

Multiscale Spatio-Temporal Enhanced Short-term Load Forecasting of Electric Vehicle Charging Stations

1st Zongbao Zhang

*Shenzhen Power Supply Bureau
Co., Ltd.*

Shenzhen, China

zbzhangwhu@163.com

2nd Jiao Hao

*Shenzhen Power Supply Bureau
Co., Ltd.*

Shenzhen, China

haojiao.sjtu@163.com

3rd Wenmeng Zhao

*Electric Power Research Institute,
CSG*

Guangzhou, China

zhaowm@csg.cn

4th Yan Liu

*Shenzhen Power Supply Bureau
Co., Ltd.*

Shenzhen, China

liuyan0829@sz.csg.cn

5th Yaohui Huang*

*Tsinghua-Berkeley Shenzhen Institute
Tsinghua University*

Shenzhen, China

yhhuang5212@gmail.com

6th Xinhang Luo

*Tsinghua-Berkeley Shenzhen Institute
Tsinghua University*

Shenzhen, China

luoxh23@mails.tsinghua.edu.cn

Abstract—The rapid expansion of electric vehicles (EVs) has rendered the load forecasting of electric vehicle charging stations (EVCS) increasingly critical. The primary challenge in achieving precise load forecasting for EVCS lies in accounting for the nonlinear of charging behaviors, the spatial interactions among different stations, and the intricate temporal variations in usage patterns. To address these challenges, we propose a Multiscale Spatio-Temporal Enhanced Model (MSTEM) for effective load forecasting at EVCS. MSTEM incorporates a multiscale graph neural network to discern hierarchical nonlinear temporal dependencies across various time scales. Besides, it also integrates a recurrent learning component and a residual fusion mechanism, enhancing its capability to accurately capture spatial and temporal variations in charging patterns. The effectiveness of the proposed MSTEM has been validated through comparative analysis with six baseline models using three evaluation metrics. The case studies utilize real-world datasets for both fast and slow charging loads at EVCS in Perth, UK. The experimental results demonstrate the superiority of MSTEM in short-term continuous load forecasting for EVCS.

Index Terms—Loading forecasting, Graph neural network, Electric vehicle charging stations, Time series forecasting.

I. INTRODUCTION

The rapid development of electric vehicles (EVs) heralds a pivotal shift in the automotive sector [1]. Recently, the number of EVs has experienced a significant upsurge. By 2021, the global cumulative count of EVs reached approximately 16.5 million, a significant leap from the 5.5 million recorded in 2018 [2]. This burgeoning growth has led to a rapid increase in the demand for efficient and reliable electric vehicle charging stations (EVCS). Efficient management plays a key role in

these stations, particularly in regulating the electric power supply. Accurate short-term forecasting of electrical loads is critical for maintaining power grid stability and enhancing the operational efficiency of EVCS. However, forecasting the electric load for EVCS is a complex task. It involves numerous intricate factors, such as consumer usage patterns, time-of-day variations, options for fast and slow charging, and the volatility in the energy market [3]. These factors interact to create a dynamic and intricate demand scenario. Therefore, developing effective predictive models is essential to making accurate predictions and managing the EV charging demands.

There are various data-driven models that have been employed to forecast the electric loading at EVCS. The statistical approaches, including Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) models and their variations, have been widely used in charging load forecasting [4]. Lu et al. [5] employ ARIMA in ultra-short-term demand response prediction. Ren et al. [6] propose a hybrid network based on seasonal-ARIMA (SARIMA) to capture the periodic features of electric loads. These models primarily focus on analyzing temporal charging patterns but often neglect the complex local correlations between charging stations within EVCS networks, which is difficult to accurately predict the charging patterns of EV charging demand. As time series analysis increasingly integrates machine learning techniques, traditional models such as Multilayer Perceptron Network (MLP), Support Vector Machines (SVM), and Tree-based algorithms have shown significant potential in identifying EVs charging patterns. Zhang et al. [7] utilize the Gradient Boosting Decision Tree (GBDT) to enhance the predictive performance. Khan et al. [8] verify the effectiveness of traditional machine learning models in load forecasting for EVs. Despite their improved reliability

(Corresponding author: Yaohui Huang, e-mail: yhhuang5212@gmail.com)

This paper was supported by the Science and Technology Project of Shenzhen Power Supply Corporation, grant number: SZKJXM20220036/09000020220301030901283.

and accuracy, these models frequently struggle to capture the nonlinear temporal dynamics of charging demand, typically requiring extensive feature engineering for improved prediction. [9]. Deep learning is a promising model for extracting complex and nonlinear temporal patterns from electric loads, which includes Recurrent Neural Networks (RNN), Graph Neural Networks (GCNs), and attention-based models [10]. Wang et al. [11] utilized a graph attention mechanism to effectively learn the energy load patterns across multiple regions. Shi et al. [12] employed GCNs to capture the spatiotemporal relationships among charging stations. However, the above models fail to account for the dynamic interconnections between charging stations and the variations in temporal details across different scales.

The major challenges in short-term load forecasting for EVCS can be summarized into three aspects: (1) *Complex Loading Patterns*. EVCS charging loads display significant nonlinear fluctuations influenced by various factors, such as weather conditions, commuting patterns, and social events. Therefore, it is a significant challenge to accurately capture these complex temporal patterns. (2) *Spatial Dynamics of EVCS*. The load trends of adjacent EVCS often exhibit similarities, highlighting the influence of spatial factors on charging behaviors. The location of each significantly affects its usage patterns. These spatial dynamics are crucial for predicting charging station load trends. (3) *Hierarchical Temporal Patterns*. The electric load exhibits distinct patterns across different temporal scales [13]. These patterns range from short-term fluctuations to long-term trends, presenting a challenge for prediction models to effectively capture and interpret these diverse temporal interactions.

To tackle the above challenges in short-term load forecasting for EVCS, we propose a novel Multiscale Spatio-temporal Enhanced Model (MSTEM). For the first challenge, MSTEM incorporates an LSTM unit to capture the nonlinear temporal dependencies from the historical load patterns. To tackle the second challenge, MSTEM incorporates GCNs to effectively consider the local correlations among charging stations. For the third challenge, MSTEM implements a multi-scale fusion strategy that captures the dynamics of the graph structure across various temporal scales. This approach enables the model to accurately discern complex load patterns at different time scales. The main contributions of this paper are summarized as follows:

- A novel neural network architecture, termed MSTEM, has been proposed for load forecasting at EVCS. MSTEM employs a multi-scale graph neural network with recurrent learning components to effectively capture linear trends and nonlinear temporal representations.
- The effectiveness of the proposed model is validated using a real-world dataset encompassing both fast and slow charging scenarios for EVCS.
- Comparative experimental results indicate that MSTEM surpasses established baseline models, including Dlinear, MLP, and variants of RNN, in terms of forecasting accuracy over multiple time steps

The rest of the article is organized as follows: Section II provides the mathematical formulation of the proposed MSTEM, followed by Section III presents the case studies and results; Section IV concludes the article.

II. THE FRAMEWORK OF MULTISCALE SPATIO-TEMPORAL ENHANCED MODEL (MSTEM)

A. Problem definition

The load forecasting at EVCS is essentially a time series forecasting problem, which aims to predict future load values based on historical observations. In this study, the objective of charging load forecasting is to utilize the observed history of the previous τ time steps to predict the values of the target time series for the next α time steps. This multi-step continuous forecasting problem can be expressed as follows:

$$[x_t, x_{t+1}, \dots, x_{t+\tau-1}] \xrightarrow{\mathbb{F}(\cdot)} [x_{t+\tau}, \dots, x_{t+\tau+\alpha-1}], \quad (1)$$

where $x_i \in \mathbb{R}^N$ represents the charging load value at the i -th time step. N is the number of target EVCS. $X_t = [x_t, x_{t+1}, \dots, x_{t+\tau-1}] \in \mathbb{R}^{\tau \times N}$ denotes the model inputs, and the subsequent sequence $Y_t = [x_{t+\tau}, \dots, x_{t+\tau+\alpha-1}] \in \mathbb{R}^{\alpha \times N}$ is the predictive target. $\mathbb{F}(\cdot)$ indicates the predictive model with learnable parameter θ .

B. The Overview of MSTEM

The proposed MSTEM, depicted in Fig. 1, consists of three sequential modules: Multiscale Graph Construction and Learning (MGCL), Temporal Enhanced Neural Network (TENN), and an output layer. The MGCL module focuses on capturing the spatio-temporal dependencies among different charging stations, utilizing a graph-based approach to model the intricate interactions and patterns in data. The TENN module combines the LSTM component with linear representations, which are derived from raw input data. This integration facilitates a comprehensive understanding of temporal patterns. The output layer incorporates output control mechanisms to refine and adjust the predicted outcomes based on the processed data and learned patterns.

C. Multiscale Graph Construction and Learning (MGCL)

For an input charging load record $X \in \mathbb{R}^{\tau \times N}$ from multiple charging stations, the multiscale load patterns can be defined as:

$$X^{s_i} = \text{ScaleTemporal}(X, s_i) \quad \forall s_i \in S, \quad (2)$$

where X^{s_i} denotes the scaled input time series. s_i is an element within the set of scale lengths, denoted as $S = \{s_1, s_2, \dots, s_k\}$. The function $\text{ScaleTemporal}(\cdot)$ samples the input data X at intervals of s_i . For instance, if $s_i = 2$, it selects every second time step from X . Then, MSTEM performs graph convolution on each scaled input. The time series of charging load can be seen as a completed graph, let A^{s_i} represents the adjacency matrix and D^{s_i} denotes the corresponding degree matrix. The graph convolution for scale s_i is:

$$H^{s_i} = \sigma(\text{BN}(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X^{s_i} W_1^{s_i})), \quad (3)$$

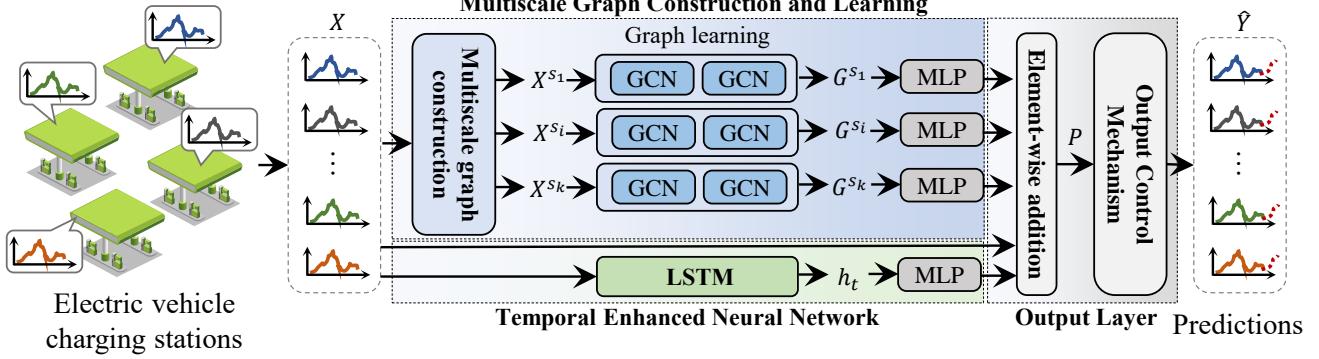


Fig. 1. The overall structure of the multiscale spatio-temporal enhanced model.

$$R^{s_i} = \text{Dropout}(H_1^{s_i}; p), \quad (4)$$

$$G^{s_i} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} R^{s_i} W_2^{s_i}, \quad (5)$$

where H^{s_i} and G^{s_i} is the hidden and final representation of graph learning at scale s_i . $\tilde{A} = A + I$ includes self-loops, I is the unit matrix. \tilde{D} is the degree matrix of \tilde{A} . W^{s_i} is the weight matrix for scale s_i , and $\sigma(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function. p is the dropout rate. BN(\cdot) represents the batch normalization layer. After processing the scaled data, MSTEM integrates the features from each scale. The features from each scale are linearly transformed to a consistent shape, and then are combined through element-wise addition:

$$G^{out} = \sum_{i=1}^k G_2^{s_i} W_L^{s_i}, \quad (6)$$

where G^{out} denotes the combined feature vector, which integrates information from all scales.

D. Temporal Enhanced Neural Network (TENN)

The MGCL module excels in learning the spatial and temporal information from the charging load data at EVCS. However, certain limitations need to be addressed. (1) The MGCL primarily focuses on capturing spatial dependencies and temporal patterns at different scales but may not effectively model long-term temporal dependencies. (2) The learning process within the MGCL module may result in information loss, particularly when handling complex nonlinear representations. To solve these issues, the TENN module integrates LSTM units to enhance the model's ability to capture long-term temporal dependencies. LSTMs are well-suited for processing time-series data and can retain information over extended periods, which is essential for accurate EVCS load forecasting. This processing can be mathematically represented as:

$$f_t = \sigma(W_f^l \cdot [h_{t-1}, x_t] + b_f^l), \quad (7)$$

$$i_t = \sigma(W_i^l \cdot [h_{t-1}, x_t] + b_i^l), \quad (8)$$

$$\tilde{c}_t = \tanh(W_C^l \cdot [h_{t-1}, x_t] + b_C^l), \quad (9)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (10)$$

$$o_t = \sigma(W_o^l \cdot [h_{t-1}, x_t] + b_o^l), \quad (11)$$

$$h_t = o_t \odot \tanh(c_t), \quad (12)$$

where f_t , i_t , and o_t are the output of the forget gate, input gate, and output gate at time step t , respectively. \tilde{c}_t and c_t are the candidate, updated cell state, respectively. h_t is the updated hidden state. \odot denotes element-wise multiplication. W^l and b^l are the weights and biases for the respective gates and states.

Additionally, the proposed MSTEM integrates a residual fusion mechanism to preserve the integrity of the original time series signal. This mechanism introduces linear characteristics into the model and plays a crucial role in stabilizing the learning process. As a result, it significantly improves the robustness and reliability of the proposed MSTEM. The linear representations within this mechanism can be expressed as:

$$L^{out} = \tilde{X}^\eta \oplus (XW^{lo}), \quad (13)$$

where \tilde{X}^η denotes a historical segment of the charging load data, capturing the sequence from the past η time steps, specifically including the data points $\{x_{t+\tau-\eta-1}, \dots, x_{t+\tau-1}\}$. W^{lo} is a learnable weighting matrix designed to modify the dimensions of X to align with the desired output shape. \oplus represents element-wise addition.

E. Output Layer

The output layer of the model combines the outputs from the MGCL and TENN modules to generate predictions, which is formulated as:

$$P = (W_1^{out} G^{out}) \oplus (W_2^{out} h_t) \oplus L^{out}, \quad (14)$$

where $P \in \mathbb{R}^{\alpha \times N}$ represents the charging load prediction of each station in the next α time steps. W^{out} is utilized to linearly transform the outputs of the aforementioned modules. To align with real-world operation, an output control mechanism is integrated to modify the extremely low predictions to zero:

$$\hat{Y} = \max(P_i - \theta, 0) \quad i \in \{t + \tau, \dots, t + \tau + \alpha - 1\}, \quad (15)$$

where \hat{Y} denotes the final predictions, and θ is a predefined threshold.

III. CASE STUDY

A. Experimental Setting

1) *Datasets*: To evaluate the accuracy of our proposed model, real-world electric vehicle charging load datasets from

Perth, UK¹ were analyzed in this study. This dataset includes the start and end times, and the total energy consumption for each charging event. To meet the requirements of load forecasting, the original dataset is transformed into an hourly resolution time series, representing the average charging load per hour. This dataset was also utilized in prior research [14], although for different research objectives. The dataset was separated into two subsets according to the charging equipment type: a fast charging dataset and a slow charging dataset. The fast charging subset comprises equipment characterized by a power output of approximately 22 kW. In contrast, the slow charging subset consists of equipment with power outputs around 7 kW or lower.

2) *Implementation Details:* The experimental setup was implemented on a server equipped with an NVIDIA A6000 GPU. The dataset was divided into three parts: the first 70% of the historical load data is used for training, 10% for validation, and the remaining 20% for testing. We train the model using the Adam optimizer and the Huber loss function, setting the learning rate at 0.001. The training process involved a batch size of 32 and is conducted over 50 epochs. For the proposed MSTEM, the scale parameters are set to {1, 5}. The architecture of each graph learning component included two sequentially stacked graph convolutional modules. The dimensions for the hidden size and graph output layers are set at 16 and 4, respectively. The dropout rate p is 0.1.

3) *Performance Metrics:* To evaluate the performance of the proposed model in charging load forecasting, three metrics are employed: mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE). Specifically, MSE emphasizes larger errors, MAE indicates the average magnitude of errors, and RMSE quantifies the standard deviation of the forecasting errors. The formulations of the above three metrics are provided below:

$$\text{MSE} = \frac{1}{N \cdot \Omega_{test}} \sum_{i=1}^N \sum_{j=1}^{\Omega_{test}} (\hat{Y}_{i,j} - Y_{i,j})^2, \quad (16)$$

$$\text{MAE} = \frac{1}{N \cdot \Omega_{test}} \sum_{i=1}^N \sum_{j=1}^{\Omega_{test}} |\hat{Y}_{i,j} - Y_{i,j}|, \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{N \cdot \Omega_{test}} \sum_{i=1}^N \sum_{j=1}^{\Omega_{test}} (\hat{Y}_{i,j} - Y_{i,j})^2}, \quad (18)$$

where \hat{Y} and Y represent the predicted and actual values, respectively. Ω_{test} denotes the number of samples in the test set. Lower metric scores indicate greater model accuracy.

B. Accuracy Comparison

In this study, we assess the performance of our proposed model using datasets encompassing both fast and slow charging loads at EVCS. These datasets are analyzed over diverse time steps (α): 6 hours, 12 hours, and 24 hours. The results of this analysis are presented in Table I for fast charging

TABLE I
RESULTS OF COMPARABLE EXPERIMENTS ON FAST CHARGING DATASET.

Step	Metrics	MA	HI	MLP	GRU	LSTM	Dlinear	MSTEM
6 h	MSE	36.89	51.37	22.19	24.15	23.63	22.49	<u>22.33</u>
	MAE	2.74	2.97	<u>2.18</u>	2.25	2.24	2.24	2.07
	RMSE	6.07	7.17	4.71	4.91	4.86	4.74	<u>4.73</u>
12 h	MSE	42.25	53.02	<u>23.03</u>	24.68	24.11	23.20	22.95
	MAE	3.16	3.06	<u>2.22</u>	2.31	2.22	2.27	2.15
	RMSE	6.50	7.28	<u>4.80</u>	4.97	4.91	4.82	4.79
24 h	MSE	29.14	54.08	<u>23.41</u>	30.54	30.47	23.53	22.61
	MAE	2.64	3.12	<u>2.27</u>	3.06	3.08	2.28	2.15
	RMSE	5.40	7.35	<u>4.84</u>	5.53	5.52	4.85	4.75

and in Table II for slow charging. The best result for each metric is emphasized in bold, while the second-highest is distinguished with an underline. To evaluate the effectiveness of the proposed model, six baseline models are selected for comparative analysis against the proposed MSTEM model, as follows:

- Moving average (MA) [15]: A classical statistical technique that computes the average of data points within a specified time window.
- Historical inertia (HI) [16]: This model utilizes the inertia or momentum inherent in historical data to make predictions.
- Multilayer perception network (MLP) [17]: A type of feedforward artificial neural network capable of modeling complex nonlinear relationships in charging load data.
- Gated recurrent unit (GRU) [18]: A variant of the traditional RNN incorporates gating mechanisms to address the vanishing gradient problem.
- Long short-term memory (LSTM) [19]: This model is distinguished by its specially designed memory cells based on RNN architecture. These cells enable the network to retain information over prolonged durations.
- Dlinear [20]: This model focuses on linear relationships in time series forecasting. It is particularly effective in scenarios where the data exhibits linear trends.

For the results from the fast charging dataset in Table I, the proposed MSTEM model demonstrates superior performance across various metrics and forecasting steps. Specifically, MSTEM achieves the lowest MSE, MAE, and RMSE scores in the 12 h and 24 h forecasting steps, with MSE scores of 22.95 and 22.61 and RMSE scores of 4.79 and 4.75, respectively. The MLP model also performs relatively higher accuracy in results, particularly in minimizing MSE, whereas the HI model generally exhibits lesser performance. Both the GRU and LSTM models demonstrate moderate effectiveness, which considers the limited electric load patterns and temporal dependencies. The Dlinear model, focusing on linear relationships, shows commendable results, particularly in shorter forecast steps. These results highlight the robust predictive accuracy and reliability of MSTEM in fast charging scenario. The consistent outperformance in MSE, MAE, and RMSE underscores its effectiveness in handling the complexities associated with fast charging load forecasting.

¹<https://data.pkc.gov.uk/>

TABLE II
RESULTS OF COMPARABLE EXPERIMENTS ON SLOW CHARGING DATASET.

Step	Metrics	MA	HI	MLP	GRU	LSTM	Dlinear	MSTEM
6 h	MSE	2.64	3.43	1.54	1.61	1.61	1.52	1.52
	MAE	0.57	0.60	0.48	0.60	0.64	0.50	0.45
	RMSE	1.63	1.85	1.24	1.27	1.27	1.23	1.23
12 h	MSE	3.08	3.64	1.67	1.89	2.03	1.64	1.66
	MAE	0.71	0.64	0.51	0.69	0.70	0.53	0.49
	RMSE	1.75	1.91	1.29	1.38	1.43	1.28	1.29
24 h	MSE	2.13	3.76	1.72	2.08	2.08	1.70	1.77
	MAE	0.60	0.67	0.52	0.71	0.71	0.53	0.50
	RMSE	1.46	1.94	1.31	1.44	1.44	1.30	1.33

Table II presents the comparative analysis results from the slow charging dataset. In the 6 h forecasting step, Dlinear and MSTEM both achieve the lowest MSE score of 1.52. MSTEM obtains the best MAE at 0.45 and tying with Dlinear for the top RMSE score of 1.23. In the 12 h and 24 h forecasting step, MSTEM maintains the lead in MAE with 0.49 and 0.50, respectively. Dlinear performs well in MSE and RMSE, which indicating its strength in shorter-term forecasts. These results highlights the effectiveness of Dlinear and MSTEM in slow charging predictions. In contrast, the HI and MA models show lower performance, and GRU and LSTM exhibit moderate results. A possible reason is that these models have a potential mismatch with the distinct characteristics of slow charging data, which typically involves longer charging cycles and less variability compared to fast charging.

Overall, the experimental results of both fast and slow charging datasets demonstrates that the proposed MSTEM outperforms other models, especially in MAE and RMSE metrics. The superior performance highlights the predictive accuracy and reliability of the proposed MSTEM.

IV. CONCLUSIONS

This paper proposes the MSTEM to enhance short-term load forecasting at EVCS. MSTEM effectively integrates a LSTM network to adeptly handle the complex temporal dynamics of EVCS electric loads, capturing nonlinear long-term dependencies. It further leverages a residual fusion approach to supplement its analysis of linear characteristics. Additionally, MSTEM accounts for the interrelationship of different charging stations and the hierarchical temporal patterns in charging loads, employing a multiscale graph neural network to discern spatial dynamics across various time scales. The efficacy of MSTEM has been validated through datasets encompassing both fast and slow charging loads at EVCS in Perth, UK. The experimental results demonstrate that the proposed MSTEM outperforms other models across multiple evaluation metrics.

REFERENCES

- [1] J. L. Richter, "A circular economy approach is needed for electric vehicles," *Nature Electronics*, vol. 5, no. 1, pp. 5–7, 2022.
- [2] Z. Lin, P. Wang, S. Ren, and D. Zhao, "Economic and environmental impacts of evs promotion under the 2060 carbon neutrality target—a cge based study in shaanxi province of china," *Applied Energy*, vol. 332, p. 120501, 2023.
- [3] C. He, J. Zhu, J. Lan, S. Li, W. Wu, and H. Zhu, "Optimal planning of electric vehicle battery centralized charging station based on ev load forecasting," *IEEE Transactions on Industry Applications*, vol. 58, no. 5, pp. 6557–6575, 2022.
- [4] J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, and L. Xi, "Review and prospect of data-driven techniques for load forecasting in integrated energy systems," *Applied Energy*, vol. 321, p. 119269, 2022.
- [5] F. Lu, J. Lv, Y. Zhang, H. Liu, S. Zheng, Y. Li, and M. Hong, "Ultra-short-term prediction of ev aggregator's demand response flexibility using arima, gaussian-arima, lstm and gaussian-lstm," in *Proceeding of the 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST)*. Guangzhou, China: IEEE, Feb 2021, pp. 1775–1781.
- [6] F. Ren, C. Tian, G. Zhang, C. Li, and Y. Zhai, "A hybrid method for power demand prediction of electric vehicles based on sarima and deep learning with integration of periodic features," *Energy*, vol. 250, p. 123738, 2022.
- [7] T. Zhang, Y. Huang, H. Liao, and Y. Liang, "A hybrid electric vehicle load classification and forecasting approach based on gbdt algorithm and temporal convolutional network," *Applied Energy*, vol. 351, p. 121768, 2023.
- [8] W. Khan, W. Somers, S. Walker, K. de Bont, J. Van der Velden, and W. Zeiler, "Comparison of electric vehicle load forecasting across different spatial levels with incorporated uncertainty estimation," *Energy*, vol. 283, p. 129213, 2023.
- [9] Y. Huang, Y. Zhao, Z. Wang, X. Liu, H. Liu, and Y. Fu, "Explainable district heat load forecasting with active deep learning," *Applied Energy*, vol. 350, p. 121753, 2023.
- [10] K. Benidis, S. S. Rangapuram, V. Flunkert, Y. Wang, D. Maddix, C. Turkmen, J. Gasthaus, M. Bohlik-Schneider, D. Salinas, L. Stella *et al.*, "Deep learning for time series forecasting: Tutorial and literature survey," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–36, 2022.
- [11] Z. Wang, X. Liu, Y. Huang, P. Zhang, and Y. Fu, "A multivariate time series graph neural network for district heat load forecasting," *Energy*, vol. 278, p. 127911, 2023.
- [12] J. Shi, W. Zhang, Y. Bao, D. W. Gao, and Z. Wang, "Load forecasting of electric vehicle charging stations: Attention based spatiotemporal multi-graph convolutional networks," *IEEE Transactions on Smart Grid*, vol. Early Access, 2023.
- [13] E. Spiliotis, F. Petropoulos, N. Kourentzes, and V. Assimakopoulos, "Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption," *Applied Energy*, vol. 261, p. 114339, 2020.
- [14] F. Mohammad, D.-K. Kang, M. A. Ahmed, and Y.-C. Kim, "Energy demand load forecasting for electric vehicle charging stations network based on convlstm and biconvlstm architectures," *IEEE Access*, vol. 11, pp. 67 350–67 369, 2023.
- [15] F. Kaytez, "A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption," *Energy*, vol. 197, p. 117200, 2020.
- [16] Y. Cui, J. Xie, and K. Zheng, "Historical inertia: A neglected but powerful baseline for long sequence time-series forecasting," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. Queensland, Australia: Association for Computing Machinery, 2021, p. 2965–2969.
- [17] G. Dudek, "Multilayer perceptron for short-term load forecasting: from global to local approach," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3695–3707, 2020.
- [18] Y. Wang, M. Liu, Z. Bao, and S. Zhang, "Short-term load forecasting with multi-source data using gated recurrent unit neural networks," *Energies*, vol. 11, no. 5, p. 1138, 2018.
- [19] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, "Long term load forecasting with hourly predictions based on long-short-term-memory networks," in *Proceedings of the 2018 IEEE Texas Power and Energy Conference*. College Station, TX, USA: IEEE, Feb 2018, pp. 1–6.
- [20] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the 37th AAAI conference on artificial intelligence*, vol. 37, no. 9. Washington DC, USA: Association for the Advancement of Artificial Intelligence, Feb 2023, pp. 11 121–11 128.