7715, 2023, doi: 10.3390/app13137696.
[21] Y. Yan, W. Li, S. Su, H. Bai, Y. Yang, S. Pan, et al., "Decentralized Wind Power Forecasting Method Based on Informer," Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), vol. 15, no. 8, pp. 679-687, 2022, doi: 10.2174/2352096515666220818122603.