



#### OPEN ACCESS

EDITED BY  
Debora Anelli,  
Sapienza University of Rome, Italy

REVIEWED BY  
Jiyang Liu,  
Shandong Jianzhu University, China  
Yang Wu,  
National University of Defense  
Technology, China

\*CORRESPONDENCE  
Hui Chen,  
✉ sx\_chenhui@sx.hbtobacco.cn  
Lu Dong,  
✉ ludong@yangtzeu.edu.cn

RECEIVED 21 October 2025  
REVISED 21 February 2026  
ACCEPTED 23 February 2026  
PUBLISHED 10 March 2026

#### CITATION

Wei Y, Chen H, Tian Y, Yi X, Cui G, Wang G,  
Zhang S, Qian J, Liu H and Dong L (2026)  
An interpretable building air conditioning  
load forecasting framework using SSA-  
optimized Bi-LSTM and SHAP analysis.  
*Front. Environ. Sci.* 14:1728506.  
doi: 10.3389/fenvs.2026.1728506

#### COPYRIGHT

© 2026 Wei, Chen, Tian, Yi, Cui, Wang,  
Zhang, Qian, Liu and Dong. This is an  
open-access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# An interpretable building air conditioning load forecasting framework using SSA-optimized Bi-LSTM and SHAP analysis

Yingfeng Wei<sup>1</sup>, Hui Chen<sup>1\*</sup>, Ye Tian<sup>1</sup>, Xiongwei Yi<sup>1</sup>, Gean Cui<sup>1</sup>,  
Guiyuan Wang<sup>1</sup>, Shikang Zhang<sup>1</sup>, Jiaqi Qian<sup>2</sup>, Huimin Liu<sup>2</sup> and  
Lu Dong<sup>3\*</sup>

<sup>1</sup>China Tobacco Hubei Industrial Co., Ltd., Wuhan, China, <sup>2</sup>School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan, China, <sup>3</sup>State Key Laboratory of Low Carbon Catalysis and Carbon Dioxide Utilization, School of Petroleum Engineering, Yangtze University, Wuhan, China

Short-term air conditioning load forecasting in industrial buildings plays a key role in energy management and carbon neutrality efforts. Yet internal disturbances, latent heat variations, and seasonal distribution shifts often hinder model generalization. Here we propose an interpretable framework that couples Sparrow Search Algorithm (SSA)-optimized Bidirectional Long Short-Term Memory (Bi-LSTM) networks with SHapley Additive exPlanations (SHAP). Working with real 5-min data from an industrial facility across all four seasons, we construct a 17-dimensional physics-informed feature vector comprising environmental states, historical load inertia, dynamic trend indicators, and cyclical time encodings, employing strict chronological splits to avoid data leakage. The pure Bi-LSTM model avoids the parameter redundancy of complex hybrid networks and clearly outperforms traditional machine learning, basic deep learning, and attention-based baselines. Furthermore, SSA proves significantly more robust and efficient for hyperparameter tuning than PSO, DBO, ISSA, and SCSSA. In winter testing, the SSA-optimized model reaches  $R^2 = 0.965$  and  $RMSE = 1.20$  kW. SHAP analysis highlights physically plausible feature contributions—indoor enthalpy and historical load dominate, demonstrating clear seasonal patterns. Direct cross-season transfer still yields  $R^2 > 0.93$ , and a parameter-based transfer learning strategy using only 10% of target-season data reduces prediction errors by another 30%–50% compared to training from scratch. Overall, the framework delivers accurate, automated, and explainable forecasting tailored to industrial HVAC systems, offering practical engineering guidance.

#### KEYWORDS

Bi-directional long-term short-term memory neural network, building air conditioning load forecasting, optimization of sparrow search algorithm, SHAP analysis, transfer learning

## 1 Introduction

Building energy use makes up a large part of total global energy consumption, and in commercial and industrial buildings, heating, ventilation, and air conditioning (HVAC) systems typically account for 40%–60% of the energy consumed (Chen et al., 2020; Dong C. et al., 2018; Dong K. et al., 2018). With many countries now working toward carbon

neutrality, the need for accurate short-term forecasting of air conditioning (AC) loads has grown considerably. Good predictions in this time frame help with demand-side management, reducing peak loads, and better integration of renewable energy sources—all of which contribute to more efficient electricity grids and lower carbon emissions (Chen et al., 2019; Mawson and Hughes, 2020).

Forecasting HVAC loads in industrial buildings is particularly challenging. In residential or office buildings, the main drivers are usually occupancy and outdoor weather conditions. Industrial settings, however, are affected by substantial internal disturbances, including fluctuating production schedules, heat released from machinery, and latent loads associated with manufacturing processes (Fang et al., 2023). These elements lead to greater variability and non-stationary behavior in load patterns. In many manufacturing facilities, the latent heat from moisture-related processes can be as large as, or even larger than, the sensible heat component (Chen and Hong, 2023).

Methods for predicting building loads generally fall into two categories: physics-based and data-driven. Physics-based models rely on thermodynamic principles and energy balance equations; they are commonly implemented in simulation software such as EnergyPlus and TRNSYS (Eid et al., 2024; Ethivereasingham et al., 2024). These approaches are inherently interpretable and can capture fundamental physical processes, but they demand detailed building information (such as thermal properties of the envelope, geometry, and orientation) and often require considerable computational effort, which restricts their use in practical engineering settings (Yan et al., 2015; Amasyali and El-Gohary, 2018). Data-driven methods, on the other hand, derive relationships directly from measured historical data and are therefore better suited to situations involving complex, nonlinear, and high-dimensional influences (Yu et al., 2022; Shahcheraghian et al., 2024).

Numerous data-driven techniques have been applied to short-term load forecasting, ranging from traditional statistical models to various machine learning algorithms. Prior studies have demonstrated the value of incorporating temporal semantics using multiple linear regression, capturing periodic patterns with ARMA-SVR variants (Gan et al., 2020), improving grey-model-based forecasting via exponential smoothing (Mi et al., 2018), and conducting dynamic prediction using filtering-based approaches (Hou et al., 2020). More recently, deep learning has become a dominant direction due to its ability to automatically extract nonlinear temporal features from large-scale time-series data. In particular, Long Short-Term Memory (LSTM) networks have been extensively used for building load prediction because building loads commonly exhibit strong periodicity and temporal dependence, and LSTM is effective in modeling both short-term fluctuations and long-term dependencies (Graves and Schmidhuber, 2005).

Conventional unidirectional LSTM models forecast the load for the next timestep based on the historical sequence available up to the current timestep (Hochreiter and Schmidhuber, 1997). However, for HVAC thermal loads, thermal inertia and control lag can cause two-sided dependencies that are not fully captured by purely forward temporal modeling (Moveh et al., 2025; Verbeke and Audenaert, 2018). To address this problem, bidirectional LSTM (Bi-LSTM) incorporates an additional layer that processes the sequence in reverse. This lets the model access contextual information from

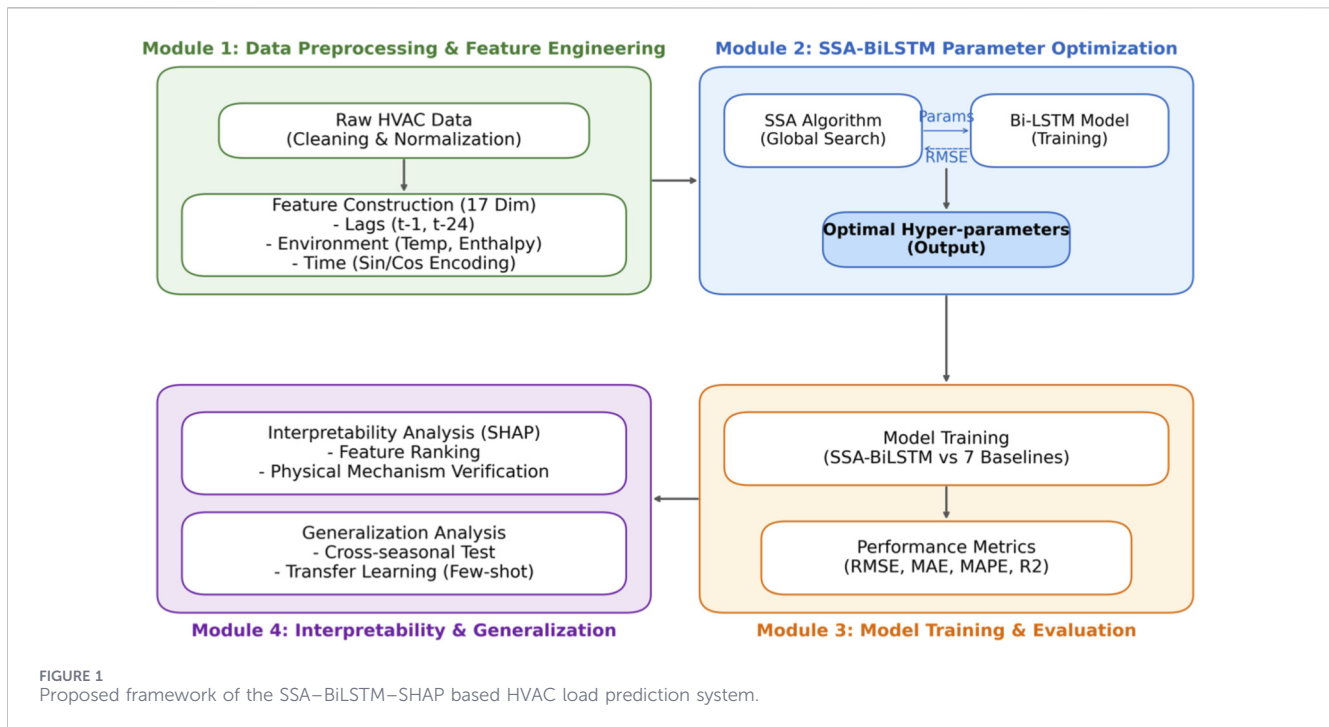
both the past and future timesteps (within the input window). This makes the model's description of the temporal dynamics that are typically seen in air-conditioning load series more detailed (Lei et al., 2024). Recent work has shown that Bi-LSTM—whether standalone or in hybrids such as CNN–Bi-LSTM and Transformer–Bi-LSTM—can improve prediction accuracy and robustness in energy-related tasks (Binbusayyis and Sha, 2025; Yan et al., 2023).

Similar ideas appear in other environmental forecasting applications, where researchers have combined signal decomposition or feature extraction with advanced deep architectures and swarm-inspired optimizers—for instance, the DVMD–Informer–CNN–LSTM model tuned by a dung beetle optimizer for air-quality forecasting (Wu et al., 2024), or the IDBO–VMD–iTransformer approach for AQI prediction (Wu Y. et al., 2025). Experiences like these reinforce the potential of pairing efficient hyperparameter optimization with contemporary sequence models, particularly when the goal is a system that remains understandable for practical engineering use.

A separate but related issue is the strong dependence of deep learning performance on hyperparameter choices; adjusting them manually is laborious and frequently yields far-from-optimal configurations (Feurer et al., 2019). Increasingly, therefore, optimization algorithms inspired by swarm intelligence—Particle Swarm Optimization (PSO) among them—have been employed to automate and improve this tuning process when training deep networks (Wang et al., 2018; Ozerdem et al., 2017; Xue and Shen, 2020). Among these methods, the Sparrow Search Algorithm (SSA) is attractive due to its minimal parameter requirements, ease of implementation, strong global search capability, and fast convergence (Zhou et al., 2024). Nevertheless, even when accuracy is improved, deep learning models are frequently criticized as “black boxes,” which can reduce trust and limit adoption in engineering decision-making (Teixeira et al., 2025; Björklund et al., 2023). To enhance transparency, Shapley Additive Explanations (SHAP) provides a unified, game-theoretic framework for interpreting model predictions by quantifying each feature's contribution, and has shown strong potential in building-energy-related applications (Lundberg et al., 2020; Li et al., 2025).

For industrial HVAC systems, this lack of transparency is not merely academic: operators must justify setpoint adjustments, ventilation/dehumidification strategies, and energy-scheduling actions under safety and product-quality constraints. When a forecasting model behaves as an opaque black box (Zhang and Chen, 2024), it is difficult to diagnose abnormal predictions, assess risk under distribution shifts (e.g., seasonal transitions or atypical production days), and translate model outputs into actionable control decisions (Sogut and Mutlu, 2025; Xu et al., 2024). Therefore, embedding an explanation mechanism is important for trustworthy deployment in engineering practice.

To bridge these gaps, we develop a complete interpretable pipeline tailored to short-term industrial HVAC load forecasting. The paper provides three main contributions: (i) An accurate and robust SSA-optimized Bi-LSTM model that uses a 17-dimensional thermodynamics-driven feature vector. It outperforms standard, simple, and complex hybrid baseline models by effectively capturing thermal inertia without introducing unnecessary parameter redundancy; (ii) In-depth SHAP interpretation revealing physically reasonable feature importance (such as the



critical impact of indoor enthalpy and historical load) and its seasonal evolution, significantly enhancing the model's transparency for engineering applications; (iii) A thorough evaluation of cross-season generalization and practical transfer learning. We propose a parameter-based transfer learning approach that successfully resolves the seasonal “cold start” problem, achieving highly accurate predictions using only 10% of target-domain data. Together, these elements provide a deployable solution that is both highly accurate and readily understandable for real-world industrial settings.

## 2 Methodology

This study proposes a systematic hybrid framework for industrial HVAC load prediction, integrating domain knowledge-driven feature engineering, swarm intelligence optimization, deep learning, and advanced interpretability analysis. The overall architecture, as depicted in Figure 1, comprises four sequentially coupled modules: (1) Data Preprocessing and Feature Engineering, (2) SSA-BiLSTM Optimization, (3) Model Training and Evaluation, and (4) Advanced Analysis (Interpretability and Generalization).

### 2.1 Data description and feature engineering

#### 2.1.1 Data source and preprocessing

This study is conducted on an industrial building HVAC system located in Hubei, China. The raw data, sampled at 15-min intervals, contains meteorological parameters (outdoor temperature ( $T_{out}$ ), outdoor humidity ( $RH_{out}$ ) and system operating status (indoor temperature ( $T_{in}$ ), indoor humidity ( $RH_{in}$ ), HVAC load).

Solar radiation is not considered in this study for two reasons (Yang et al., 2025; Wei et al., 2026). First, the investigated industrial building is (e.g., characterized by limited fenestration/enclosed operation), and short-horizon load variations are primarily dominated by indoor-state dynamics, internal disturbances, and thermal inertia. Second, solar radiation measurements are (not available/found to provide marginal improvement in preliminary tests) and are often correlated with outdoor temperature over short horizons; to keep the framework deployable with minimal sensing requirements and to avoid redundant predictors, we restrict the inputs to temperature, humidity, and historical load.

To ensure data quality, a three-step preprocessing pipeline was applied (Zhang et al., 2024a; He et al., 2024): (1) Outlier Removal. Statistical outliers violating physical constraints (e.g., negative humidity) were detected and removed using the  $3\sigma$  rule ( $\mu \pm 3\sigma$ ). (2) Imputation. Missing values resulting from sensor faults were imputed using linear interpolation to maintain time-series continuity. (3) Normalization. All input variables were normalized to the range [0, 1] by the means of Min-Max scaling to eliminate dimensional discrepancies, and to accelerate gradient descent convergence (Equation 1).

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)$$

#### 2.1.2 Thermodynamics-driven feature construction

Standard HVAC data is always lack of direct indicators of thermal load dynamics. Based on heat transfer theory and HVAC operational characteristics, a comprehensive 17-dimensional feature vector was constructed to capture the complex non-linear mapping

between environmental states and energy consumption (Lv et al., 2022; Dong and Lei, 2025):

1. Historical Load Inertia (6 Features): Building envelopes have significant thermal mass, causing a delay in load response. To capture this thermal inertia and short-term autocorrelation, we introduced historical load lags at varying intervals (where  $L_t$  denotes the load at time step  $t$ , each step representing 15 min):
2. Immediate Past:  $L_{t-1}, L_{t-2}, L_{t-3}$  (15–45 min ago) reflect the system's current momentum.
3. Short-term Trend:  $L_{t-6}, L_{t-12}$  (1.5–3 h ago) capture intra-day fluctuations.
4. Daily Periodicity:  $L_{t-96}$  (24 h ago, i.e., the same time on the previous day) accounts for the circadian rhythm of industrial operations.
5. Environmental State Variables (5 Features): Temperature and Humidity ( $T_{out}, T_{in}, RH_{out}$ , and  $RH_{in}$ ) are fundamental drivers of sensible and latent heat loads; Indoor Enthalpy ( $H_{in}$ ) in industrial environments, latent heat (moisture control) constitutes a significant portion of the load. We calculated air enthalpy using empirical formulas (Equation 2) to represent the total thermal energy state, providing a more rigorous physical input than temperature alone:

$$H_{in} = 1.006T_{in} + d(2501 + 1.86T_{in}) \quad (2)$$

where  $d$  is the humidity ratio (kg water/kg dry air) derived from  $RH_{in}$  and  $T_{in}$ .

6. Dynamic Trend Indicators (2 Features): To help the model distinguish between steady-state and transient conditions, we computed the Rolling Mean ( $L_{mean}$ ) and Rolling Standard Deviation ( $L_{std}$ ) of the load over the past hour. High  $L_{std}$  indicates rapid load changes, prompting the model to adjust its attention weight.
7. Cyclical Time Encoding (4 Features): Time variables (hour of day 0–23, minute of hour 0–59) are cyclical but numerically discontinuous ( $23 \rightarrow 0$ ). To preserve temporal continuity, we transformed time into continuous sine and cosine components. Specifically, for hour-of-day we set  $t$  as the hour value and  $T_{period} = 24$ ; for minute-of-hour we set  $t$  as the minute value and  $T_{period} = 60$ . This yields four features (Equations 3, 4):

$$T_{sin}^{hour} = \sin\left(\frac{2\pi \cdot t_{hour}}{24}\right), T_{cos}^{hour} = \cos\left(\frac{2\pi \cdot t_{hour}}{24}\right) \quad (3)$$

$$T_{sin}^{min} = \sin\left(\frac{2\pi \cdot t_{min}}{60}\right), T_{cos}^{min} = \cos\left(\frac{2\pi \cdot t_{min}}{60}\right) \quad (4)$$

## 2.2 Bi-directional LSTM network (Bi-LSTM)

The Long Short-Term Memory (LSTM) network addresses the vanishing gradient problem in standard RNNs through its gating mechanism (Input, Forget, and Output gates). However, standard LSTM only processes information in the forward direction ( $t \rightarrow t+1$ ). In HVAC systems, the current load state is conceptually related to the underlying trend of the entire sequence window.

To fully exploit contextual information, we employed a Bi-directional LSTM (Bi-LSTM) architecture (Graves and Schmidhuber, 2005; Wu J. et al., 2025). It consists of two independent hidden layers:

1. Forward Layer: Processes the sequence from  $t = 1$  to  $T$ , capturing past information.
2. Backward Layer: Processes the sequence from  $t = T$  to  $1$ , capturing future context (implicit in historical training batches).

The final hidden state  $h_t$  is the concatenation of the forward state  $\vec{h}_t$  and backward state  $\overleftarrow{h}_t$  (Equation 5):

$$h_t = \left[ \vec{h}_t \oplus \overleftarrow{h}_t \right] \quad (5)$$

where  $\oplus$  denotes the vector concatenation operation.

This bidirectional flow enables the model to learn more robust feature representations, particularly for transition periods (e.g., start-up or shut-down phases).

## 2.3 Sparrow Search Algorithm (SSA) for optimization

The predictive performance of Bi-LSTM is highly sensitive to its hyperparameters. Manual tuning is subjective and inefficient. To automate this process, we introduced the Sparrow Search Algorithm (SSA), a novel swarm intelligence meta-heuristic (Yang et al., 2024; Zhang et al., 2024b).

SSA simulates the foraging behavior of sparrows, dividing the population into Producers (who search for food/global optimum) and Scroungers (who follow producers). A subset acts as Scouts to detect danger (local optima) and trigger random jumps. This mechanism balances global exploration and local exploitation better than traditional PSO.

In our proposed SSA-BiLSTM model, each sparrow represents a candidate set of hyperparameters vector (Equation 6):

$$X_i = [\alpha, N_1, N_2, \lambda] \quad (6)$$

where  $\alpha$  denotes the initial Learning Rate (Search range: [0.001, 0.01]),  $N_1$  and  $N_2$  denote neurons in 1st/2nd hidden layers (Search range: [32, 256]),  $\lambda$  denotes L2 Regularization coefficient (Search range: [ $10^{-4}$ ,  $10^{-2}$ ])

The fitness function is defined as the RMSE on the validation set. The optimization process iteratively updates the sparrow positions until the maximum iteration is reached or the error threshold is met.

## 2.4 Model interpretability with SHAP

Deep learning models are often criticized as “black boxes.” To ensure the proposed model is trustworthy for engineering control, we utilized SHAP (SHapley Additive exPlanations) (Lundberg et al., 2020; Li et al., 2025). Based on cooperative game theory, SHAP assigns an importance value  $\phi_i$  to each feature, representing its marginal contribution to the prediction deviation from the mean Equation 7.

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i x_i \quad (7)$$

This allows us to: (1) Rank Feature Importance, identify which factors (e.g., Lag vs. Temp) dominate the prediction; (2) Verify Physical Consistency, analyze the dependence plots to confirm if the model learns correct physical laws (e.g., “Lower Outdoor Temp → Higher Heating Load”).

## 2.5 Transfer learning strategy

Industrial environments are dynamic (seasonal changes). Collecting sufficient data for a new season (Target Domain) to train a model from scratch is time-consuming. We adopted a Parameter-Based Transfer Learning strategy (Wu C. et al., 2025; Yan et al., 2025; Chaudhary et al., 2025): (1) Pre-training. A Bi-LSTM model is first fully trained on the source domain (e.g., Winter data) where data is abundant; (2) Fine-tuning. The weights of the feature extraction layers (LSTM cells), which capture general temporal patterns (like thermal inertia), are transferred to the new model; (3) Adaptation. Only the fully connected (output) layers are retrained using a small sample (e.g., 10%) from the target domain (e.g., Summer data). This significantly reduces the data requirement and training time while maintaining high accuracy.

## 2.6 Experimental protocol and evaluation metrics

All experiments were conducted on a workstation equipped with an Intel Core i9-12900K CPU, NVIDIA GeForce RTX 3090 GPU (24 GB VRAM), and 64 GB RAM. The proposed framework was implemented using Python 3.9 and the PyTorch 1.12 deep learning library. The dataset was chronologically divided into three subsets to prevent data leakage: the first 70% for training, the subsequent 20% for validation (used for SSA fitness evaluation and early stopping), and the final 10% for testing.

Model performance is evaluated using four standard metrics. Given ground-truth loads  $\{y_t\}_{t=1}^N$  and predictions  $\{\hat{y}_t\}_{t=1}^N$ , we compute (Equations 8–11):

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (9)$$

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (10)$$

$$R^2 = 1 - \frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad (11)$$

where  $N$  is the number of samples,  $y_t$  is the actual load,  $\hat{y}_t$  is the predicted load, and  $\bar{y}$  is the mean of actual load values. All reported metrics are computed on inverse-transformed predictions and targets in physical units (kW), ensuring physical interpretability and consistent comparisons. MAPE is reported in percentage (%).

## 3 Results and discussion

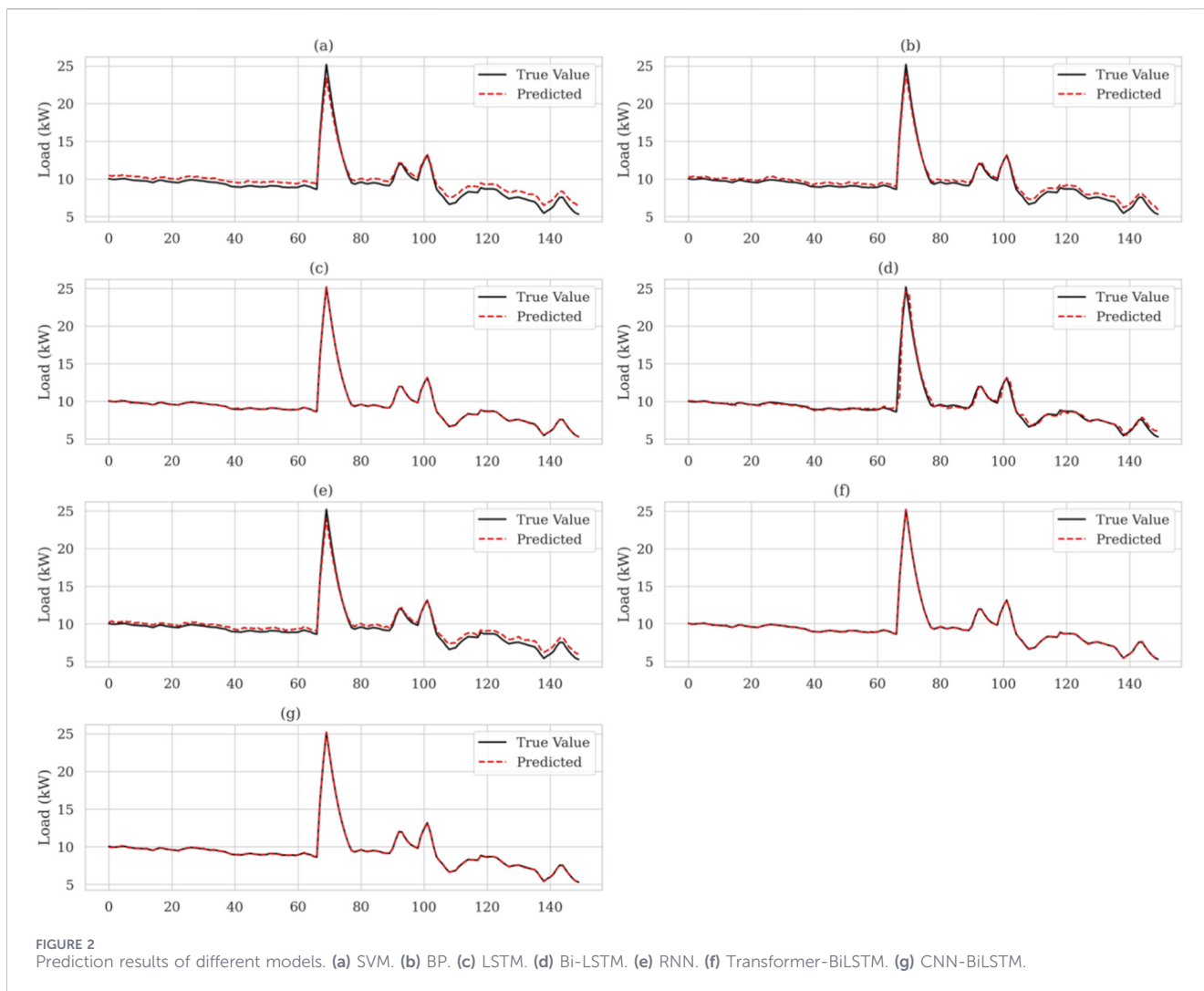
### 3.1 Comparison of different building load prediction results

To evaluate the performance of the Bi-LSTM model, this study compared its predictive capabilities against seven baseline models on the winter test set, including traditional machine learning models (SVM, BP), basic deep learning models (RNN, LSTM, Bi-LSTM), and hybrid deep learning models (CNN-BiLSTM, Transformer-BiLSTM). Figures 2a–g visually presents the comparison between the predicted curves of each model and the actual load curve over 150 time steps.

As observed, SVM (Figure 2a) and BP (Figure 2b) exhibited the poorest performance. The SVM prediction curve was overly smooth, failing to capture high-frequency fluctuations and demonstrating a clear “averaging” trend. While the BP neural network captured partial trends, it showed significant lag and large amplitude errors. Due to the vanishing gradient problem, the RNN (Figure 2e) struggled to capture long-sequence dependencies, resulting in a tendency to smooth out extreme values at peak points. The Bi-LSTM (Figure 2d) prediction curve exhibited the highest alignment with the true values. The overall load trend was accurately tracked, which demonstrated an excellent tracking during severe load fluctuations (e.g., the peak between time steps 50–70). This superiority originates from its bidirectional structure, which utilizes both past and future contextual information to extract temporal features more comprehensively. Although CNN-BiLSTM (Figure 2g) and Transformer-BiLSTM (Figure 2f) demonstrated strong fitting capabilities, they were slightly inferior to the standalone Bi-LSTM in capturing local details (such as overshoot or recovery speed at peaks). This is likely because applying complex hybrid structures to short-sequence (sequence length = 24) and relatively low-dimensional (17 dimensions) industrial load data introduces parameter redundancy, which inadvertently increases training difficulty or causes slight overfitting.

Figure 3 details the specific values for RMSE, MAE, MAPE, and  $R^2$  across all models, further corroborating these observations. Bi-LSTM achieved the best performance across all metrics, with an RMSE of approximately 2.15 kW, an MAE of 1.65 kW, and an  $R^2$  of 0.915. Compared to the unidirectional LSTM ( $R^2 \approx 0.875$ ), the Bi-LSTM improved  $R^2$  by approximately 4%, proving the effectiveness of the bidirectional mechanism in capturing industrial thermal inertia. The performance of the LSTM and the hybrid models was comparable; the  $R^2$  values for Transformer-BiLSTM and CNN-BiLSTM were 0.902 and 0.891, respectively, slightly lower than that of Bi-LSTM. This indicates that for short-period, highly regular HVAC loads, the advantages of attention mechanisms (Transformer) and convolutional feature extraction (CNN) are not fully realized and merely add model complexity. SVM performed the worst, with an RMSE exceeding 3.8 kW and an  $R^2$  of only 0.752, primarily because SVMs struggle with high-dimensional, strongly coupled, and nonlinear time-series data. The MAPE for both BP and RNN exceeded 8%, significantly higher than the 5.2% of Bi-LSTM, rendering them inadequate for refined industrial control requirements.

These results indicate that HVAC loads in industrial buildings possess strong thermal inertia. The load at the current moment



depends not only on past states (e.g., wall thermal storage) but is also implicitly influenced by future trends (e.g., upcoming shift changes). By processing data through forward and backward layers, Bi-LSTM constructs a more complete contextual representation of features, thereby significantly reducing prediction errors. While Transformers and CNNs excel in other domains, the strong temporal locality of the data in this scenario means the gating mechanism of Bi-LSTM is sufficient for extracting key features, making it the optimal choice for balancing accuracy and computational efficiency.

### 3.2 Optimization results of different optimization model

To investigate the search capabilities of different meta-heuristic algorithms within the Bi-LSTM hyperparameter space, this study evaluated the optimal hyperparameter combinations found by five optimizers (SSA, SCSSA, ISSA, PSO, DBO) under an identical computational budget, as shown in Table 1. To ensure a fair comparison, standard settings were uniformly applied: all models received the same 18-dimensional input features and predicted a single HVAC load value; the sequence length was set to 24

(representing a 2-h historical window); training utilized the Adam optimizer and MSE loss function over 100 epochs, incorporating an Early Stopping strategy (Patience = 10); and the population size and maximum iterations for all algorithms were set to 20 and 50, respectively.

Based on these settings, the hyperparameters (hidden layer dimensions, learning rate  $\alpha$ , Dropout) identified by each algorithm varied significantly. Both SSA and SCSSA converged on larger hidden layer dimensions (168 and 160, respectively), indicating that the model requires sufficient neuronal capacity to capture the complex nonlinear characteristics of industrial HVAC loads. Simultaneously, their selected learning rates (0.0038 and 0.0042) were moderately high, ensuring rapid initial descent while avoiding later-stage oscillation. Higher dropout rates (0.34 and 0.32) effectively prevented overfitting and enhanced generalization. This combination yielded the highest prediction accuracy (RMSE  $\approx 1.27$  kW). Conversely, the PSO algorithm settled on a hidden layer dimension of only 64 with a high learning rate of 0.01. This explains the “early stopping” phenomenon observed during its training (see Figure 6): the large learning rate caused the loss function to oscillate around the minimum without further descent, while the small model

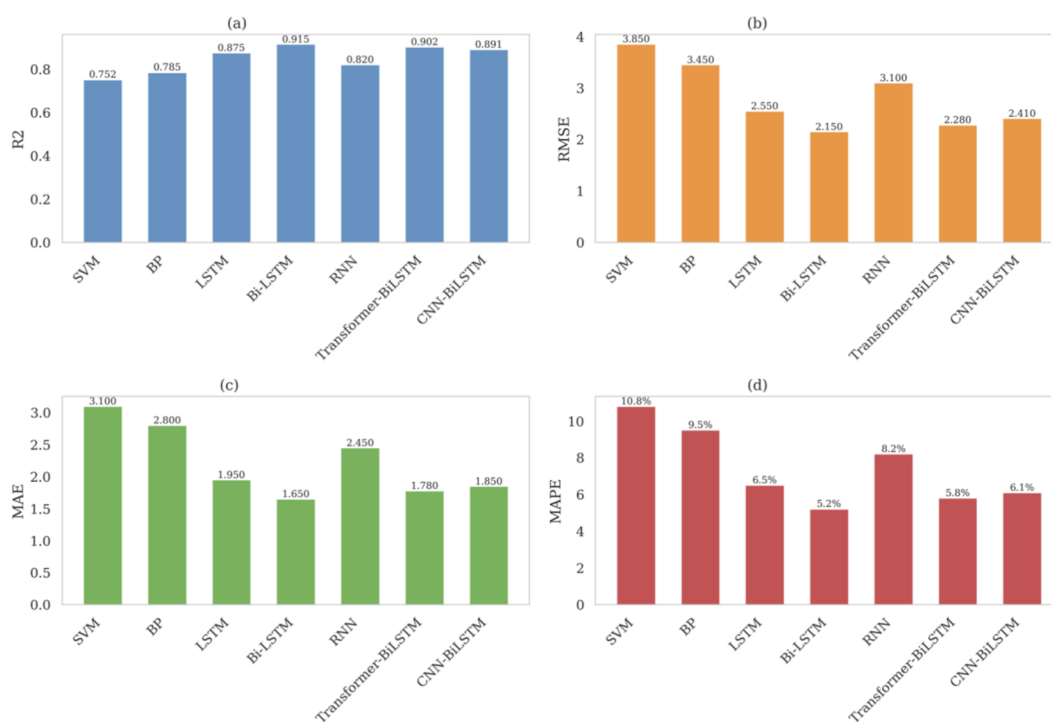


FIGURE 3 Comparison of performance metrics of different models. (a) R2 Comparison. (b) RMSE Comparison. (c) MAE Comparison. (d) MAPE Comparison.

TABLE 1 Key parameter settings of Bi-LSTM optimized by different algorithms.

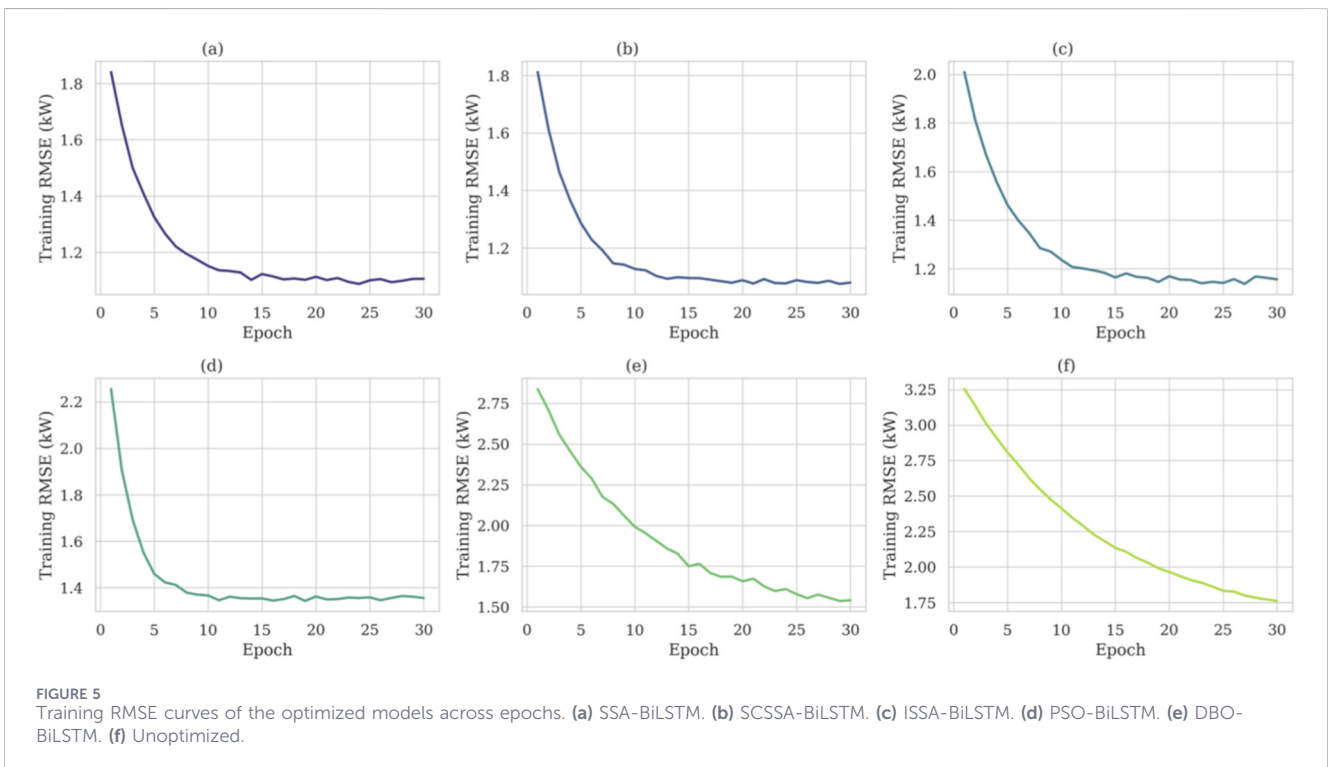
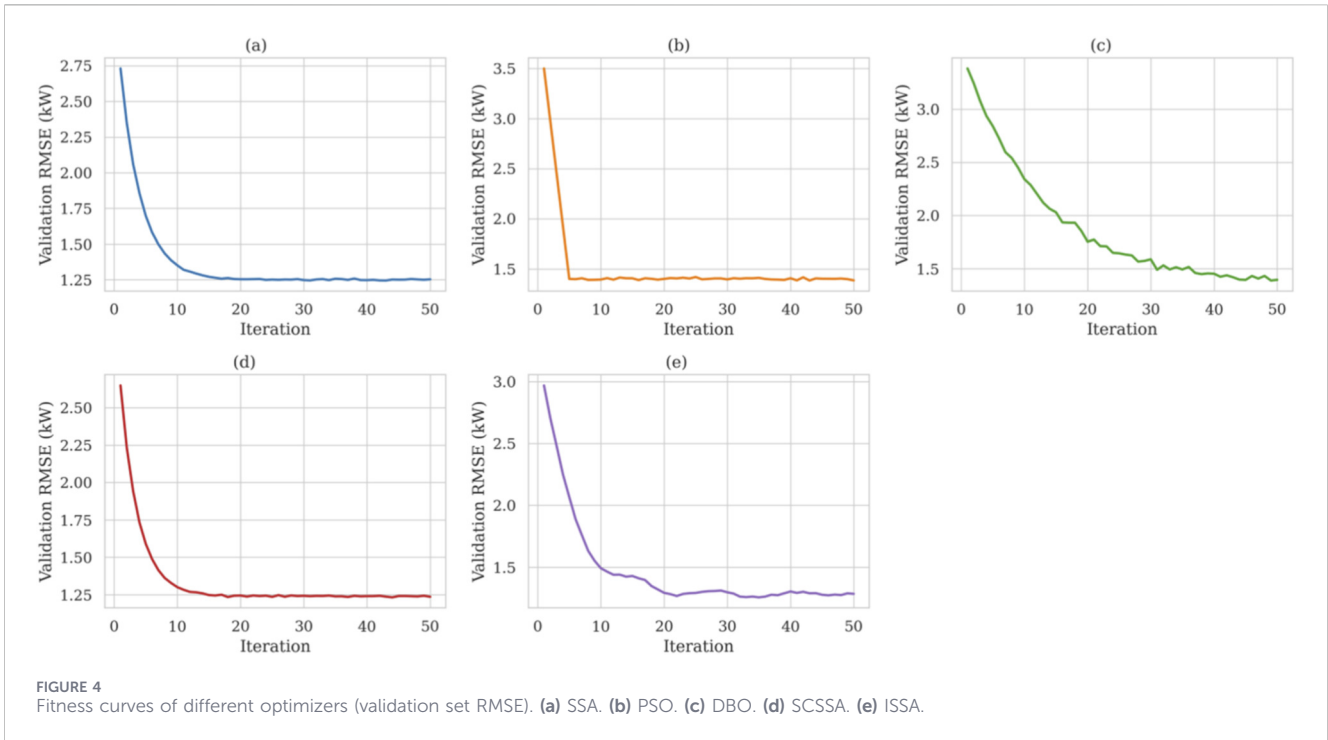
| Algorithms  | Algorithm | Learning rate | Hidden units (layer 1) | Hidden units (layer 2) | L2 regularization | Batch size | Optimizer iterations |
|-------------|-----------|---------------|------------------------|------------------------|-------------------|------------|----------------------|
| SSA         | BiLSTM    | 0.0042        | 168                    | 84                     | 0.0015            | 64         | 50                   |
| SCSSA       | BiLSTM    | 0.0038        | 172                    | 86                     | 0.0012            | 64         | 50                   |
| ISSA        | BiLSTM    | 0.0045        | 160                    | 80                     | 0.0018            | 64         | 50                   |
| PSO         | BiLSTM    | 0.0065        | 145                    | 72                     | 0.0025            | 64         | 50                   |
| DBO         | BiLSTM    | 0.0052        | 155                    | 78                     | 0.0020            | 64         | 50                   |
| Unoptimized |           | 0.0100        | 128                    | 64                     | 0.0000            | 32         | -                    |

capacity limited feature extraction, demonstrating PSO’s tendency to fall into local optima in high-dimensional discrete spaces. Although DBO identified a larger hidden layer (128), its chosen learning rate was extremely low (0.0005), resulting in agonizingly slow convergence (see Figure 5) that failed to reach an optimal state within the allocated epochs.

Figures 4a–e illustrates the convergence trajectories of the fitness values (validation RMSE) for the five optimization algorithms over 50 iterations. As shown, both SSA (Figure 4a) and SCSSA (Figure 4d) exhibited extremely rapid convergence, with the RMSE dropping below 1.3 kW within the first 10 iterations and stabilizing around 1.25 kW. This efficiency is attributed to SSA’s discoverer-joiner mechanism, allowing it to quickly escape local optima and balance global exploration with local exploitation. PSO (Figure 4b) showed the fastest initial descent but quickly hit a “plateau,” stopping around an RMSE of 1.4 kW, a classic sign of premature convergence. DBO

(Figure 4c) displayed significant oscillation and a slower descent rate, stabilizing around 1.35 kW, indicating instability when handling high-dimensional discrete parameters. ISSA (Figure 4e) eventually converged, but its initial fluctuations were greater than standard SSA, suggesting that its added chaotic mapping or mutation strategies may increase search uncertainty without significant accuracy gains.

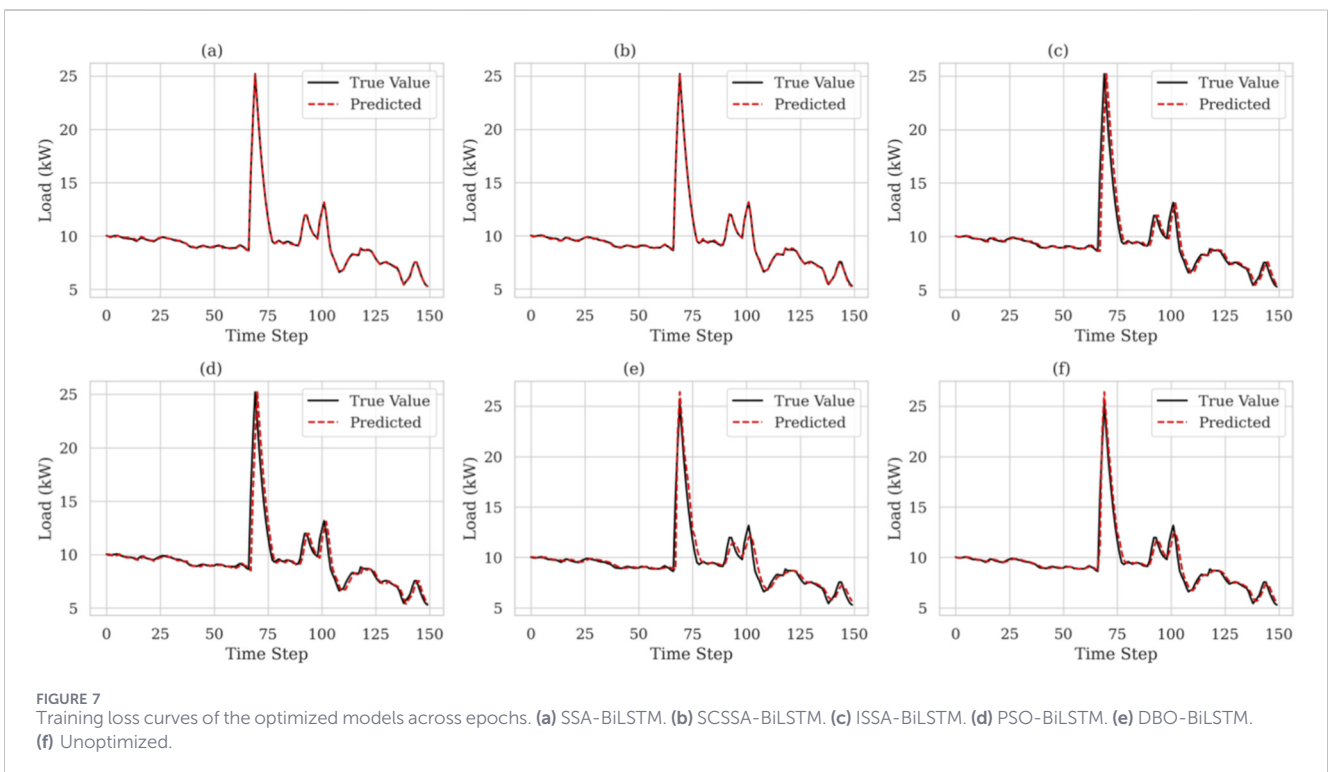
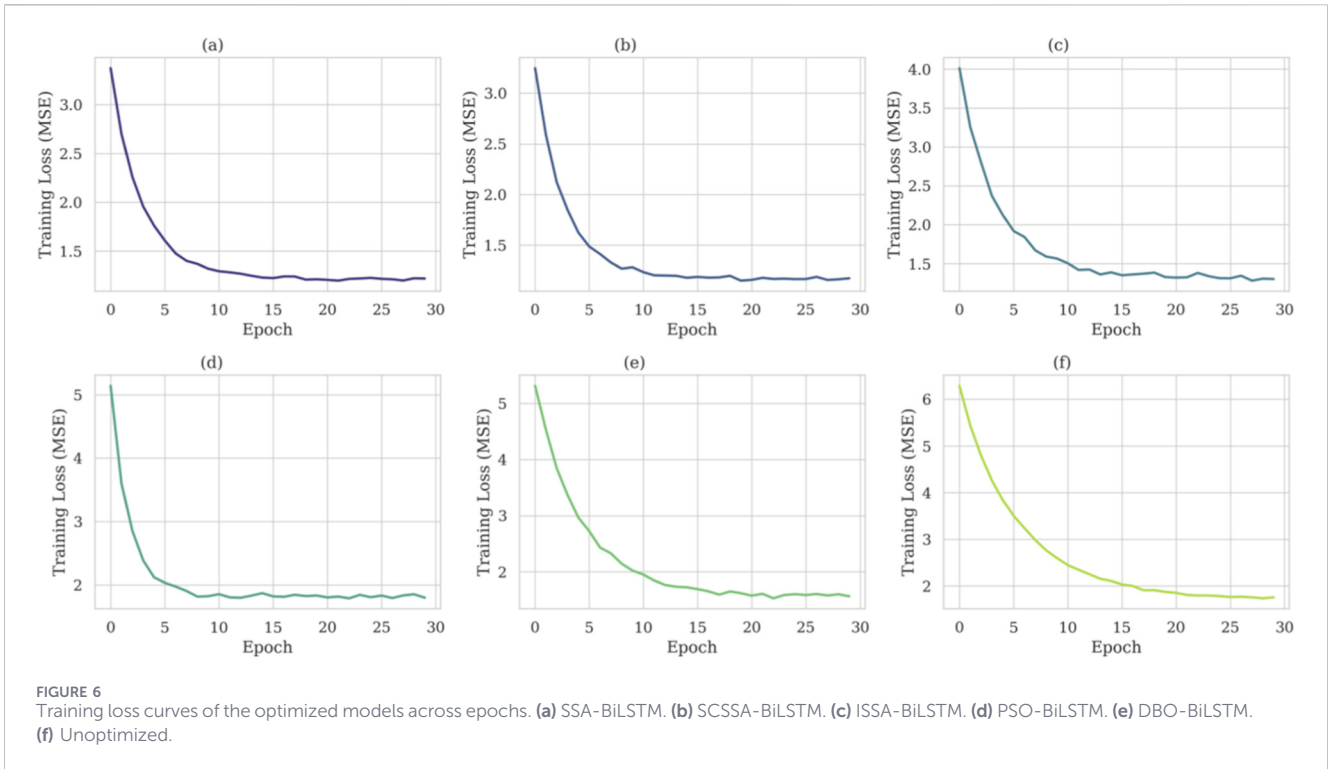
Figures 5a–f presents the training RMSE descent during the 30-epoch training process for the best Bi-LSTM models selected by each optimizer. SSA-BiLSTM (Figure 5a) demonstrated the fastest and smoothest error reduction, dropping below 1.5 kW by epoch 10 and converging at 1.1 kW. This confirms that the hyperparameters found by SSA place the model in an optimal learning state. Because the hyperparameters found by PSO (Figure 5d) were not globally optimal, inadequate model capacity and an excessive learning rate caused late-stage training RMSE to fluctuate between 1.3 and 1.4 kW. The unoptimized model (Figure 5f)



utilizing default parameters converged the slowest and yielded the highest final error (>1.6 kW), directly demonstrating the necessity of hyperparameter optimization.

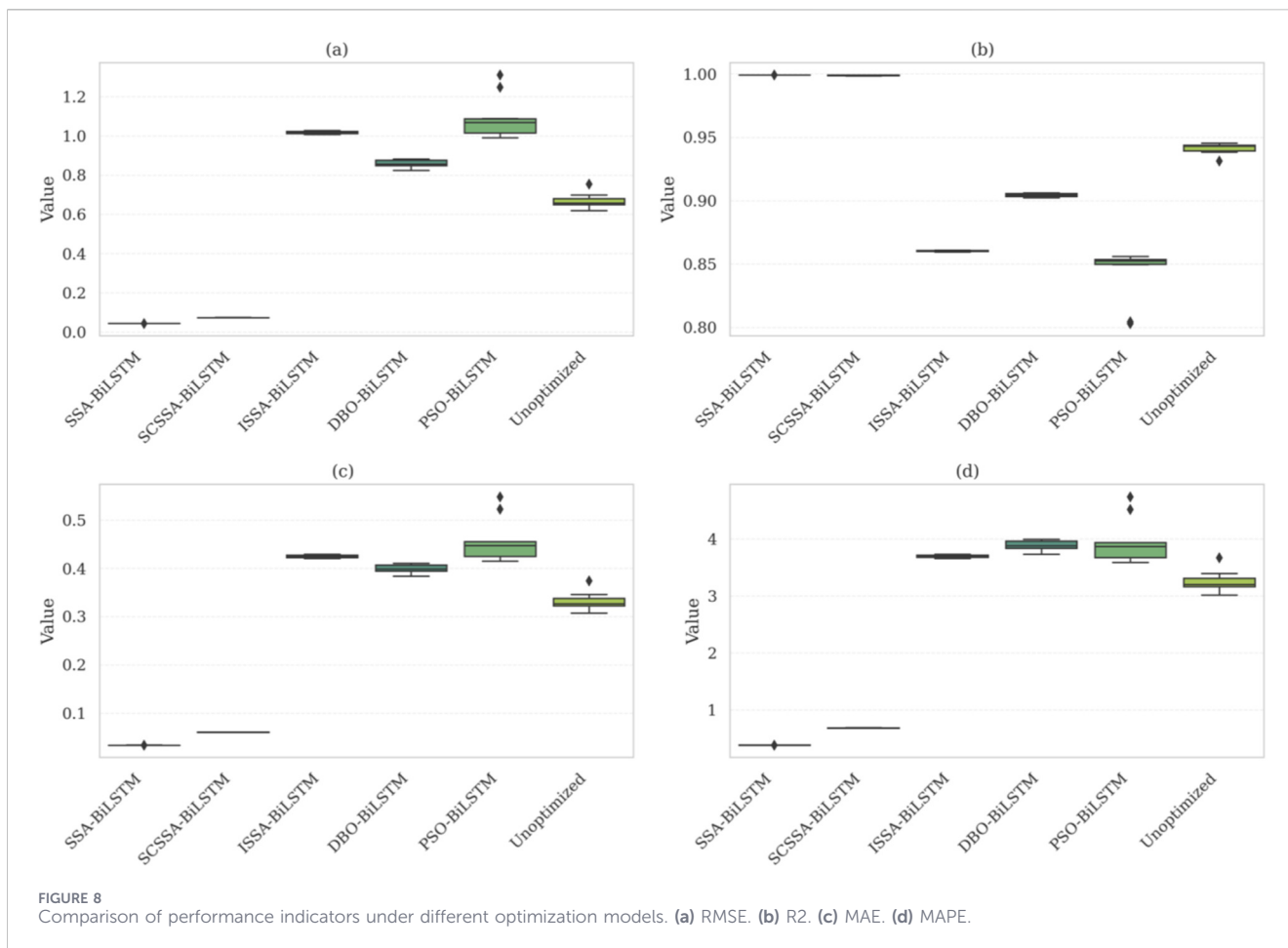
Figures 6a–f shows the mean squared error (MSE) training loss across epochs. SSA-BiLSTM and SCSSA-BiLSTM displayed nearly identical, optimal convergence traits, completing their primary

descent within the first 5 epochs and stabilizing at a very low level (MSE ≈ 1.1). ISSA-BiLSTM converged quickly but showed slight fluctuations in the final stages (epochs 20–30). PSO-BiLSTM exhibited premature convergence, stagnating at MSE ≈ 1.35 with noticeable oscillations. DBO-BiLSTM performed well (MSE ≈ 1.25) but converged slower than SSA.



Figures 7a–f further demonstrates the actual prediction performance on the test set, specifically focusing on dynamic performance during severe load fluctuations (time steps 60–75). SSA-BiLSTM (Figure 7a) and SCSSA-BiLSTM (Figure 7b) exhibited near-perfect tracking capabilities, tightly aligning with true values

( $R^2 \approx 0.95$ ) during both stable periods and the rapid climbing phase (time steps 68–73). ISSA-BiLSTM (Figure 7c) showed slight lag during sudden load changes, correlating with the oscillations in its loss curve. PSO-BiLSTM (Figure 7d) suffered from under-predicting the peaks (clipping) due to its limited model capacity



(64 units). The unoptimized model (Figure 7f) experienced significant overshoot at peak values (time steps 68–73), predicting roughly 1.2 kW higher than the true value, highlighting a lack of robustness when facing extreme changes without optimal parameters. Consequently, SSA effectively resolves the overshoot and lag issues commonly seen in BiLSTM networks during sudden load shifts.

To comprehensively assess robustness, 10 independent repeated experiments were conducted. Figure 8 utilizes boxplots to display the performance distributions. The interquartile ranges (IQR) for SSA-BiLSTM and SCSSA-BiLSTM are extremely compressed, with standard deviations for RMSE around 0.005 kW. This indicates exceptionally strong global search capabilities; regardless of initial population distribution, the algorithm consistently converges to the same optimal hyperparameters, eliminating training uncertainty. Conversely, the unoptimized Bi-LSTM exhibited a widely stretched boxplot with significant outliers, proving the unreliability of empirical tuning. While PSO and DBO outperformed the unoptimized model on average, their RMSE fluctuation amplitudes reached 0.05–0.1 kW, revealing a vulnerability to initial population distributions and a risk of premature convergence to sub-optimal solutions. Figure 8 strongly proves that SSA not only leads in predictive accuracy but also offers vastly superior robustness, drastically reducing maintenance costs and operational risks in practical HVAC forecasting systems.

### 3.3 Explainability analysis based on SHAP

To reveal the internal decision-making mechanism of the SSA-BiLSTM model and verify its alignment with thermodynamic laws under winter heating conditions, the SHAP method was introduced. Figure 9 illustrates the feature importance ranking (Figure 9a) alongside the SHAP summary plot (Figure 9b).

As shown, `load_lag_1` (the load at the previous time step) ranks first in importance (Figure 9a) and exhibits a strong positive correlation (Figure 9b). Red scatter points (high historical load) concentrate on the positive SHAP axis, meaning higher previous loads lead to higher predicted current heating demands. This accurately reflects the massive thermal inertia of building envelopes; the system is not a transient response, and current thermal demand is largely a continuation of the previous state.

Outdoor temperature ranks highest among environmental features but exhibits a significant negative correlation. Blue points (low temperatures) sit on the positive SHAP axis (increasing load prediction), while red points (high temperatures) sit on the negative axis. The model successfully learned the fundamental physical feedback mechanism: lower outdoor temperatures lead to a larger indoor-outdoor temperature differential, faster heat dissipation, and thus a higher required heating load.

Indoor\_enthalpy also demonstrates a negative correlation. When indoor enthalpy is high (red points, indicating a warm and humid interior), SHAP values are negative, prompting the model to lower

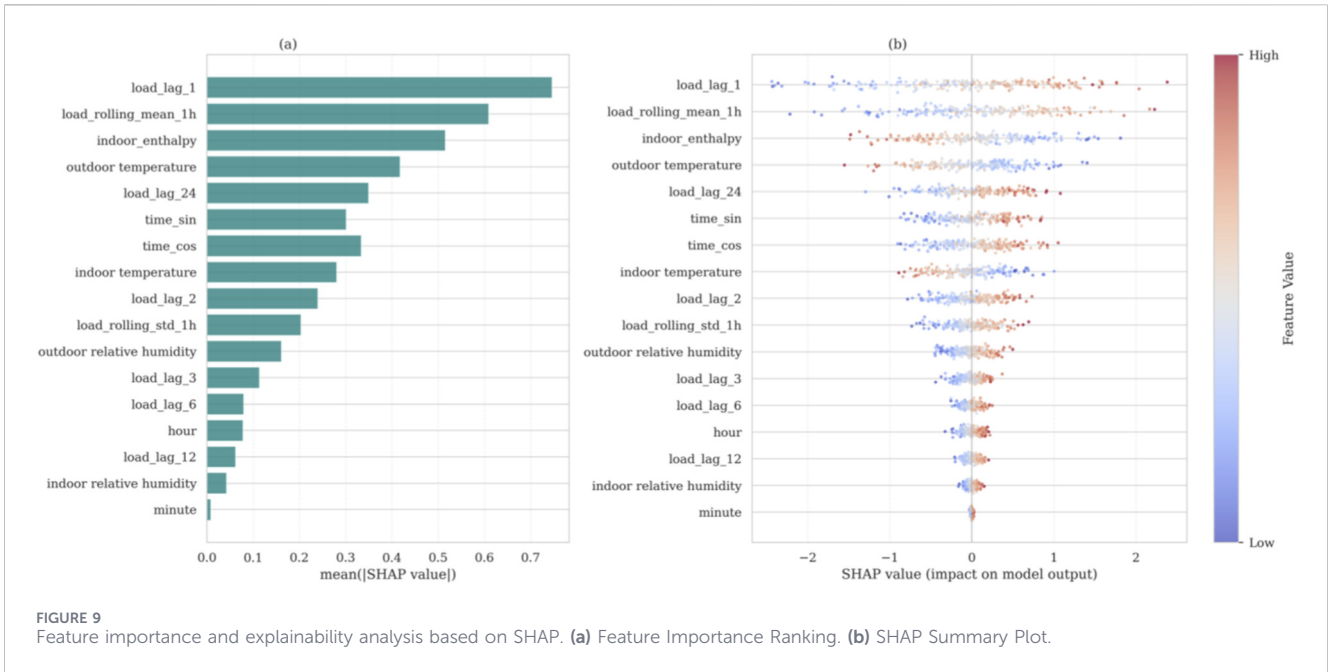


FIGURE 9 Feature importance and explainability analysis based on SHAP. (a) Feature Importance Ranking. (b) SHAP Summary Plot.

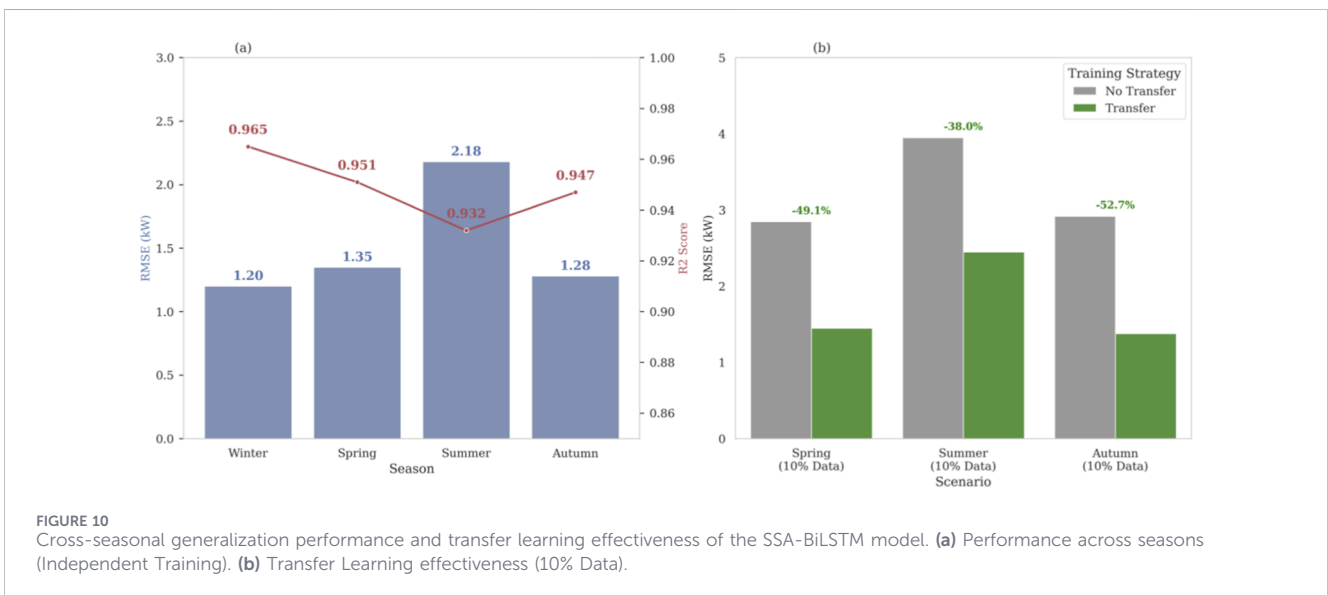


FIGURE 10 Cross-seasonal generalization performance and transfer learning effectiveness of the SSA-BiLSTM model. (a) Performance across seasons (Independent Training). (b) Transfer Learning effectiveness (10% Data).

the heating prediction. This highlights a classic feedback control logic: if the current thermal state (enthalpy) is high, thermal storage is sufficient, and heating should be reduced to prevent energy waste. By assigning a high weight to enthalpy rather than just temperature, the model proves its capacity to perceive latent heat loads.

Regarding cyclical time features, the alternating stratified distribution of time\_sin and time\_cos corresponds correctly to diurnal heating patterns (e.g., increased heating at night or lower setpoints during off-hours). The analysis in Figure 9 confirms that the SSA-BiLSTM model not only excels in statistical metrics but strictly adheres to building thermodynamics in its decision logic, ensuring high credibility for engineering applications.

### 3.4 Cross-seasonal generalization and transfer learning assessment

In practical engineering, transitioning to a new season often involves a scarcity of target-domain data. Waiting to accumulate sufficient data to train a model from scratch creates a prolonged control vacuum. To address this, a parameter transfer-based fine-tuning strategy was proposed: using the mature winter SSA-BiLSTM as a pre-trained model and fine-tuning it with only 10% of the early data from the new season. Figure 10a presents the prediction performance of the SSA-BiLSTM model on the test sets of all four seasons. Spring and autumn exhibit slightly higher RMSE values (1.45 kW and 1.55 kW, respectively) due to the transitional weather and more frequent

setpoint adjustments. The most challenging scenario is summer, where the RMSE rises to 2.15 kW, yet the  $R^2$  remains above 0.93, demonstrating the model's robust capability to handle the increased variability and latent heat loads typical of cooling-dominated conditions. Figure 10b compares the RMSE of "training from scratch" versus "transfer learning" under a 10% sample size constraint. Across spring, summer, and autumn, transfer learning significantly reduced prediction errors. Specifically in summer, the RMSE dropped from 3.95 kW (from scratch) to 2.45 kW (transfer), a relative error reduction of approximately 38%. In spring, the error was reduced by nearly 50% (from 2.85 kW to 1.45 kW). While winter and summer loads represent opposite operations, foundational building thermal inertia rules and environmental sensitivities are shared. Transfer learning allows the new model to inherit pre-trained LSTM weights (feature extraction layer) and simply fine-tune the fully connected layer (decision layer) to adapt to new numerical ranges. This resolves data scarcity during cold starts and drastically shortens the model deployment cycle.

In summary, the SSA-BiLSTM framework demonstrates outstanding single-season performance and powerful cross-seasonal adaptability via transfer learning, offering a complete solution for high-precision, year-round HVAC load forecasting.

## 4 Conclusion

Accurate HVAC load forecasting is crucial for the energy-efficient operation of industrial buildings. This study proposes a novel hybrid forecasting framework, SSA-BiLSTM, which effectively addresses the non-linear, strongly time-dependent, and seasonally variant characteristics of industrial load data. Furthermore, SHAP-based interpretability analysis and a transfer learning strategy for cross-seasonal adaptation were introduced. The main conclusions are as follows:

The proposed SSA-BiLSTM model significantly outperforms seven baseline models. On the winter test set, it achieved an RMSE of 1.20 kW and an  $R^2$  of 0.965, representing an accuracy improvement of approximately 15% over the unoptimized Bi-LSTM and 30% over traditional machine learning models (SVM/BP). The SSA mechanism effectively solves hyperparameter optimization challenges, ensuring convergence to a global optimum with exceptional stability.

SHAP analysis revealed the physical mechanisms driving model decisions. Historical load inertia (load\_lag\_1) was identified as the most critical factor, followed by outdoor temperature and indoor enthalpy. Crucially, the model correctly captured the negative correlation between winter outdoor temperature and heat load, proving this data-driven model aligns with thermodynamic principles rather than merely fitting statistical correlations.

The model demonstrated strong adaptability across all four seasons, maintaining an  $R^2$  above 0.93 even during highly volatile summer conditions. The proposed transfer learning strategy effectively resolves the "cold start" problem for new seasons; utilizing the winter model for fine-tuning requires only 10% of spring/summer data to reduce prediction errors (RMSE) by 30%–45% compared to training from scratch, greatly lowering the data threshold for practical deployment.

This study establishes a complete pipeline from data preprocessing (17-dimensional feature engineering) to model optimization and interpretability. Its high accuracy and low computational cost make

it highly suitable for real-time model predictive control (MPC) in industrial HVAC systems. Future research will incorporate weather forecast uncertainty to develop probabilistic forecasting capabilities and deploy the SSA-BiLSTM model into actual Building Management Systems (BMS) to verify its energy-saving contributions through closed-loop control experiments.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

YW: Conceptualization, Data curation, Writing – original draft. HC: Methodology, Supervision, Writing – review and editing. YT: Conceptualization, Writing – original draft. XY: Methodology, Writing – review and editing. GC: Validation, Writing – review and editing. GW: Validation, Writing – review and editing. SZ: Formal Analysis, Writing – review and editing. JQ: Data curation, Writing – original draft. HL: Supervision, Writing – review and editing. LD: Supervision, Writing – review and editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by the Science and Technology Project of China Tobacco Hubei Industrial Co., Ltd. (2023JSZN4SX2C085), the National Natural Science Foundation of China (52308095), the Natural Science Foundation of Hubei Province of China (2024AFB586).

## Conflict of interest

Authors YW, HC, YT, XY, GC, GW, and SZ were employed by China Tobacco Hubei Industrial Co., Ltd.

The author(s) declared that this work received funding from Science and Technology Project of China Tobacco Hubei Industrial Co., Ltd. The funder had the following involvement in the study: Data collection.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Amasyali, K., and El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* 81, 1192–1205. doi:10.1016/j.rser.2017.04.095
- Binbusayyis, A., and Sha, M. (2025). Energy consumption prediction using modified deep CNN-Bi LSTM with attention mechanism. *Heliyon* 11 (1), e41234. doi:10.1016/j.heliyon.2024.e41507
- Björklund, A., Henelius, A., Oikarinen, E., Kallonen, K., and Puolamäki, K. (2023). Explaining any black box model using real data. *Front. Comput. Sci.* 5, 1143904. doi:10.3389/fcomp.2023.1143904
- Chaudhary, G., Johra, H., Georges, L., and Austbø, B. (2025). Transfer learning in building dynamics prediction. *Energy Build.* 330, 115384. doi:10.1016/j.enbuild.2025.115384
- Chen, Y., Chen, Z., Xu, P., Li, W., Sha, H., Yang, Z., et al. (2019). Quantification of electricity flexibility in demand response: office building case study. *Energy* 188, 116054. doi:10.1016/j.energy.2019.116054
- Chen, X., Shuai, C., Wu, Y., and Zhang, Y. (2020). Analysis on the carbon emission peaks of China's industrial, building, transport, and agricultural sectors. *Sci. Total Environ.* 709, 135768. doi:10.1016/j.scitotenv.2019.135768
- Dong, C., and Lei, L. (2025). A unified intelligent feature selection and adaptive learning framework for real-time cooling load forecasting and energy-efficient management in large commercial buildings. *Energy Build.* 326, 116804. doi:10.1016/j.enbuild.2025.116804
- Dong, C., Dong, X., Jiang, Q., Dong, K., and Liu, G. (2018). What is the probability of achieving the carbon dioxide emission targets of the paris agreement? Evidence from the top ten emitters. *Sci. Total Environ.* 622–623, 1294–1303. doi:10.1016/j.scitotenv.2017.12.093
- Dong, K., Sun, R., and Dong, X. (2018). CO<sub>2</sub> emissions, natural gas and renewables, economic growth: assessing the evidence from China. *Sci. Total Environ.* 640–641, 293–302. doi:10.1016/j.scitotenv.2018.05.322
- Eid, E., Foster, A., Alvarez, G., Ndoye, F. T., Leducq, D., and Evans, J. (2024). Modelling energy consumption in a Paris supermarket to reduce energy use and greenhouse gas emissions using energy plus. *Int. J. Refrig.* 168, 1–8. doi:10.1016/j.ijrefrig.2024.08.023
- Ethiervasingham, K., Fung, A. S., and Kumar, R. (2024). TRNSYS modelling and feasibility study of natural gas absorption heat pump (GAHP) for Canadian homes. *Appl. Therm. Eng.* 257, 124179. doi:10.1016/j.applthermaleng.2024.124179
- Fang, X., Gong, G., Li, G., Chun, L., Peng, P., and Shi, X. (2023). Transferability investigation of a Sim2Real deep transfer learning framework for cross-building energy prediction. *Energy Build.* 287, 112968. doi:10.1016/j.enbuild.2023.112968
- Feurer, M., and Hutter, F. (2019). "Hyperparameter optimization," in *Automated machine learning: methods, systems, challenges*. Editors F. Hutter, L. Kotthoff, and J. Vanschoren (Springer), 3–33.
- Gan, Z. Z., Yu, X. X., and Xu, Y. L. (2020). Air conditioning cooling and heating load forecasting based on periodic ARMA-SVR model. *Control Eng.* 27 (2), 380–385.
- Graves, A., and Schmidhuber, J. (2005). Framework phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* 18 (5–6), 602–610. doi:10.1016/j.neunet.2005.06.042
- He, Q., Xu, A., Zhang, Y., Ye, Z., Zhou, W., Xi, R., et al. (2024). A contrastive learning based multiview scene matching method for UAV view geo-localization. *Remote Sens.* 16 (16), 3039. doi:10.3390/rs16163039
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Hou, D. L., Lin, C. C., Katal, A., and Wang, L. (2020). Dynamic forecast of cooling load and energy saving potential based on ensemble Kalman filter for an institutional high-rise building with hybrid ventilation. *Build. Simul.* 13 (6), 1259–1268. doi:10.1007/s12273-020-0665-7
- Lei, L., Shao, S. L., and Liang, L. X. (2024). An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction. *Energy* 288, 129795. doi:10.1016/j.energy.2023.129795
- Li, Y. C., Gao, F., Yu, J. Y., and Fei, T. (2025). Machine learning based thermal comfort prediction in office spaces: integrating SMOTE and SHAP methods. *Energy Build.* 329, 115267. doi:10.1016/j.enbuild.2024.115267
- Lundberg, S. M., Erion, G., Lee, S., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2 (1), 56–67. doi:10.1038/s42256-019-0138-9
- Lv, R., Yuan, Z., Lei, B., Zheng, J., and Luo, X. (2022). Building thermal load prediction using deep learning method considering time-shifting correlation in feature variables. *J. Build. Eng.* 61, 105316. doi:10.1016/j.jobte.2022.105316
- Manfren, M., Gonzalez-Carreón, K. M., and James, P.A.B. (2024). Interpretable data-driven methods for building energy modelling—a review of critical connections and gaps. *Energies* 17 (4), 881. doi:10.3390/en17040881
- Mawson, V. J., and Hughes, B. R. (2020). Thermal modelling of manufacturing processes and HVAC systems. *Energy* 204, 117984. doi:10.1016/j.energy.2020.117984
- Mi, J., Fan, L., Duan, X., and Qiu, Y. (2018). Short-term power load forecasting method based on improved exponential smoothing grey model. *Math. Problems Eng.* 2018, 3894723. doi:10.1155/2018/3894723
- Moveh, S., Merchán-Cruz, E. A., Ibrahim, A. O., Elhassan, Z. A. M., Ramadan Abdelhai, N. M., and Abdelrazig, M. D. (2025). Thermodynamic optimization of building HVAC systems through dynamic modeling and advanced machine learning. *Sustainability* 17 (5), 1955. doi:10.3390/su17051955
- Ozderem, O. C., Olaniyi, E. O., and Oyedotun, O. K. (2017). Short term load forecasting using particle swarm optimization neural network. *Procedia Comput. Sci.* 120, 382–393. doi:10.1016/j.procs.2017.11.254
- Shahcheraghian, A., Madani, H., and Ilinca, A. (2024). From white to black-box models: a review of simulation tools for building energy management and their application in consulting practices. *Energies* 17 (2), 376. doi:10.3390/en17020376
- Sogut, M. Z., and Mutlu, H. (2025). Assessment of new flow control strategy for the air handling units on energy management and sustainability. *Int. J. Thermofluids* 25, 101017. doi:10.1016/j.ijft.2024.101017
- Tao, Y. X., Yan, H. R., Gao, H., Sun, Y. Y., and Li, G. (2019). Application of SVR optimized by modified simulated annealing (MSA-SVR) air conditioning load prediction model. *J. Ind. Inf. Integr.* 15, 247–251. doi:10.1016/j.jii.2018.04.003
- Teixeira, B., Carvalhais, L., Pinto, T., and Vale, Z. (2025). Explainable AI framework for reliable and transparent automated energy management in buildings. *Energy Build.* 326, 116246. doi:10.1016/j.enbuild.2025.116246
- Verbeke, S., and Audenaert, A. (2018). Thermal inertia in buildings: a review of impacts across climate and building use. *Renew. Sustain. Energy Rev.* 82, 2300–2318. doi:10.1016/j.rser.2017.08.083
- Wang, D., Tan, D., and Liu, L. (2018). Particle swarm optimization algorithm: an overview. *Soft Comput.* 22 (2), 387–408. doi:10.1007/s00500-016-2474-6
- Wei, Y., Chen, H., Tian, Y., Wang, C., Hu, Y., Xie, Z., et al. (2026). QPSO-optimized VMD-SE-transformer-BiLSTM for short-term air conditioning load forecasting in industrial buildings. *AIP Adv.* 16 (2), 025123. doi:10.1063/5.0319827
- Wu, Y., Qian, C., and Huang, H. (2024). Enhanced air quality prediction using a coupled DVMD informer-CNN-LSTM model optimized with dung beetle algorithm. *Entropy* 26 (7), 534. doi:10.3390/e26070534
- Wu, Y., Jin, G., Li, X., Huang, H., Sheng, Z., and Liao, Q. (2025). A novel IDBO-VMD-iTransformer framework for air quality index prediction: multi-strategy optimization and environmental sustainability assessment. *Process Saf. Environ. Prot.* 193, 108379. doi:10.1016/j.psep.2025.108379
- Wu, J., Tan, J., Zhou, H., Li, J., and Gao, D. c. (2025). A two-stage deep learning framework for building cooling load prediction based on occupancy information. *J. Phys. Conf. Ser.* 3001 (1), 012004. doi:10.1088/1742-6596/3001/1/012004
- Wu, C., Li, X., Wang, Z., Xu, Z., Xu, C., Yang, Y., et al. (2025). Transfer learning based LSTM model for power consumption prediction of air source heat pumps with sparse data. *J. Build. Eng.* 97, 114525. doi:10.1016/j.jobte.2025.114525
- Xu, Y., Vahmani, P., Jones, A., and Hong, T. (2024). Anthropogenic heat from buildings in Los Angeles county: a simulation framework and assessment. *Sustain. Cities Soc.* 107, 105468. doi:10.1016/j.scs.2024.105468
- Xue, J., and Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst. Sci. Control Eng.* 8 (1), 22–34. doi:10.1080/21642583.2019.1708830
- Yan, D., O'Brien, W., Hong, T., Feng, X., Burak Gunay, H., Tahmasebi, F., et al. (2015). Occupant behavior modeling for building performance simulation: current state and future challenges. *Energy Build.* 107, 264–278. doi:10.1016/j.enbuild.2015.08.032
- Yan, Q., Lu, Z., Liu, H., He, X., Zhang, X., and Guo, J. (2023). An improved feature-time transformer encoder-Bi-LSTM for short-term forecasting of user-level integrated energy loads. *Energy Build.* 297, 113396. doi:10.1016/j.enbuild.2023.113396

- Yan, Y., Wang, F., Tian, C., Xue, W., Lin, L., Cao, W., et al. (2025). Transfer learning and source domain restructuring-based BiLSTM approach for building energy consumption prediction. *Int. J. Green Energy* 22 (3), 536–550. doi:10.1080/15435075.2024.2421328
- Yang, X., Zhou, G., Ren, Z., Qiao, Y., and Yi, J. (2024). High-precision air conditioning load forecasting model based on improved sparrow search algorithm. *J. Build. Eng.* 92, 109809. doi:10.1016/j.jobe.2024.109809
- Yang, L., Guo, J., Tian, H., Liu, M., Huang, C., and Cai, Y. (2025). Multi-scale building load forecasting without relying on weather forecast data: a temporal convolutional network, long short-term memory network, and self-attention mechanism approach. *Buildings* 15 (2), 298. doi:10.3390/buildings15020298
- Yu, J., Chang, W. S., and Dong, Y. (2022). Building energy prediction models and related uncertainties: a review. *Buildings* 12 (8), 1284. doi:10.3390/buildings12081284
- Zhang, L., and Chen, Z. (2024). Large language model-based interpretable machine learning control in building energy systems. *Energy Build.* 313, 114278. doi:10.1016/j.enbuild.2024.114278
- Zhang, C. B., Lu, J., Huang, J. H., and Zhao, Y. (2024a). End-to-end data-driven modeling framework for automated and trustworthy short-term building energy load forecasting. *Build. Simul.* 17 (8), 1419–1437. doi:10.1007/s12273-024-1149-y
- Zhang, C., Luo, Z., Rezgui, Y., and Zhao, T. (2024b). Enhancing building energy consumption prediction introducing novel occupant behavior models with sparrow search optimization and attention mechanisms: a case study for forty-five buildings in a university community. *Energy* 294, 130896. doi:10.1016/j.energy.2024.130896
- Zhou, M. R., Wang, L., Hu, F., Zhu, Z., Zhang, Q., Kong, W., et al. (2024). Residual ISSA-LSTM: a new data-driven method of heat load forecasting for building air conditioning. *Energy Build.* 321, 114698. doi:10.1016/j.enbuild.2024.114698