Proceedings of the 4th International Conference on Functional Materials and Civil Engineering
DOI: 10.54254/2755-2721/2026.35335

Deep Learning-Based Load Forecasting Approach for Smart Grids

Qiuhui Jiang

Faculty of Artificial Intelligence, Shanghai University of Electric Power, Shanghai, China
jhq1027sqt@gmail.com

Abstract. The high penetration of renewable energy and the large-scale integration of electric vehicles have significantly enhanced the nonlinearity and non-stationarity of the smart grid load sequence. Traditional time series models struggle to meet the prediction requirements in complex scenarios. This paper focuses on three kinds of deep learning load forecasting methods: recurrent neural network variants, CNN hybrid architectures, and Transformers, and systematically conducts a comparative analysis of them from three dimensions: prediction accuracy, long-term dependency modeling ability and computational efficiency. The results show that the three types of architectures each have distinct advantages, and there is no universal optimal solution: the recurrent variant balances accuracy and efficiency in stationary short-term prediction, the CNN hybrid architecture is more robust in strong coupling scenarios such as extreme weather, and the Transformers' trend modeling ability is outstanding but the computational cost for long sequences is large. The pros and cons of each method are highly dependent on the stationarity of the specific scenario, the forecasting horizon, and resource constraints. The model selection must be based on the task characteristics rather than the pursuit of general solutions.

Keywords: Smart grid, Load forecasting, Deep learning, Comparative analysis

1. Introduction

The global energy system is undergoing a profound transformation. The high penetration of renewable energy, the large-scale integration of electric vehicles, and the widespread deployment of distributed energy resources are reshaping the operating paradigm of the power grid at an unprecedented speed. This transformation accelerates decarbonization and makes the load characteristics of the power grid more complex: the minute-level fluctuations of intermittent photovoltaic output, the random fluctuations of electric vehicle charging demand, the dynamic adjustment of demand response strategy, and the uncertainty of meteorological conditions and social behavior, which significantly enhances the nonlinearity and non-stationarity of the load sequence [1]. As a result, traditional time series models (such as ARIMA and exponential smoothing) struggle to capture complex meteorological coupling and spatiotemporal dependencies due to linear assumptions and shallow feature extraction capabilities, which restricts the prediction accuracy under extreme weather [2].

DOI: 10.54254/2755-2721/2026.35335

Deep learning provides a new paradigm for solving the above challenges. Deep networks automatically extract implicit representations of load data through hierarchical nonlinear transformations, and they have significant advantages in characterizing the interactions between multi-scale temporal dependencies and external factors. Biswal et al. systematically reviewed the application of Long Short-Term Memory (LSTM), CNN and other infrastructures in smart grid load forecasting, and confirmed the superior performance of deep learning models in processing high-dimensional nonlinear data [1]. Ali et al. systematically reviewed the improvement trend of deep learning models in short-term load forecasting accuracy compared with traditional time series methods, and confirmed the significant advantages of hybrid architectures in dealing with high-dimensional nonlinear load data [3]. Aiming at the problem of high-precision identification and prediction of low-voltage load characteristics, Cui et al. constructed a hybrid deep learning framework combining improved CNN, LSTM and Transformers, and achieved load identification performance significantly better than the comparison model on real-world datasets [4].

Most existing reviews tend to evaluate the algorithmic performance of various deep learning models separately. In general, there is little systematic correlation among data characteristics, model mechanisms, and application scenarios. Especially among the three mainstream architectures of recurrent neural network variants, convolutional hybrid architectures, and Transformer models, there is still a lack of systematic horizontal comparison in terms of prediction accuracy, long-term dependency modeling capabilities, and computational overhead. In view of this, this paper focuses on the above three types of architectures, and selects two representative empirical cases in recent years to analyze each type. The comparative analysis is carried out based on the three dimensions of prediction accuracy, long-term dependency capture and computational efficiency, aiming to provide a clear reference framework for model selection in the field of smart grid load forecasting.

2. Background and methodological motivation of modern load forecasting

2.1. Core challenges in load forecasting

The high penetration of renewable energy and the large-scale integration of electric vehicles in modern power systems have significantly enhanced the nonlinearity and nonstationarity of load series. Traditional time series models are limited by linear assumptions and shallow feature extraction capabilities, and they struggle to accurately describe the complex dynamic characteristics of power grid load [2].

The key to load forecasting is to capture temporal dependencies across multiple time scales, and minute-level fluctuations, diurnal cycles and even seasonal trends are nested with each other. The standard recurrent neural network is prone to the problem of vanishing gradients when dealing with long sequences, which makes it difficult to effectively transmit distant historical information to the current prediction [1].

The load is affected by the coupling of multi-modal factors such as meteorological conditions, temporal features and economic signals. These data are significantly different in time granularity, numerical units and information structure. It is difficult for a single model architecture to extract local fluctuation patterns and global contextual information at the same time, which requires the prediction model to have hierarchical multi-scale feature learning capabilities [3].

In interconnected power systems, the load dynamics of a single node not only depend on local factors, but also may be causally related to the operating state of the remote node through the grid topology. This requires that the model can transcend spatial distances and temporal spans, and simultaneously analyze the complex interactions among multiple nodes [5].

DOI: 10.54254/2755-2721/2026.35335

2.2. Applicability of deep learning

The reason why deep learning is widely recognized in the field of load forecasting is its inherent advantages in multiple dimensions. Specifically, the deep neural network can automatically extract high-order features from load data through multi-layer nonlinear transformation, thereby significantly reducing the dependence on artificial feature engineering. Recurrent Neural Networks (RNNs) and their variants (such as LSTM, Gated Recurrent Unit (GRU)) deal with long-term dependencies in time series data through specially designed gating mechanisms, which makes them perform well in capturing the time correlation of load series. In contrast, convolutional neural networks are better at identifying local patterns, while the attention mechanism and Transformer model further improve the modeling ability of long-range dependencies. It is precisely these technical characteristics that are highly compatible with the core challenges faced by load forecasting, prompting deep learning to become the mainstream method in this field.

3. Three categories of deep learning architectures for smart grid load forecasting

3.1. Recurrent Neural Networks and their variants

Recurrent Neural Networks (RNNs) have become the first widely used deep learning method in the field of load forecasting due to their recurrent hidden state mechanism designed for sequence data [6]. Through the hidden state transfer between time steps, the historical information can be encoded into the current prediction. However, the standard RNN faces the problem of vanishing or exploding gradients in long sequence training, and it is difficult to learn distant temporal dependencies [7].

Long Short-Term Memory networks (LSTMs) achieve selective memory and forgetting of historical information through forgetting gate, input gate and output gate, so that the gradient can propagate stably [8]. Gated Recurrent Units (GRUs) further combine the forgetting gate and the input gate into a single update gate, which has a simpler structure and lower training overhead [3]. Bidirectional RNNs capture contextual information through both forward and backward paths, and have stronger feature extraction capabilities [7].

In terms of prediction accuracy, the recurrent variant is reliable in the short-term prediction of stable power consumption mode. In the empirical comparison of the Greek power system, the mean absolute percentage error (MAPE) of GRU in the short time domain of 1 hour is as low as 1.17%, and the accuracy and training efficiency are better than LSTM [9]. The stability of Hybrid Bidirectional Long Short-Term Memory (BiLSTM) in multi-step prediction is better than that of unidirectional models [7]. However, in extreme events, the error of a single recurrent architecture increases significantly.

In terms of long-term dependency modeling, LSTM and GRU are far superior to standard RNN, but information attenuation is still inevitable when the sequence is too long. In terms of computational efficiency, GRU parameters are reduced by about one-third compared with LSTM, and training efficiency is higher [3]. The bidirectional structure of BiLSTM doubles the computational cost, but provides a balance between accuracy and efficiency [7].

3.2. Convolutional Neural Networks and hybrid architectures

Convolutional Neural Networks (CNNs) automatically extract the local fluctuation and periodic characteristics in the load sequence by sliding the convolution kernel along the time axis, which have the advantages of parameter sharing, high training efficiency and translation invariance.

However, a single CNN is limited by the fixed receptive field, and it is difficult to model the long-term evolution across multiple days [3]. Therefore, CNNs are often combined with recurrent networks, such as using a one-dimensional convolutional layer to extract local features and then feeding them into LSTM or GRU layers for temporal modeling.

In terms of prediction accuracy, the CNN hybrid architecture is more robust in extreme scenarios. On the urban load dataset of eastern Mongolia, the hybrid model of CNN and extended multi-layer LSTM achieved MAPE, RMSE, and MAE for single-step, four-step, and thirty-two-step predictions that were lower than those of independent LSTM and GRU [10]. The CNN-GRU model integrates multi-scale one-dimensional convolutional layers, which can simultaneously capture spatio-temporal features at different time granularities [11]. In multi-area load forecasting of interconnected power grids, the CNN input channel can be extended to a graph structure to capture the spatial correlation between substations [12]. In addition, the MAPE of CNN-GRU during hurricanes is 11.68%, which is lower than that of pure recurrent architectures [12]. The GRU-Moving Average (GRU-MA) model combined with multi-head attention also verifies the long-term dependency capture potential of the hybrid architecture.

In terms of long-term dependency modeling, the long-term ability of the CNN hybrid architecture depends more on the back-end recurrent network, and the front-end convolution serves for local pattern extraction. In terms of computational efficiency, the computational cost of CNN hybrid architecture is lower than that of Transformers and slightly higher than GRU, which is a practical choice in resource-constrained scenarios.

3.3. Attention mechanism and transformer models

The attention mechanism allows the model to dynamically allocate weights and focus on the most critical time steps. Transformers are completely based on self-attention and have two significant advantages: it is not limited by sequence distance and can stably capture long-range dependencies; the computations can be fully parallelized and the training time is short. However, the complexity of standard self-attention increases quadratically with the sequence length [13].

Transformers possess unique strengths in capturing overall trends. In day-ahead forecasting on the Greek power system, they achieved accuracy comparable to LSTM but produced predictions closer to the original data distribution [14]. However, Transformers are more sensitive to high-frequency noise, and in certain scenarios their performance may not surpass that of hybrid architectures.

Long-term dependency modeling remains the most prominent strength of Transformers, as they theoretically enable direct connections between any two time steps. Although training is highly parallelizable, standard self-attention requires $O(L^2)$ memory. Informer reduces this complexity to $O(L \log L)$ via ProbSparse sparse attention and further lowers inference cost through self-attention distillation and a generative decoder [13]. Applied to load data from a substation in Nanchang, Informer used 1440 time steps as input and achieved superior prediction accuracy and efficiency compared to RNN and LSTM [10].

DOI: 10.54254/2755-2721/2026.35335

3.4. Comparative analysis and discussion

Table 1. Performance comparison of time series forecasting architectures

Key Dimension	Prediction Accuracy	Long-term Dependency Modeling	Computational Efficiency
Recurrent Variants	Accurate for short-term, high error on extremes	Moderate to strong	Relatively high
CNN Hybrid Architectures	Strong capability in multi-step prediction	Moderate, often dependent on recurrent/attention back-end	Moderate
Transformer Variants	Strong trend capture, noise-sensitive	Strong	Relatively low

Table 1 provides a concise comparison of the three architectures discussed above in terms of prediction accuracy, long-term dependency modeling, and computational efficiency.

Recurrent variants excel in short-term forecasting (e.g., GRU achieves a low 1-hour MAPE of 1.17% [9]), but their performance deteriorates during extreme events. CNN-based hybrid architectures demonstrate strong robustness under extreme conditions (e.g., MAPE=11.68% during hurricanes [12]). Transformers are powerful in capturing trends [14], yet sensitive to noise. Regarding long-term dependencies, Transformers outperform gated RNNs, which in turn surpass CNN hybrids, though at higher computational cost. In terms of computational efficiency, GRU is optimal [3], while Informer mitigates Transformer's complexity through sparse mechanisms [13].

3.5. Application scenarios and comprehensive comparison

Each model excels in specific application domains: recurrent variants (such as LSTM and GRU) are well-suited for high-quality, stationary data in short-term forecasting, offering low computational overhead and ease of deployment [3, 7, 9], but suffer reduced accuracy under non-stationary conditions. CNN-based hybrid architectures perform better in scenarios with multi-scale periodicity, significant external factor coupling, or extreme weather anomalies, showing superior robustness [12] and multi-scale feature extraction capability compared to pure recurrent models [11], although their long-term modeling capacity remains limited. Transformers (and Informer) are best suited for long-sequence forecasting requiring capture of ultra-long-term trends, offering the strongest global modeling ability and excellent training parallelism [13, 14], but they struggle with short-term fluctuations and demand substantial resources. Model selection should be guided by data stationarity, prediction time horizon, degree of external coupling, and resource constraints.

4. Open challenges and emerging trends

The three types of models compared above, namely recurrent variants, CNN hybrid architectures, and Transformers, face common challenges in real-world deployment, including data scarcity, efficiency limitations in long sequences, and privacy risks. This section explores emerging technical approaches that address these three dimensions.

The performance of the aforementioned recurrent variants and hybrid architectures is highly dependent on abundant historical samples, but in practical scenarios such as new charging stations and newly connected photovoltaic nodes, the available data is often extremely limited. Zero-shot learning and few-shot meta-learning frameworks provide new ideas for this. The core logic is not to train from scratch but to learn to adapt quickly to new tasks, which can effectively overcome the

DOI: 10.54254/2755-2721/2026.35335

dependence of recurrent variants and hybrid architectures on large-scale historical data. Moon et al.'s review integrates more than 150 peer-reviewed studies from 2019 to mid-2025, systematically summarizes the progress of cold-start applications of zero-shot and few-shot meta-learning frameworks in three scenarios of photovoltaic, wind power and load forecasting, and points out that the combination of transfer learning and spatiotemporal graph neural networks can effectively transfer the knowledge of existing energy assets to a data-sparse environment [15]. The engineering value lies in providing reliable forecasts for new facilities from the first day of operation.

The parallel computing advantage of Transformers is offset by the quadratic complexity of self-attention when the input sequence is long, resulting in a sharp increase in memory requirements, which is a core shortcoming of Transformers in long-sequence scenarios. In recent years, the structured State Space Model (SSM) has received significant attention as an alternative. Its representative architecture Mamba reduces the complexity to linear scale through a selective scanning mechanism, which significantly reduces the inference overhead while maintaining modeling ability comparable to Transformers [16], and provides a feasible path to break through the bottleneck of Transformers' computational efficiency. Menati et al. introduced this paradigm into the power system domain and proposed the PowerMamba multivariate time series forecasting model, which integrates the traditional state space model with deep learning methods. Compared with existing models, the prediction error is reduced by 7% on average and the model parameters are reduced by 43% in load and electricity price prediction tasks [16]. The state space model is reopening the possibility of processing longer sequences with fewer resources.

Cross-regional collaboration can improve prediction quality at individual nodes, but once user-side data is aggregated, privacy leakage risks emerge. This issue is particularly prominent when CNN hybrid architectures and Transformers process multi-node spatio-temporal data. Approaches such as graph convolution often require centralized data aggregation. Federated learning allows participants to train models locally and only share encrypted parameter updates, which alleviates this issue at the architectural level. In a systematic review of the application of federated learning in smart grids, Zhang et al. pointed out that federated learning achieves accuracy comparable to traditional centralized methods in load forecasting tasks [17]. Ding et al. further proposed an industrial federated forecasting framework that integrates meta-learning with differential privacy allocation. Compared with independent training, the MAPE is reduced by 1.99 percentage points on real industrial datasets and by 2.63 percentage points in new operational scenarios [18]. These works show that the federated framework is gradually offering practical privacy protection solutions while maintaining the benefits of collaboration.

5. Conclusion

The results show that there is no uniform ranking of the performance of these three architectures, but there is a clear task dependence.

Specifically, in the short-term prediction of stable electricity consumption scenarios, recurrent variants such as LSTM and GRU can achieve reliable accuracy with low computational overhead by effectively modeling temporal dependencies with gating mechanisms. When external meteorological factors intensify or when extreme weather events occur, the CNN hybrid architecture extracts multi-scale local features through front-end convolution, showing obvious advantages in robustness. The advantage of Transformer is that its self-attention mechanism can directly capture the global temporal relationship, so it has unique value in trend modeling. However, when the sequence length increases significantly, the computational pressure caused by its quadratic complexity cannot be ignored. This situation indicates that model selection should not aim for a universally optimal

DOI: 10.54254/2755-2721/2026.35335

solution. Instead, it must accurately assess the data characteristics and constraints of the target scenario, including whether the data are stationary or non-stationary, whether the forecasting horizon is short-term or long-term, and whether computational resources are abundant or limited.

Based on the above analysis, the follow-up research can be carried out in three directions. Existing methods generally rely on abundant historical samples, while cold-start problems in new nodes or data-sparse scenarios still need more practical solutions. Few-shot learning and transfer strategies may be the key to breaking through this bottleneck. The efficiency problem of long sequence prediction has not been fundamentally solved, and exploring linear-complexity alternatives for sequence modeling is a direction worthy of attention. Multi-region collaborative prediction brings data privacy constraints while improving accuracy. How to maintain collaborative gain under the premise that data is available but not visible is a challenge that must be faced when moving from single-node algorithm research to actual deployment of interconnected systems. Only when the prediction model is no longer just a tool for accuracy competitions, but is placed under realistic constraints such as data scarcity, limited computing power, and privacy protection, can this field truly mature.

It should be pointed out that this paper is mainly based on literature case analysis and lacks a unified experimental platform. The datasets and evaluation metrics of different papers are not completely consistent, which limits horizontal comparison. In the future, the comparability and reproducibility of different studies can be enhanced by establishing a standardized evaluation framework and cross-dimensional fairness benchmarks.

References

[1] Biswal, B., Deb, S., Datta, S., Ustun, T. S., & Cali, U. (2024). Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques. *Energy Reports, 12*, 3654–3670. https://doi.org/10.1016/j.egyr.2024.09.056

[2] Haben, S., Arora, S., Giasemidis, G., Voss, M., & Greetham, D. (2021). Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy, 304*(Suppl C), 117798. https://doi.org/10.1016/j.apenergy.2021.117798

[3] Ali, S., Bogarra, S., Riaz, M. N., Phyo, P. P., Flynn, D., & Taha, A. (2024). From time-series to hybrid models: Advancements in short-term load forecasting embracing smart grid paradigm. *Applied Sciences (Switzerland), 14*(11), 4442. https://doi.org/10.3390/app14114442

[4] Cui, Z., Wang, W., Jiang, W., Guo, J., & Liu, Y. (2024). High-precision identification and prediction of low-voltage load characteristics in smart grids based on hybrid deep learning framework. *International Journal of Low-Carbon Technologies, 19*, 2656–2666. https://doi.org/10.1093/ijlct/ctae221

[5] Mansoor, H., Gull, M. S., Rauf, H., Shaikh, I. U. H., Khalid, M., & Arshad, N. (2024). Graph convolutional networks based short-term load forecasting: Leveraging spatial information for improved accuracy. *Electric Power Systems Research, 230*, 110263. https://doi.org/10.1016/j.epsr.2024.110263

[6] Rodrigues, F., Cardeira, C., Calado, J. M. F., & Melicio, R. (2023). Short-term load forecasting of electricity demand for the residential sector based on modelling techniques: A systematic review. *Energies, 16*(10), 4098. https://doi.org/10.3390/en16104098

[7] Adekunle, A. A., Fofana, I., Picher, P., Rodriguez-Celis, E. M., Arroyo-Fernandez, O. H., & Zemouri, R. (2025). Optimizing deep learning predictive models: A comprehensive review of RNN and its variant architectures. *Applied Soft Computing, 185*(Part B), 114015. https://doi.org/10.1016/j.asoc.2025.114015

[8] Liu, H., Li, Z., Li, C., Shao, L., & Li, J. (2024). Research and application of short-term load forecasting based on CEEMDAN-LSTM modeling. *Energy Reports, 12*, 2144–2155. https://doi.org/10.1016/j.egyrep.2024.08.035

[9] Zelios, V., Mastorocostas, P., Kandilogiannakis, G., Kesidis, A., Tselenti, P., & Voulodimos, A. (2025). Short-term electric load forecasting using deep learning: A case study in Greece with RNN, LSTM, and GRU networks. *Electronics (Switzerland), 14*(14), 2820. https://doi.org/10.3390/electronics14142820

[10] Wang, C., Li, X., Shi, Y., Jiang, W., Song, Q., & Li, X. (2024). Load forecasting method based on CNN and extended LSTM. *Energy Reports, 12*, 2452–2461. https://doi.org/10.1016/j.egyrep.2024.07.030

[11] Hasanat, S. M., Ullah, K., Yousaf, H., Munir, K., Abid, S., Bokhari, S. A. S., Aziz, M. M., Naqvi, S. F. M., & Ullah, Z. (2024). Enhancing short-term load forecasting with a CNN-GRU hybrid model: A comparative analysis. *IEEE Access, 12*, 184132–184141. https://doi.org/10.1109/ACCESS.2024.3511653

[12] Unlu, A., & Peña, M. (2025). Comparative analysis of hybrid deep learning models for electricity load forecasting during extreme weather. *Energies, 18*(12), 3068. https://doi.org/10.3390/en18123068

[13] Yang, C., & Shu, Z. (2024). Long-term rolling prediction of transformer power load capacity based on the informer model. *Journal of Physics: Conference Series, 2782*(1), 012023. https://doi.org/10.1088/1742-6596/2782/1/012023

[14] Spatha, M. M., & Sotiropoulos, D. N. (2024). "Transforming" electricity demand prediction. In *2024 15th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE. https://doi.org/10.1109/IISA62523.2024.10786682

[15] Moon, J. (2026). Learning from scarcity: A review of deep learning strategies for cold-start energy time-series forecasting. *CMES - Computer Modeling in Engineering and Sciences, 146*(1). https://doi.org/10.32604/cmes.2025.071052

[16] Menati, A., Doudi, F., Kalathil, D., & Xie, L. (2026). PowerMamba: A deep state space model and comprehensive benchmark for time series prediction in electric power systems. *IEEE Transactions on Power Systems, 41*(3), 1812–1825. https://doi.org/10.1109/TPWRS.2025.3647539

[17] Zhang, Z., Rath, S., Xu, J., & Xiao, T. (2026). Federated learning for smart grid: A survey on applications and potential vulnerabilities. *ACM Transactions on Cyber-Physical Systems, 10*(1), 1–26. https://doi.org/10.1145/3760788

[18] Ding, M., Cai, Z., Chen, B., & Han, J. (2026). A federated collaborative short-term industrial load forecasting method integrating meta-learning and differentiated privacy allocation. *Frontiers in Energy Research, 14*. https://doi.org/10.3389/fenrg.2026.1778611