



ARTICLE

Grid-Supplied Load Prediction under Extreme Weather Conditions Based on CNN-BiLSTM-Attention Model with Transfer Learning

Qingliang Wang¹, Chengkai Liu¹, Zhaohui Zhou¹, Ye Han¹, Luebin Fang², Moxuan Zhao³ and Xiao Cao^{3,*}

¹Power Dispatch Control Center, State Grid Fuzhou Power Supply Company, State Grid Fujian Electric Power Co., Ltd., Fuzhou, 350013, China

²Science and Technology Office, State Grid Fuzhou Power Supply Company, State Grid Fujian Electric Power Co., Ltd., Fuzhou, 350013, China

³New Energy Research Center, China Electric Power Research Institute Co., Ltd., Nanjing, 530004, China

*Corresponding Author: Xiao Cao. Email: caoxiao@epri.sgcc.com.cn

Received: 21 May 2025; Accepted: 11 July 2025; Published: 27 October 2025

ABSTRACT: Grid-supplied load is the traditional load minus new energy generation, so grid-supplied load forecasting is challenged by uncertainties associated with the total energy demand and the energy generated off-grid. In addition, with the expansion of the power system and the increase in the frequency of extreme weather events, the difficulty of grid-supplied load forecasting is further exacerbated. Traditional statistical methods struggle to capture the dynamic characteristics of grid-supplied load, especially under extreme weather conditions. This paper proposes a novel grid-supplied load prediction model based on Convolutional Neural Network-Bidirectional LSTM-Attention mechanism (CNN-BiLSTM-Attention). The model utilizes transfer learning by pre-training on regular weather data and fine-tuning on extreme weather samples, aiming to improve prediction accuracy and robustness. Experimental results demonstrate that the proposed model outperforms traditional statistical methods and existing machine learning models. Through comprehensive experimental validation, the attention mechanism demonstrates exceptional capability in identifying and weighting critical temporal features across different timescales, which significantly contributes to enhanced prediction performance and stability under diverse weather conditions. Moreover, the proposed approach consistently exhibits strong generalization capabilities across multiple test cases when applied to different regional power grids with distinct operational patterns and varying load characteristics, showcasing its practical adaptability to real-world scenarios. This study provides a practical solution for enhancing grid-supplied load forecasting capabilities in the face of increasingly complex and unpredictable weather patterns.

KEYWORDS: Grid-supplied load forecasting; extreme weather; transfer learning

1 Introduction

With the expansion of the power system and the increasing frequency of extreme weather events, the impact of extreme weather on grid-supplied load is becoming increasingly significant. This places higher demands on the stable operation of the grid and load forecasting [1,2]. Grid-supplied load, defined as the total electric energy delivered by the grid to users via the public network, encompasses actual user power consumption and transmission losses, thereby reflecting the grid's supply capacity. In contrast, common load merely accounts for the total power consumed by user terminals, without distinguishing the power sources. Under extreme weather conditions such as typhoons, heavy rainfall, and extremely high temperatures,



common load fluctuations are transmitted through the distribution and transmission hierarchy, exacerbating the source-load spatial mismatch caused by cross-regional weather differences. This, in turn, affects the efficiency of grid scheduling and energy management, posing a threat to the stability of regional power supply and people's livelihood security [3,4]. To ensure that the power system can implement appropriate safety protection measures promptly when extreme weather events occur, it is crucial to predict these events in advance and provide accurate grid-supplied load prediction results [5].

Statistical forecasting methods, such as time series analysis and regression analysis, are commonly used [6,7]. However, with the ongoing advancement of science and technology and the refinement of forecasting theory, these statistical methods have struggled to meet the accuracy requirements of load forecasting [8–10]. Although the time series method can rapidly predict the continuous change in load over a future period using limited data, it is complex to model, demands relatively smooth original time series data, and is only appropriate for short-term predictions in scenarios with minor load fluctuations [11]. The regression analysis method has a straightforward principle, fast computation, is applicable to historical data, and can achieve a good fit [12]. Nevertheless, it is highly reliant on historical data and has a relatively simplistic model structure, making it unable to accurately address complex issues [13]. In general, the aforementioned methods have notable deficiencies in capturing the dynamically changing characteristics of grid-supplied loads and adapting to external factors like extreme weather conditions. This results in limitations in the accuracy and reliability of forecast results under such conditions.

Machine learning is rooted in rigorous mathematical theory, with algorithms trained on large datasets to generate predictions [14]. Unlike traditional statistical methods, machine learning can effectively model nonlinear input-output relationships, identify latent patterns in historical data, and through iterative optimization weight parameters improve prediction accuracy [15,16]. In machine learning applications, support vector machines (SVMs) and random forests (RFs) are frequently used for urban load forecasting [17,18]. SVMs can efficiently handle small sample sizes and manage nonlinear relationships and high-dimensional data [19]. However, SVMs are sensitive to parameter selection, perform poorly with highly volatile time series, and converge slowly when processing large datasets [20–22]. Random forests show strong resistance to overfitting, can process high-dimensional nonlinear data, and support parallel computation, although they have significant computational demands and limited interpretability [23,24]. Additionally, the machine learning-based approach has a limitation in modeling the dynamic change characteristics of grid-supplied loads. Especially when facing external disturbances like extreme weather conditions, the model's capacity to capture the time-varying characteristics of the system is constrained, making it difficult to accurately predict grid-supplied loads under such conditions.

Compared to traditional prediction models, deep learning models can better capture the complex interdependencies within data, thus improving the prediction accuracy and robustness of the models [25]. The current deep learning-based methods include Temporal Convolutional Networks (TCN), Recurrent Neural Networks (RNN), and Transformer architectures. Specifically, RNN-based models have enhanced the dynamic modeling of time series by introducing gating mechanisms or simplifying the architecture, significantly improving their capacity to model temporal dynamics [26–28]. TCNs utilize causal convolutional structures to facilitate the efficient extraction of temporal features. However, their performance heavily relies on large-scale training datasets; in situations with limited data, this reliance may result in a decrease in prediction performance due to inadequate pattern learning [29,30]. The Transformer model, characterized by its self-attention mechanism, is proficient in establishing long-range dependencies across sequential data. Nevertheless, when applied to power load data—which features high-frequency sampling and complex temporal-spatial correlations—it faces the challenge of exponentially increasing computational complexity, restricting its practical application in real-time prediction tasks [31,32]. Kim et al. use a transfer learning

approach to forecast solar photovoltaic power generation between different countries [33]. Jia et al. designed a model integrating CNN and LSTM for grid short-term load forecasting [34]. Monia et al. used the complex functionality of CNN-BiLSTM combined with Bayesian-based optimization techniques to accurately predict power system loads in real time [35]. Despite the outstanding performance of deep learning methods in load forecasting, existing models still have inherent limitations in representing the dynamic time-varying characteristics of grid-supplied loads and are not sufficiently responsive to exogenous variables such as extreme weather conditions. This leads to significant limitations in the accuracy and reliability of their forecasts under such conditions.

Although all of the aforementioned methods have yielded corresponding research results in the field of load forecasting, grid-supplied load forecasting is significantly influenced by complex factors [36–38], including dynamic time-varying characteristics and extreme weather conditions. Traditional forecasting models struggle to meet the requirements for high-precision and high-reliability grid-supplied load forecasting due to their inadequate ability to model nonlinear relationships, low sensitivity to external variables, and limited capacity to capture time-series features [39,40]. While a single deep learning algorithm excels in feature extraction and complex relationship modeling, it is susceptible to overfitting risks when dealing with a large number of input variables [41–43]. Moreover, its forecasting performance deteriorates significantly in data-scarce scenarios or under abnormal climate conditions. Additionally, existing research primarily focuses on conventional load forecasting and has not yet fully investigated the challenges of grid-supplied load forecasting under extreme weather conditions.

Therefore, this paper introduces a comprehensive methodology for incorporating extreme weather conditions into the process of predicting grid-supplied loads. Specifically, a new model is proposed to account for extreme weather impacts on electric grids, which utilizes meteorological factors and historical grid-supplied load data to identify extreme weather events. Furthermore, a model based on CNN-BiLSTM-Attention is developed to predict grid-supplied loads, employing a “pre-training-fine-tuning” strategy derived from transfer learning. This strategy involves pre-training the model using regular weather samples and subsequently fine-tuning the model parameters with extreme weather samples. Ultimately, the prediction model is adaptively employed, based on the identified extreme weather processes, to forecast the grid-supplied load.

2 Impact of Extreme Weather on Grid-Supplied Loads

2.1 Impact of Extreme Weather

Extreme weather events exert significant and distinct influences on grid-supplied load patterns, categorized primarily by their impact on demand characteristics. High-load weather events occur during periods of intense heat, for example, heat waves lead to an exponential surge in electricity demand due to the intensive use of refrigeration equipment, particularly during afternoon peak hours. Or heatwaves, as well as during cold snaps, heavy rain, snow, or freezing conditions. These conditions trigger a concentrated surge in customer-side electricity demand. Specifically, sustained high temperatures drive an exponential increase in consumption due to the widespread, intensive operation of cooling equipment, leading to pronounced peaks during afternoon hours. Conversely, cold snaps and related inter weather significantly elevate heating loads, resulting in overall steeper load curves characterized by heightened demand levels.

Low load weather events manifest under conditions of abnormal temperature deviations relative to seasonal norms, such as unusually warm winter periods or unexpectedly cool summer periods. These deviations cause a substantial reduction in traditional customer-side heating or cooling demand. The altered thermal environment leads to inverse fluctuations in temperature-sensitive loads, causing the typical daily

load curve to flatten considerably. This flattening often results in a single, less pronounced peak or the absence of a distinct peak pattern entirely, reflecting significantly suppressed overall consumption.

Meanwhile, sudden load change weather events are closely associated with rapid meteorological transitions, such as the onset of summer rainfall, typhoons, or intense cold snaps. These events induce dramatic, dynamic fluctuations in user-side load. The key driver is the instantaneous and severe shift in temperature, humidity, and other meteorological factors during such severe weather. This abrupt change provokes rapid responses from temperature-sensitive loads (e.g., air conditioning and heating systems), causing the grid-supplied load curve to exhibit jagged, high-frequency oscillations as demand adjusts sharply to the changing environment.

2.2 Defining Conditions for Extreme Weather

- (1) This is shown in Fig. 1, a regional load-equivalent meteorological condition index I is classified as the weather process of high load if it satisfies Eq. (1):

$$\begin{cases} I \geq I_{\beta 82lh} \text{ (Summer)} \\ I \leq I_{\beta 82zh} \text{ (Winter)} \end{cases} \quad (1)$$

where $I_{\beta 82lh}$ represents the regional load-equivalent meteorological condition index corresponding to the 82nd percentile of the temperature-sensitive load heating/cooling ratio. Here, β is the temperature—sensitive load heating/cooling ratio. It reflects the relative proportion of heating and cooling demands in the load due to temperature changes and is a key coefficient for measuring the relationship between temperature and heating/cooling demands of the load, β_{82zh} denotes the 82nd percentile value of historical β samples, and $I_{\beta 82zh}$ is the index value associated with β_{82zh} .

- (2) This is shown in Fig. 2, the weather process of low load is defined when the monthly average daily temperature (T) in a region deviates by $\geq 10\%$ from the climatological average for the same period:

$$\begin{cases} T \leq 0.9T_s \text{ (Summer)} \\ T \geq 1.1T_w \text{ (Winter)} \end{cases} \quad (2)$$

where T_s is the average summer perennial temperature and T_w is the average winter perennial temperature.

- (3) This is shown in Fig. 3, the weather process of sudden load changes is identified if the 4-h variation amplitude of the temperature-sensitive load heating/cooling ratio ($\Delta\beta$) satisfies Eq. (3):

$$\Delta\beta \geq \Delta\beta_{82th} \quad (3)$$

where $\Delta\beta$ is the 4-h variation amplitude of the temperature-sensitive load heating/cooling ratio. $\Delta\beta_{82th}$ represents the 82nd percentile of historical $\Delta\beta$ samples.

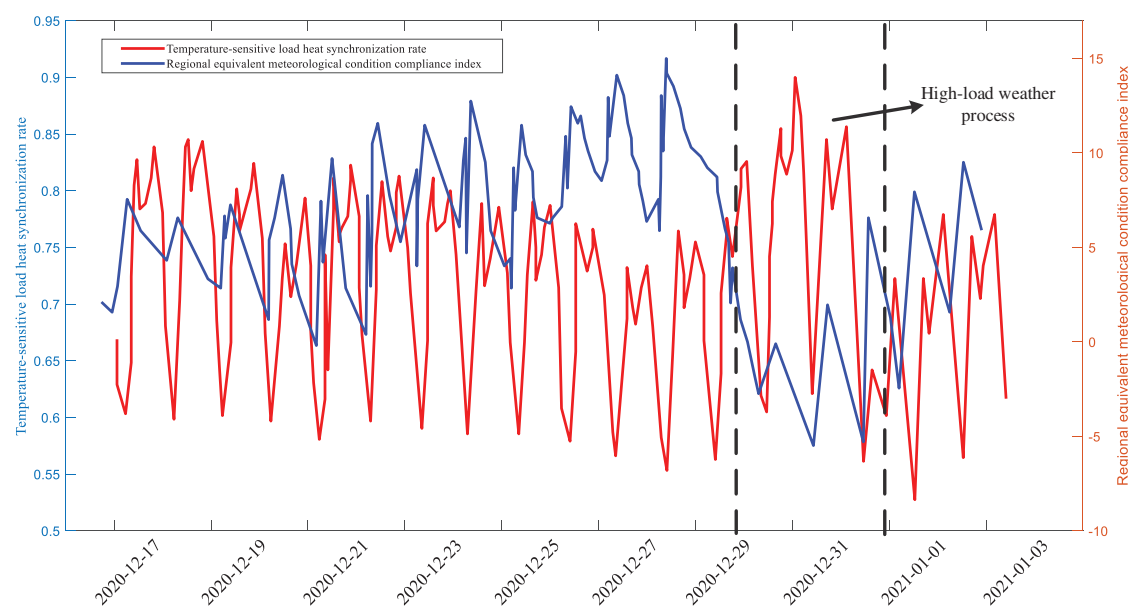


Figure 1: The weather process of high load

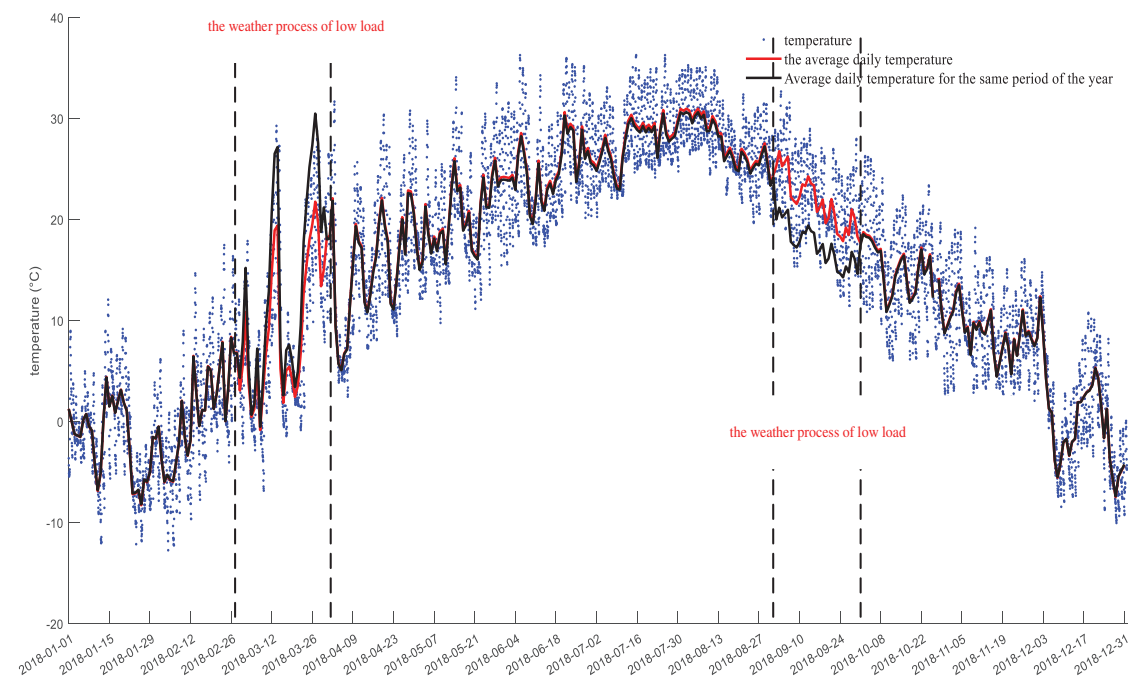


Figure 2: The weather process of low load

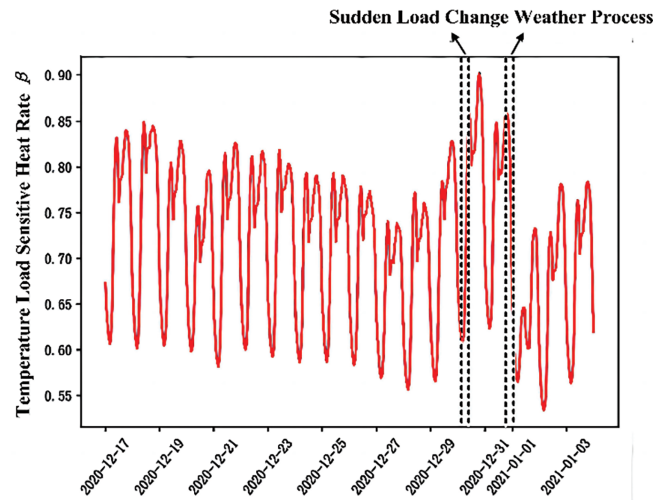


Figure 3: The weather process of sudden load

3 Construction of Grid-Supplied Load Forecasting Method under Extreme Weather Conditions

The overall process depicted in Fig. 4 begins with the collection of regular weather samples from the source domain, which serve as the foundation for pre-training an initial prediction model. Upon completion of the pre-training phase, a pre-trained prediction model is obtained. Subsequently, extreme weather samples from the target domain are utilized to fine-tune this pre-trained model, resulting in a fine-tuned prediction model. Following this, a decision-making mechanism is activated, which relies on a predefined standard for extreme weather. If the prevailing weather conditions meet this standard, the system will generate extreme weather forecast results. Conversely, if the conditions do not meet the standard, conventional weather forecast results will be generated.

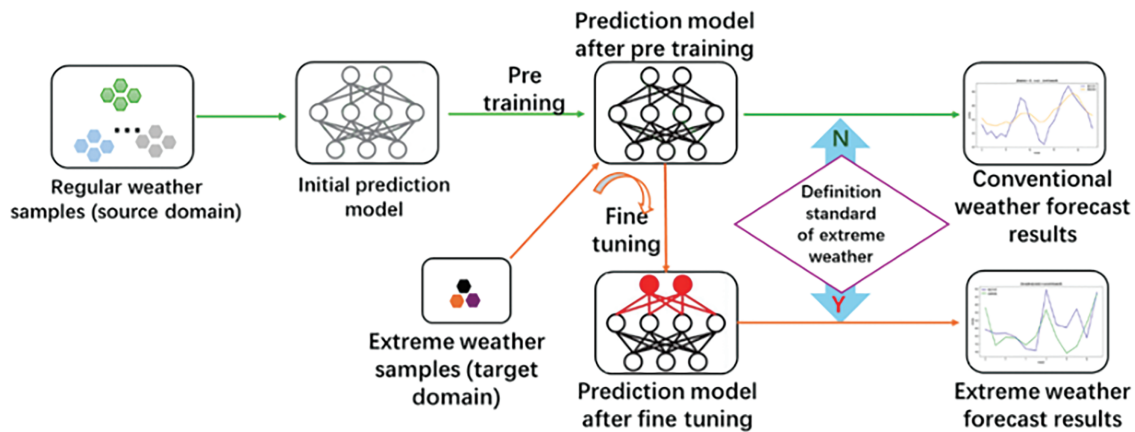


Figure 4: General structure

3.1 CNN-BiLSTM-Attention Structure

In order to predict the grid-supplied loads more accurately, this paper proposes a grid-supplied loads prediction model based on CNN-BiLSTM-Attention, and its algorithm flow is shown in Fig. 5.

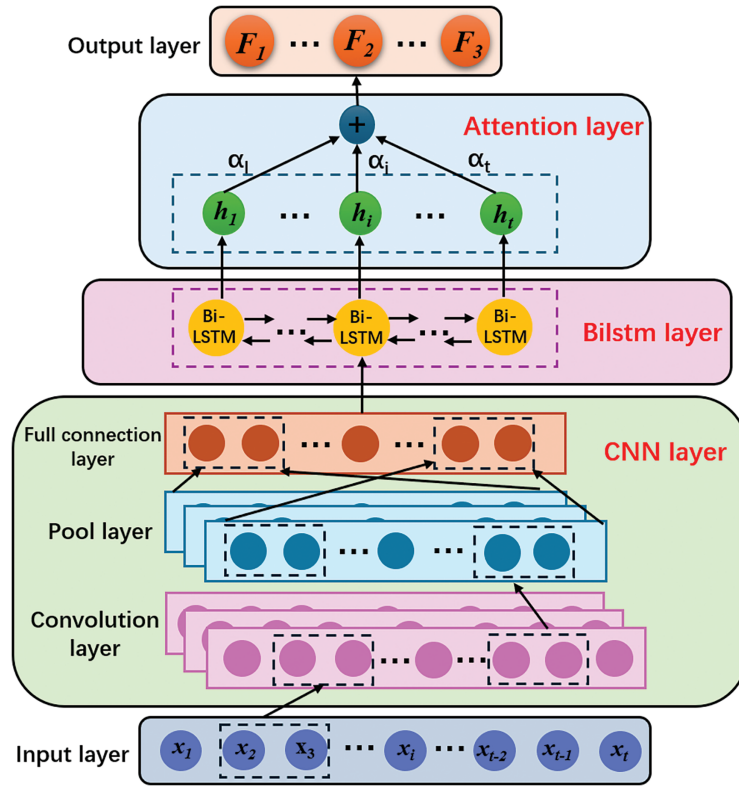


Figure 5: CNN-BiLSTM-Attention structure

As shown in Fig. 5, the proposed method first preprocesses the input data using Pearson correlation coefficient analysis in the input layer. The CNN then extracts features from the processed grid-supplied load data, which are subsequently fed into the BiLSTM network for training to capture intrinsic relationships within the data. The Attention layer dynamically assigns different weights to each feature, generating prediction values while evaluating their similarity with the time-series data. The attention scores are then normalized through a SoftMax function for weight calculation, ultimately producing the grid-supplied load forecasting results. This hierarchical architecture enables the model to effectively handle various sequential data processing tasks.

3.1.1 Convolutional Neural Network (CNN) Layer

The structure of Convolutional Neural Network (CNN) usually consists of an input layer, a convolutional layer, a pooling layer, a fully-connected layer, and an output layer, which are combined to achieve learning and prediction from complex data.

The convolutional layer extracts the features of the input data through a convolutional kernel. The size, step size and padding of the convolution kernel define the scope of the convolution operation and the size of the output feature map. The convolution operation generates a feature map by calculating the dot product of the convolution kernel weights and the input data, thus extracting local features of the input data. The generalized formula for the convolution layer is:

$$Z_{i,j} = \sum_{m,n} y_{i+m,j+n} * f_{m,n} \quad (4)$$

where $Z_{i,j}$ denotes the value of the output feature map, $y_{i+m,j+n}$ denotes the value of the input feature map, and $f_{m,n}$ denotes the value of the convolution kernel at position (m, n) .

3.1.2 Bidirectional Long Short-Term Memory Network (BiLSTM) Layer

The LSTM unit has 3 gates, namely the forgetting gate, the input gate and the output gate. The forgetting gate forgets a certain percentage of the past information; the input gate records part of the current moment's input information into the cell state; and the output gate encodes the hidden state vector and the cell state selectively as the input to the LSTM unit at the next moment.

The output of the current moment may be related not only to past information but also to future information. However, LSTM cannot encode information from backward to forward, whereas BiLSTM can better capture the effects of bidirectional sequences by reversing the time series, consisting of forward and backward LSTMs. The BiLSTM output expression is as follows:

$$h_t = \text{concat}(h_{tf}, h_{tb}) \quad (5)$$

where h_t denotes the hidden state vector of BiLSTM; *concat* denotes the splicing operation in the output dimension; h_{tf} , h_{tb} denote the hidden state vectors of forward and backward LSTM, respectively.

3.1.3 Attention Layer

Temporal Pattern Attention (TPA) mechanism is an improved attention mechanism specialized in dealing with multivariate time series forecasting problems. Different from traditional attention mechanisms, TPA is able to simultaneously extract complex nonlinear relationships between time series and correlations among different variables. The core idea is to extract features from the hidden state matrix through convolutional operations, so as to capture the interdependencies among multivariate variables. The working principle of TPA can be divided into several key steps: first, the original time series are processed using a bidirectional long short-term memory (BiLSTM) network to obtain the hidden state matrix. Next, multiple one-dimensional convolution kernels are applied to the row vectors of the hidden state matrix to extract the temporal pattern matrix. Each convolution kernel performs a convolution operation along the time step to capture the features of different time steps. Then, the attention weights are computed by the attention mechanism function and the temporal pattern matrix is weighted and summed to obtain an attention vector containing multivariate association information. Finally, the attention vector is linearly mapped to the hidden state vector of BiLSTM and then summed to obtain the final prediction value.

3.1.4 Model Feature Training

However, directly employing a CNN-BiLSTM-Attention architecture for forecasting grid-supplied load based solely on meteorological data reveals a significant degradation in predictive performance during specific temporal segments, particularly when load trends undergo transitions. The hourly distribution heatmap of actual grid-supplied load values and their corresponding predictions, illustrated in Fig. 6, unequivocally demonstrates substantial discrepancies in load value distributions during the periods of 5:00–7:00 a.m., 12:00–1:00 p.m., and 6:00–7:00 p.m. Consequently, a specialized modification of the prediction model is warranted. This paper augments the original model by incorporating time-segment dummy variables, explicitly modeling the systematic prediction errors observed during the aforementioned three periods. These dummy variables are introduced as additional regressors, enabling the CNN-BiLSTM-Attention model to learn the incremental impact of these specific time segments on load, thereby rectifying systematic overestimation or underestimation of predicted values.

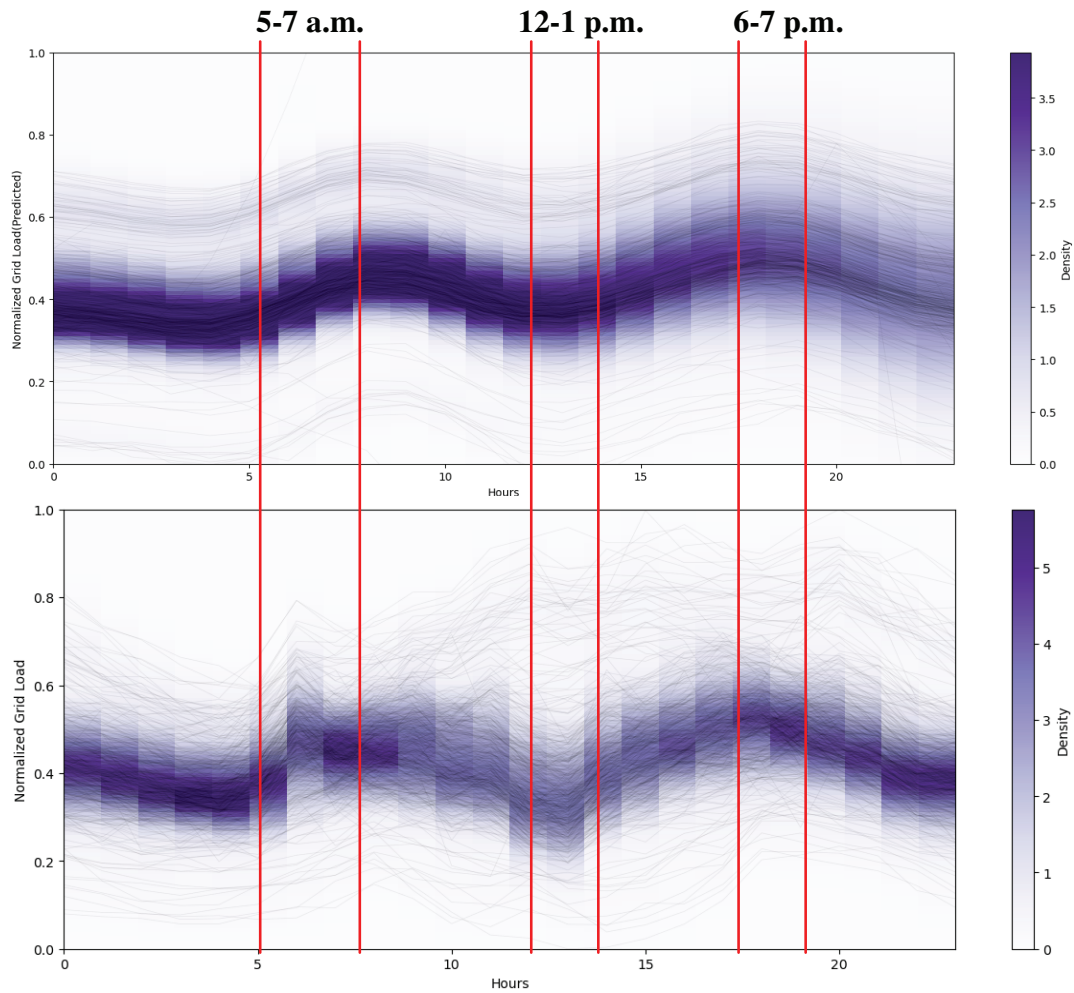


Figure 6: The hourly distribution heatmap of actual grid-supplied load values and their corresponding predictions

Therefore, the regression equation used in this paper is shown below.

$$\begin{aligned} \text{GridLoad} = & \alpha_0 + \alpha_1 \text{FLoad} + \alpha_2 \text{FH} + \alpha_3 \text{FT} + \alpha_4 \text{FWS} + \alpha_5 \text{FAP} \\ & + \alpha_6 \text{FP} + \alpha_7 \text{FV} + \sum_{j=2}^{24} \varphi_j \text{Hour}_j + \sum_{k=2}^7 \beta_k \text{Weekday}_k \\ & + \sum_{s=2}^{73} \gamma_s \text{Season}_s + \sum_{y=2015}^{2017} \delta_y \text{Year}_y + \sum_{i=1}^3 \sigma_i x_i(t) \end{aligned}$$

here, the Latin letters denote the respective regression coefficients for each corresponding term. FLoad represents the forecasted grid-supplied load. FH, FT, FWS, FAP, FP, and FV represent humidity, temperature, wind speed, atmospheric pressure, precipitation, and visibility, respectively. Hour, Weekday, Season, and Year represent the corresponding textual variables. Finally, $x(t)$ denotes the dummy variable for the i th time period, defined as follows:

$$x_i(t) = \begin{cases} 1, & \text{If time } t \text{ falls within the target, period} \\ 0, & \text{other situation} \end{cases}$$

3.2 Transfer Learning

Transfer learning, a machine learning paradigm, that leverages knowledge from a source task to enhance performance on a related but distinct target task, addresses critical limitations of traditional approaches. Unlike conventional methods that assume identical training-test data distributions, transfer learning enables effective adaptation to scenarios with scarce target domain samples or distribution shifts. The flow chart is shown in Fig. 7, by transferring features learned from abundant source domain data (e.g., “normal weather” patterns), this approach reduces dependence on limited target data while accelerating model convergence and strengthening generalization capabilities—particularly valuable for extreme weather forecasting where suddenness, nonlinearity, and data sparsity challenge traditional load prediction models.

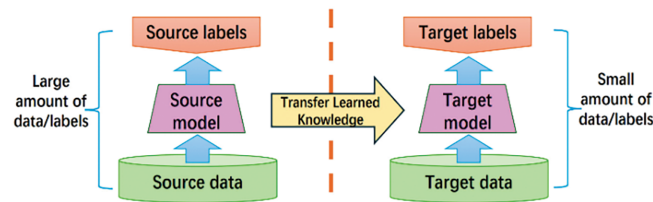


Figure 7: Transfer learning

To further ensure robustness during fine-tuning with scarce target samples, we implement a targeted strategy to mitigate overfitting risks. Specifically, instead of retraining the entire model, only high-level parameters (the Attention layer and latter layers of the BiLSTM network) are updated during adaptation. This selective refinement drastically reduces learnable parameters exposed to limited extreme weather data. Concurrently, we enforce a constant low learning rate (Section 3.2) to constrain update magnitudes, preserving universal spatiotemporal patterns from pre-training while adapting to extreme dynamics. Consequently, the transfer learning model achieves $MAE = 7.23$ and $R^2 = 0.8894$, substantially outperforming direct training on extreme data ($MAE = 10.79$, $R^2 = 0.7973$)—validating its efficacy in preventing overfitting and enhancing generalization under data scarcity. This methodology enhances prediction robustness by systematically combining historical pattern recognition with targeted adaptation to rare meteorological events.

4 Experiments and Analysis of Results

4.1 Data Collection and Preprocessing

This study utilized meteorological data and grid-supplied loads data from a specific location in the southeastern coast of China in 2018, as well as grid-supplied loads under extreme weather conditions. The dataset for pre-training on normal weather included approximately 300 days of hourly grid-supplied load (traditional load minus new energy generation), six-dimensional meteorological parameters (temperature, humidity, wind speed, air pressure, precipitation, visibility), and temporal features (hour, week, season). Normal weather was defined as periods without extreme weather triggers, covering typical load patterns in all four seasons. For fine-tuning on extreme weather, about 20 days of data were selected based on the regional load-equivalent meteorological index (Eqs. (1)–(3)), encompassing high-load weather (sustained $\geq 35^\circ\text{C}$ heatwaves), low-load weather (winter temperatures deviating $\geq 10\%$ from annual mean), and sudden load change weather (4-h temperature-sensitive load ratio variation $\geq 82\text{nd}$ percentile). Meteorological parameters and grid-supplied loads were normalized prior to model input to address data volatility, implemented in Python on an Intel Core i7-14700K platform.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

where $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of the load dataset, scaling values to the range $[0, 1]$.

Furthermore, we conducted a comprehensive analysis of the annual grid-supplied load data and meteorological records for 2018, identifying eight extreme weather days that satisfy the classification criteria outlined in Sections 2.1 and 2.2. These selected dates—specifically 14, 18, 21, 29, and 30 January; 1 March; and 1 and 31 December 2018—constitute our validation set for model verification. The January and December dates correspond to extreme cold wave events, while March 1 represents the unique regional meteorological phenomenon known in China as the return of the humid southerly winds.

4.2 Experimental Design and Evaluation Criteria

In order to evaluate the forecasting model performance and improve the training process, this paper uses several error evaluation criteria. Specifically, the RMSE is obtained by taking the square root of the average value of the squared differences for each data point. In addition, the gradient of the function related to the RMSE decreases as the error decreases, which promotes function convergence. The function can rapidly achieve its minimum value when the learning rate remains constant. However, when outliers exist within the sample, the RMSE assigns a higher weight to these outliers in the sense that the squared differences are averaged and then the square root is taken. The MAE measures the average absolute error between the predicted value and the true value. The smaller the value of MAE, the better the model performance. This is because a smaller MAE value means that the model predicts less error. The input value exhibits a stable gradient regardless of its magnitude, ensuring that it does not result in the gradient explosion issue. Therefore, it possesses a fairly robust solution. While the MAE curve maintains continuity, it lacks differentiability at $x = 0$. Additionally, the MAE gradient remains uniform in the majority of cases, leading to significant gradients for even minor loss values. This characteristic hampers function convergence and impedes model learning. The R-square (R^2) represents the portion of the variation in the dependent variable that can be explained by the model. It usually takes a value in the range of 0 to 1. The closer it is to 1, the better the model's ability to explain the data. And the predicted values built by the model fit the actual observations better.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (8)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (9)$$

4.3 Comparison of Experimental Results

4.3.1 Advanced CNN-BiLSTM-Attention Model Result Comparison

This section elaborates on the superior performance of the proposed advanced CNN-BiLSTM-Attention model in grid-supplied load forecasting, particularly under extreme weather conditions, as evidenced by Figs. 8 and 9, and Table 1. Across the eight extreme weather days used for validation (including 14, 18, 21, 29, 30 January; 1 March; and 1, 31 December 2018), the results consistently demonstrate a clear pattern. The “Corrected Load” curve from the improved model consistently aligns more closely with the “True Load” curve compared to the “Predicted Load” curve from the original model. This improvement is particularly

noticeable during periods of significant load fluctuations or pronounced peaks and troughs, which are characteristic of extreme weather events. For instance, in several subplots (e.g., 14 January, 21 January, 1 March) the original model's predictions ("Predicted Load") often exhibit noticeable deviations or delayed responses to rapid changes in demand. In contrast, the "Corrected Load" from the advanced model closely tracks the actual load, accurately capturing both the magnitude and timing of these dynamic shifts. These visual comparisons strongly attest to the significantly enhanced prediction accuracy and robustness of the advanced CNN-BiLSTM-Attention model when dealing with challenging extreme weather scenarios.

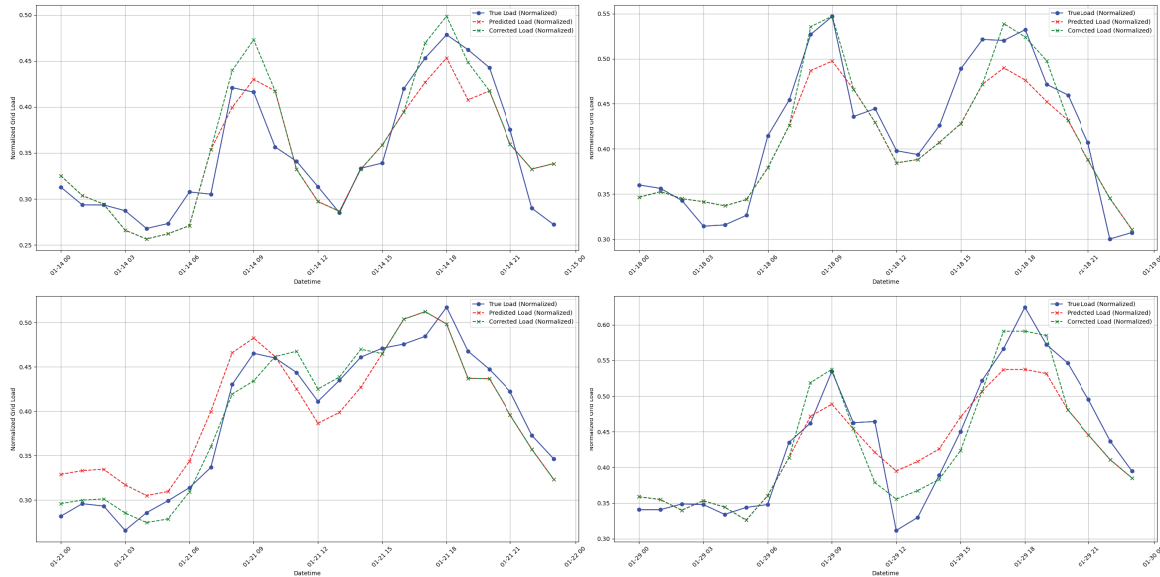


Figure 8: Comparison of grid-supplied load forecasts: original vs. modified CNN-BiLSTM (Part 1)

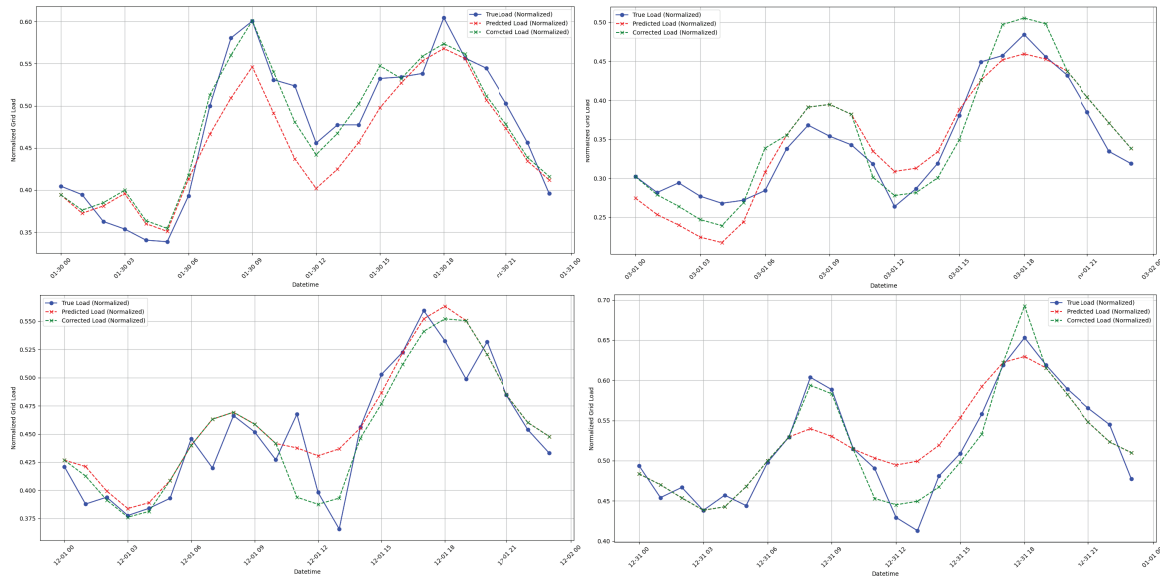


Figure 9: Comparison of grid-supplied load forecasts: original vs. modified CNN-BiLSTM (Part 2)

Table 1: Comparison of prediction error metrics between the original and modified models over eight days

		CNN-BiLSTM-attention	CNN-BiLSTM-attention (Advanced)
Case 1 (1-14)	MAE	0.0316	0.0255
	RMSE	0.0401	0.0329
	R²	80.4%	86.8%
Case 2 (1-18)	MAE	0.0247	0.0202
	RMSE	0.0340	0.0266
	R²	74.02%	84.09%
Case 3 (1-21)	MAE	0.0240	0.0234
	RMSE	0.0301	0.0297
	R²	79.69%	80.14%
Case 4 (1-29)	MAE	0.0263	0.0235
	RMSE	0.0301	0.0274
	R²	79.59%	83.12%
Case 5 (1-30)	MAE	0.0257	0.0212
	RMSE	0.0294	0.0242
	R²	79.14%	84.24%
Case 6 (3-1)	MAE	0.0262	0.0199
	RMSE	0.0300	0.0231
	R²	79.59%	85.12%
Case 7 (12-1)	MAE	0.0284	0.0229
	RMSE	0.0321	0.0283
	R²	78.56%	81.34%
Case 8 (12-31)	MAE	0.0301	0.0229
	RMSE	0.0375	0.0301
	R²	74.02%	80.45%

Table 1 quantifies the prediction performance of the original CNN-BiLSTM-Attention model vs. the CNN-BiLSTM-Attention (advanced) model across eight specific extreme weather dates. For all eight cases, the “CNN-BiLSTM-Attention (advanced)” model consistently outperforms the original “CNN-BiLSTM-Attention” model. Specifically, the advanced model consistently yields lower MAE values, indicating a smaller average absolute difference between the predicted and actual loads. For example, in Case 1 (14 January), the MAE decreased from 0.0316 to 0.0255. Similar reductions are observed across all cases, with the advanced model achieving a lowest MAE of 0.0199 (Case 6, 1 March) compared to 0.0262 for the original model on the same day. Moreover, the R^2 values for the advanced model are consistently higher, indicating that a larger proportion of the variance in grid-supplied load can be explained by the model. For example, in Case 1, the R^2 increased from 80.4% to 86.8%. The advanced model achieves R^2 values above 80% in most cases, reaching up to 86.8%, which demonstrates its superior ability to fit and explain the data compared to the original model, which typically ranged from 74% to 80%.

Furthermore, in terms of training time comparison, although the CNN-BiLSTM-Attention model requires longer training time than the basic model, it is relatively shorter compared to the Transformer model. We conducted a training duration comparison between the CNN-BiLSTM-Attention model and the Transformer model. The CNN-BiLSTM-Attention model averages 1.252 s per epoch, while the Transformer

model averages 2.173 s per epoch. Additionally, thanks to the early stopping mechanism applied in the CNN-BiLSTM-Attention model, its training steps are also shorter than those of the Transformer model. Moreover, this paper proposes transfer training for extreme weather conditions, which can avoid the excessively long training time caused by model retraining.

4.3.2 Comparison of Prediction Models for Extreme Weather

This control experiment also used CNN bilstm attention as the initial model, first training the source domain and then fine-tuning the parameters in the target domain. Finally, it was compared with direct prediction in the target domain. The experiment proved that transfer learning has significant advantages in the application of extreme weather conditions in the grid-supplied load. By utilizing a large amount of data under normal weather conditions for source domain training, the model can learn effective feature representations. Under extreme weather conditions, transfer learning not only effectively solves the problem of data scarcity by fine-tuning pre trained models, but also significantly improves the prediction accuracy and interpretability of the models. Compared with models trained directly on extreme weather data, transfer learning performs better in error metrics such as MAE, RMSE, and MAPE, and has higher R^2 values (Table 2, Figs. 10–12), indicating its stronger ability to fit and interpret data. Although transfer learning performs slightly better than source domain training on certain error metrics, its overall performance in the target domain is superior to direct training, especially in situations where data is scarce. Therefore, transfer learning is an effective method to solve the problem of extreme weather prediction for grid-supplied load, which can improve the robustness and reliability of the model.

Table 2: Comparison of error metrics and R^2 values for different training approaches

	Source domain performance	Target domain (Transfer learning)	Target domain (Direct training)
MAE	3.94	7.23	10.79
RMSE	6.26	8.75	11.85
MAPE	1.55%	2.84%	3.96%
R^2	0.9013	0.8894	0.7973

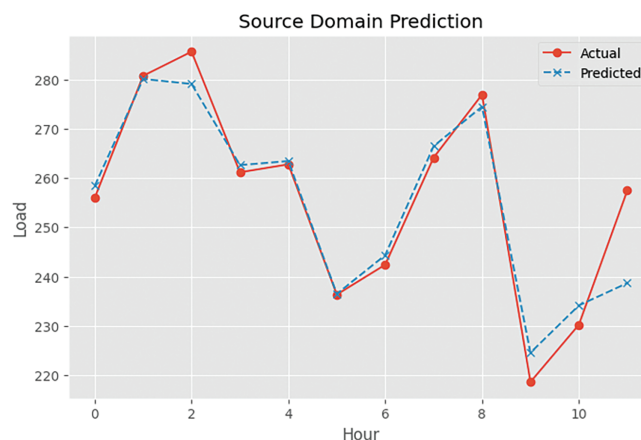


Figure 10: Source domain prediction

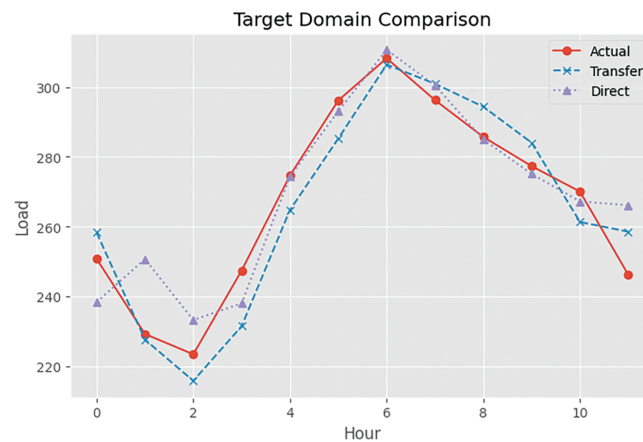


Figure 11: Target domain comparison

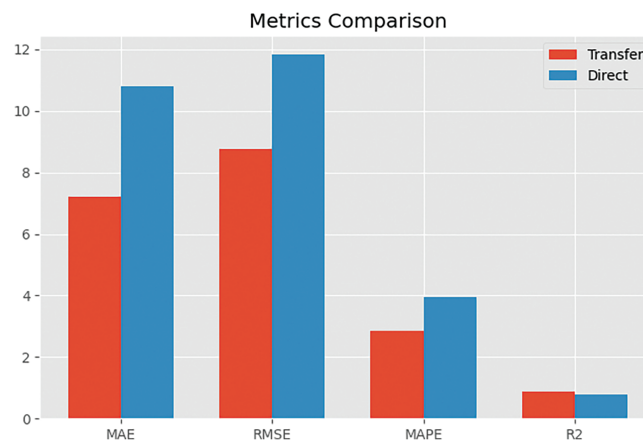


Figure 12: Metrics comparison

As illustrated in Table 2, the transfer learning strategy exhibits remarkable advantages in the target domain (extreme weather scenarios): compared with direct training on extreme weather data, the transfer learning model achieves a 33.0% reduction in MAE (from 10.79 to 7.23), a 26.2% decrease in RMSE (from 11.85 to 8.75), and a 28.3% drop in MAPE (from 3.96% to 2.84%), while the R^2 value increases by 11.6% (from 0.7973 to 0.8894). These metrics clearly demonstrate that transfer learning effectively mitigates the negative impact of scarce extreme weather samples on model performance. Fig. 10 presents the source domain prediction results, where the predicted values closely align with the actual grid-supplied load under regular weather, with a high R^2 of 0.9013, indicating that the pre-trained model has learned robust general spatiotemporal patterns from abundant normal weather data, laying a solid foundation for subsequent fine-tuning. In the target domain comparison (Fig. 11), the prediction curve of the transfer learning model shows a tighter fit to the actual load, especially during periods of sharp load fluctuations caused by extreme weather, where it accurately captures the magnitude and timing of load changes, whereas the directly trained model exhibits larger deviations and delayed responses to such fluctuations. The metrics comparison in Fig. 12 further visualizes these improvements: the bar heights for MAE, RMSE, and MAPE are significantly lower for transfer learning than for direct training, while the R^2 bar is notably higher, providing intuitive evidence for

the superiority of the transfer learning strategy in balancing prediction accuracy and generalization under extreme weather conditions.

5 Conclusion

In conclusion, given the inadequate accuracy of grid-supplied load forecasting in an expanding power system characterized by increasingly frequent extreme weather conditions, this paper proposes a CNN-BiLSTM-Attention model integrated with transfer learning. The model leverages pre-training on approximately 300 days of normal weather data and fine-tuning on 20 days of extreme weather samples from a region in the southeastern coast of China, where data were strictly screened based on regional load-equivalent meteorological indices to cover high-load, low-load, and sudden load change scenarios. Experimental results demonstrate superior accuracy over traditional methods, with significant improvements in MAE, RMSE, MAPE, and R^2 metrics. The regional case study, contextualized to the southeastern coast's typhoon-prone climate and subtropical load dynamics, highlights the model's adaptability to climate-specific patterns, with multivariate meteorological inputs (including temperature, humidity, visibility, etc.) ensuring comprehensive capture of weather-load interactions. The regional case study in Fujian Province, contextualized to its typhoon-prone climate, highlights the model's adaptability to climate-specific load dynamics. Multivariate meteorological inputs ensure comprehensive capture of weather-load interactions, while the industry-standard definition of grid-supplied load enhances practical relevance. Building on the validated transfer learning strategy, future research may expand to cross-regional datasets and integrate off-grid energy dynamics to further enhance generalizability. This work advances both methodological innovation and operational resilience for power systems in evolving climatic landscapes.

Acknowledgement: The authors would like to thank the China Electric Power Research Institute and the State Grid Company Limited for providing useful infrastructure and support for carrying out various studies related to this paper.

Funding Statement: The research is supported by the Science and Technology Project of State Grid Fujian Electric Power Co., Ltd. (Project No. B31300240001), with the project title “Research on Key Technologies for Load Forecasting and Regulation Capability Evaluation of Regional Power Grid Taking into Account Wide Area Distributed New Energy Access”.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Qingliang Wang and Zhaohui Zhou; data curation, Zhaohui Zhou and Ye Han; methodology, Chengkai Liu; software, Chengkai Liu; validation, Chengkai Liu; formal analysis, Chengkai Liu; investigation, Luebin Fang and Moxuan Zhao; writing—original draft preparation, Xiao Cao; writing—review and editing, Qingliang Wang, Chengkai Liu, Zhaohui Zhou, Ye Han, Luebin Fang, Moxuan Zhao and Xiao Cao; visualization, Chengkai Liu; supervision, Qingliang Wang and Zhaohui Zhou. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Mansouri A, Abolmasoumi AH, Ghadimi AA. Weather sensitive short term load forecasting using dynamic mode decomposition with control. *Electr Power Syst Res.* 2023;221(4):109387. doi:10.1016/j.epsr.2023.109387.
2. Trakas DN, Hatziaargyriou ND. Resilience constrained day-ahead unit commitment under extreme weather events. *IEEE Trans Power Syst.* 2020;35(2):1242–53. doi:10.1109/tpwrs.2019.2945107.

3. Yu GZ, Lu L, Tang B, Wang SY, Chung CY. Ultra-short-term wind power subsection forecasting method based on extreme weather. *IEEE Trans Power Syst.* 2023;38(6):5045–56. doi:10.1109/tpwrs.2022.3224557.
4. Chen J, Cheng Z, Wang H, Huangfu C, Kong X, Lu Z. Coordinated optimal operation of integrated electrical and transportation network considering source-load uncertainties in severe weather scenarios. *Sustain Energy Grids Netw.* 2024;39(4):101401. doi:10.1016/j.segan.2024.101401.
5. Deng X, Ye A, Zhong J, Xu D, Yang W, Song Z, et al. Bagging-XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Rep.* 2022;8(11):8661–74. doi:10.1016/j.egy.2022.06.072.
6. Kim N, Park H, Lee J, Choi JK. Short-term electrical load forecasting with multidimensional feature extraction. *IEEE Trans Smart Grid.* 2022;13(4):2999–3013. doi:10.1109/tsg.2022.3158387.
7. Jiang L, Wang X, Li W, Wang L, Yin X, Jia L. Hybrid multitask multi-information fusion deep learning for household short-term load forecasting. *IEEE Trans Smart Grid.* 2021;12(6):5362–72. doi:10.1109/TSG.2021.3091469.
8. Guo Y, Li Y, Qiao X, Zhang Z, Zhou W, Mei Y, et al. BiLSTM multitask learning-based combined load forecasting considering the loads coupling relationship for multienergy system. *IEEE Trans Smart Grid.* 2022;13(5):3481–92. doi:10.1109/tsg.2022.3173964.
9. Raza MQ, Mithulananthan N, Li J, Lee KY. Multivariate ensemble forecast framework for demand prediction of anomalous days. *IEEE Trans Sustain Energy.* 2020;11(1):27–36. doi:10.1109/tste.2018.2883393.
10. Waheed W, Xu Q, Aurangzeb M, Iqbal S, Dar SH, Elbarbary ZS. Empowering data-driven load forecasting by leveraging long short-term memory recurrent neural networks. *Heliyon.* 2024;10(24):e40934. doi:10.1016/j.heliyon.2024.e40934.
11. Kong X, Li C, Zheng F, Wang C. Improved deep belief network for short-term load forecasting considering demand-side management. *IEEE Trans Power Syst.* 2020;35(2):1531–8. doi:10.1109/tpwrs.2019.2943972.
12. Cheng L, Zang H, Xu Y, Wei Z, Sun G. Probabilistic residential load forecasting based on micrometeorological data and customer consumption pattern. *IEEE Trans Power Syst.* 2021;36(4):3762–75. doi:10.1109/TPWRS.2021.3051684.
13. Cao Z, Wan C, Zhang Z, Li F, Song Y. Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting. *IEEE Trans Power Syst.* 2020;35(3):1881–97. doi:10.1109/tpwrs.2019.2946701.
14. Mahadevkar SV, Khemani B, Patil S, Kotecha K, Vora DR, Abraham A, et al. A review on machine learning styles in computer vision—techniques and future directions. *IEEE Access.* 2022;10(1):107293–329. doi:10.1109/access.2022.3209825.
15. Gaboitaolelwe J, Zungeru AM, Yahya A, Lebekwe CK, Vinod DN, Salau AO. Machine learning based solar photovoltaic power forecasting: a review and comparison. *IEEE Access.* 2023;11:40820–45. doi:10.1109/access.2023.3270041.
16. Mestav KR, Wang X, Tong L. A deep learning approach to anomaly sequence detection for high-resolution monitoring of power systems. *IEEE Trans Power Syst.* 2023;38(1):4–13. doi:10.1109/TPWRS.2022.3168529.
17. Wang Z, Ku Y, Liu J. The power load forecasting model of combined SaDE-ELM and FA-CAWOA-SVM based on CSSA. *IEEE Access.* 2024;12:41870–82. doi:10.1109/access.2024.3377097.
18. Bareth R, Kochar M, Yadav A. Comparative analysis of different machine learning models for load forecasting. In: *IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT)*; 2023 Mar 11–12; Male, Maldives. doi:10.1109/GlobConHT56829.2023.10087406.
19. Liao R, Liao J, Zhang Y, Li G, Li X, Wu Y. Sag prediction of high-voltage transmission lines based on PSO-SVM. In: *2020 35th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*; 2020 Oct 16–18; Zhanjiang, China. p. 772–7. doi:10.1109/yac51587.2020.9337698.
20. He Y, Shi C, Guo X, He W, Han T. Photovoltaic power prediction algorithm based on parameter optimization of multi-kernel SVM. *Acta Energiæ Solaris Sin.* 2024;45(9):394–404. doi:10.19912/j.0254-0096.tynxb.2023-0826.
21. VanDeventer W, Jamei E, Thirunavukkarasu GS, Seyedmahmoudian M, Soon TK, Horan B, et al. Short-term PV power forecasting using hybrid GASVM technique. *Renew Energy.* 2019;140(7386):367–79. doi:10.1016/j.renene.2019.02.087.

22. Pan M, Li C, Gao R, Huang Y, You H, Gu T, et al. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization. *J Clean Prod.* 2020;277(1):123948. doi:10.1016/j.jclepro.2020.123948.
23. Hao J, Zhu C, Guo X. Wind power short-term forecasting model based on the hierarchical output power and poisson re-sampling random forest algorithm. *IEEE Access.* 2020;9:6478–87. doi:10.1109/access.2020.3048382.
24. Madhu Malini MK, Iswariya B, Prasad H, Sudhakar TD. Load forecasting using random forest regression algorithm in machine learning. In: 2024 International Conference on Science Technology Engineering and Management (ICSTEM); 2024 Apr 26–27; Coimbatore, India. doi:10.1109/ICSTEM61137.2024.10560982.
25. Eren Y, Küçükdemiral İ. A comprehensive review on deep learning approaches for short-term load forecasting. *Renew Sustain Energy Rev.* 2024;189(1):114031. doi:10.1016/j.rser.2023.114031.
26. Bento PMR, Pombo JAN, Calado MRA, Mariano SJPS. Stacking ensemble methodology using deep learning and ARIMA models for short-term load forecasting. *Energies.* 2021;14(21):7378. doi:10.3390/en14217378.
27. Wang T, Lai CS, Ng WWY, Pan K, Zhang M, Vaccaro A, et al. Deep autoencoder with localized stochastic sensitivity for short-term load forecasting. *Int J Electr Power Energy Syst.* 2021;130(11):106954. doi:10.1016/j.ijepes.2021.106954.
28. Somu N, Raman MRG, Ramamritham K. A deep learning framework for building energy consumption forecast. *Renew Sustain Energy Rev.* 2021;137:110591. doi:10.1016/j.rser.2020.110591.
29. Bedi J, Toshniwal D. Deep learning framework to forecast electricity demand. *Appl Energy.* 2019;238(9):1312–26. doi:10.1016/j.apenergy.2019.01.113.
30. Yaprakdal F. An ensemble deep-learning-based model for hour-ahead load forecasting with a feature selection approach: a comparative study with state-of-the-art methods. *Energies.* 2023;16(1):57. doi:10.3390/en16010057.
31. Wang C, Wang Y, Ding Z, Zheng T, Hu J, Zhang K. A transformer-based method of multienergy load forecasting in integrated energy system. *IEEE Trans Smart Grid.* 2022;13(4):2703–14. doi:10.1109/TSG.2022.3166600.
32. L'Heureux A, Grolinger K, Capretz MAM. Transformer-based model for electrical load forecasting. *Energies.* 2022;15(14):4993. doi:10.3390/en15144993.
33. Jiao X, Li X, Lin D, Xiao W. A graph neural network based deep learning predictor for spatio-temporal group solar irradiance forecasting. *IEEE Trans Ind Inf.* 2022;18(9):6142–9. doi:10.1109/tii.2021.3133289.
34. Jia L, Li G, Zhang Z, Wang Y, Sun Y, Li S. Deep learning-based short-term load forecasting for power grids. In: 2024 4th International Conference on Energy, Power and Electrical Engineering (EPEE); 2024 Sep 20–22; Wuhan, China. p. 188–91. doi:10.1109/EPEE63731.2024.10875100.
35. Bartouli M, Helali A, Hassen F. Applying Bayesian optimized CNN-BiLSTM to real-time load forecasting model for smart grids. In: 2024 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET); 2024 Apr 27–29; Hammamet, Tunisia. doi:10.1109/IC_ASET61847.2024.10596257.
36. Huang J, Yang M, Liu Y, Chen W, Rong F. Multi-scale dilated convolutional residual network based model for short-term load forecasting. In: 2024 3rd Asian Conference on Frontiers of Power and Energy (ACFPE); 2024 Oct 25–27; Chengdu, China. p. 374–8. doi:10.1109/ACFPE63443.2024.10801096.
37. Priyadharshini B, Ganapathy V, Sudhakara P. An optimal model to meet the hourly peak demands of a specific region with solar, wind, and grid supplies. *IEEE Access.* 2020;8:13179–94. doi:10.1109/access.2020.2966021.
38. Biswal B, Deb S, Datta S, Ustun TS, Cali U. Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques. *Energy Rep.* 2024;12:3654–70. doi:10.1016/j.egyr.2024.09.056.
39. Wan C, Cao Z, Lee WJ, Song Y, Ju P. An adaptive ensemble data driven approach for nonparametric probabilistic forecasting of electricity load. *IEEE Trans Smart Grid.* 2021;12(6):5396–408. doi:10.1109/tsg.2021.3101672.
40. Papalexopoulos AD, Hesterberg TC. A regression-based approach to short-term system load forecasting. *IEEE Trans Power Syst.* 1990;5(4):1535–47. doi:10.1109/59.99410.
41. Jin X, Dong Y, Wu J, Wang J. An improved combined forecasting method for electric power load based on autoregressive integrated moving average model. In: 2010 International Conference of Information Science and Management Engineering; 2010 Aug 7–8; Shaanxi, China. p. 476–80. doi:10.1109/ISME.2010.124.
42. Amjady N, Keynia F, Zareipour H. Short-term load forecast of microgrids by a new bilevel prediction strategy. *IEEE Trans Smart Grid.* 2010;1(3):286–94. doi:10.1109/tsg.2010.2078842.
43. Teeraratkul T, O'Neill D, Lall S. Shape-based approach to household electric load curve clustering and prediction. *IEEE Trans Smart Grid.* 2018;9(5):5196–206. doi:10.1109/tsg.2017.2683461.