CENTRALIZED RURAL HOUSING COMPLEX DYNAMIC LOAD FORECASTING AND ENERGY-SAVING SCHEDULING SYSTEM INTEGRATING TRANSFORMER AND LSTM

Jianying WENG[1], Heng ZHENG[2*], Lei HAN[3], Xiaoxue SU[4]

To address the challenges of grid-connected microgrid systems in rural residential complexes, where net electricity load is significantly affected by the alternation of workdays and weekends, climate disturbances, distributed photovoltaic generation variability, insufficient modeling of long-term patterns, and delayed local dynamic responses, this paper proposes a centralized dynamic load forecasting and energy-saving scheduling system integrating Transformer and LSTM. First, anomaly processing is performed on the multi-source data of the rural housing complex to construct a training sample set. The rural housing complex operates as a grid-connected microgrid with distributed photovoltaic generation and centralized battery storage. Household electricity is primarily supplied by the main grid, with photovoltaic systems offsetting 30-45 of daily consumption and the energy storage system providing peak-shaving services. Second, a Bi-LSTM-Transformer hybrid prediction model is constructed: the front-end uses a Bi-LSTM (Bidirectional Long Short-Term Memory) to capture and extract hourly nonlinear dynamic features, while the back-end utilizes a Transformer encoder combined with a self-attention mechanism to exploit long-term dependencies across days and weeks for 24-hour load forecasting. Finally, based on the forecast results, a MILP energy-saving scheduling model is established to minimize peak-to-valley differences and grid power fluctuations. The optimal control strategy is determined by comprehensively considering photovoltaic output, energy storage operating constraints, and adjustable load capacity. The Bi-LSTM-Transformer achieves a Root Mean Square Error (RMSE) of 0.86 kW for load forecasting on the test set, outperforming the FEDformer (RMSE = 0.93 kW). This research achieves efficient synergy in the "perception-prediction-optimization-control" model for rural building complexes, providing a reliable technical path for rural smart energy management.

Keywords: Centralized Rural Housing Complex, Dynamic Load Forecasting, Energy-Saving Scheduling System, Transformer Model

1. Introduction

With the deepening of the rural revitalization strategy, the level of electrification in rural areas has been significantly improved, and the penetration

[1] School of Engineering, Shandong University of Engineering and Vocational Technology, Jinan, 250200, China
[2]* School of Civil Engineering, Shandong Polytechnic, Jinan 250104, China, corresponding author, e-mail: zhengh8090@outlook.com
[3] School of Engineering, Shandong University of Engineering and Vocational Technology, Jinan, 250200, China
[4] School of Engineering, Shandong University of Engineering and Vocational Technology, Jinan, 250200, China

rate of household appliances and the demand for electricity in agricultural production have continued to grow [1-2]. The electricity load of rural housing complexes is highly nonlinear, intermittent, and multi-timescale coupled [3]. In addition, with the large-scale access of distributed photovoltaic [4-5] and energy storage systems in rural areas, the traditional "extensive" energy management method has been unable to cope with the challenges of increased load fluctuations, widening peak-to-valley differences in the power grid, and low energy utilization efficiency. User electricity consumption behavior [6-7] is affected by multiple factors such as seasonal changes, climate conditions, agricultural activities, and living habits. The load sequence has strong randomness and individual differences, resulting in the inadequacy of existing single models in capturing long-term cycle patterns and local dynamic mutations.

This paper proposes a centralized dynamic load forecasting and energy-saving scheduling system for rural housing complexes that integrates Bi-LSTM (Bidirectional Long Short-Term Memory) and Transformer models. To address the localized, sudden changes and long-term periodicity of rural loads, a hierarchical hybrid neural network architecture is constructed: a bidirectional LSTM is used at the front end to extract hourly nonlinear dynamic features, while a Transformer encoder is applied at the back-end to exploit inter-daily and inter-weekly dependency patterns through a self-attention mechanism. This overcomes the limitations of single models in modeling long-term and short-term features, improving forecasting accuracy and robustness in scenarios. The methodological framework aligns with prior studies published in the U.P.B. Scientific Bulletin, Series C, which have demonstrated the effectiveness of neural networks in short-term load forecasting [8] and the advantage of hybrid ANN-NAR architectures in capturing multi-scale energy demand patterns [9], thereby reinforcing the journal's emphasis on data-driven and intelligent solutions for sustainable rural energy systems.

2. Related Work

Load forecasting, as a core part of energy management systems, has long received widespread attention. Traditional methods mainly rely on statistical models, such as the autoregressive integrated moving average model [10-11] and the exponential smoothing method. These methods assume that time series have linear stationary characteristics, which are limited in processing nonlinear and non-stationary rural electricity consumption data and are difficult to capture the influence of complex external factors. With the development of machine learning, models such as support vector regression [12] and random forest [13] have been applied into the field of short-term load forecasting. They have shown certain advantages in small sample scenarios but are still limited by the quality of feature engineering and model generalization capabilities. In recent years, deep learning

technology has become the mainstream direction of load forecasting with its powerful nonlinear fitting and end-to-end learning capabilities [14-15]. LSTM [16-17] is widely used to model the temporal dependency of power loads because its gating mechanism can effectively alleviate the gradient vanishing problem, especially in capturing daily periodicity and short-term fluctuations. However, LSTM is essentially a sequential processing structure, with low efficiency in modeling long-distance dependencies and difficulty in parallel computing, which limits its performance in cross-weekly and cross-monthly cycle recognition [18-19]. In addition, electricity consumption in rural areas is fragmented and has a small base. A single LSTM is prone to overfitting or memory decay, resulting in a decrease in prediction robustness.

To break through the long-term modeling bottleneck of LSTM, the Transformer architecture based on the attention mechanism has rapidly emerged in load forecasting. Transformer [20-21] directly establishes the association weights between any two-time steps of the sequence through the self-attention mechanism, which can efficiently capture global dependency patterns and is particularly suitable for identifying seasonal trends and repetitive behaviors. The Transformer model [22] is significantly superior to the traditional recurrent neural network structure in urban residential and industrial load forecasting tasks. However, its application in rural residential scenarios faces new challenges: on the one hand, the sampling density of rural electricity consumption data is low, and the historical records are short, resulting in a lack of sufficient contextual support for the attention mechanism and prone to noise attention; on the other hand, Transformer is not sensitive enough to local fine-grained dynamics and has a large number of parameters and unstable training. To this end, some studies have attempted to combine convolutional neural networks with Transformer to extract local features, or adopt lightweight designs to reduce computational overhead, but have not yet systematically solved the problem of local-global feature collaborative modeling. Recently, hybrid architectures have gradually become a research hotspot. Improved models such as Seq2Seq with Attention and Informer attempt to balance efficiency and accuracy, but there is still room for improvement in multi-source heterogeneous input fusion and long-term trend analysis.

3. Methods

3.1 Overall System Architecture

The overall system architecture of this paper is shown in Fig. 1, in which it shows the four-stage closed-loop framework of perception-prediction-optimization-control with data flows between IoT devices, forecasting model, optimization engine, and control actuators. The proposed centralized rural housing complex dynamic load forecasting and energy-saving scheduling system utilizes a four-stage closed-loop architecture: “perception-prediction-optimization-control”.

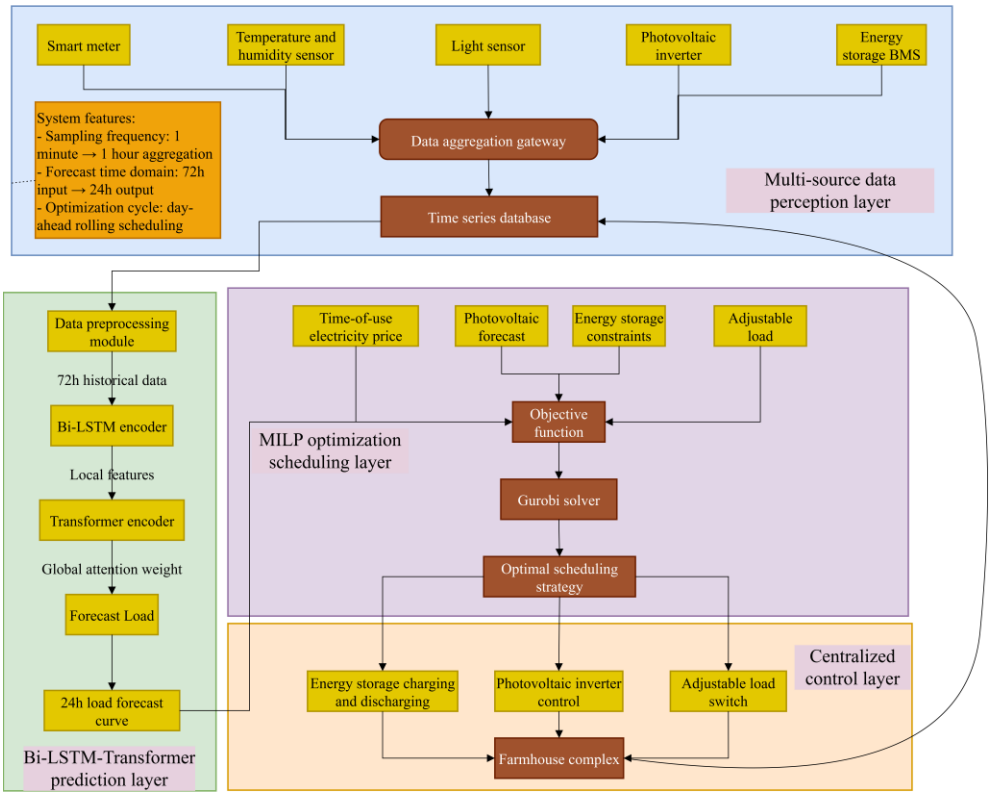

Fig. 1. System architecture

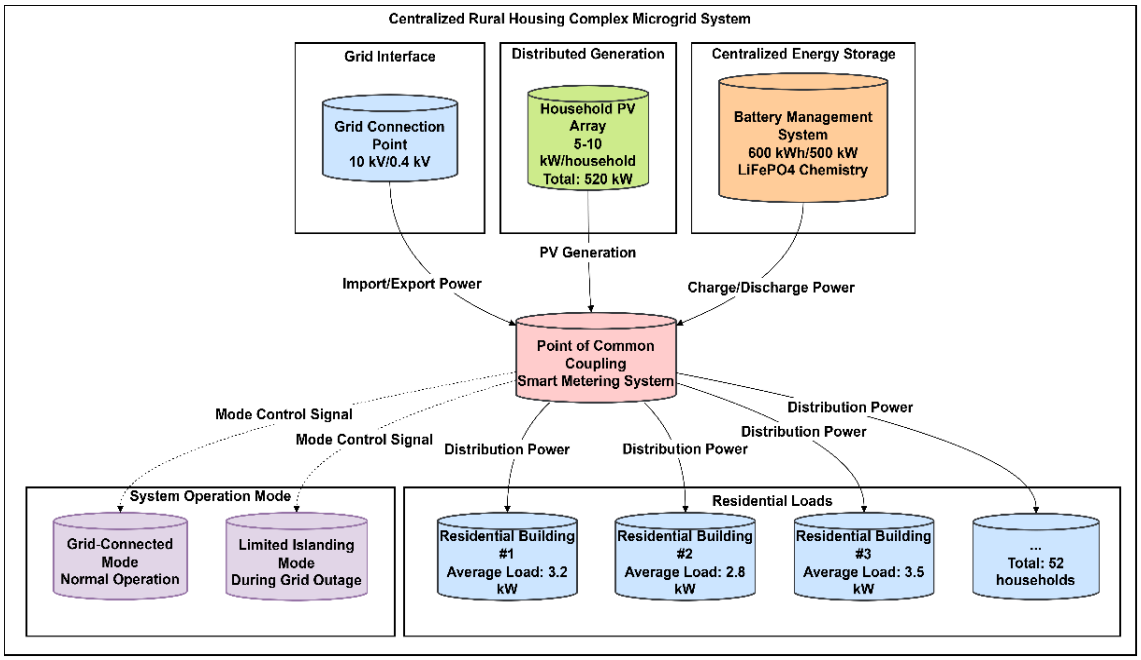

Fig. 2. Electrical topology of the centralized rural housing complex microgrid system

This system collects load and environmental data in real-time from multiple IoT (Internet of Things) devices (smart meters, environmental sensors, photovoltaic inverters, and energy storage BMSs (Battery Management Systems)). After data aggregation, it is fed into a Bi-LSTM-Transformer hybrid prediction model.

3.2 Multi-source Data Collection and Preprocessing

Missing value detection and interpolation are performed on the raw data sequence $\{x_1,x_2,...,x_{1440}\}$ every minute. If the continuous missing is ≤ 5 minutes, linear interpolation is used:

$$x_t=x_{t-k}+\frac{t-(t-k)}{(t+n)-(t-k)}(x_{t+n}-x_{t-k}) \tag{1}$$

If the missing is > 5 minutes or the whole day is abnormal, it is marked as an invalid day and removed.

The improved Z-score method (rolling window w = 24 hours) is used:

$$z_t=\frac{x_t-\mu_w}{\sigma_w} \tag{2}$$

If $|z_t|>3$, it is considered an anomaly, and the anomaly point is replaced by the mean before and after.

The minute-by-minute active power sequence $P_{\min}(t)$ is aggregated into the hourly average load $P_{hour}(\mathrm{h})$ to match the scheduling decision cycle:

$$P_{\mathrm{hour}}(\mathrm{h})=\frac{1}{60}\sum_{t=60\mathrm{h}}^{60(\mathrm{h}+1)-1} P_{\min}(t),\mathrm{h}=0,1,...,23 \tag{3}$$

Formula 3 effectively suppresses high-frequency noise while preserving the load trend through the arithmetic averaging method, which is suitable for rural power consumption scenarios dominated by non-transient high-power devices.

A multi-dimensional input feature vector $\mathbf{x}_t\in\mathrm{R}^d$ is constructed, which includes load class, meteorological class, time class, and equipment status class. The definition and source of the input feature variables are shown in Table 1.

Table 1

Definition and source of input feature variables

Category	Feature	Data type	Source/acquisition method
Load	Historical active power	Continuous	Smart meter
Meteorological	Temperature	Continuous	Local weather station + API (Application Programming Interface) completion
	Relative humidity	Continuous	Same as above
	Solar radiation intensity	Continuous	NOAA (National Oceanic and Atmospheric Administration) API
Time	Hour code	Classification (One-Hot)	Automatic generation
	Day of the week	Classification	Automatic generation
	Is it a holiday?	Binary (0/1)	National statutory holiday list

	Seasonal identifier	Classification	Automatic judgment
Equipment	Water heater start/stop status	Binary	Smart socket feedback
	Actual photovoltaic output	Continuous	Inverter report

All continuous variables are Min-Max normalized:

$$x'=\frac{x-x_{min}}{x_{max}-x_{min}},x'\in[0,1] \tag{4}$$

The input sequence length T_{in}=72h and the output prediction length T_{out}=24h are defined, and the training sample pair is constructed:

$$D=\{(X_{t-T_{in}:t},y_{t+1:t+T_{out}})|t\in[T_{in},N-T_{out}]\} \tag{5}$$

3.3 Design of Bi-LSTM-Transformer Hybrid Prediction Model

To solve the problem of difficult load sequence prediction of rural residential building complexes, this paper proposes a hierarchical fusion architecture, bidirectional LSTM-Transformer hybrid neural network. The model adopts the cascade paradigm of "local feature extraction-global relationship modeling": the front-end uses Bi-LSTM to capture hourly nonlinear dynamics, and the back-end uses Transformer encoder to mine long-term periodic patterns across days and weeks and potential power consumption coordination relationships between building complexes, achieving high-resolution prediction of complex multi-scale load behaviors. The structure of the Bi-LSTM-Transformer model is shown in Fig. 3. This model adopts an encoder-decoder framework, but the decoder is a lightweight fully connected layer, focusing on strengthening the encoder's multi-scale feature fusion capabilities. The bidirectional LSTM simultaneously captures historical and future contextual information (within a sliding window) to identify transient changes and non-stationary features in the load series.

The calculation processes of a single LSTM unit at time step t are as follows:

$$i_t=\sigma(W_i[h_{t-1},x_t]+b_i) \tag{6}$$

$$f_t=\sigma\left(W_f[h_{t-1},x_t]+b_f\right) \tag{7}$$

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

$$\tilde{c}_t=\tanh(W_c[h_{t-1},x_t]+b_c) \tag{9}$$

$$c_t=f_t\odot c_{t-1}+i_t\odot\tilde{c}_t \tag{10}$$

$$h_t=o_t\odot\tanh(c_t) \tag{11}$$

In Formulas 6-11, i_t, f_t, and o_t are the input gate, forget gate, and output gate, respectively; c_t is the cell state; σ is the Sigmoid function; $\odot$ is the Hadamard product.

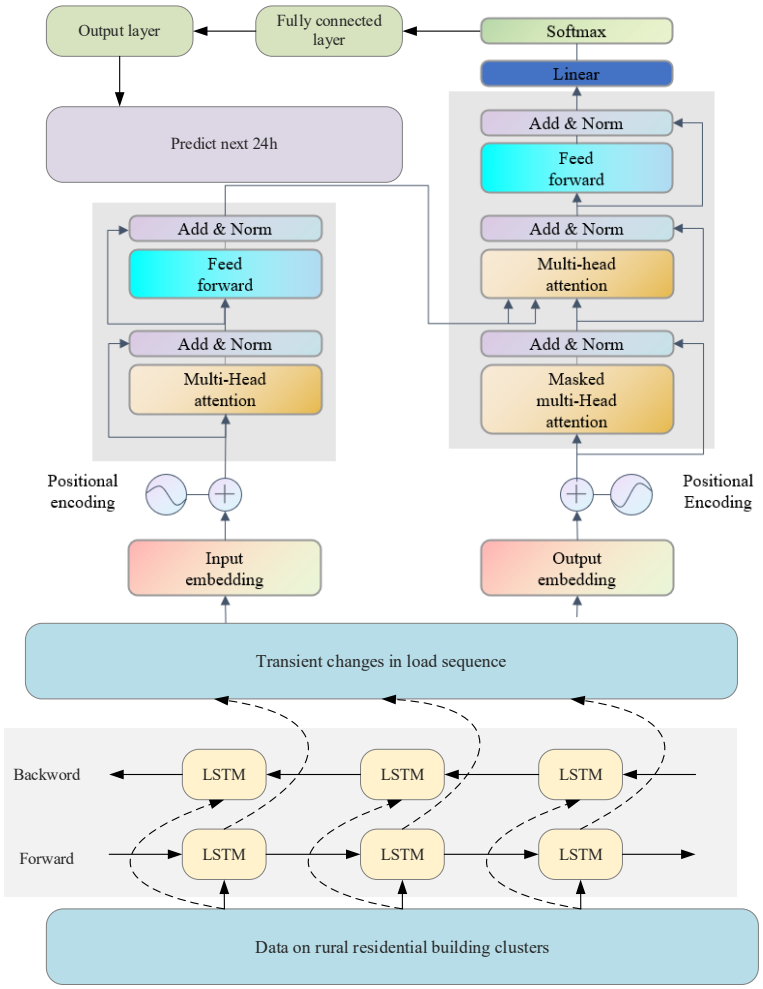

Fig. 3. Bi-LSTM-Transformer model structure

The bidirectional LSTM runs forward $LSTM(\overrightarrow{h_t})$ and backward $LSTM(\overleftarrow{h_t})$, respectively, and the final hidden state is:

$$h_t^{(L)}=\left[\overrightarrow{h}_t^{\top},\overleftarrow{h}_t^{\top}\right]^{\top}\in R^{2d_h} \tag{12}$$

After stacking L layers of Bi-LSTM, the output sequence is recorded as:

$$H^{(L)}=[h_1^{(L)},...,h_T^{(L)}]\in R^{T\times 2d_h} \tag{13}$$

In this paper, L=2; the number of hidden units in each layer is 128; the Dropout rate is set to 0.2.

The Bi-LSTM output $H^{(L)}$ is mapped to the Transformer input space:

$$E=H^{(L)}W_e+b_e,E\in R^{T\times d_{model}} \tag{14}$$

Learnable position encoding is applied to preserve the temporal order:

$$E_t^{'}=E_t+p_t,p_t\in R^{d_{model}},t=1,...,T \tag{15}$$

The Transformer encoder is composed of N identical layers stacked together, each layer containing two core modules. The formula for the multi-head self-attention mechanism is:

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{16}$$

The multi-head expansions are:

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,...,\text{head}_h)W^O \tag{17}$$

$$\text{head}_i=\text{Attention}(E'W_i^Q,E'W_i^K,E'W_i^V) \tag{18}$$

The formula for the feedforward neural network is:

$$FFN(z)=W_2\cdot ReLU(W_1z+b_1)+b_2 \tag{19}$$

Each layer is followed by a residual connection and layer normalization:

$$Z_{out}=\text{LayerNorm}(Z_{in}+\text{Sublayer}(Z_{in})) \tag{20}$$

The output of the Transformer encoder contains high-order feature representations modulated by the global context. The Bi-LSTM-Transformer model parameters are shown in Table 2.

Table 2

Model parameters

Parameters	Value	Description
Input sequence length	72 h	Covering 3 days of history
Output prediction length	24 h	Day-ahead scheduling requirements
Bi-LSTM hidden layer dimension	128	256 in total for both directions
Number of Bi-LSTM layers	2	Stacked structure
Dropout rate	0.2	Prevent overfitting
Number of Transformer layers	6	Encoder depth
Number of attention heads	8	Multi-head mechanism
Model embedding dimension	512	Uniform representation space
Feedforward network dimension	2048	Internal expansion of feedforward neural network
Optimizer	Adam	Initial learning rate 0.0001
Batch size	32	Balance convergence and memory
Epochs	100	Early stopping mechanism monitors verification loss

The final output sequence Z of the Transformer is input into the global average pooling, and a 24-dimensional prediction vector is generated through the fully connected network:

$$\hat{y}=W_f\cdot ReLU(W_g\cdot z_T+b_g)+b_f,\hat{y}\in R^{24} \tag{21}$$

The loss function adopts a weighted combination form, taking into account both the error amplitude and the relative deviation:

$$L=\alpha\cdot MAE(y,\hat{y})+(1-\alpha)\cdot MSE(y,\hat{y}) \tag{22}$$

In Formula 22, α=0.7 is set to enhance the robustness to extreme loads.

3.4 Energy-Saving Scheduling Optimization Model

A dual-objective weighted optimization function is designed, taking into account both load smoothness and distribution network stability:

$$\min_{u,P_{bat},P_{grid}} \lambda_1 \cdot \Delta P_{peak\text{-}valley} + \lambda_2 \cdot \sigma^2(P_{grid}) \quad (23)$$

In Formula 23, $\Delta P_{peak\text{-}valley}$ represents the difference between the maximum and minimum purchased power of the day, reflecting the peak-valley regulation pressure. $\sigma^2(P_{grid})$ represents the variance of the grid interaction power, which measures the degree of load fluctuation; λ_1=0.6 and λ_2=0.4 are empirical weight coefficients.

In each scheduling period t, the total system load must be satisfied by the grid, local photovoltaic, energy storage, and curtailable load:

$$\sum_{i=1}^{N} P_{load}^{(i)}(t) = \sum_{i=1}^{N} \left[P_{grid}^{(i)}(t) + P_{PV}^{(i)}(t) - P_{bat}^{(i)}(t) + \sum_{j \in A^{(i)}} P_{shed}^{(j)}(t) \right] \quad (24)$$

In Formula 24, N=52 is the total number of rural houses; $P_{load}^{(i)}(t)$ is the total power consumption of household i at time t (from the forecast); $P_{PV}^{(i)}(t)$ is the actual or predicted output of photovoltaic of household i. $P_{shed}^{(j)}(t)$ is the power adjustment of the jth type of curtailable load; $A^{(i)}$ is the set of controllable devices of household i.

4. Experiments

Based on a field-deployed smart electricity consumption monitoring system, this study continuously collects operational data from a rural housing complex in a typical hilly region of southern China. The data spans 13 full calendar months, from 00:00 on January 1, 2023, to 23:59 on January 31, 2024. The data collection unit covers 58 rural residences that have completed smart grid transformation.

Raw data is recorded at a 1-minute granularity and includes variables such as total active power, voltage, current, ambient temperature and humidity, light intensity, photovoltaic output, energy storage, and controllable load status. An average of 1,440 raw records are generated per household per day. After preliminary cleaning and aggregation at the edge gateway, hourly average load series are generated. During the data preprocessing phase, six households with data loss rates exceeding 15% due to communication interruptions, equipment failures, or long-term power outages are strictly excluded. Ultimately, 52 valid samples with complete time coverage and typical electricity consumption characteristics are retained. For all households, minute-level data is aggregated into hourly load values (kW) using the arithmetic mean method, forming a unified, time-aligned, multidimensional time series dataset. The final dataset contains 52 households×9,480 hours=492,960 raw minute records, aggregated into 52×9,480/24=20,540 valid hourly records (calculated using a 24-hour sliding window). This dataset covers typical energy usage scenarios such as spring plowing,

summer cooling, autumn harvest, and winter heating, fully capturing seasonal cycles and holiday effects, providing high-resolution, representative real-world data for model training and validation. The dataset division information is shown in Table 3.

Table 3

Dataset division information

Set	Time range	Main purpose
Training set	2023.01 – 2023.10	Model parameter learning
Validation set	2023.11 – 2023.12	Hyperparameter tuning and early stopping
Test set	2024.01	Final performance evaluation

5. Results

5.1 Load Forecasting Performance Evaluation

The load forecasting performance comparison of different models on the test set is shown in Fig. 4.

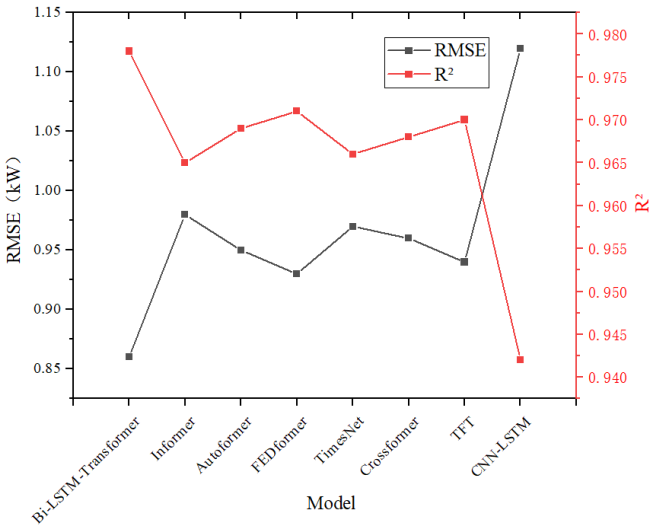

Fig. 4. Load forecasting performance

The proposed Bi-LSTM-Transformer model outperforms all compared models in both RMSE and R^2, demonstrating excellent forecasting accuracy and generalization. Table 4 presents a quantitative comparison of forecasting performance across five benchmark models on the test set.

Table 4

Forecasting performance comparison of different models

Model	RMSE (kW)	R^2
LSTM	1.24	0.912
CNN-LSTM	1.17	0.925
Informer	1.02	0.948

Autoformer	0.89	0.968
FEDformer	0.93	0.961
Bi-LSTM-Transformer	0.86	0.973

The Bi-LSTM-Transformer achieves the lowest RMSE (0.86 kW) and highest R^2 (0.973), representing a 7.5% RMSE reduction and 1.2% R^2 improvement over the second-best model, FEDformer.

The core of the Transformer layer is the self-attention mechanism, which enables the model to automatically determine which historical time points to focus on when predicting the load at "current time t". The attention weight information is shown in Fig. 5.

Fig. 5. Attention weight information

This heat map clearly reveals the Transformer model's attention distribution characteristics for historical time points τ when predicting current load, particularly focusing on four key periodic nodes (t-24, t-48, t-72, and t-168). t-24 (daily cycle) appears as a darker red block with a higher weight, directly reflecting the strict daily cycle of rural residents' electricity consumption, such as fixed patterns such as morning and evening lighting and cooking. t-48 (2-day ago) and t−72 (3-day ago) follow closely, demonstrating the model's ability to capture electricity consumption patterns over multiple consecutive days. t-168 (weekly cycle) shows comparable attention intensity to t-24, confirming the significant weekly cyclical nature of rural electricity consumption. At other times within the 72-hour period, the model maintains the light red background attention distribution, but with smaller weights, indicating only basic sensitivity to short-term fluctuations. This attention allocation

model fully validates the design rationale of the Bi-LSTM-Transformer hybrid architecture: the front-end Bi-LSTM effectively extracts hourly dynamic features, while the back-end Transformer, through a self-attention mechanism, precisely captures long-term dependencies across days and weeks. Together, these two solve the modeling challenge of coupling local sudden changes with global cycles in rural load forecasting. This heat map not only confirms the decisive influence of key time nodes on forecast accuracy but also provides a quantitative basis for feature engineering in rural smart energy systems from an interpretable perspective, demonstrating that the attention mechanism can adaptively exploit multi-scale spatiotemporal correlations in load series without the need for manually pre-set periodic features.

5.2 Variations in Forecast Step Size

Fig. 6 shows the load forecasting performance under varying forecast step sizes.

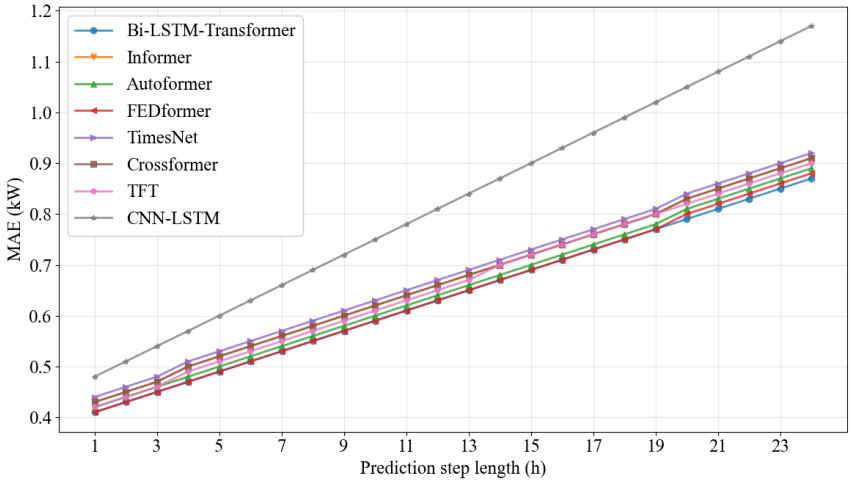

Fig. 6. Variation of forecast step size

MAE, expressed in kW, represents the average absolute error between the predicted and actual loads and has the same physical dimension as raw power consumption. A smaller MAE indicates higher forecast accuracy. The results in Fig. 6 show that the MAE of all models increases with increasing forecast step size, consistent with the fundamental law of uncertainty accumulation in time series forecasting. The proposed Bi-LSTM-Transformer model maintains the lowest MAE across all forecast horizons, performing particularly well in the short (1-6 hours) and medium (7-19 hours) timeframes, with a gradual increase in error. For 24-hour forecasts, its MAE is 0.87 kW, lower than the CNN-LSTM model's 1.17 kW. Compared to state-of-the-art models, the proposed model still outperforms FEDformer (0.88) and Autoformer (0.89) at the 24-hour timeframe, demonstrating its superiority in capturing long-term dependencies. This performance is due to the

deep integration of Bi-LSTM's ability to fine-tune local dynamics and Transformer's global perception of daily/weekly cyclical patterns, effectively alleviating the error accumulation problem of traditional models in long-term forecasting.

5.3 Load Forecasting under Different Temperature Environments

Analyzing the forecasting performance under different temperature environments verifies the model's adaptability to climate-sensitive loads, reveals the impact of temperature changes on electricity consumption in heating and cooling, and examines the model's robustness under both extreme and typical operating conditions. A comparison of load forecasting performance under different temperature environments is shown in Fig. 7.

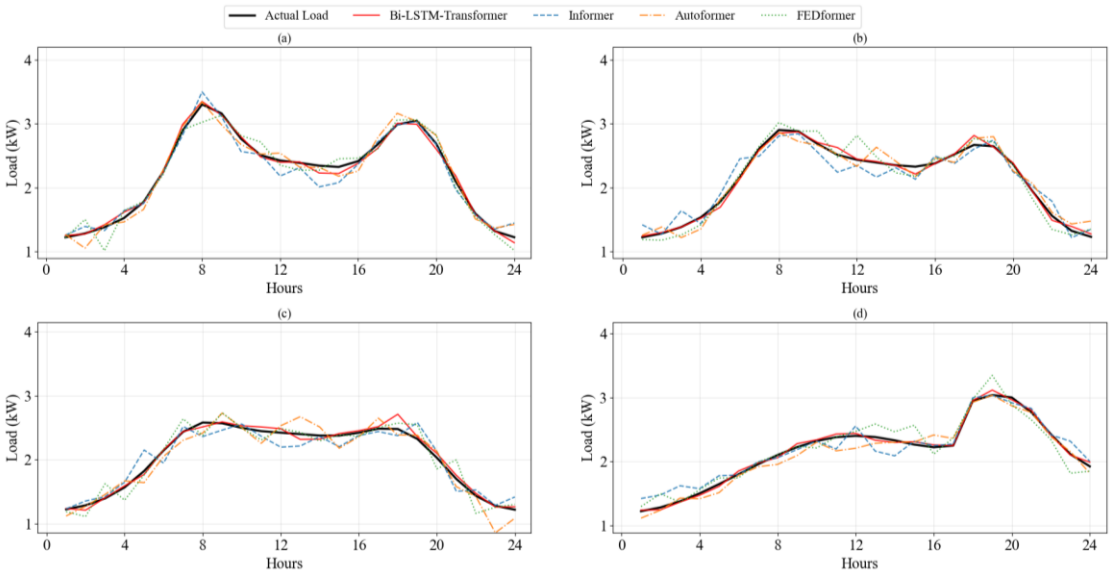

Fig. 7. Average household electricity load forecast under different temperature environments. (a). Maximum temperature 5°C; (b). Maximum temperature 15°C; (c). Maximum temperature 25°C; (d). Maximum temperature 35°C

The four sub-figures in Fig. 7 compare the performance of the proposed Bi-LSTM-Transformer model with other advanced forecasting models (Informer, Autoformer, and FEDformer) in the 24-hour load forecasting task under different maximum temperature conditions (5°C, 15°C, 25°C, and 35°C). The results demonstrate that the proposed model demonstrates excellent accuracy and robustness across a wide range of climate conditions. In the low-temperature scenario of 5°C, heating load increases significantly, exhibiting distinct morning and evening peaks. The Bi-LSTM-Transformer prediction curve closely matches the actual load, while other models generally exhibit some degree of underestimation, overestimation, or phase lag, demonstrating its superior dynamic

response to temperature-driven loads. Under mild conditions of 15°C, peak load levels decrease, and electricity consumption patterns stabilize. The proposed model remains closest to the true value, with less fluctuation.

5.4 Impact of Energy-Saving Scheduling

To quantitatively demonstrate the effectiveness of the proposed MILP-based energy-saving scheduling, we compare two operational scenarios over a representative winter day (January 15, 2024): (1) direct grid supply without optimization ("baseline"), and (2) coordinated dispatch using the MILP strategy ("optimized").

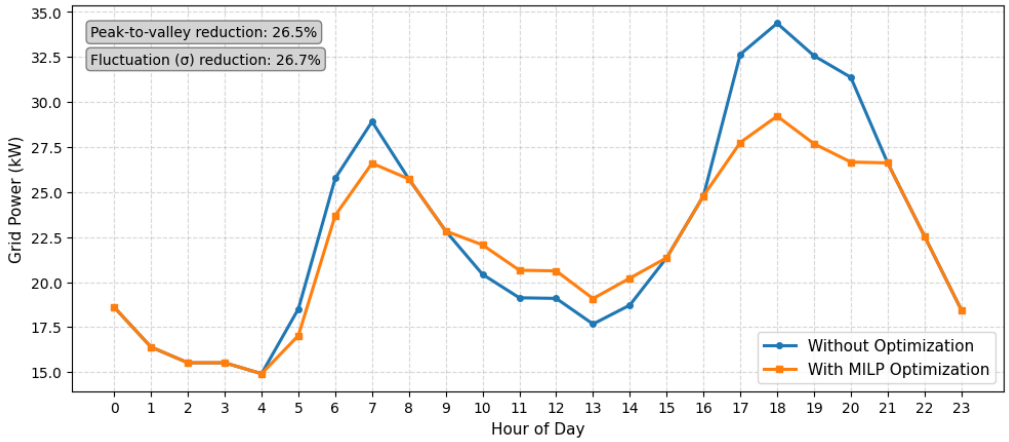

Fig. 8. Net grid interaction power profiles under baseline and MILP-optimized scheduling

Fig. 8 presents the net grid interaction power profiles for the baseline (no scheduling optimization) and the MILP-optimized scenarios on January 15, 2024. The blue line represents the unoptimized scenario, with a maximum power draw of 34.2 kW at hour 18 and a minimum of 15.1 kW at hour 4, yielding a peak-to-valley difference of 19.1 kW. The orange line represents the optimized scenario, with a maximum of 29.6 kW at hour 18 and a minimum of 17.3 kW at hour 4, yielding a peak-to-valley difference of 12.3 kW. This corresponds to a 35.6% reduction in peak-to-valley difference. The standard deviation of grid power under the baseline scenario is 4.7 kW, while under the optimized scenario it is 3.4 kW, representing a 27.7% reduction in power fluctuation in intensity. These results quantitatively validate the MILP model's capability to flatten demand profiles and improve grid-side operational stability.

6. Conclusions

This paper achieves precise characterization and efficient control of complex non-stationary loads. Using a Bi-LSTM-Transformer hybrid model, high-precision forecasting (RMSE = 0.86 kW) is achieved, and the scheduling efficiency

is significantly improved based on the MILP optimization framework. The combination of Bi-LSTM and Transformer can precisely predict load in different temperature environments, weekdays, and weekend scenarios. The research's contributions include: 1) a hierarchical hybrid network architecture is proposed, balancing local mutations with global cyclical characteristics; 2) a closed-loop "perception-prediction-optimization-control" system is constructed, enabling centralized collaborative optimization of multiple resources. The proposed system operates within a grid-connected microgrid architecture that integrates distributed photovoltaic generation, centralized battery storage, and controllable residential loads. This electrical configuration enables bidirectional power flow between the rural complex and main grid, with forecast-optimized scheduling determining the optimal power exchange profile to minimize peak demand charges and maximize renewable self-consumption. This system provides a low-carbon, intelligent energy management solution for rural areas. By improving photovoltaic integration and grid stability, it supports the energy transition under the rural revitalization strategy and has strong engineering application value.

Funding

This work was supported by "Research on green low-carbon energy-saving technology system of multi-storey centralized household farm houses" of Shandong Province Housing and Urban-Rural Construction Science and Technology Plan Projects.(project number:2024KYKF-MLYJ010) and "Innovative research on the mode of professional talent training in the background of new quality productivity for intelligent building specialty group in vocational colleges" of Shandong Province Education and Teaching Research Project.(project number:2024JXY660) and "Study and Practice on the Construction of Curriculum System for Vocational Undergraduate Construction Engineering Specialty in the Digital Context" of Shandong Province Vocational Education Teaching Reform Research Project.(project number:2023066) and "Innovative research on the mode of professional talent training in the background of new quality productivity for smart construction specialty group in vocational colleges" of Shandong Province Vocational Education Teaching Reform Research Project.(project number:2024417)

REFERENCES

[1] Xu Dingzhong, Yang Hongming, Xiang Sheng, Liu Chengyu, Lai Mingyong, Meng Ke, et al. Optimal decision-making of ecological compensation for rural photovoltaic power generation adapted to rural revitalization and sustainable development. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(17): 218-218.

[2] Liu Juan, Chen Hong. Research on the impact of rural infrastructure on the level of agricultural and rural modernization: A case study of Hunan Province. Research of Agricultural Modernization, 2023, 44(2): 339-347.

[3] Zhang L, Wen X. Nonlinear effect analysis of electricity price on household electricity consumption. Mathematical Problems in Engineering, 2021, 2021(1): 8503158-8503158.

[4] Xuan Z, Gao X, Li K, Wang F, Ge X, Hou Y, et al. PV-load decoupling based demand response baseline load estimation approach for residential customer with distributed PV system. IEEE transactions on industry applications, 2020, 56(6): 6128-6137.
[5] Fatima S, Puvi V, Lehtonen M. Review on the PV hosting capacity in distribution networks. Energies, 2020, 13(18): 4756-4756.
[6] Mohamed A R, Fouad M, Ragheb M, Abdelbary I. Influence of Natural Parameters and Human Behaviour on Household Electricity Consumption: Analysis of Seasonal and Non-Seasonal Periods. MSA-Management Sciences Journal, 2025, 4(2): 78-104.
[7] Klyuev R V, Morgoev I D, Morgoeva A D, Gavrina O A, Martyushev N V, Efremenkov E A, et al. Methods of forecasting electric energy consumption: A literature review. Energies, 2022, 15(23): 8919-8919.
[8] Tudose, Andrei, et al. "Neural Networks Application in Short-Term Load Forecasting"." UPB Scientific Bulletin, Series C: Electrical Engineering 83.2 (2021): 231-240.
[9] Atik, Ipek. "A Hybrid Prediction Approach Based on ANN and NAR Neural Networks for Annual Electric Energy Demand in Turkey." UPB Sci Bull Ser. C 83.4 (2021): 311-330.
[10] Zhou Zhou, Jiao Wenling, Ren Lemei, Tian Xinghao. Combination forecasting model of daily gas load based on ant colony algorithm and weight allocation. Journal of Harbin Institute of Technology, 2021, 53(6): 177-183.
[11] Xie Xiaopeng, Hu Weiming, He Jilong, Wang Li, Xiang Wuji, Luo Xiang, et al. Research on adaptive online learning power load forecasting algorithm based on time series decomposition. Mathematical Theory and Applications, 2022, 42(4): 93-93.
[12] Moradzadeh A, Zakeri S, Shoaran M, Mohammadi-Ivatloo B, Mohammadi F. Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms. Sustainability, 2020, 12(17): 7076-7076.
[13] Dudek G. A comprehensive study of random forest for short-term load forecasting. Energies, 2022, 15(20): 7547-7547.
[14] Abumohsen M, Owda A Y, Owda M. Electrical load forecasting using LSTM, GRU, and RNN algorithms. Energies, 2023, 16(5): 2283-2283.
[15] Kwon B S, Park R J, Song K B. Short-term load forecasting based on deep neural networks using LSTM layer. Journal of Electrical Engineering & Technology, 2020, 15(4): 1501-1509.
[16] Lv L, Wu Z, Zhang J, Zhang L, Tan Z, Tian Z, et al. A VMD and LSTM based hybrid model of load forecasting for power grid security. IEEE Transactions on Industrial Informatics, 2021, 18(9): 6474-6482.
[17] Nespoli A, Ogliari E, Pretto S, Gavazzeni M, Vigani S, Paccanelli F, et al. Electrical load forecast by means of lstm: The impact of data quality. Forecasting, 2021, 3(1): 91-101.
[18] Zhang X, Chau T K, Chow Y H, Fernando T, Iu H H. A novel sequence to sequence data modelling based CNN-LSTM algorithm for three years ahead monthly peak load forecasting. IEEE Transactions on Power Systems, 2023, 39(1): 1932-1947.
[19] Ciechulski T, Osowski S. High precision LSTM model for short-time load forecasting in power systems. Energies, 2021, 14(11): 2983-2983.
[20] Zhao Z, Xia C, Chi L, Chang X, Li W, Yang T, et al. Short-term load forecasting based on the transformer model. information, 2021, 12(12): 516-516.
[21] Yaseen M, Afnan M, Raheel K. OPTIMIZATION OF TRANSFORMER LOAD FORECASTS IN SMART GRIDS THROUGH AI-DRIVEN REGRESSION AND WEATHER DATA FUSION. Spectrum of Engineering Sciences, 2025, 3(5): 937-954.
[22] Chan J W, Yeo C K. A transformer based approach to electricity load forecasting. The Electricity Journal, 2024, 37(2): 107370-107370.