

Article

A Method for Load Forecasting of Distribution Transformers Based on Parameter-Efficient Fine-Tuning of Large Language Models in Power System

Weijian Zhang ^{1,*}, Shanfeng Liu ^{2,*}, Miaomiao Li ², Hua Bao ³, Shaoguang Yuan ² and Xiawei Cheng ²

¹ State Grid Henan Electric Power Company, Zhengzhou 450052, China

² Electrical Power Research Institute of Henan Electric Power Corporation, Zhengzhou 450052, China; limiaomiao2599@163.com (M.L.); yuansg091005@foxmail.com (S.Y.); chengxiawei06@126.com (X.C.)

³ School of Artificial Intelligence, Anhui University, Hefei 230039, China; baohua@ahu.edu.cn

* Correspondence: zhangweijian2051@163.com (W.Z.); 13838513134@163.com (S.L.)

Abstract

Distribution grids are essential to power systems, as they ensure the safe and reliable operation of power systems. It is important to forecast the operational status of each distribution transformer. Consequently, there is an urgent need to develop load forecasting models that exhibit robust generalization and exceptional accuracy. However, distribution transformers are numerous, exhibit strong inherent features. Single-device load samples are scarce, with widespread few-shot issues. Traditional models rely on massive amounts of data and computational resources, making it challenging to balance computational costs, generalization ability, and accuracy. To address these issues, this paper analyzes the inherent features of distribution transformers and fine-tunes large language models using parameter-efficient fine-tuning methods to reduce training costs while preserving their generalization capabilities and accuracy. Experimental results demonstrate that this method achieves significantly higher accuracy than baseline models in few-shot scenarios.

Keywords: load forecasting for distribution transformers; few-shot learning; parameter-efficient fine-tuning methods; large language models



Academic Editors: Franco F. Yanine and Sarat Kumar Sahoo

Received: 27 July 2025

Revised: 20 August 2025

Accepted: 25 August 2025

Published: 18 September 2025

Citation: Zhang, W.; Liu, S.; Li, M.; Bao, H.; Yuan, S.; Cheng, X. A Method for Load Forecasting of Distribution Transformers Based on Parameter-Efficient Fine-Tuning of Large Language Models in Power System. *Processes* **2025**, *13*, 2986. <https://doi.org/10.3390/pr13092986>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a critical component of the power system, the distribution network covers a large area and is responsible for the important task of power transmission [1,2]. Among its components, distribution transformers serve as key hubs in the distribution network [3]. Their safe and stable operation directly impacts the reliability of the power supply. Accurately predicting the operational status of each distribution transformer and proactively identifying potential hazards is crucial for ensuring the safe and reliable operation of the distribution network [4,5].

Distribution transformers are numerous, widely distributed, and operate in complex environmental conditions [6–9]. When the power load exceeds the rated capacity of the distribution transformer, the internal temperature of the transformer rises rapidly, increasing its lifespan loss, reducing operational reliability, and posing a threat to the safe and reliable power supply of the distribution network [10–12]. Researchers have focused on power load forecasting to prevent overloading events and further optimize grid dispatch strategies. Traditional power load forecasting methods primarily include time series analysis [13], regression analysis [14], Kalman filtering [15], exponential smoothing [16], and

gray forecasting methods [17]. Traditional power load forecasting methods are constructed based on the statistical patterns of historical load data, offering advantages such as low computational cost and high interpretability. They have been widely applied in regional-level and system-level power load forecasting. However, the power load of distribution transformers is highly influenced by user electricity consumption habits, and the nonlinear characteristics of distribution transformer power loads are significantly more pronounced than those at regional and system levels. This results in a significant reduction in the forecasting accuracy of traditional methods when forecasting the power load of distribution transformers [18,19].

With the continuous improvement of computer performance, artificial intelligence (AI) technology has gradually emerged in multiple fields [20]. In recent years, AI technology has become the primary research method in power load forecasting. AI technology primarily includes methods such as support vector machines, random forests, and deep learning. Among these, deep learning methods, with their unique model structure and excellent nonlinear feature extraction capabilities, have gradually become a research hotspot among scholars in recent years and achieved breakthrough progress in multiple fields [21]. In recent years, load forecasting methods for distribution transformers have primarily focused on improvements to Long Short-Term Memory (LSTM), Transformer, and Convolutional Neural Network (CNN) models, as well as their integration with other models. As an improved version of Recurrent Neural Networks (RNNs), LSTM effectively addresses the gradient vanishing and explosion issues of traditional RNNs through special gating mechanisms (such as forget gate, input gate, and output gate) [22]. Specifically, the forget gate filters out redundant temporal features to retain key information. The input gate updates the parameters of memory units. The output gate integrates the current input with the memory state to selectively generate prediction results, thereby avoiding training instability caused by parameter sharing. To fully exploit the temporal features in historical load data, the reference [23] further extended this framework by constructing a Bi-directional Long Short-Term Memory (Bi-LSTM) model, which includes two independent LSTM units: one processes data in chronological order, and the other processes data in reverse order, thereby comprehensively capturing the temporal patterns in load data. To reduce the computational cost of the model and improve its prediction efficiency, the reference [24] constructed a Gate Recurrent Unit (GRU) model, a simplified variant of LSTM, which contains only two components: a reset gate and an update gate. The update gate regulates the relationship between the current input and historical information, controlling the degree of information retention. The reset gate determines whether to ignore certain historical features and adjusts their influence on the current state. This design enables GRU to dynamically adjust its memory and update processes, thereby better adapting to different time series data. The above methods use a recursive structure and must be calculated sequentially. To solve this problem, the reference [25] constructs a Temporal Convolutional Network (TCN) model based on the CNN network structure. This model comprises convolutional layers and residual connections, which can flexibly control the size of the receptive field by adjusting the network depth to extract features of different scales from historical load data. To fully consider the impact of external factors such as weather on power load, the reference [26] employs a Transformer model to construct a load forecasting model. The Transformer model fundamentally redefines traditional sequence processing architectures by abandoning the inherent frameworks of CNN and LSTM, instead relying entirely on self-attention mechanisms to model global dependencies within temporal data. The core of this model lies in the self-attention mechanism, which dynamically assigns differentiated weights to each element in the input sequence. It also quantifies the dependencies between different positions. Building upon this, the model integrates multiple self-attention modules through

a multi-head self-attention mechanism, enabling parallel analysis of raw data features across multiple subspaces. This not only enhances feature expression capabilities but also significantly improves prediction accuracy.

The above methods rely heavily on massive training data and feature engineering during training, and the load forecasting models are often limited to a single objective [27,28]. However, there are numerous distribution transformers, and their installation environments vary greatly, as do load sequences, which exhibit high heterogeneity. Missing data can lead to overfitting or underfitting of prediction models, and the above methods experience a sharp decline in prediction accuracy under few-shot conditions [29,30]. To address these issues, scholars have explored transfer learning algorithms to improve the accuracy of models in few-shot scenarios [31–33]. The reference [34] employs multiple kernel learning to extract the spatiotemporal correlation of power loads, thereby reducing the complexity of the prediction model and significantly alleviating the overfitting problem in few-shot scenarios. The reference [35] quantifies the spatiotemporal correlation of power load by sharing model parameters and fully leveraging prior knowledge, enabling the model to achieve effective transfer in few-shot scenarios. However, the aforementioned methods require a high degree of similarity between the source domain features and the target domain features [36,37]. When the data features of the source domain and target domain differ significantly, negative transfer issues may arise.

To address the aforementioned issues, this paper proposes a method for load forecasting of distribution transformers based on a parameter-efficient fine-tuning method of large language models. This paper leverages the superior generalization ability of large language models to quantify the spatiotemporal correlation and inherent features of distribution transformers, thereby enhancing the generalization ability of the load forecasting model in few-shot scenarios. A parameter-efficient fine-tuning method is proposed to reduce the computational cost of the load forecasting model effectively. In summary, the main contributions of this paper are as follows:

- (1) This paper proposes a method for load forecasting of distribution transformers based on parameter-efficient fine-tuning method of large language models, effectively improving the prediction accuracy and generalization ability in few-shot scenarios.
- (2) This paper thoroughly examines the spatiotemporal correlation and inherent features of power loads in distribution transformers. It employs the Linde–Buzo–Gray algorithm and the dynamic mode decomposition algorithm to separate and analyze the spatiotemporal correlation and inherent characteristics of power loads, thereby enhancing the generalization capability of the prediction model.
- (3) This paper presents a parameter-efficient fine-tuning method based on simultaneous perturbation stochastic approximation, which effectively reduces the computational cost of the model and enhances its efficiency.

2. Load Forecasting Method for Distribution Transformers Based on Large Language Models

Existing load forecasting methods for distribution transformers mainly rely on much historical load data. However, there are many distribution transformers in distribution networks, and each distribution transformer may not have sufficient training data, which can lead to overfitting or underfitting of the model. Relevant scholars have used transfer learning algorithms to solve few-shot issues and fully use existing a priori knowledge of distribution transformers to construct transfer learning models. However, this method requires a strong correlation between the target and source domains, but the inherent features of distribution transformers cannot meet this requirement. Therefore, this paper

fully uses the generalization ability of a large language model (LLM) to propose a load forecasting model for distribution transformers based on LLM.

Owing to the rapid development of computer technology, AI technology has made significant strides. In 2020, OpenAI announced the release of the GPT-3 model, which is primarily based on the Transformer model and relies on massive amounts of training data and extensive computational resources to optimize model parameters. This model has fully demonstrated that large language models possess exceptional generalization capabilities. After pre-training, the parameters of LLMs contain a wealth of prior knowledge, enabling them to adapt quickly to new scenarios. In recent years, to further explore the application potential of LLMs in various scenarios, ultra-large-scale parameter AI models such as LLaMA and Flamingo have emerged. This paper selects the LLaMA model as the primary subject for analysis. The LLaMA model is an improvement over the Transformer model, constructed using only the Decoder model. It employs the Causal Mask method to enhance the model's expressive capabilities while preventing information leakage. The partial structure of the LLaMA model is shown in Figure 1.

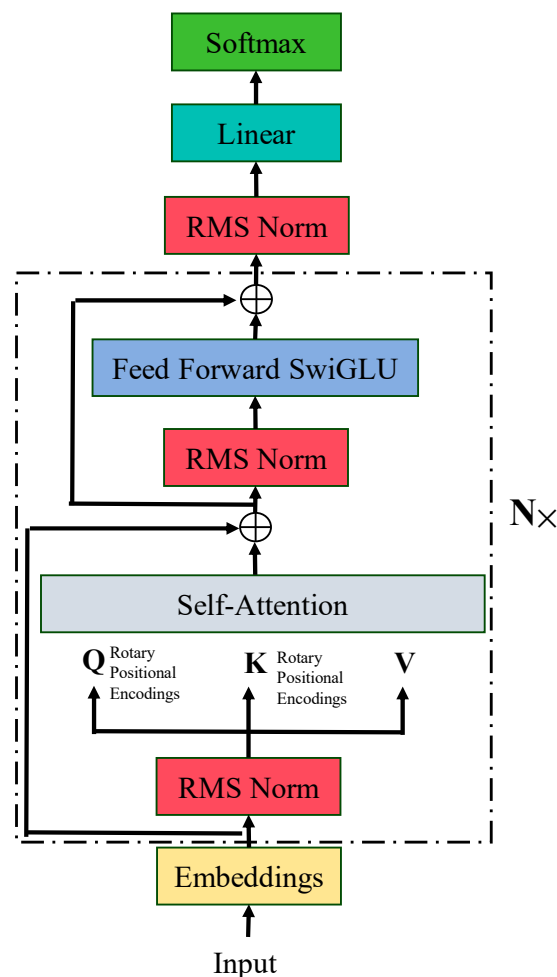


Figure 1. Partial structure of LLaMA model.

The LLaMA model uses RMSNorm instead of LayerNorm to scale input vectors, making them invariant to scaling. This accelerates model convergence and reduces computational complexity during training. The Rotary Position Embedding (RoPE) method is used to encode the position of input vectors. This encoding method cleverly combines absolute position encoding and relative position encoding, making the position encoding sensitive to the dimension of the input vectors. When the LLaMA model processes long-sequence data, as the dimension increases, the influence of RoPE rotation position

encoding gradually weakens, thereby reducing the interference of redundant features. The LLaMA model replaces the ReLU activation function with the SwiGLU activation function, enabling the model to characterize complex data features. Additionally, the smooth and continuous characteristics of the SwiGLU activation function help mitigate the issue of gradient vanishing during training. The SwiGLU activation function is illustrated in Figure 2.

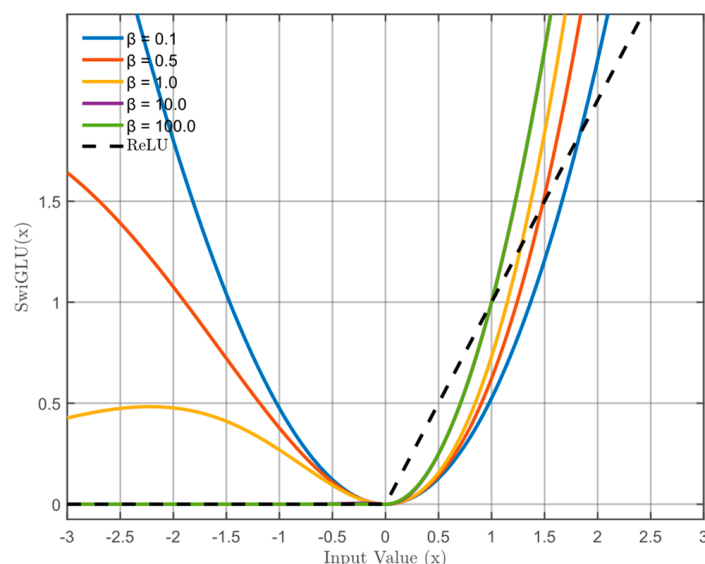


Figure 2. Comparison of SwiGLU activation function and ReLU activation function.

As can be seen from the figure above, the SwiGLU activation function is based on the Swish activation function and the GLU activation function, which gives the SwiGLU activation function better nonlinear feature capture capabilities and filters out secondary information through a gating mechanism, enabling the LLaMA model to quantify the long-range dependencies inherent in high-dimensional sequential data. The ReLU activation function lacks a gating mechanism. It has a gradient of zero in the negative interval, which prevents the model from effectively extracting key features and reduces its generalization ability.

The LLaMA model, with its distinctive model structure, exhibits excellent generalization capabilities and can rapidly adapt to new scenarios, yielding remarkable results across multiple fields. However, the LLaMA model has a large number of parameters, and directly fine-tuning its parameters would incur high computational costs. To adapt to the characteristics of the distribution transformer load forecasting scenario, this paper leverages the superior generalization ability of the LLaMA model to deeply explore the spatiotemporal correlations and inherent features of the distribution transformer. The distribution transformer load data is encoded, and the spatiotemporal correlations and inherent features of the distribution transformer loads are integrated. The main model structure proposed in this paper is shown in Figure 3.

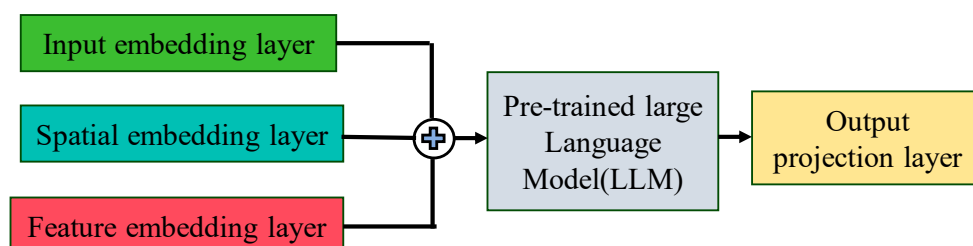


Figure 3. Load forecasting model architecture for distribution transformers.

3. A Method for Feature Extraction of Distribution Transformers Based on the Linde–Buzo–Gray and Dynamic Mode Decomposition Algorithms

Distribution transformers have relatively low voltage ratings, and the load composition of each distribution transformer varies. Users' electricity consumption habits easily influence power loads and exhibit distinct inherent features. Additionally, under similar environmental temperature conditions, users typically exhibit similar electricity consumption habits. For example, during high-temperature weather, users typically turn on their air conditioners to cool down. This causes the air conditioning load of each distribution transformer to rise rapidly. Therefore, under similar conditions, distribution transformers located near each other exhibit temporal and spatial correlations. To fully explore and quantify the temporal and spatial correlations and inherent features of distribution transformers, this paper employs the Linde–Buzo–Gray (LBG) algorithm and the dynamic mode decomposition (DMD) algorithm to investigate data features and enhance the accuracy of distribution transformer load forecasting model.

3.1. The Quantification Method of Spatiotemporal Correlation

The primary factors influencing the spatiotemporal correlation of distribution transformers include the load composition of transformers and the temperature of the power supply area [21]. Blindly assuming that transformers located nearby have spatiotemporal correlation can lead to feature redundancy, thereby reducing the predictive accuracy of the model. For example, during late-night hours when electricity prices are lower, many factories reduce production costs by conducting production activities during this time. Such industrial loads remain at relatively high levels during late-night hours, while most residential loads cease during this period, resulting in lower levels of residential loads. Therefore, this study employs the Linde–Buzo–Gray (LBG) algorithm to identify distribution transformers with spatiotemporal correlation [6].

This paper selects variance, mean, linearity, and other related features of distribution transformer load time sequence data as clustering features. The main features selected in this paper are shown in Table 1.

Table 1. Partial features of time series power load data.

Features	Description
Mean	Mean
Var	Variance
Linearity	Strength of linearity
ACF1	First order of autocorrelation
Peak	Strength of peaks
Curvature	Strength of curvature
Entropy	Spectral entropy
Spikiness	Strength of spikiness

During the clustering process, the LBG algorithm categorizes the sequential load data of distribution transformers by measuring the similarity of different characteristics of the sequential load data of different distribution transformers [6]. Distribution transformer load curves with similar characteristics, such as load curve change patterns, peak times, and load composition, are grouped into the same category, fully considering the spatiotemporal correlation between distribution transformers, thereby laying a solid foundation for improving model accuracy.

3.2. The Quantification Method for Inherent Features of Distribution Transformers

To fully capture the inherent features of distribution transformers, this paper employs the dynamic mode decomposition (DMD) algorithm to analyze the sequential power load data. The DMD algorithm can extract the potential patterns contained in the sequential data and decompose the nonlinear complex features into multiple superimposed linear features. To capture the potential patterns in the sequential data, this paper converts the original sequential data. This paper uses the Hankel operator to construct a multi-dimensional augmented data matrix [33]. The primary process is as follows.

$$O_{it} = \begin{bmatrix} \mathbf{x}_{it}(1) & \mathbf{x}_{it}(2) & \cdots & \mathbf{x}_{it}(K+1-c) \\ \mathbf{x}_{it}(2) & \mathbf{x}_{it}(3) & \cdots & \mathbf{x}_{it}(K+2-c) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{it}(c) & \mathbf{x}_{it}(c+1) & \cdots & \mathbf{x}_{it}(K) \end{bmatrix} \quad (1)$$

$$= [\mathbf{o}_1, \mathbf{o}_2, \cdots, \mathbf{o}_{K+1-c}],$$

In the above equation, distribution transformers power load data $\mathbf{x}_{it} \in \mathbb{R}^K$, $\mathbf{x}_{it}$ represents i th distribution transformers historical load data. Within a short period, load data for distribution transformers exhibits local correlation, and current load data can be obtained by a linear combination of previous time points [33], as shown below.

$$\mathbf{x}_{it}(a) = \sum_{w=1}^q \alpha_w \mathbf{x}_{it}(a-w) + e_a \approx \sum_{w=1}^q \alpha_w \mathbf{x}_{it}(a-w), \quad (2)$$

In the above equation, e_a represents residual error. This paper refers to reference [33] and ignores residual error. Therefore, vectors at adjacent time steps satisfy a linear relationship. The elements in the multi-dimensional augmented data matrix are as follows:

$$\begin{cases} O_1 = [\mathbf{o}_1, \cdots, \mathbf{o}_{K-c}] \in \mathbb{R}^{c \times (K-c)} \\ O_2 = [\mathbf{o}_2, \cdots, \mathbf{o}_{K-c+1}] \in \mathbb{R}^{c \times (K-c)}. \end{cases} \quad (3)$$

After ignoring the residual error, the above equation can be converted to

$$O_2 \approx \mathcal{A}O_1, \quad (4)$$

In the above equation, linear operator $\mathcal{A} \in \mathbb{R}^{c \times c}$. Linear operators contain the main characteristics of sequential load data for distribution transformers, such as oscillation mode, attenuation trend, or growth trend. Due to the high dimensionality of multi-dimensional augmented data matrices, in order to reduce computational complexity and avoid overfitting, this paper refers to reference [33] to perform singular value decomposition (SVD) on multi-dimensional augmented data matrices. The primary process is as follows.

$$O_1 \approx U_r \Sigma_r V_r^\dagger \quad (5)$$

In the above equation, left singular vectors $U_r \in \mathbb{R}^{c \times r}$ represents the typical features of power load fluctuations in distribution transformers. Right singular vectors $V_r \in \mathbb{R}^{(K-c) \times r}$ represents the evolution coefficient of each typical fluctuation pattern, such as periodic weighting. Diagonal matrix $\Sigma_r = \text{diag}[\sigma_1, \cdots, \sigma_r] \in \mathbb{R}^{r \times r}$ represents the importance of each pattern of change. $V_r^\dagger$ denotes the conjugate transpose of V_r . In order to ignore noise and

minor fluctuations, this paper refers to reference [33] to project high-dimensional operators onto low-dimensional subspaces. The primary process is as follows.

$$\tilde{\mathcal{A}} = U_r^\dagger \mathcal{A} U_r \approx U_r^\dagger O_2 V_r \Sigma_r^{-1}, \quad (6)$$

In the above equation, project matrix O_2 onto subspace U_r to obtain the next state in the low-dimensional space. This avoids dependence on the original high-dimensional operator. Finally, through the above derivation, we can obtain the following equation.

$$\tilde{\mathcal{A}} = Q \Lambda Q^\dagger, \quad (7)$$

In the above equation, $Q \in \mathbb{R}^{r \times r}$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r) \in \mathbb{R}^{r \times r}$. The above equation performs feature decomposition on the operator after dimension reduction, breaks down the typical fluctuation features and time-varying patterns of the sequential load data of distribution transformers, and enhances the generalization ability of the load forecasting model in few-shot scenarios. When $|\lambda_i| \approx 1$, it indicates that this type of fluctuation pattern hardly decays, such as industrial loads that are at high levels during working days and working hours. When $|\lambda_i| < 1$, it indicates that this type of fluctuation pattern gradually decays over time, such as short-term peak loads.

4. A Parameter-Efficient Fine-Tuning Method Based on the Simultaneous Perturbation Stochastic Approximation Algorithm

LLMs have a large number of model parameters. During the training phase, LLM often requires millions of dollars in computational costs to mine and quantify general knowledge. Although LLMs can quickly adapt to different new scenarios, they have significant limitations in highly specialized distribution transformer load forecasting scenarios. This is because the training data for distribution transformer load forecasting scenarios is exceptionally scarce. The issue of data imbalance results in significant limitations for LLM in highly specialized scenarios. Therefore, this paper needs to fine-tune LLM while strictly limiting computational costs. Existing fine-tuning techniques for LLM primarily fall into two categories: full fine-tuning and partial fine-tuning (PEFT). Full fine-tuning requires updating all parameters of the LLM, resulting in high computational costs and limited scalability. Additionally, full fine-tuning may cause the complex structure of the LLM to over-adapt to the distribution transformer load forecasting scenario during fine-tuning, leading to catastrophic forgetting. Therefore, this paper selects PEFT technology for fine-tuning the LLM.

In the distribution transformer load forecasting model, this paper freezes the parameters of the LLM and adapts it to the characteristics of the distribution transformer load forecasting scenario by constructing an input embedding layer, a spatial embedding layer, a feature extraction layer, and an output projection layer. The parameters of the input embedding layer, spatial embedding layer, feature extraction layer, and output projection layer are fine-tuned using the error backpropagation algorithm. To improve computing efficiency and reduce training costs, this paper uses a linear layer as the basic structure of the input embedding layer, spatial embedding layer, feature extraction layer, and output projection layer [33]. In the input embedding layer, the sequential data of the distribution transformer load is cleaned and normalized. The equation for the input embedding layer is shown below.

$$E_{it}^1 = \mathbf{x}_{it}^{n-p} W_1^\top + \mathbf{b}_1, \quad (8)$$

In the above equation, E_{it}^1 is part of the input vector of the LLM, and W_1 and $\mathbf{b}_1$ are learnable parameters. These parameters enable the LLM to quickly adapt to the specific characteristics of load forecasting scenarios for distribution transformers.

The load composition of distribution transformers varies significantly, and adjacent distribution transformers may exhibit different temporal fluctuation patterns. For example, the peak times and peak-to-valley differences in the load curves of distribution transformers with a high proportion of industrial loads may differ markedly from those of distribution transformers with a high proportion of residential loads, even if the two transformers are located in close proximity. There is no significant spatiotemporal correlation between the two distribution transformers. To solve the above problems, this paper uses the LBG algorithm to screen out distribution transformers with spatiotemporal correlation. In order to enable the load forecasting model to consider the spatiotemporal correlation between distribution transformers fully, this paper uses a weighted average method to aggregate the load data of similar distribution transformers. The specific equation is shown below.

$$\mathbf{x}_{it}^{\text{agg}} = \frac{\mathbf{x}_{it}^n + \sum_{j \in \mathcal{N}^i} \mathbf{x}_{jt}^n}{B + 1} \quad (9)$$

In the above equation, $\mathbf{x}_{it}^{\text{agg}} \in \mathbb{R}^K$ represents the aggregated load sample. This paper randomly selects B distribution transformers from N similar distribution transformers, and these samples constitute the set $\mathcal{N}^i$. $\mathbf{x}_{jt}$ represents the transformer related to the i -th transformer $\mathbf{x}_{it}$, and $\mathbf{x}_{jt}^n$ represents the normalized data. In the spatial embedding layer, this paper uses the aggregated features as part of the input vector for the LLM, enabling the LLM to consider the spatiotemporal correlation between distribution transformers sufficiently [33]. The specific equation is shown below.

$$E_{it}^2 = \mathbf{x}_{it}^{\text{agg}} W_2^\top + \mathbf{b}_2, \quad (10)$$

In the above equation, W_2 and $\mathbf{b}_2$ are learnable parameters. These parameters enable LLM to capture the spatiotemporal correlations that exist between distribution transformers quickly. E_{it}^2 is part of the input vector of the LLM.

This paper uses the DMD algorithm to extract the inherent features of distribution transformers. In the feature extraction layer, the features extracted by the DMD algorithm are used as part of the input vector for LLM [33]. The specific equation is shown below.

$$E_{it}^3 = \mathbf{v}_{it} W_3^\top + \mathbf{b}_3, \quad (11)$$

In the above equation, $\mathbf{v}_{it} = (\lambda_1; \dots; \lambda_r; \sigma_1; \dots; \sigma_r)$ represents the main fluctuation pattern of the load data, and W_3 and $\mathbf{b}_3$ are learnable parameters.

To enable the LLM to sufficiently consider the spatiotemporal correlation and inherent features of distribution transformers, this paper concatenates the above feature vectors as input vectors for the LLM. The input vectors are shown in the following equation.

$$E_{it} = \begin{bmatrix} E_{it}^1 \\ E_{it}^2 \\ E_{it}^3 \end{bmatrix} \quad (12)$$

In the above equation, E_{it} represents the input vectors for the LLM. In the output projection layer, this paper uses a linear layer to convert the abstract vector output by the

LLM into distribution transformer load data [33]. The expression of the output projection layer is shown below.

$$\hat{\mathbf{y}}_{it} = \mathbf{h}_{it} W_4^\top + \mathbf{b}_4, \quad (13)$$

In the above equation, W_4 and $\mathbf{b}_4$ are learnable parameters, $\hat{\mathbf{y}}_{it}$ represents the load forecasting value of the distribution transformer, and $\mathbf{h}_{it}$ represents the embedding of the last hidden layer of LLM. LLM structures are complex and have numerous model parameters. Traditional error backpropagation algorithms require the usage of all LLM parameters, greatly increasing computational costs.

In order to reduce the training cost of the load forecasting model for distribution transformers and improve inference efficiency, this paper selects the simultaneous perturbation stochastic approximation (SPSA) algorithm to enhance the traditional error backpropagation algorithm [33]. The SPSA algorithm is based on the Simultaneous Perturbation Gradient Approximation method and the Stochastic Newton method. It uses a Gain Sequence and Perturbation Distribution to ensure the convergence of the algorithm. The SPSA algorithm efficiently estimates gradients by simultaneously randomly perturbing all parameters. It only requires two function evaluations to achieve gradient estimation, significantly reducing computational costs. The SPSA algorithm is insensitive to model parameter dimensions and can solve high-dimensional optimization problems with constant computational costs. Based on the SPSA algorithm, LLM can quickly adapt to the characteristics of power distribution transformer load forecasting scenarios at a low price.

The improved optimization algorithm can efficiently optimize the learnable parameters in the input embedding layer, spatial embedding layer, feature extraction layer, and output projection layer. This paper calculates the first-order gradients of W_4 and $\mathbf{b}_4$ [33]. The specific equation is shown below.

$$\begin{cases} \nabla W_4 := \frac{\partial L}{\partial W_4} = \frac{\partial L}{\partial \hat{\mathbf{y}}_{it}} \frac{\partial \hat{\mathbf{y}}_{it}}{\partial W_4} \\ \nabla \mathbf{b}_4 := \frac{\partial L}{\partial \mathbf{b}_4} = \frac{\partial L}{\partial \hat{\mathbf{y}}_{it}} \frac{\partial \hat{\mathbf{y}}_{it}}{\partial \mathbf{b}_4} \end{cases} \quad (14)$$

In the above equation, L represents the loss function. Next, this paper adopts the SPSA algorithm [33] to calculate the zero-order gradients for $\{W_j, \mathbf{b}_j; j = 1, 2, 3\}$. The specific equation is shown below.

$$\hat{\nabla} \Theta' = \frac{L(\Theta' + \epsilon \mathbf{Z}) - L(\Theta' - \epsilon \mathbf{Z})}{2\epsilon} \mathbf{Z} \quad (15)$$

In the above equation, $\Theta' = \{W_j, \mathbf{b}_j; j = 1, 2, 3\}$. $\mathbf{Z}$ represents a vector chosen randomly from the standard Gaussian distribution, and ϵ represents perturbation scale parameter. After obtaining the gradient, this paper uses the adaptive moment estimation (Adam) method to optimize the model parameters. In the t -th iteration, the main optimization process of the model parameters is as follows [33].

$$\begin{cases} \mathcal{M}_{(t)} = \beta_1 \mathcal{M}_{(t-1)} + (1 - \beta_1) \nabla \Theta_{(t-1)} \\ \mathcal{V}_{(t)} = \beta_2 \mathcal{V}_{(t-1)} + (1 - \beta_2) (\nabla \Theta_{(t-1)})^2 \\ \hat{\mathcal{M}}_{(t)} = \frac{\mathcal{M}_{(t)}}{1 - \beta_1^{(t)}} \\ \hat{\mathcal{V}}_{(t)} = \frac{\mathcal{V}_{(t)}}{1 - \beta_2^{(t)}} \\ \Theta_{(t)} = \Theta_{(t-1)} - \alpha \hat{\mathcal{M}}_{(t)} / (\sqrt{\hat{\mathcal{V}}_{(t)}} + \gamma), \end{cases} \quad (16)$$

In the above equation, $\mathcal{M}_{(t)}$ and $\mathcal{V}_{(t)}$ represent the biased first moment estimate and second raw moment estimate, respectively; $\hat{\mathcal{M}}_{(t)}$ and $\hat{\mathcal{V}}_{(t)}$ represent the biased-corrected

first moment estimate and second raw moment estimate, respectively; β_1 and β_2 represent the hyperparameters that control the exponential decay rates; and α and γ represent the learning rate and small constant for numerical stability, respectively.

5. Experiments and Results

To validate the effectiveness of the proposed method, this paper selects LSTM [10], N-BEATS [35], and Autoformer [38] as baseline models, and LLaMa-7B as the primary LLM. The baseline models are trained on the same aggregated dataset as the proposed model. The training data set and validation data set use the same ratio. The models are fine-tuned using the same target data. To fully characterize the error characteristics of the load forecasting model, this paper selects relative Mean Absolute Error (rMAE), Mean Absolute Percentage Error (MAPE) and Normalized Root Mean Square Error (NRMSE) as the primary error indicators. The original definition of MAPE is shown below.

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\widehat{L}_i - L_i}{L_i} \right| \quad (17)$$

MAPE converts the prediction error into a relative percentage, reflecting the degree of deviation between the predicted results and the actual results. It is calculated by dividing the absolute error between the predicted value and the actual value of the distribution transformer load data by the actual value. Next, the average value of all samples is taken. In summary, MAPE emphasizes the relative error in prediction results. This metric standardizes prediction errors from 0% to 100%, facilitating comparisons between multiple prediction models and enabling power companies to select models with superior prediction performance more easily. The NRMSE indicator is an improvement on the RMSE indicator. First, the sum of the squares of all errors is calculated and then divided by the sum of the squares of the actual values. NRMSE is a dimensionless indicator that can be applied to distribution transformers with different load levels. The three indicators comprehensively reflect the error characteristics of the forecast results. This paper selects the power load dataset of a province in China as the main research object. The data set used in this paper has a time granularity of 1 h. Two distribution transformers were selected to verify the effectiveness of the model in a few-shot scenario. Since this research involves confidential data, the relevant data has been anonymized in this paper. For missing values and outliers in the data set, the interquartile range (IQR) method was used to identify outliers, and linear interpolation was used to supplement the missing values and outliers. The linear interpolation method is shown below.

$$L(t) = L_i + \frac{L_{i+1} - L_i}{t_{i+1} - t_i} (t - t_i) \quad (18)$$

In the above equation, t_i and t_{i+1} represent two adjacent time points. $L(t)$ represents missing or abnormal power load data. The prediction results of each method are shown in Figure 4.

As seen from the above figure, the method in this paper effectively predicts changes in the load of distribution transformers and accurately reflects peak load times. Such prediction results can help power companies anticipate possible transformer overload events and arrange remedial measures. Autoformer uses an encoder–decoder architecture, with Series Decomposition Block and Auto-Correlation as its core. This model structure enables Autoformer to capture long-term dependencies in sequential load data effectively. However, there are few load samples for the first distribution transformer, making it

impossible to effectively fine-tune the model parameters of Autoformer, which results in low prediction accuracy.

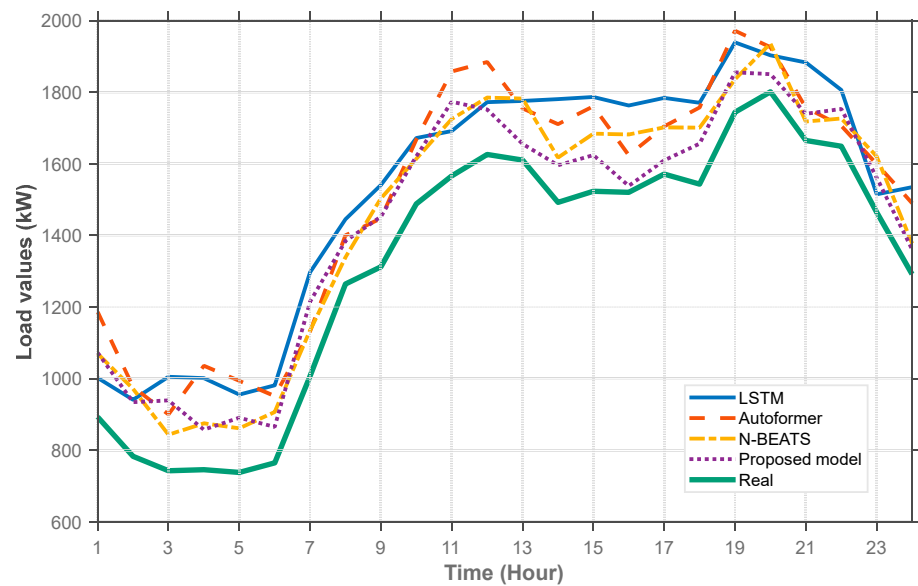


Figure 4. Load forecasting results for each method on first distribution transformer.

LSTM uses a chain structure to capture the fluctuation patterns in sequential power load data through a special gate mechanism. The forget gate in LSTM effectively avoids the problems of gradient disappearance and gradient explosion during training. However, LSTM cannot be effectively fine-tuned with the limited load data of the first distribution transformer, and its prediction accuracy is lower than that of the method proposed in this paper.

The model and the benchmark models in this paper were trained and verified using high-quality data sets, eliminating the impact of missing values and outliers on the model. Based on the power load forecasting curve of the model in this paper, the model accurately reflects the fluctuation characteristics of real data and has good ability to repeat the data dynamic features, which fully verifies the effectiveness of the model in this paper. The load forecasting results of the second distribution transformer are shown in Figure 5.

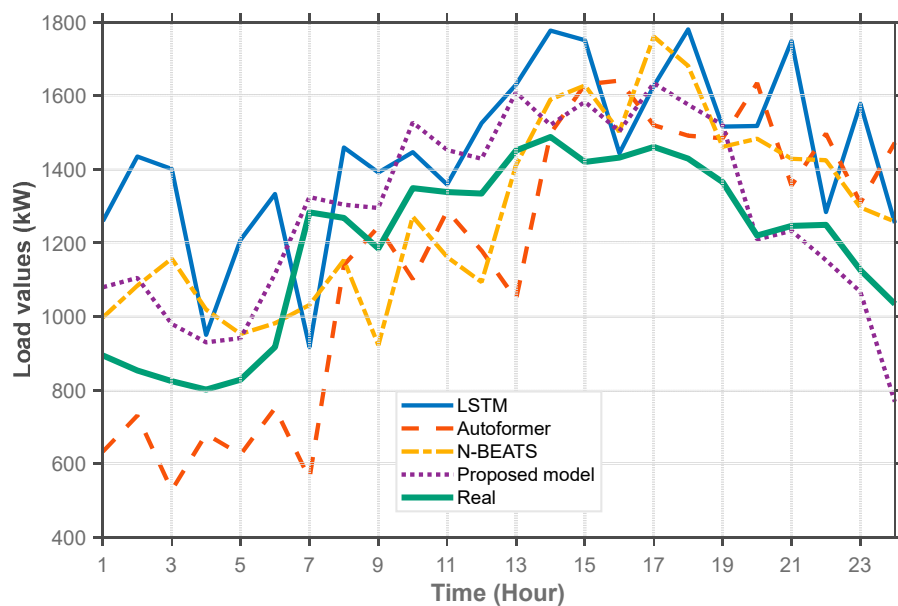


Figure 5. Load forecasting results for each method on second distribution transformer.

As can be seen from the figure above, the prediction accuracy of N-BEATS significantly decreases on the second distribution transformer. N-BEATS adopts a two-layer architecture consisting of Stack and Block, where each Block comprises multiple fully connected layers, and each Stack comprises multiple Blocks. The outputs of multiple Stacks are accumulated to form the final prediction result of the N-BEATS model. Reference [35] uses the Model-Agnostic Meta-Learning (MAML) optimization algorithm to optimize N-BEATS.

The MAML algorithm mainly includes inner-layer optimization and outer-layer optimization. In the inner-layer optimization stage, the model parameters are fine-tuned for a single prediction object. In the outer-layer optimization stage, the meta-model parameters are optimized using the features of multiple prediction objects. The inner-layer optimization and outer-layer optimization are repeated until the meta-model converges. At this point, N-BEATS can quickly adapt to new prediction objects. The new prediction objects must be highly correlated with the training data to ensure high prediction accuracy. If the correlation decreases, the prediction accuracy of N-BEATS will significantly decline. Therefore, N-BEATS achieves high prediction accuracy on the first transformer but shows a significant drop in prediction accuracy on the second transformer. The method proposed in this paper achieves high prediction accuracy on both the first and second distribution transformers. The prediction errors of various methods are shown in Table 2, which fully validates the effectiveness of the proposed method.

Table 2. Prediction error of each model.

Model	The 1st Transformer				The 2nd Transformer			
	MAPE	rMAE	R ²	NRMSE	MAPE	rMAE	R ²	NRMSE
LSTM	16.77	0.15	0.66	15.04	25.21	0.23	−0.9	25.89
Autoformer	15.77	0.14	0.70	14.15	18.29	0.17	−0.3	21.35
N-BEATS	11.18	0.10	0.84	10.09	15.73	0.15	0.26	16.01
Proposed model	10.19	0.08	0.88	9.12	11.18	0.10	0.61	11.54

The rMAE indicator is dimensionless, allowing power companies to select appropriate prediction models from different data sets. As can be seen from Table 2, the prediction errors of each model on the second distribution transformer increased to varying degrees. This is because the power load of the second distribution transformer is less regular, and the dependency between the time series data has undergone dynamic changes. The model proposed in this paper, leveraging the exceptional generalization capabilities of LLM, can quickly adapt to new data features, fully validating the advantages of the proposed model.

In order to fully verify the effectiveness of each component of the model in this paper, ablation experiments were conducted on two transformers to validate the effectiveness of each component of the model. The spatial embedding layer and feature extraction layer were selected as the main objects for analysis. The results of the ablation experiments are shown in Table 3.

As can be seen from the table above, after losing the spatial embedding layer or feature extraction layer, the prediction accuracy decreased to varying degrees, which fully verifies the effectiveness of the spatial embedding layer and feature extraction layer in improving the accuracy of distribution transformers.

To validate the training cost of the proposed method, this paper compares the memory and time used by different baseline models during the training stage of the first distribution transformer. The main results are shown in Table 4.

Table 3. Results of ablation experiment.

Variants	The 1st Transformer			The 2nd Transformer		
	MAPE	rMAE	NRMSE	MAPE	rMAE	NRMSE
Feature extraction layer	10.92	0.08	9.36	11.93	0.13	12.77
Spatial embedding layer	11.02	0.09	9.84	12.51	0.15	13.18
Proposed model	10.19	0.08	9.12	11.18	0.10	11.54

Table 4. Memory and time used by each model during training stage.

Model	Mem (MiB)	Time (s)
LSTM	216	106.4
Autoformer	2950	542.7
N-BEATS	1128	1671.1
Proposed model	492	852.3

This paper accesses LLaMa-7B through cloud services. The memory and time used in the training stage are mainly used to optimize the parameters of the input embedding layer, spatial embedding layer, feature extraction layer, and output projection layer. As seen from the table above, this paper effectively reduces training costs. It enables LLaMa-7B to quickly adapt to the characteristics of power distribution transformer load forecasting scenarios with only a small amount of memory.

6. Conclusions

Facing the large number of distribution transformers, the load data of different distribution transformers show different inherent features, and the few-shot problem is common. This paper proposes a method for load forecasting distribution transformers based on parameter-efficient fine-tuning of large language models. This paper utilizes the superior generalization ability of large language model (LLM) to enhance the accuracy of load forecasting model in few-shot scenarios. Through the Linde–Buzo–Gray algorithm, the spatiotemporal correlation between distribution transformers is analyzed. Key features are extracted from the original load data based on the dynamic mode decomposition algorithm. The proposed model captures the inherent features of each distribution transformer, converting time series load data into low-dimensional dynamic features, combining singular value decomposition to extract the main fluctuation patterns, and finally generating an embedded representation suitable for LLMs through linear transformation, which significantly improves the generalization ability of the model in few-shot scenarios. Since the proposed model accesses LLaMa-7B through cloud services, the computing cost of the proposed model is significantly reduced, and the memory of the proposed model is only 492 MiB. This paper proposes a parameter-efficient fine-tuning method that freezes the LLM parameters and adapts the LLM to the characteristics of the distribution transformer load forecasting scenario by constructing an input embedding layer, a spatial embedding layer, a feature extraction layer, and an output projection layer. The simultaneous perturbation stochastic approximation algorithm is used to fine-tune the parameters of the input embedding layer, spatial embedding layer, feature extraction layer, and output projection layer.

The experimental results demonstrate that the proposed method achieves high prediction accuracy in few-shot scenarios. It can rapidly capture and infer the potential fluctuation patterns of different distribution transformers based on limited data. This method significantly enhances the grid company's ability to identify potential risks in distribution grids and effectively prevents distribution transformer overload incidents. In future work, we will further explore the impact of enterprise working schedules, emergency events, typhoon disaster event, power grid expansion and other sudden events on the load of distribution transformers, continue to optimize the sensitivity of the prediction model to various types of events, and enhance the edge deployment performance of the prediction model.

Author Contributions: Conceptualization, W.Z., S.L., M.L., H.B., S.Y. and X.C.; methodology, W.Z., S.L., M.L., H.B., S.Y. and X.C.; software, W.Z., S.L., M.L., H.B., S.Y. and X.C.; writing—original draft, W.Z., S.L., M.L., H.B., S.Y. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the State Grid Henan Electric Power Company Science and Technology Project: Key Technologies for Lightweight Training of Multi-modal Large Models for Power Grids (Project Number: 521702240015).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: We would like to express our sincere gratitude to the reviewers for your professional opinions and valuable suggestions, and our heartfelt thanks to the editorial board of the journal for their efficient work in supporting the rigorous presentation of the scholarly results. We would like to pay tribute to the academic community.

Conflicts of Interest: Author Weijian Zhang was employed by the State Grid Henan Electric Power Company. Authors Shanfeng Liu, Miaomiao Li, Shaoguang Yuan and Xiawei Cheng were employed by the Electrical Power Research Institute of Henan Electric Power Corporation. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Liu, Y.; Dutta, S.; Kong, A.W.K.; Yeo, C.K. An Image Inpainting Approach to Short-Term Load Forecasting. *IEEE Trans. Power Syst.* **2023**, *38*, 177–187. [\[CrossRef\]](#)
2. Quilumba, F.L.; Lee, W.-J.; Huang, H.; Wang, D.Y.; Szabados, R.L. Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. *IEEE Trans. Smart Grid* **2015**, *6*, 911–918. [\[CrossRef\]](#)
3. Liu, B.; Nowotarski, J.; Hong, T.; Weron, R. Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. *IEEE Trans. Smart Grid* **2017**, *8*, 730–737. [\[CrossRef\]](#)
4. Chen, Q.; Yang, G.; Geng, L. Short Term Power Load Combination Forecasting Method Based on Feature Extraction. *IEEE Access* **2024**, *12*, 51619–51629. [\[CrossRef\]](#)
5. Acquah, M.A.; Jin, Y.; Oh, B.-C.; Son, Y.-G.; Kim, S.-Y. Spatiotemporal Sequence-to-Sequence Clustering for Electric Load Forecasting. *IEEE Access* **2023**, *11*, 5850–5863. [\[CrossRef\]](#)
6. Grabner, M.; Wang, Y.; Wen, Q.; Blažič, B.; Štruc, V. A Global Modeling Framework for Load Forecasting in Distribution Networks. *IEEE Trans. Smart Grid* **2023**, *14*, 4927–4941. [\[CrossRef\]](#)
7. Son, J.; Cha, J.; Kim, H.; Wi, Y.-M. Day-Ahead Short-Term Load Forecasting for Holidays Based on Modification of Similar Days' Load Profiles. *IEEE Access* **2022**, *10*, 17864–17880. [\[CrossRef\]](#)
8. Madhukumar, M.; Sebastian, A.; Liang, X.; Jamil, M.; Shabbir, M.N.S.K. Regression Model-Based Short-Term Load Forecasting for University Campus Load. *IEEE Access* **2022**, *10*, 8891–8905. [\[CrossRef\]](#)
9. Jiang, Z.; Wu, H.; Zhu, B.; Gu, W.; Zhu, Y.; Song, Y.; Ju, P. A Bottom-up Method for Probabilistic Short-Term Load Forecasting Based on Medium Voltage Load Patterns. *IEEE Access* **2021**, *9*, 76551–76563. [\[CrossRef\]](#)
10. Rubasinghe, O.; Zhang, X.; Chau, T.K.; Chow, Y.H.; Fernando, T.; Iu, H.H.-C. A Novel Sequence to Sequence Data Modelling Based CNN-LSTM Algorithm for Three Years Ahead Monthly Peak Load Forecasting. *IEEE Trans. Power Syst.* **2024**, *39*, 1932–1947. [\[CrossRef\]](#)

11. Oza, A.; Patel, D.K.; Ranger, B.J. Fusion ConvLSTM-Net: Using Spatiotemporal Features to Increase Residential Load Forecast Horizon. *IEEE Access* **2025**, *13*, 12190–12202. [\[CrossRef\]](#)
12. Zhou, Y.; Wang, M. Unifying Load Disaggregation and Prediction for Buildings with Behind-the-Meter Solar. *IEEE Trans. Power Syst.* **2025**, *40*, 1335–1347. [\[CrossRef\]](#)
13. González, J.P.; Roque, A.M.S.M.S.; Pérez, E.A. Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting. *IEEE Trans. Power Syst.* **2018**, *33*, 545–556. [\[CrossRef\]](#)
14. Jiahao, L.; Yuan, G.; Dunnan, L. Research on power medium-and long-term load forecasting based on forecasting empirical partial least squares regression. In Proceedings of the 2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), Guangzhou, China, 14–16 May 2021; pp. 395–398. [\[CrossRef\]](#)
15. Sharma, S.; Majumdar, A.; Elvira, V.; Chouzenoux, É. Blind Kalman Filtering for Short-Term Load Forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 4916–4919. [\[CrossRef\]](#)
16. Smyl, S.; Dudek, G.; Pelka, P. ES-dRNN: A Hybrid Exponential Smoothing and Dilated Recurrent Neural Network Model for Short-Term Load Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 11346–11358. [\[CrossRef\]](#)
17. Li, J.; Feng, S.; Zhang, T.; Ma, L.; Shi, X.; Zhou, X. Study of Long-Term Energy Storage System Capacity Configuration Based on Improved Grey Forecasting Model. *IEEE Access* **2023**, *11*, 34977–34989. [\[CrossRef\]](#)
18. Qu, K.; Si, G.; Shan, Z.; Wang, Q.; Liu, X.; Yang, C. Forwardformer: Efficient Transformer with Multi-Scale Forward Self-Attention for Day-Ahead Load Forecasting. *IEEE Trans. Power Syst.* **2024**, *39*, 1421–1433. [\[CrossRef\]](#)
19. Liu, H.; Wang, Y.; Wei, C.; Li, J.; Lin, Y. Two-Stage Short-Term Load Forecasting for Power Transformers Under Different Substation Operating Conditions. *IEEE Access* **2019**, *7*, 161424–161436. [\[CrossRef\]](#)
20. Sajjad, M.; Khan, Z.A.; Ullah, A.; Hussain, T.; Ullah, W.; Lee, M.Y.; Baik, S.W. A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. *IEEE Access* **2020**, *8*, 143759–143768. [\[CrossRef\]](#)
21. Rasheed, T.; Bhatti, A.R.; Farhan, M.; Rasool, A.; El-Fouly, T.H.M. Improving the Efficiency of Deep Learning Models Using Supervised Approach for Load Forecasting of Electric Vehicles. *IEEE Access* **2023**, *11*, 91604–91619. [\[CrossRef\]](#)
22. Masood, Z.; Gantassi, R.; Choi, Y. Enhancing Short-Term Electric Load Forecasting for Households Using Quantile LSTM and Clustering-Based Probabilistic Approach. *IEEE Access* **2024**, *12*, 77257–77268. [\[CrossRef\]](#)
23. Liu, Q.; Cao, J.; Zhang, J.; Zhong, Y.; Ba, T.; Zhang, Y. Short-Term Power Load Forecasting in FGSM-Bi-LSTM Networks Based on Empirical Wavelet Transform. *IEEE Access* **2023**, *11*, 105057–105068. [\[CrossRef\]](#)
24. Hu, Y.; Pan, L.; Yang, X.; Wei, X.; Xie, D. Short-Term Power Load Dynamic Scheduling Based on GWO-TCN-GRU Optimization Algorithm. *IEEE Access* **2024**, *12*, 144919–144931. [\[CrossRef\]](#)
25. Tan, D.; Tang, Z.; Zhou, F.; Xie, Y. A Novel Hybrid Model Based on EMD-Improved TCN-Improved TST for Short-Term Railway Traction Load Forecasting. *IEEE Trans. Transp. Electr.* **2025**, *11*, 6418–6427. [\[CrossRef\]](#)
26. 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*, 2703–2714. [\[CrossRef\]](#)
27. Zeng, P.; Sheng, C.; Jin, M. A learning framework based on weighted knowledge transfer for holiday load forecasting. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 329–339. [\[CrossRef\]](#)
28. Zhang, Z.; Zhao, P.; Wang, P.; Lee, W.-J. Transfer Learning Featured Short-Term Combining Forecasting Model for Residential Loads With Small Sample Sets. *IEEE Trans. Ind. Appl.* **2022**, *58*, 4279–4288. [\[CrossRef\]](#)
29. Wu, D.; Lin, W. Efficient Residential Electric Load Forecasting via Transfer Learning and Graph Neural Networks. *IEEE Trans. Smart Grid* **2023**, *14*, 2423–2431. [\[CrossRef\]](#)
30. Zhao, P.; Cao, D.; Wang, Y.; Chen, Z.; Hu, W. Gaussian Process-Aided Transfer Learning for Probabilistic Load Forecasting Against Anomalous Events. *IEEE Trans. Power Syst.* **2023**, *38*, 2962–2965. [\[CrossRef\]](#)
31. Tian, Y.; Sehovac, L.; Grolinger, K. Similarity-Based Chained Transfer Learning for Energy Forecasting with Big Data. *IEEE Access* **2019**, *7*, 139895–139908. [\[CrossRef\]](#)
32. Forootani, A.; Rastegar, M.; Zareipour, H. Transfer Learning-Based Framework Enhanced by Deep Generative Model for Cold-Start Forecasting of Residential EV Charging Behavior. *IEEE Trans. Intell. Veh.* **2024**, *9*, 190–198. [\[CrossRef\]](#)
33. Zhou, Y.; Wang, M. Empower Pre-Trained Large Language Models for Building-Level Load Forecasting. *IEEE Trans. Power Syst.* **2025**, *40*, 4220–4232. [\[CrossRef\]](#)
34. Wu, D.; Wang, B.; Precup, D.; Boulet, B. Multiple Kernel Learning-Based Transfer Regression for Electric Load Forecasting. *IEEE Trans. Smart Grid* **2020**, *11*, 1183–1192. [\[CrossRef\]](#)
35. Oreshkin, B.N.; Carpo, D.; Chapados, N.; Bengio, Y. Meta-learning framework with applications to zero-shot time-series forecasting. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), Online, 2–9 February 2021; Volume 35, pp. 9242–9250.
36. Tiwari, R.S.; Sharma, J.P.; Gupta, O.H.; Sufyan, M.A.A. Extension of pole differential current based relaying for bipolar LCC HVDC lines. *Sci. Rep.* **2025**, *15*, 16142. [\[CrossRef\]](#) [\[PubMed\]](#)

37. Wang, Z.; Hou, H.; Wei, R.; Li, Z. A Distributed Market-Aided Restoration Approach of Multi-Energy Distribution Systems Considering Comprehensive Uncertainties from Typhoon Disaster. *IEEE Trans. Smart Grid* **2025**, *16*, 3743–3757. [[CrossRef](#)]
38. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.