RLIDT: A Novel Reinforcement Learning-Infused Deep Transformer Model for Multivariate Electricity Load Forecasting

1st Seyed Mohammad Jafar Jalali
School of Computer Science
Edith Cowan University
Joondalup, Australia
s.jalali@ecu.edu.au

2nd Fateme Fahiman
School of Computing and Information Systems
University of Melbourne
Melbourne, Australia
fateme.fahiman@gmail.com

3rd Mahdi Khodayar
Department of Computer Science
University of Tulsa
Tulsa, Oklahoma, United States
mahdi-khodayar@utulsa.edu

***Abstract*—Electricity load forecasting plays a crucial role in the management of electricity power grids, enhancing operational efficiency, ensuring network reliability, facilitating infrastructure planning, and promoting energy sustainability. As the complexity of energy consumption patterns increases, traditional forecasting techniques struggle to accommodate the intricate and nonlinear temporal dynamics characteristics present within the data. This paper introduces a novel hybrid model called RLIDT that merges reinforcement learning (RL) with the deep transformer architecture to address the complicated challenges associated with load forecasting effectively. The integration of RL for hyperparameter optimization within the transformer framework not only utilizes their respective advantages but also provides a dynamic, adaptive model that exhibits versatility, robustness, and enhanced predictive accuracy through continuous learning. The experimental investigations conducted on real-world datasets have clearly demonstrated the remarkable advantage of the proposed RLIDT model when compared to traditional methods.**

***Index Terms*—Electricity load forecasting, Deep Transformer Neural Network, Reinforcement learning, Time series forecasting, Hyperparameter tuning**

I. Introduction

Electricity load forecasting is one of the most critical tasks for power system management, as it has wide-ranging involvement in the efficiency of operations, economic feasibility, and long-term sustainability of energy systems [1], [2]. The accuracy of these forecasts is critical for grid reliability, facilitating the integration of sustainable energy sources, and optimising generation and distribution strategies [3]. Traditional forecasting techniques, mainly dependent on statistical time series models, have played a fundamental role but often face challenges in capturing the complex, non-linear patterns which is inherent in electricity consumption patterns [4], [5].

Recent advancements in the field of machine learning have sparked a new era of forecasting methodologies [6], [7]. Transformer-based models, notably, have demonstrated promise in addressing the complexities of load forecasting due to their capability to capture long-term dependencies within time series data [8]. The work by Ran et al. utilises the transformer architecture in conjunction with the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) technique for the purpose of short-term load forecasting. This study effectively showcased the model's ability to capture intricate temporal dependencies accurately. [9]. Zhang et al. [10] , in a similar manner, improve the transformer's capabilities by utilising a time augmentation approach. This particular methodology serves to enrich the input data, as a result leading to more accurate short-term forecasts. In the domain of load forecasting for end-user transformers, Chen et al. [11] apply deep learning techniques to address the fluctuations introduced by electric heating loads effectively. This study demonstrates a novel perspective on the application of transformer models in a more granular setting. By utilising the transformers-based approach, Wang et al. expand the scope of applications to include multi-energy load forecasting within an integrated energy system. Through this study, they illustrate the flexibility and resilience of the architecture, especially when confronted with complex multi-dimensional data [12].

In [13], the authors examined the challenges related to day-ahead load forecasting and introduced a novel network architecture, Forwardformer. This innovative architecture is an extension of the traditional transformer model. The Forwardformer is designed with the intention of enhancing computational efficiency and improving prediction accuracy. This is achieved through the utilisation of a multi-scale forward self-attention mechanism, accompanied by a unique correction structure consisting of an encoder and dual decoders. The effectiveness of the Forwardformer method was validated by conducting experiments on datasets sourced from both China and America. These experiments demonstrated a notable reduction in runtime and improved level of accuracy in the domain of forecasting, specifically noticeable during weekends and holidays, providing a new and effective solution to the DALF challenge.

In addition to these transformer-centric approaches, Reinforcement Learning (RL) presents a decision-making framework that dynamically adjusts to variations in electricity demand by learning optimal actions through interactions with the environment [14]. This study delves into the intersection of Reinforcement Learning and transformer models, termed RLIDT, aiming to tackle the challenges in load forecasting through an innovative methodological fusion. While the literature shows individual progress made by each approach in forecasting, the full potential of integrating RL for hyperpa-

979-8-3503-7115-4/24/$31.00 ©2024 IEEE

rameter optimization within the transformer architecture has not yet been fully realized.

This study contributes significantly to the existing body of knowledge, with key innovations including:

- The introduction of RLIDT, a novel RL algorithm that targets the distinctive features of time series data in electricity load forecasting, integrating deep transformer models. This method enhances the model's ability to effectively manage temporal dynamics in load patterns. Notably, the RL component in RLIDT is specifically tasked with optimizing the hyperparameters of the transformer, enabling it to adapt more effectively to the evolving nature of time series data.

- Our study demonstrates innovation through the integration of the deep transformer model with an RL-based approach to advance the process of sequential decision-making in time series forecasting. This unique combination facilitates the management of time-dependent information through an enhanced strategy that progressively refines prediction accuracy, particularly by optimizing the transformer's configuration for better adaptability to changing patterns in the data.

RLIDT poses a unique capability to proficiently handle multivariate time series data, crucial for electricity load forecasting. Its architectural design allows for efficient processing and analysis of numerous correlated time-dependent variables simultaneously. This capability enables RLIDT to capture intricate interrelationships among diverse factors, thereby enhancing its forecasting accuracy. The proficiency of the model in effectively handling multivariate inputs represents a crucial advancement, offering a more comprehensive and authentic methodology for predicting electricity load in scenarios where various factors impact the load. The integration of RL for hyperparameter optimization within the transformer framework represents a significant step forward in developing more dynamic and responsive forecasting models for complex time series data.

The rest of the paper is organized as follows: Section II presents our methodology, including data preprocessing, the RL algorithm, and the transformer model adaptation. Section III describes our experimental setup and presents the results, highlighting the model's performance, interprets these findings, discusses their broader implications, and considers potential limitations. Finally, the conclusion in Section IV reflects on the impact of our work and proposes directions for future research.

II. Problem Formulation of the Proposed Model

This study addresses the electricity load forecasting problem by proposing a hybrid model, RLIDT, which synergizes Reinforcement Learning (RL) with deep transformer architectures. This model aims to harness temporal pattern recognition and sequential decision-making to enhance forecasting accuracy.

A. Data Description

Let us assume a time series x_t $1 \leq t \leq T$ where each time step measurement x_t at time t is a 5-dimensional vector that includes the electricity load and four additional measurements such as temperature, humidity, precipitation, and wind speed. Given a historical time window of measurements $< x_{t-m}, x_{t-m+1}, ..., x_t >$ at each time t, the objective is to estimate the one hour ahead load value corresponding to time $t+1$, denoted by y_t (i.e., the output of load forecast at t). In this problem setting, we consider the input historical time window to be the hourly measurements of one week, hence, the input measurements length is equal to $m = 7 \times 24 = 168$. The dataset is segmented into training $\mathcal{D}_{train}$, validation $\mathcal{D}_{val}$, and testing $\mathcal{D}_{test}$ subsets.

B. Hybrid RLIDT Model

Our hybrid RLIDT model integrates a transformer network with a RL framework to optimize the transformer's hyperparameters. The model is characterized by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, encompassing:

- **State Space** $\mathcal{S}$: Each state s_t includes an encoded context from the transformer model, aggregating historical data across a window of $m+1$ steps, from $t-m$ to t.
- **Action Space** $\mathcal{A}$: Actions a_t consist of possible hyperparameter settings for the transformer model, influencing its forecasting performance.
- **Transition Dynamics** $\mathcal{T}$: The state transition function, encapsulated by the transformer model, processes the current state to predict the next hour's load.
- **Reward Function** $\mathcal{R}$: The reward r_t is calculated as the negative Mean Squared Error (MSE) between the transformer's predictions and the actual load values, encouraging hyperparameter choices that reduce forecasting errors.
- **Discount Factor** γ: The discount factor γ underscores the importance of immediate forecasting accuracy.

C. Transformer Network for Load Forecasting

The transformer model in our hybrid approach is essential for predicting future electricity load values. Its structure comprises an encoder and a decoder, each with specific functions:

- **Encoder**: The encoder function **Enc** takes the input sequence $\mathbf{X}$ and transforms it into a set of contextual representations $\mathbf{Z}$. This is formally expressed as:

$$\mathbf{Z} = \textbf{Enc}(\mathbf{X})$$

  where $\mathbf{X} = [x_{t-m}, \ldots, x_t]$ spans the historical window of $m+1$ time steps.
- **Decoder**: The decoder function **Dec** uses the contextual representations $\mathbf{Z}$ to forecast the load for the next hour $t+1$. The forecasted load $\hat{y}_{t+1}$ is given by:

$$\hat{y}_{t+1} = \textbf{Dec}(\mathbf{Z})$$

Self-Attention Mechanism: The self-attention mechanism in the encoder allows the model to weigh different parts of the input sequence while generating the encoded representations. This is mathematically represented as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are query, key, and value matrices, respectively, and d_k is the dimension of the keys.

D. DRL for Hyperparameter Optimization

The deep reinforcement learning (DRL) component is tasked with optimizing the hyperparameters of the transformer to enhance forecast accuracy. This process involves adjusting hyperparameters such as the learning rate, the number of layers, the size of the attention heads, and the dropout rate. The optimization process can be described as follows:

- **Policy Learning**: The policy $\pi(a|s;\theta)$ represents the probability of choosing action a (a hyperparameter configuration) given state s (the current model state), parameterized by θ. The policy is learned by maximizing the expected reward:

$$\theta^* = \arg\max_{\theta} \mathbb{E}_{\pi(a|s;\theta)}[R(s,a)]$$

where $R(s,a)$ is the reward function.

- **Hyperparameter Actions**: The action space $\mathcal{A}$ includes possible configurations of hyperparameters such as:
  - Learning Rate (lr)
  - Number of transformer Layers (n_{layers})
  - Size of Attention Heads (n_{heads})
  - Dropout Rate (dropout)
- **Reward Function**: The reward function is the negative MSE between the predicted and actual load values:

$$R(s,a) = -\text{MSE}(\hat{y}_{t+1}, y_{t+1}) = -\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{t+1}^{(i)} - y_{t+1}^{(i)})^2$$

where N is the number of predictions, and $\hat{y}_{t+1}^{(i)}, y_{t+1}^{(i)}$ are the predicted and actual loads, respectively.
- **Policy Gradient Update**: The parameters of the policy network are updated using the policy gradient method:

$$\theta_{\text{new}} = \theta_{\text{old}} + \alpha \nabla_{\theta} \mathbb{E}_{\pi(a|s;\theta)}[R(s,a)]$$

where α is the learning rate.

This approach enables the DRL agent to iteratively learn the optimal configuration of the transformer's hyperparameters, thereby improving the model's ability to forecast electricity load accurately. The policy gradient update ensures that the hyperparameter adjustments are made in the direction of increasing the expected reward, which is tied to forecasting accuracy.

E. Training Procedure

The training of the RLIDT model involves the simultaneous learning of both the transformer network and the deep reinforcement learning (DRL) agent, aimed at optimizing load forecasting. This training process is iterative and adaptive, responding to the evolving dynamics of the data:

- **Transformer Training**: The transformer is trained to minimize the forecasting error. The parameters θ_{Trans} of the transformer are updated based on the gradient of the loss function $\mathcal{L}$, which measures the discrepancy between the predicted and actual load values. The update rule for the transformer parameters is:

$$\theta_{\text{Trans}}^{(\text{new})} = \theta_{\text{Trans}}^{(\text{old})} - \eta_{\text{Trans}} \nabla_{\theta_{\text{Trans}}} \mathcal{L}(\theta_{\text{Trans}})$$

where η_{Trans} is the learning rate for the transformer, and the loss function $\mathcal{L}$ is typically the Mean Squared Error (MSE):

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{t+1}^{(i)} - y_{t+1}^{(i)})^2$$

- **DRL Agent Training**: The DRL agent updates the policy network parameters θ_π to optimize the selection of hyperparameters for the transformer. The update follows the policy gradient method, where the expected reward guides the learning:

$$\theta_{\pi}^{(\text{new})} = \theta_{\pi}^{(\text{old})} + \eta_\pi \nabla_{\theta_\pi} \mathbb{E}\left[r_t \nabla_{\theta_\pi} \log \pi(a_t \mid s_t; \theta_\pi)\right]$$

where η_π is the learning rate for the policy network, r_t is the reward at time t, a_t is the action (hyperparameter configuration) chosen at time t, and s_t represents the current state of the transformer model.

This iterative approach ensures that the transformer model is continually refined based on the prediction error, while the DRL agent adjusts its policy based on the reward feedback. This dual learning mechanism allows the hybrid RLIDT model to adapt and improve its forecasting accuracy over time, effectively responding to changes and trends in electricity load data.

F. Evaluation Metrics

The evaluation of performance is conducted through employing of established forecasting metrics as follows:

- Mean Absolute Error (MAE): $\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|\hat{y}_t - y_t|$.
- Root Mean Squared Error (RMSE): $\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}$.
- Mean Absolute Percentage Error (MAPE): $\text{MAPE} = \frac{100\%}{T}\sum_{t=1}^{T}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$.

III. Numerical Results

Our study aimed to compare the performance of our proposed RLIDT model with various state-of-the-art neural network models for time-series forecasting. To facilitate this comparison, we utilized a dataset acquired from Kaggle [15], specifically focusing on the city of Tocumen in Panama City. This dataset comprises 8760 hourly samples, reflecting a full year's data. We also performed a normalization process in order to ensure the comparability and uniformity of the dataset. We also transformed the normalized dataset into sequences, each consisting of 168 consecutive hours (one week), to predict the electricity demand for the subsequent hour.

In the next step, we divide the dataset into training, validation, and test sets, maintaining an 80-10-10 split, respectively. The comparison involved the following models: multi-layer perceptron (MLP), recurrent neural network (RNN), long short-term memory (LSTM) network, gated recurrent unit (GRU), temporal convolutional network (TCN), deep transformer and proposed RLIDT. It should be also mentioned that to accommodate the input requirements of the neural network models, especially the transformer, the data was reshaped into a format suitable for each model. This involved adjusting the shape to fit the dimensions of batch size, sequence length, and the number of features.

In the experimental configuration for evaluating the RLIDT model, a variety of cutting-edge Python libraries were utilised to facilitate the development and examination of our proposed RLIDT model alongside other deep neural network models. Among the libraries that played a crucial role in this context, Pytorch stood out as it served as the principal framework for the implementation and training of both the deep transformer model and the components related to reinforcement learning. The dynamic computation graph and extensive ecosystem of PyTorch positioned it as a highly suitable option for this investigation. We conducted a comparison between our proposed model and the baseline models, namely MLP, RNN, LSTM, GRU, and TCN. The best combination of hyperparameters that we selected for the models rather than our proposed model is are selected based on trail and error as well as performing a greedy search in order to obtain the best possible accuracy from them. Furthermore, PyTorch was utilised to ensure a uniform platform throughout all experiments. Our experiments were executed on a high-capacity computing system equipped with cutting-edge graphics processing unit (GPU) acceleration to manage the computationally intensive tasks of training and assessing deep learning models. This configuration guaranteed the effective processing of large datasets and complex neural network architectures, thus allowing us to carry out comprehensive and reliable assessments of the RLIDT model's performance in electricity load forecasting. The following table presents the results of our proposed RLIDT framework in comparison to alternative contenders with respect to diverse assessment criteria.

TABLE I
COMPARISON OF DIFFERENT MODELS ON TEST SET USING MAE, RMSE, AND MAPE.

Model	MAE	RMSE	MAPE
MLP	0.8279	0.9924	1.0366
RNN	0.1939	0.2801	0.9549
LSTM	0.2421	0.3462	1.0673
GRU	0.2066	0.3054	0.9571
TCN	0.2706	0.3532	0.8929
Transformer	0.2267	0.3319	1.0571
Proposed RLIDT	**0.1081**	**0.1772**	**0.5588**

The performance metrics displayed in the table deliver a thorough and comprehensive examination of the forecasting precision and effectiveness of different machine learning and deep learning algorithms. The combination of the reinforcement learning and transformer model, as proposed (RLIDT), showcases a remarkably higher level of efficacy in contrast to conventional and profound learning models such as MLP, RNN, LSTM, GRU, TCN, and the standard transformer model. This is apparent from the considerably reduced MAE value of 0.1081, RMSE value of 0.1772, and MAPE value of 55.8%. The significant decrease in these measures of error suggests an enhanced level of precision in forecasting and an enhanced understanding of the underlying patterns in the data.

The capability of the RLIDT to acquire intricate temporal interconnections and generate more precise forecasts is potentially an outcome of its sophisticated structure, which brings together the advantages of transformer models with the principles of reinforcement learning. In comparison, conventional machine learning algorithms such as the MLP, despite their straightforwardness and simplicity of execution, demonstrate noticeably elevated levels of inaccuracies, as evidenced by the utmost MAE, RMSE, and especially MAPE, which stands at more than 103%.This implies an inadequate match for intricate time series forecasting undertakings. The other deep learning models such as RNN, LSTM, GRU, and TCN, demonstrate diverse levels of effectiveness. The RNN and GRU models perform comparatively better than LSTM and TCN within this particular framework. The divergence in results can be attributed to the contrasting methodologies employed in processing sequential data and the varying capabilities to capture temporal association. The standard transformer model, despite surpassing the MLP, falls behind in comparison to the proposed RLIDT. This suggests that the integration of reinforcement learning substantially improves the forecasting capability. Overall, the above-mentioned findings emphasize the significance of selecting suitable methodologies for the prediction of time series, specifically in situations that require the highest level of precision and accuracy.

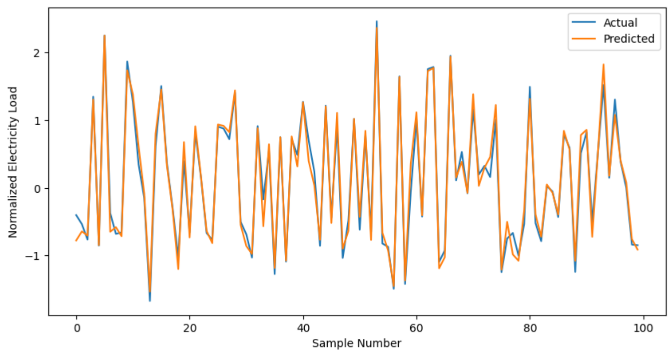

Fig. 1. The last 100 samples of test set for actual vs. predicted load values generated by our proposed RLIDT

Fig. 1 depicts the predicted values adhering closely to the pattern and fluctuations of the actual data. This close correspondence indicates the effectiveness of the model in capturing both the inherent underlying pattern and the seasonal fluctuations within the dataset. The precision of our suggested framework is further underscored during periods of sudden changes or fluctuations in the data, as the anticipated trajectory demonstrates an ability to conform and react to these fluctuations with minimal delay. However, there may be certain regions in which the anticipated values deviate marginally from the real values, a phenomenon frequently observed in the field of time series load forecasting. These inconsistencies could be attributed to inherent fluctuations present within the dataset, unforeseen and sudden shifts in the trend, or limitations in the model's capacity to capture periods characterised by high volatility accurately.

IV. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this study, we introduced RLIDT, an innovative model that integrates deep transformer architecture with reinforcement learning techniques, specifically tailored for electricity load forecasting. Our approach leverages the advanced capabilities of transformer architectures in capturing complex temporal relationships within electricity consumption data. Additionally, the incorporation of RL not only adjusts to evolving patterns in the data but also optimizes the hyperparameters of the transformer model, thereby significantly enhancing prediction accuracy. The superiority of RLIDT over traditional neural network models, such as MLP, RNN, LSTM, GRU, TCN, and a standard transformer, is demonstrated by the experimental results, which have been validated on a dataset consisting of 8760 hourly samples. The adaptability and precision of our model are particularly remarkable when it comes to handling fluctuating loads, which are a vital element of modern energy systems.

Future research endeavours may be directed towards various areas in order to enhance and broaden the functionalities of RLIDT. One possible approach involves investigating the incorporation of external variables, such as economic indicators, energy policies like electrification measures, and the growth of behind-the-meter solar power, which possesses the ability to exert a substantial impact on the demand for electricity. Additionally, adapting the model for real-time forecasting in smart grid environments could yield significant knowledge for regulating demand-side management. Similarly, exploring the scalability of RLIDT for larger, more diverse datasets as well as its utilisation in diverse geographical areas, would also prove advantageous. Furthermore, the expansion of the model to not only predict load but also forecast prices and renewable energy production has the potential to transform it into a more all-encompassing instrument for the management of energy. Finally, examining the interpretability of the model and providing insights into the decision-making process of the reinforcement learning component may potentially enhance trustworthiness and reliability in practical applications. Since the computational time for using the RL algorithm in the training process of deep transformers is a bit high, it is highly suggested to explore different techniques such as experience replay, batch training, parallelization, and parameter sharing to expedite the training process.

Acknowledgment

This research is supported by the National Science Foundation under grant ECCS-2223628.

References

[1] R. Wazirali, E. Yaghoubi, M. S. S. Abujazar, R. Ahmad, and A. H. Vakili, "State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques," *Electric Power Systems Research*, vol. 225, p. 109792, 2023.

[2] M. Saffari, M. Khodayar, M. E. Khodayar, and M. Shahidehpour, "Behind-the-meter load and pv disaggregation via deep spatiotemporal graph generative sparse coding with capsule network," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[3] S. M. J. Jalali, P. Arora, B. Panigrahi, A. Khosravi, S. Nahavandi, G. J. Osório, and J. P. Catalão, "An advanced deep neuroevolution model for probabilistic load forecasting," *Electric Power Systems Research*, vol. 211, p. 108351, 2022.

[4] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustainable cities and society*, vol. 35, pp. 257–270, 2017.

[5] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 1104–1122, 2017.

[6] S. M. J. Jalali, S. Ahmadian, A. Khosravi, M. Shafie-khah, S. Nahavandi, and J. P. Catalão, "A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8243–8253, 2021.

[7] A. Dogra, A. Anand, and J. Bedi, "Consumers profiling based federated learning approach for energy load forecasting," *Sustainable Cities and Society*, vol. 98, p. 104815, 2023.

[8] Z. Zhang, X. Wang, and Y. Gu, "Sageformer: Series-aware graph-enhanced transformers for multivariate time series forecasting," *arXiv preprint arXiv:2307.01616*, 2023.

[9] P. Ran, K. Dong, X. Liu, and J. Wang, "Short-term load forecasting based on ceemdan and transformer," *Electric Power Systems Research*, vol. 214, p. 108885, 2023.

[10] G. Zhang, C. Wei, C. Jing, and Y. Wang, "Short-term electrical load forecasting based on time augmented transformer," *International Journal of Computational Intelligence Systems*, vol. 15, no. 1, p. 67, 2022.

[11] Q. Chen, M. Xia, T. Lu, X. Jiang, W. Liu, and Q. Sun, "Short-term load forecasting based on deep learning for end-user transformer subject to volatile electric heating loads," *IEEE Access*, vol. 7, pp. 162 697–162 707, 2019.

[12] C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, and K. Zhang, "A transformer-based method of multienergy load forecasting in integrated energy system," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2703–2714, 2022.

[13] K. Qu, G. Si, Z. Shan, Q. Wang, X. Liu, and C. Yang, "Forwardformer: Efficient transformer with multi-scale forward self-attention for day-ahead load forecasting," *IEEE Transactions on Power Systems*, 2023.

[14] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 209–220, 2020.

[15] E. A. Madrid, "Short-term electricity load forecasting, panama," 2020. [Online]. Available: https://www.kaggle.com/datasets/ernestojaguilar/shortterm-electricity-load-forecasting-panama/