energies

MDPI

Article

A Virtual Power Plant Load Forecasting Approach Using COM Encoding and BiLSTM-Att-KAN

Yong Zhu [1], Liangyi Pu [1], Di Yang [1], Tun Kang [2], Chao Liang [1], Mingzhi Peng [3,*] and Chao Zhai [4]

[1] Chongqing Huizhi Energy Co., Ltd., Chongqing 400000, China; yydlzy@163.com (Y.Z.); 15399692688@163.com (L.P.); a13983886054@126.com (D.Y.); 15720588750@163.com (C.L.)
[2] SPIC Chongqing Co., Ltd., Chongqing 400000, China; kangtun@163.com
[3] School of Future Technology, China University of Geosciences, Wuhan 430074, China
[4] School of Automation, China University of Geosciences, Wuhan 430074, China; zhaichao@amss.ac.cn
* Correspondence: 18827405428@163.com

Abstract

Virtual Power Plant (VPP) is capable of aggregating and intelligently coordinating diverse distributed energy resources, among which the accuracy of load forecasting is a key factor in ensuring their regulation capability. To address the periodicity and complex nonlinear fluctuations of electricity load data, this study introduces a Cyclic Order Mapping (COM) encoding method, which maps weekly and intraday sequences into continuous ordered variables on the unit circle, thereby effectively preserving load periodic features. On the basis of the COM encoding, a novel forecasting model is proposed by integrating Bidirectional Long Short-Term Memory (BiLSTM) networks, an efficient self-attention mechanism, and the Kolmogorov–Arnold Network (KAN). This model is termed BiLSTM-Att-KAN. Comparative and ablation experiments were conducted to assess the scientific validity and predictive accuracy of the proposed approach. The results confirm its superiority, achieving a Root Mean Square Error (RMSE) of 141.403, a Mean Absolute Error (MAE) of 106.687, and a coefficient of determination (R^2) of 0.962. These findings demonstrate the effectiveness of the proposed model in enhancing load forecasting performance for VPP applications.

Keywords: virtual power plants; COM encoding; electrical load forecasting; BiLSTM-Att-KAN

check for updates

Academic Editor: Marcin Kaminski

Received: 12 September 2025
Revised: 18 October 2025
Accepted: 22 October 2025
Published: 24 October 2025

Citation: Zhu, Y.; Pu, L.; Yang, D.; Kang, T.; Liang, C.; Peng, M.; Zhai, C. A Virtual Power Plant Load Forecasting Approach Using COM Encoding and BiLSTM-Att-KAN. *Energies* **2025**, *18*, 5598. https://doi.org/10.3390/en18215598

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

The efficient utilization of energy and the pursuit of sustainable development are fundamental objectives of the modern electrical industry. As a key technology driving digital transformation, the Virtual Power Plant (VPP) integrates distributed renewable energy sources, energy storage systems, and flexible loads. By intelligently optimizing dispatch strategies and market trading mechanisms, a VPP enhances grid flexibility, reduces carbon emissions, and facilitates the transition of electrical systems toward cleaner, low-carbon, and intelligent development [1–4]. Nevertheless, in the context of intelligent VPP operation, accurate electrical load forecasting is a critical prerequisite for ensuring efficient dispatch and effective market responsiveness [5]. Consequently, the development of high-precision load forecasting models constitutes a core technological foundation for enabling VPPs to flexibly aggregate distributed resources, support carbon reduction, and improve economic performance [6].

In VPP systems, electrical load fluctuations exhibit distinct periodic characteristics across both short- and long-term time scales. At the short-term level, with daily monitoring cycles, load curves demonstrate clear diurnal periodicity, characterized by pronounced

day–night variations that closely correspond to human activity patterns [7]. At the long-term scale, seasonal temperature changes lead to a dual-peak pattern, driven by summer cooling demand and winter heating loads [8]. Beyond these periodic features, however, the randomness of human behavior and the nonlinear response to external disturbances introduce additional complexity to the load profile. As a result, effectively extracting and modeling the multi-scale, composite temporal patterns of load data while mitigating noise interference remains a key challenge for improving the accuracy of load forecasting in VPP applications.

Electrical load forecasting models can generally be classified into mechanistic models and data-driven models. Mechanistic approaches require substantial domain expertise and detailed system knowledge, which often limits their applicability. With the rapid advancement of artificial intelligence, machine learning and deep learning techniques have emerged as powerful alternatives for load forecasting. These data-driven approaches can automatically extract complex patterns from large-scale historical datasets, thereby improving the forecasting accuracy of fluctuating load series [9]. In general, data-driven models can be divided into static models, represented by machine learning methods [10], and dynamic models, represented by deep learning methods [11]. Commonly applied machine learning algorithms for load forecasting include BP neural networks (BPNN) [12], support vector regression (SVR) [13], and random forests (RF) [14], whose effectiveness has been widely demonstrated [15]. However, such models rely on a static modeling framework and often assume that data are independent and identically distributed. As a result, they fail to capture temporal dependencies and struggle to adapt to the time-varying characteristics of electrical load. In recent years, recurrent neural networks (RNNs) [16] have gained prominence due to their recurrent architecture, which allows them to retain historical information in temporal tasks and thereby overcome the limitations of static models. However, when dealing with long-term sequential tasks, increasing the depth of RNN layers leads to difficulties during backpropagation. Owing to the chain rule, this process is highly prone to exponential gradient explosion or vanishing, which in turn degrades predictive performance. To mitigate this issue, long short-term memory (LSTM) networks [17] and gated recurrent unit (GRU) [18,19] introduce gating mechanisms to dynamically filter and update historical load information in real time. Building on these developments, bidirectional models based on LSTM and GRU, such as bidirectional long short-term memory (BiLSTM) networks [20] and bidirectional gated recurrent unit (Bi-GRU) [21], have been extensively studied and validated for time-series load forecasting. However, the sequential nature of recurrent architectures fundamentally limits their capacity to model long-term dependencies, often resulting in information decay. To address this, the Transformer was developed, utilizing a self-attention mechanism to directly capture distant temporal relationships [22,23]. Despite its proficiency in modeling global context, the standard Transformer is hampered by the quadratic computational complexity of its self-attention mechanism and may be less effective at resolving fine-grained local features [24]. With continued research, numerous hybrid forecasting models have been developed to overcome the limitations of single models by performing multi-level extraction of time-series features [25,26]. While existing hybrid models improve predictive accuracy, they are often constrained by their reliance on traditional Multi-Layer Perceptron (MLP) with fixed activation functions for the final nonlinear mapping. This design can limit their ability to capture complex nonlinearities within the data. The recently proposed Kolmogorov-Arnold Network (KAN) [27] overcomes this issue by replacing these fixed functions with learnable splines, which substantially improves the model's nonlinear fitting capacity.

Therefore, this paper proposes the BiLSTM-Att-KAN, a novel hybrid model designed to overcome the aforementioned challenges. This architecture synergistically integrates

three specialized components: the BiLSTM component for capturing local short-term dependencies and incorporating a comprehensive contextual representation by processing the sequence in both forward and backward directions; an efficient self-attention mechanism for modeling global long-term dependencies; and a KAN as the final mapping layer to resolve complex nonlinearities. This integrated design significantly enhances the model's overall predictive accuracy.

The main contributions of this paper can be summarized as follows:

(1) This paper proposes a cyclic order mapping (COM) encoding method, which explicitly preserves the gradual periodic variation patterns of electrical load by mapping weekly and intraday time sequences to continuous ordered vectors on the unit circle. Moreover, owing to its constant dimensionality, COM encoding substantially reduces feature space complexity and mitigates overfitting problems caused by high-dimensional sparsity.

(2) This paper constructs a BiLSTM-Att-KAN ensemble model that integrates BiLSTM, an efficient self-attention mechanism, and KAN. Specifically, the primary BiLSTM captures short-term dependencies and local features from raw load data, while the self-attention mechanism extracts long-term dependencies and global structures. A secondary BiLSTM then fuses these multi-scale temporal features to further enhance dynamic representation. Finally, KAN maps the refined features into accurate forecasting results. The synergistic interaction of these components effectively resolves the difficulty of jointly modeling short- and long-term dependencies and significantly improves predictive performance.

(3) This paper replaces conventional fully connected layers with KAN, which enhances the model's nonlinear fitting capability and improves the reliability of forecasting results. By transforming complex temporal features into accurate forecasting results, KAN effectively overcomes the limitations of conventional architectures and ensures high-quality load forecasting.

The remainder of this paper is organized as follows: Section 2 introduces the concept and structure of VPP, Section 3 presents the proposed COM-based BiLSTM-Att-KAN forecasting model, Section 4 describes the experimental design and results; and Section 5 concludes the whole paper.

2. Virtual Power Plant

A VPP represents an integrated system enabled by advanced information and communication technologies in fusion with intelligent control strategies. Its primary function is the centralized aggregation, coordinated optimization, and flexible regulation of diverse distributed energy resources (DERs) [28], the structure of a VPP is shown in Figure 1. These DERs mainly include distributed renewable energy sources, controllable loads on the demand side, and various energy storage systems. By integrating geographically dispersed, heterogeneous, and inherently fluctuating resources into a single virtual entity with enhanced predictability, dispatchability, and rapid responsiveness, a VPP enables the efficient aggregation and intelligent management of multiple distributed resources.

This virtualization and integration capability enables the VPP to proactively respond to real-time grid demands by providing critical ancillary services to the electrical system. Such services include peak shaving and valley filling, balancing intraday load fluctuations, supplying reserve capacity, and enhancing system resilience against unexpected events. Through these functions, the VPP plays a vital role in supporting the secure, stable, and economically efficient operation of the electrical grid.

In the daily management and refined dispatch practices of electrical systems, accurate load forecasting not only provides essential data support and a scientific basis for dispatch-

ing decisions but also constitutes a crucial prerequisite for achieving deep collaborative interaction and efficient matching optimization among four key parties: the generation side, the grid, the load side, and energy storage. The accuracy of load forecasting directly determines the capability of a VPP for optimal coordinated control of aggregated resources, serving as a key input factor for enhancing overall energy utilization efficiency and system operational flexibility.

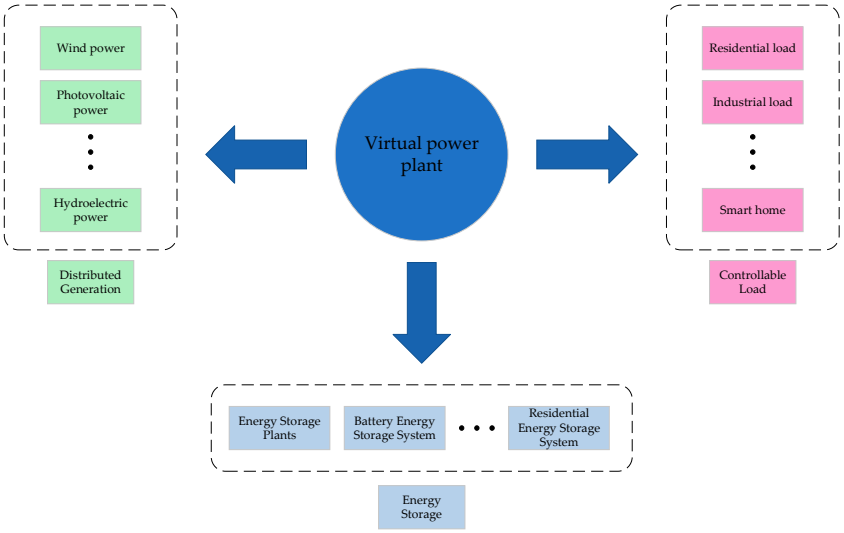


Figure 1. The structure of virtual power plant.

3. Methodology

To enhance the accuracy of load forecasting for virtual power plants, this study proposes an integrated methodology comprising innovations in both feature engineering and model architecture. First, the COM encoding method is introduced at the feature engineering stage. This approach addresses the endpoint discontinuity problem inherent in traditional temporal encoding by mapping discrete time indices onto a continuous unit circle. Building on this feature representation, the integrated BiLSTM-Att-KAN forecasting model is constructed. This architecture synergistically integrates a BiLSTM network to capture local short-term dependencies with an efficient self-attention mechanism to establish global long-term dependencies. Finally, leveraging the powerful nonlinear mapping capabilities of the KAN, the model effectively fits the fused multi-scale temporal features to generate the final prediction results. The overall workflow of this methodology is illustrated in Figure 2.

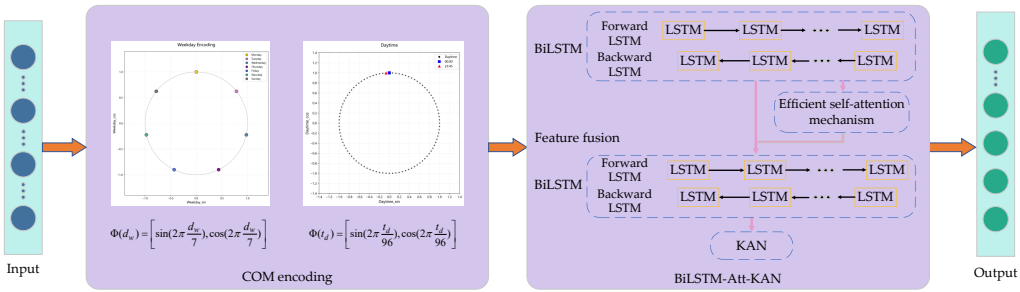


Figure 2. Model workflow.

3.1. COM Encoding

This paper proposes an innovative encoding method called COM encoding for periodic feature sequences. The core design principle is to project discrete temporal variables onto

continuous ordered variables on the unit circle, thereby explicitly preserving the periodic patterns inherent in electrical load data. By mapping discrete temporal points, such as weekly sequence numbers and intraday time indices onto continuous variables on the unit circle, the method achieves seam-less integration of periodic characteristics. In doing so, it effectively resolves the endpoint discontinuity problem inherent in conventional discrete encoding approaches, ensuring the complete representation of periodic features. Moreover, this encoding approach not only maintains the sequential relationships within the time series but also precisely expresses periodic characteristics through vector distances and angles. As a result, Monday and Sunday, as well as the final and initial sampling times within a day, are naturally connected in the vector space, forming a closed-loop structure. This design enables the continuous representation of temporal features within geometric space, overcoming the limitations of traditional methods where periodic characteristics are fragmented into isolated points. Consequently, it provides a feature representation for time-series modeling that possesses both continuity and geometric interpretability. The specific formula is given as follows:

$$\Phi(d_w) = \left[\sin(2\pi\frac{d_w}{7}), \cos(2\pi\frac{d_w}{7})\right] \tag{1}$$

$$\Phi(t_d) = \left[\sin(2\pi\frac{t_d}{96}), \cos(2\pi\frac{t_d}{96})\right] \tag{2}$$

where $\Phi()$ denotes the COM encoding function, d_w represents the week ordinal, indicating the position of the date within a week, sin() is the sine function, capturing the phase of the time point within the cycle, cos() is the cosine function, providing orthogonal supplementary information on the phase, t_d denotes the intraday time ordinal, indicating the sampling moment within a single day.

3.2. Electrical Load Forecasting Model

3.2.1. BiLSTM-Att-KAN

This paper proposes an integrated model named BiLSTM-Att-KAN, which integrates BiLSTM, an efficient self-attention mechanism, and KAN. The model structure is illustrated in Figure 3. First, the input data passes through a primary BiLSTM layer that captures short-term dependencies and local features within the sequence. As a result, the raw data is transformed into more representative feature variables, providing high-quality local temporal information for subsequent processing. Subsequently, the extracted basic features are fed into an efficient self-attention mechanism to capture global relationships and establish long-range dependencies in the sequence. This mechanism provides a global perspective and structured understanding of the sequence, significantly enhancing the model's ability to comprehend complex sequential relationships. Additionally, feature fusion is performed between the outputs of the BiLSTM layer and the self-attention mechanism to prevent information fragmentation. By fusing the complementary advantages of BiLSTM and the self-attention mechanism, a more comprehensive and powerful feature foundation is established for subsequent processing. The fused features are then processed by a secondary BiLSTM layer to further integrate temporal information, ensuring the model effectively captures dynamic temporal patterns in the fused features. This provides highly refined and discriminative temporal features for the final forecast. Finally, KAN replaces the traditional fully connected layer to perform nonlinear mapping. Leveraging KAN's expressive efficiency and function approximation capability, the complex time-series features extracted

and refined by all preceding layers are accurately transformed into forecasting results. The specific working principle can be summarized as follows:

$$X_1 = \text{BiLSTM}(X) \tag{3}$$

$$X_2 = \text{Attention}(X_1) \tag{4}$$

$$X_3 = \text{c}[X_1 \parallel X_2] \tag{5}$$

$$X_4 = \text{BiLSTM}(X_3) \tag{6}$$

$$Y = \text{KAN}(X_4) \tag{7}$$

where X denotes the model input, X_1 denotes the output of the primary BiLSTM layer, BilSTM() denotes the computational unit of the BiLSTM network, X_2 denotes the output of the efficient self-attention mechanism, Attention() denotes the computational unit of the efficient self-attention mechanism, X_3 denotes the output after feature fusion; denotes the computational unit responsible for feature fusion, X_4 denotes the output of the secondary BiLSTM layer, Y denotes the output of the integrated forecasting model, KAN() denotes the computational unit of the KAN module.

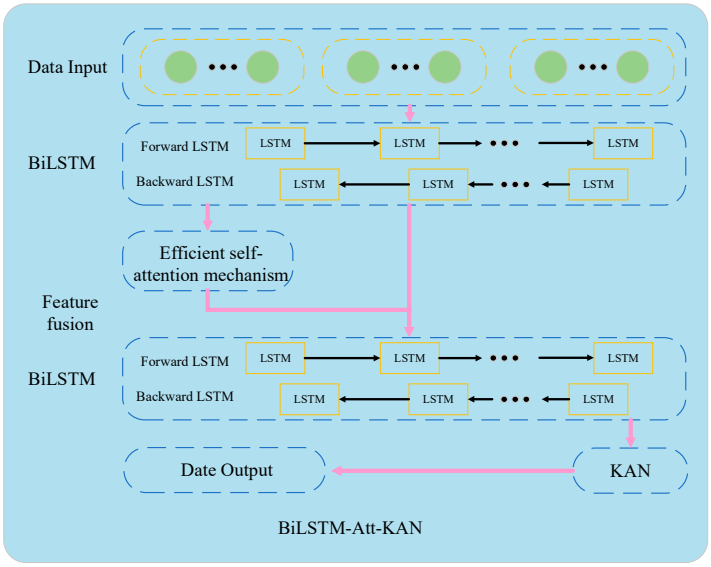


Figure 3. The structure of BiLSTM-Att-KAN model.

3.2.2. Bidirectional Long Short-Term Memory Network

LSTM is an advanced recurrent neural network architecture designed to overcome the vanishing gradient problem inherent in traditional RNNs when processing long sequences. This is achieved through a sophisticated gating mechanism consisting of input, forget, and output gates [29]. To capture a more complete contextual understanding, this study utilizes the BiLSTM. The BiLSTM enhances the standard LSTM by processing sequences in both forward and backward directions using two independent LSTM layers [30]. This dual-pass structure allows the model to integrate both historical and subsequent information, thereby generating more comprehensive feature representations and significantly improving predictive accuracy [31]. In the proposed BiLSTM-Att-KAN architecture, BiLSTM is strategically employed in a two-stage process. Initially, a primary BiLSTM layer processes the raw input sequence to extract local features and short-term dependencies. Following feature fusion by the self-attention mechanism, a secondary BiLSTM layer re-processes the enriched feature set to perform deeper temporal integration and capture complex dynamic patterns. The

fundamental structures of the LSTM and BiLSTM units are illustrated in Figure 4, and the working principle of these models can be summarized as follows:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{8}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{9}$$

$$\widetilde{C}_t = \tan h(W_C.[h_{t-1}, x_t] + b_C) \tag{10}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{11}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{12}$$

$$h_t = o_t * \tanh(C_t) \tag{13}$$

$$\overrightarrow{h}_t = \mathrm{LSTM}\left(x_t, \overrightarrow{h}_{t-1}\right) \tag{14}$$

$$\overleftarrow{h}_t = \mathrm{LSTM}\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{15}$$

$$y_t = g\left(\mathrm{W}_y\left[\overrightarrow{h}_t \parallel \overleftarrow{h}_t\right] + \mathrm{b}_y\right) \tag{16}$$

where W_f, W_i, W_c and W_o respectively denote the weight matrices of each gate control unit, b_f, b_i, b_c and b_o respectively denote the biases corresponding to each gate control unit, f_t, i_t, $\widetilde{C}_t$, C_t, o_t and h_t denote the output of the forget gate, the output of the input gate, candidate cell state, cell state, the output of the output gate, and hidden state, respectively, x_t denotes the node input at time step t, h_{t-1} denotes the hidden state of the cell at the previous time step, σ denotes the sigmoid activation function, $\overrightarrow{h}_t$ denotes the forward hidden state, which captures the historical context of the sequence up to time step t, $\overleftarrow{h}_t$ denotes the backward hidden state, which captures the future context of the sequence from time step t onward. LSTM() denotes the computational unit of the long short-term memory network, y_t denotes the node output at time step t, $g()$ denotes the activation function applied at time step t, W_y denotes the weight matrix of the corresponding output node, b_y denotes the bias vector of the corresponding output node.

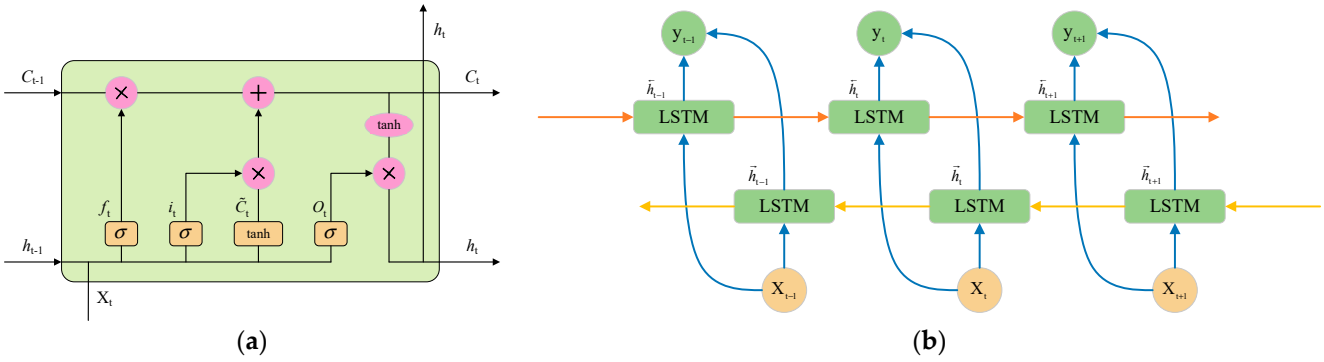


Figure 4. The structure for (**a**) LSTM, (**b**) BiLSTM.

3.2.3. Efficient Self-Attention Mechanism

Based on the traditional self-attention mechanism [32], this paper proposes an efficient self-attention mechanism that generates query, key, and value vectors simultaneously through a single linear projection, significantly reducing the number of parameters and computational complexity. This mechanism adopts a multi-head attention architecture, which divides the feature space into multiple subspaces and independently computes at-

tention weights in each subspace, thereby capturing dependencies from different aspects of the sequence. The attention scores are calculated using scaled dot products and normalized via the Softmax function. Finally, the weighted sum of the value vectors is used to obtain a context-aware feature representation. An output projection layer further integrates the multi-head information and maintains consistent input and output dimensions to ensure seamless integration with subsequent network layers. This design not only preserves the feature extraction capability but also greatly improves computational efficiency, making it particularly suitable for modeling long sequential time-series data. The structure is illustrated in Figure 5, and the specific formulas are as follows:

$$\mathrm{QKV} = \mathrm{X} \cdot \mathrm{W}_{qkv} \tag{17}$$

$$\mathrm{Q}, \mathrm{K}, \mathrm{V} = \mathrm{split}(\mathrm{QKV}, 3) \tag{18}$$

$$\mathrm{Q}_h = \mathrm{reshape}(\mathrm{Q}, [B, T, N, d]) \tag{19}$$

$$\mathrm{K}_h = \mathrm{reshape}(\mathrm{K}, [B, T, N, d]) \tag{20}$$

$$\mathrm{V}_h = \mathrm{reshape}(\mathrm{V}, [B, T, N, d]) \tag{21}$$

$$\mathrm{S} = \frac{\mathrm{Q}_h {\mathrm{K}_h}^{\mathrm{T}}}{\sqrt{d}} \tag{22}$$

$$\mathrm{Y} = \mathrm{ReLU}(\mathrm{CombineHeads}(\mathrm{Softmax}(\mathrm{S}))) \tag{23}$$

where QKV denotes the projection matrix for the input matrix X, W_{qkv} denotes the projection weight matrix, Q, K and V denote the query, key, and value obtained after three-way partitioning, respectively, Q_h, K_h and V_h denote the query, key, and value obtained after multi-head partitioning, respectively, B denotes the batch size, reshape() denotes the multi-head partition function, T denotes the time step length, N denotes the number of heads, d denotes the dimensionality of each head, S denotes the attention score, K_h^T denotes the transpose of the key matrix, Y denotes the output matrix, Softmax() denotes the normalization function, CombineHeads() denotes the multi-head concatenation function, ReLU() denotes the activation function.

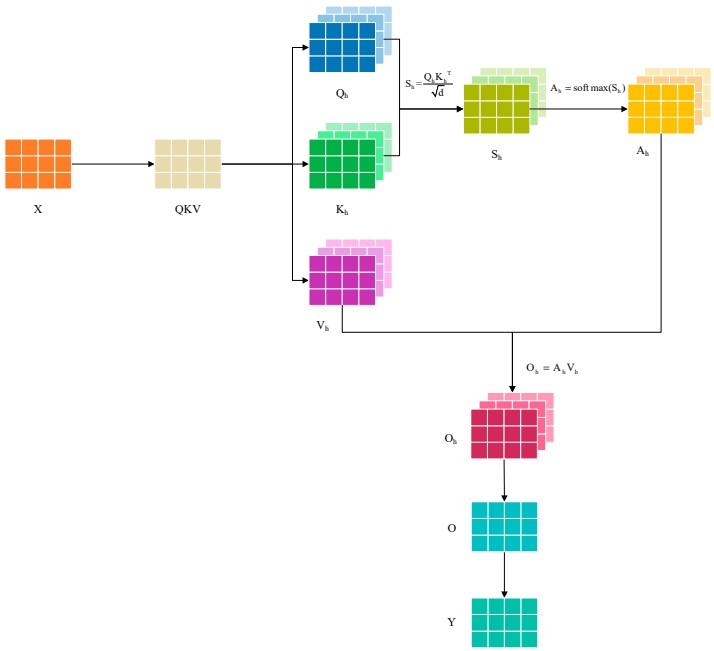


Figure 5. The structure of Efficient self-attention mechanism.

3.2.4. KAN

KAN is a neural network architecture based on the Kolmogorov-Arnold representation theorem, which states that any multivariate continuous function can be represented as a composition of univariate functions [33]. Unlike traditional MLP, KAN integrates univariate functions to approximate multivariate continuous functions. For example, while MLP use fixed activation functions applied to neurons, KAN employs learnable activation functions along the weights [34]. As a result, KAN offers higher parameter efficiency, improved interpretability, and stronger nonlinear fitting capabilities compared to MLP. Furthermore, nodes in KAN simply sum the incoming signals without applying additional nonlinear transformations. The structure of KAN is illustrated in Figure 6, and its working principle can be summarized as follows:

$$x_{l+1} = \underbrace{\begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix}}_{\Phi_l} x_l \tag{24}$$

$$\mathrm{KAN}(x) = (\mathbf{\Phi}_{l-1}\mathbf{\Phi}_{l-2}\ldots\mathbf{\Phi}_1\mathbf{\Phi}_0)x \tag{25}$$

where X_{l+1} denotes the input of layer $l+1$, X_l denotes the input of layer l, $\phi_{l,j,i}()$ denotes the activation function connecting the i-th neuron of layer l to the j-th neuron of layer $l+1$, Φ_l denotes the B-spline function matrix corresponding to layer l, KAN(x) denotes the output of the KAN layer, x denotes the input of the KAN layer.

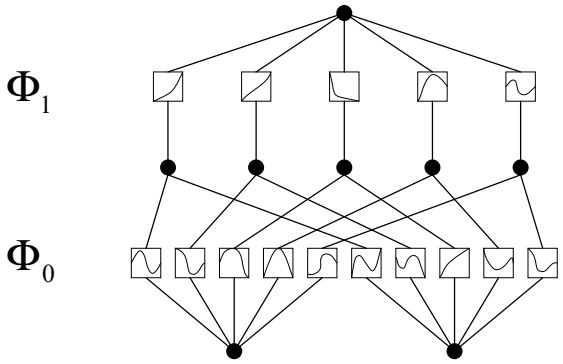


Figure 6. The frame of Efficient self-attention mechanism.

4. Experimental Procedures, Results and Analysis

4.1. Experimental Procedures

This paper evaluates the proposed method using the power load data of consumers within a VPP system. The experimental procedure, outlined in Figure 7, is detailed as follows:

(1) Data Collection and Feature Extraction. Historical load data were collected from the power load data acquisition platform. The periodic characteristics extracted from the electrical load data are encoded using the COM encoding to preserve intrinsic temporal patterns and facilitate subsequent modeling.

(2) Dataset Partitioning. To rigorously evaluate model performance, the processed dataset is partitioned into a training set and a test set, with 80% of the data allocated for training and the remaining 20% reserved for testing.

(3) Model Construction and Training. A BiLSTM-Att-KAN integrated forecasting model is constructed and trained using the training dataset. The integration of BiLSTM, an

efficient self-attention mechanism, and KAN enables effective multi-scale temporal feature learning for electrical load forecasting.

(4) Model Evaluation. BiLSTM-Att-KAN integrated model is evaluated using the test set to comprehensively assess its forecasting performance and validate its effectiveness in electrical load forecasting.

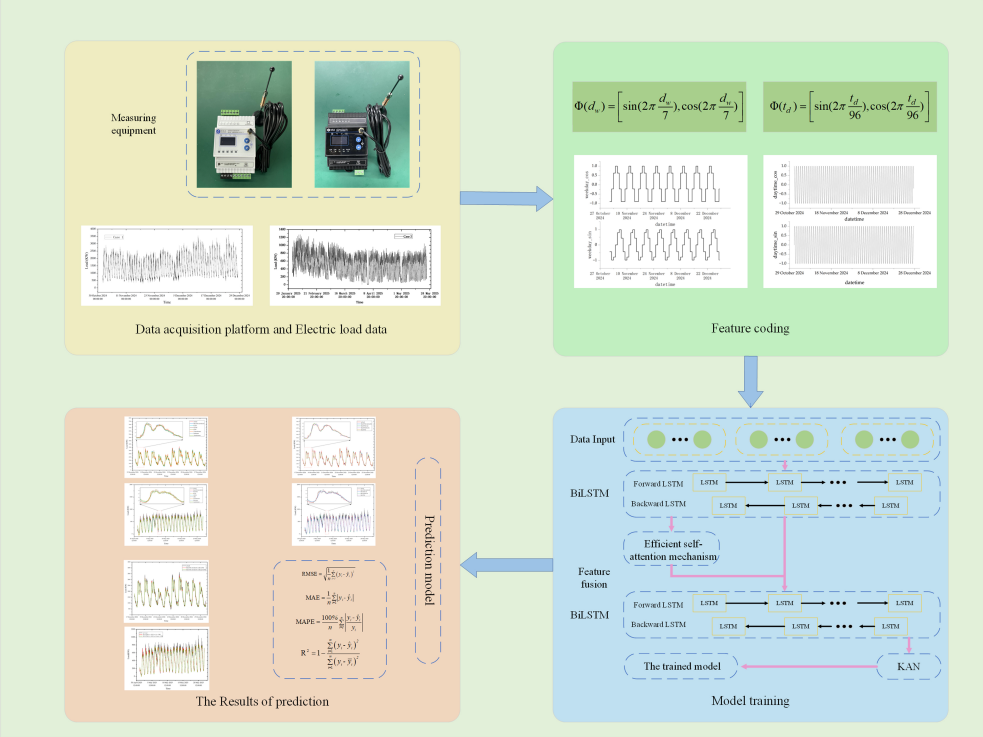


Figure 7. Flowchart of the integrated forecasting model.

4.2. Feature Extraction and Encoding

Electrical load data contain abundant temporal information, and effectively extracting these features is essential for constructing an accurate electrical load forecasting model. This paper emphasizes the extraction of significant periodic characteristics from the load data to enhance forecasting performance. Figures 8 and 9 presents the time-series plot of historical load sample points obtained from the electrical load data acquisition platform. As shown in Figure 8, the electrical load data exhibit distinct cyclical variations. On a daily scale, the load demonstrates significant alternating peaks and troughs. Analyzing its intraday patterns reveals a clear sequence: the load typically reaches a relatively low point around midnight, then gradually increases and reaches a prominent peak at approximately 10:00 a.m. Subsequently, the load declines, forming a pronounced trough around noon, before rising again to achieve a second peak at approximately 5:00 p.m. During the nighttime, the load decreases continuously until the early hours of the following morning, thereby completing a full daily cycle. This highly repetitive intraday fluctuation pattern, which is closely aligned with human activity rhythms, highlights the strong diurnal periodicity inherent in the electrical load data. Simultaneously, the electrical load data exhibit regular and repetitive patterns on a weekly scale. To verify this characteristic, a comparative analysis was conducted using data from identical workday types across two consecutive weeks. Specifically, the observation period from 1 November (Friday) to 7 November (Thursday) 2024, was compared with the period from 8 November to 14 November 2024. The results demonstrate that, when considering a week as a complete cycle, the overall trends of load variations exhibit a high degree of consistency, thereby confirming the stable weekly periodicity inherent in the electrical load data. Specifically, the weekly variation

pattern of the electrical load data can be described as follows: From Friday to Sunday, the daily peak loads exhibit a gradual decreasing trend. From Monday to Tuesday, the daily peak loads rebound significantly and increase progressively. On Wednesday, the daily peak loads generally experience a noticeable decline, forming a relative trough within the weekly load pattern. On Thursday, the daily peak load shows a marked recovery, culminating in the weekly peak load on Friday. This weekly cyclical pattern of load fluctuations provides strong evidence of the stable weekly periodicity inherent in the electrical load data, reflecting the systemic influence of societal production and human activity rhythms on electricity consumption patterns. Based on this analysis, this paper adopts COM encoding to perform feature representation of the extracted daily and weekly sequence indices. The encoded feature variables are shown in Figure 10.

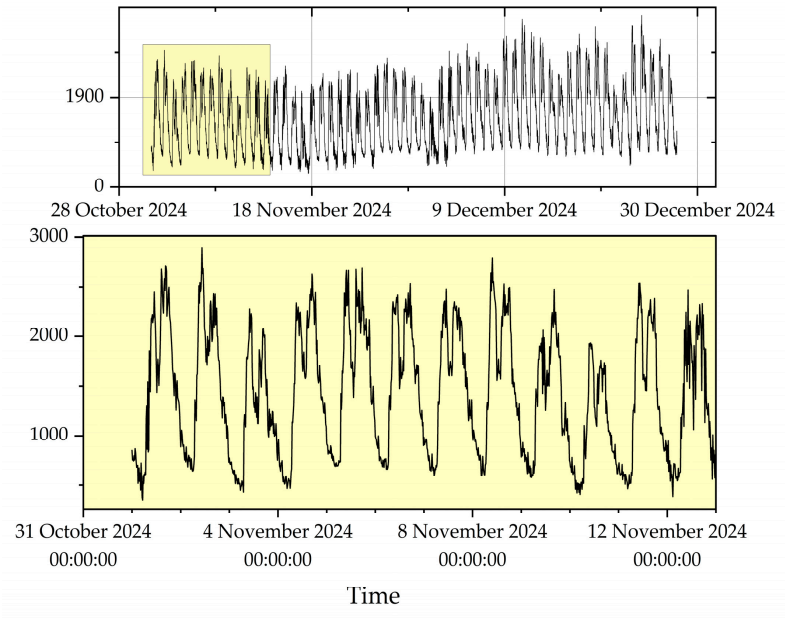


Figure 8. Electrical load sample of Case 1.

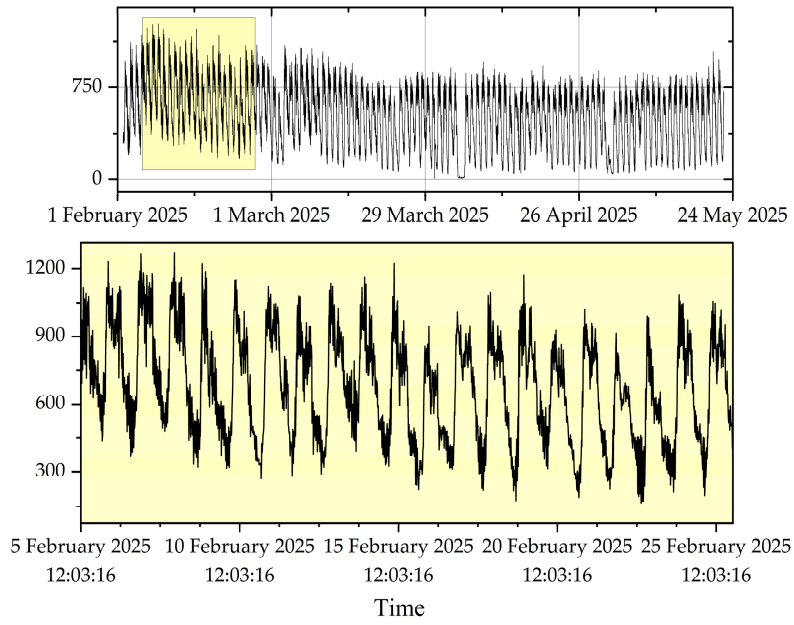
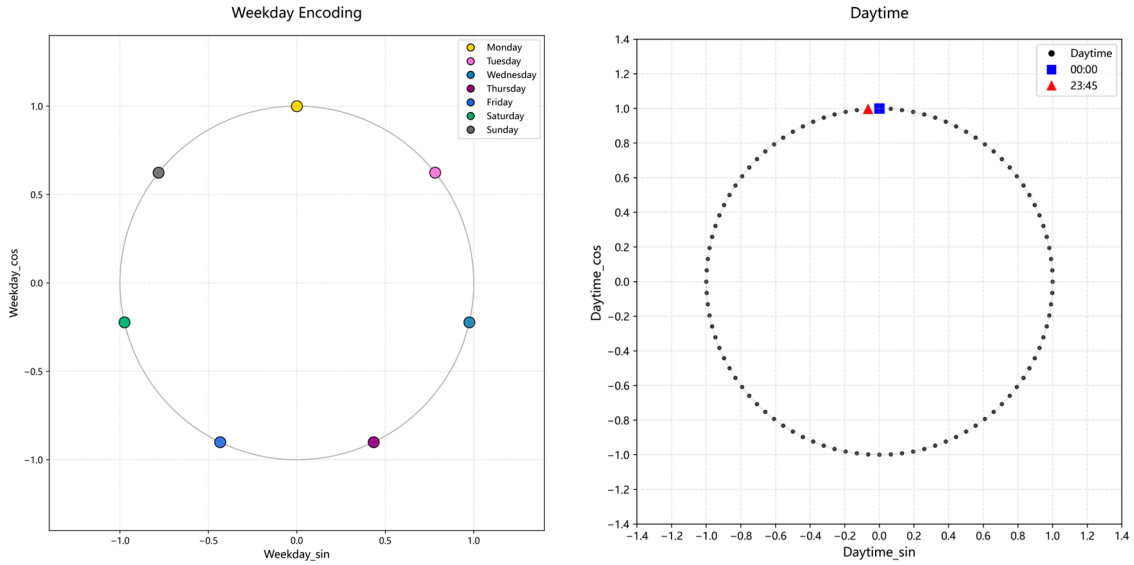


Figure 9. Electrical load sample of Case 2.



(**a**) (**b**)

Figure 10. Feature encoding for (**a**) Weekday encoding, (**b**) Daytime encoding.

4.3. Dataset Partitioning and Evaluation Indicators

This paper is based on two power load datasets, designated Case 1 and Case 2, containing 5500 and 10,500 samples, respectively. Both datasets feature a 15 min sampling interval, a 500-step input window, and were partitioned into an 80% training set and a 20% test set. The model's predictive performance is assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). Their respective calculation formulas are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{26}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{27}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{28}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{29}$$

where y_i, $\hat{y}_i$ and $\overline{y}$ denote the true value, the predicted value, and the average of the true values for the i-th sample, respectively, RMSE quantifies the square root of the mean squared difference between predicted and true values, MAE measures the average absolute difference between the predicted and true values, MAPE assesses the average absolute percentage error between predicted and actual values, and R^2 evaluates the model's ability to explain the variance of the target variable.

4.4. Comparative Analysis of Forecasting Models

To evaluate its accuracy and robustness, the proposed BiLSTM-Att-KAN model was benchmarked against several baseline models, including BiLSTM, KAN, LSTM, GRU, Transformer, Informer, and Autoformer, across two distinct case studies, Case 1 and Case 2. The performance metrics for each model are summarized in Table 1 and visualized in

Figure 11. The experimental results demonstrate that the proposed model achieves optimal performance across all key metrics. This superiority is attributed to the effective synergy within the BiLSTM-Att-KAN architecture, which cohesively integrates temporal feature extraction, an attentional focus on key information, and nonlinear fitting capabilities. This synergistic integration ultimately manifests as significantly improved prediction accuracy. For instance, under the experimental conditions of Case 1, the BiLSTM-Att-KAN model achieved RMSE, MAE, MAPE, and R^2 values of 141.403, 106.687, 6.958%, and 0.962, respectively. Furthermore, an analysis of the benchmark models reveals their respective inherent limitations. RNN variants, such as LSTM, GRU, and BiLSTM, exhibited higher overall errors, which can be attributed to their reliance on fixed activation functions. This structural constraint limits their capacity to fully model the complex nonlinear relationships present in the power load data. In contrast, the standalone KAN model achieved superior performance compared to the RNN variants by leveraging its learnable activation functions to more accurately capture the nonlinear mapping between input features and load values. However, its performance remains suboptimal because it lacks an intrinsic architecture specifically designed for modeling temporal dependencies. Conversely, the Transformer-based models outperformed the RNN variants, underscoring the efficacy of the self-attention mechanism for capturing long-term dependencies. Nevertheless, their global focus can result in a diminished capacity to model local-level features, a traditional strength of recurrent architectures.

Table 1. Baseline model comparison results.

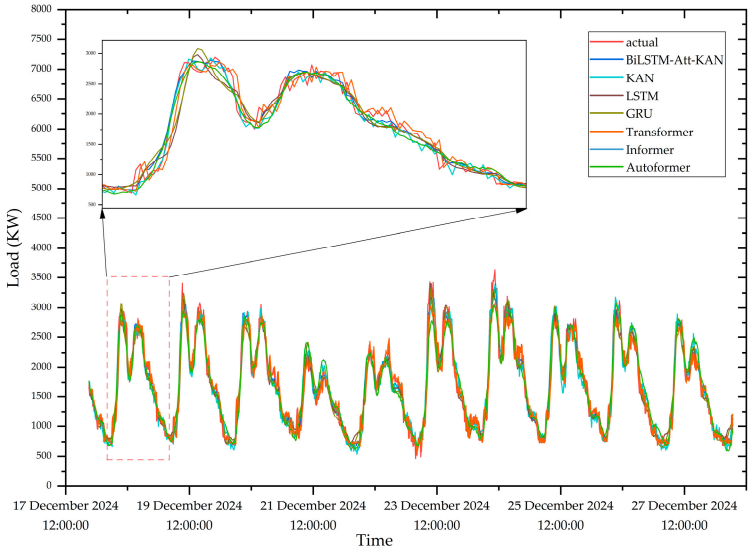
Case	Models	RMSE	MAE	MAPE (%)	R^2
Case 1	BiLSTM-Att-KAN	141.403	106.687	6.958	0.962
	BiLSTM	172.957	130.656	8.526	0.944
	KAN	166.437	129.631	8.455	0.948
	LSTM	181.989	136.529	8.814	0.938
	GRU	186.851	139.494	8.899	0.934
	Transformer	174.614	132.177	8.429	0.943
	Informer	183.387	139.976	8.959	0.937
	Autoformer	185.136	141.282	9.044	0.935
Case 2	BiLSTM-Att-KAN	53.489	39.145	10.329	0.950
	BiLSTM	71.391	54.758	17.030	0.910
	KAN	65.560	50.111	16.489	0.924
	LSTM	73.304	54.304	15.585	0.906
	GRU	75.397	58.498	18.821	0.900
	Transformer	66.025	49.783	15.831	0.923
	Informer	64.594	48.683	15.177	0.927
	Autoformer	65.183	49.153	15.426	0.925

4.5. Model Interpretability Analysis

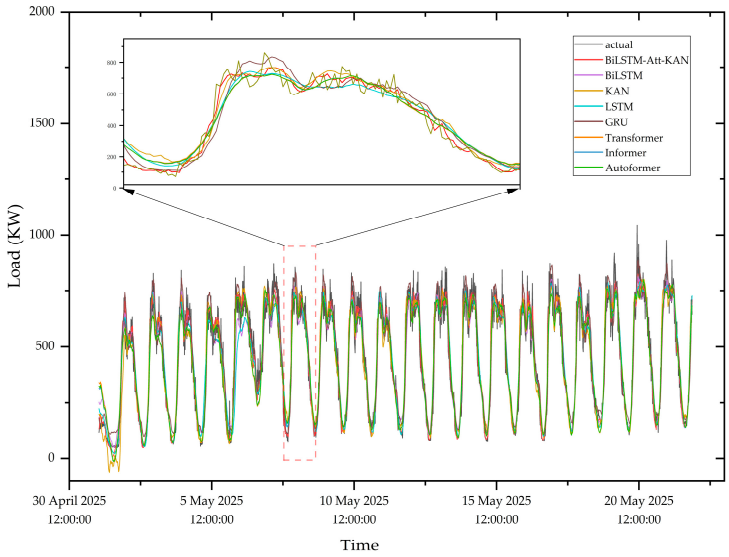
4.5.1. Ablation Experiments

Furthermore, an ablation study was conducted to rigorously validate the contribution of each key component and demonstrate the overall superiority of the proposed model, with results summarized in Table 2 and visualized in Figure 12. The findings from Case 1 clearly illustrate a powerful synergistic effect within the architecture. The study reveals that the efficient self-attention mechanism excels at establishing global, long-term dependencies, which complements the local, short-term features captured by the BiLSTM's recurrent structure. The synergy arises when these distinct types of temporal information are fused, creating a far more comprehensive feature representation. Concurrently, the KAN module addresses the inherent limitations of fixed activation functions in recurrent networks,

providing an enhanced capability to model complex nonlinearities. Therefore, the optimal performance of the BiLSTM-Att-KAN model stems from this structured approach: it first generates a high-quality feature set by integrating both global and local dependencies, which is then precisely mapped to the final output by the KAN module.



(**a**)

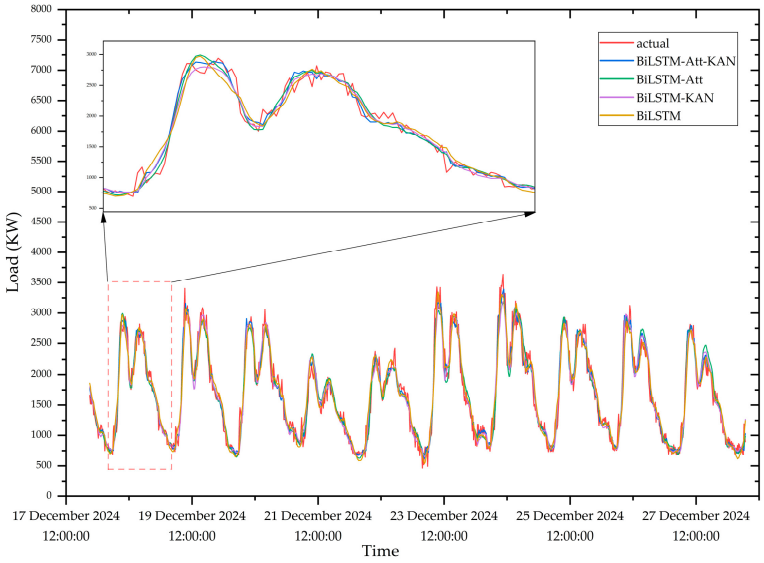


(**b**)

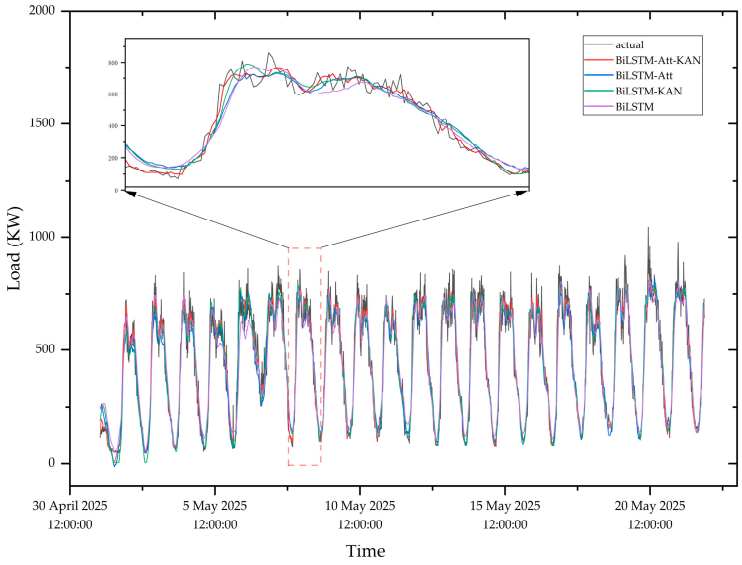
Figure 11. Baseline model comparison results for (**a**) Case 1, (**b**) Case 2.

Table 2. Forecasting results for ablation experiments.

Case	Models	RMSE	MAE	MAPE (%)	R^2
Case 1	BiLSTM-Att-KAN	141.403	106.687	6.958	0.962
	BiLSTM-Att	159.029	119.926	7.635	0.952
	BiLSTM-KAN	161.354	121.226	7.761	0.951
	BiLSTM	172.957	130.656	8.526	0.944
Case 2	BiLSTM-Att-KAN	53.489	39.145	10.329	0.950
	BiLSTM-Att	60.766	45.526	13.297	0.935
	BiLSTM-KAN	60.062	45.223	13.616	0.937
	BiLSTM	71.391	54.758	17.030	0.910



(**a**)



(**b**)

Figure 12. Ablation study results for (**a**) Case 1, (**b**) Case 2.

4.5.2. Visualizing Attention and Key Feature Contributions

An interpretability analysis was conducted to elucidate the internal mechanisms of the BiLSTM-Att-KAN model. The attention heatmaps presented in Figures 13 and 14 visually confirm the model's ability to capture data periodicity, evidenced by prominent diagonal patterns. Variations among the heads highlight the multi-head architecture's strength in learning diverse dependency patterns concurrently, leading to a more robust understanding of temporal dynamics. The SHAP analysis, shown in Figures 13d and 14d, quantifies feature importance, identifying the historical load sequence as the primary driver due to its strong autocorrelation, with intra-day and weekly time encodings serving as critical secondary features for discerning daily and weekly patterns. Collectively, these analyses demonstrate that the BiLSTM-Att-KAN model is not only accurate but also interpretable. Its predictions are grounded in a logical understanding of key drivers as validated by SHAP and an effective capture of periodic trends as visualized by the attention maps, making it a reliable tool for VPP to optimize resource scheduling and anticipate load fluctuations.

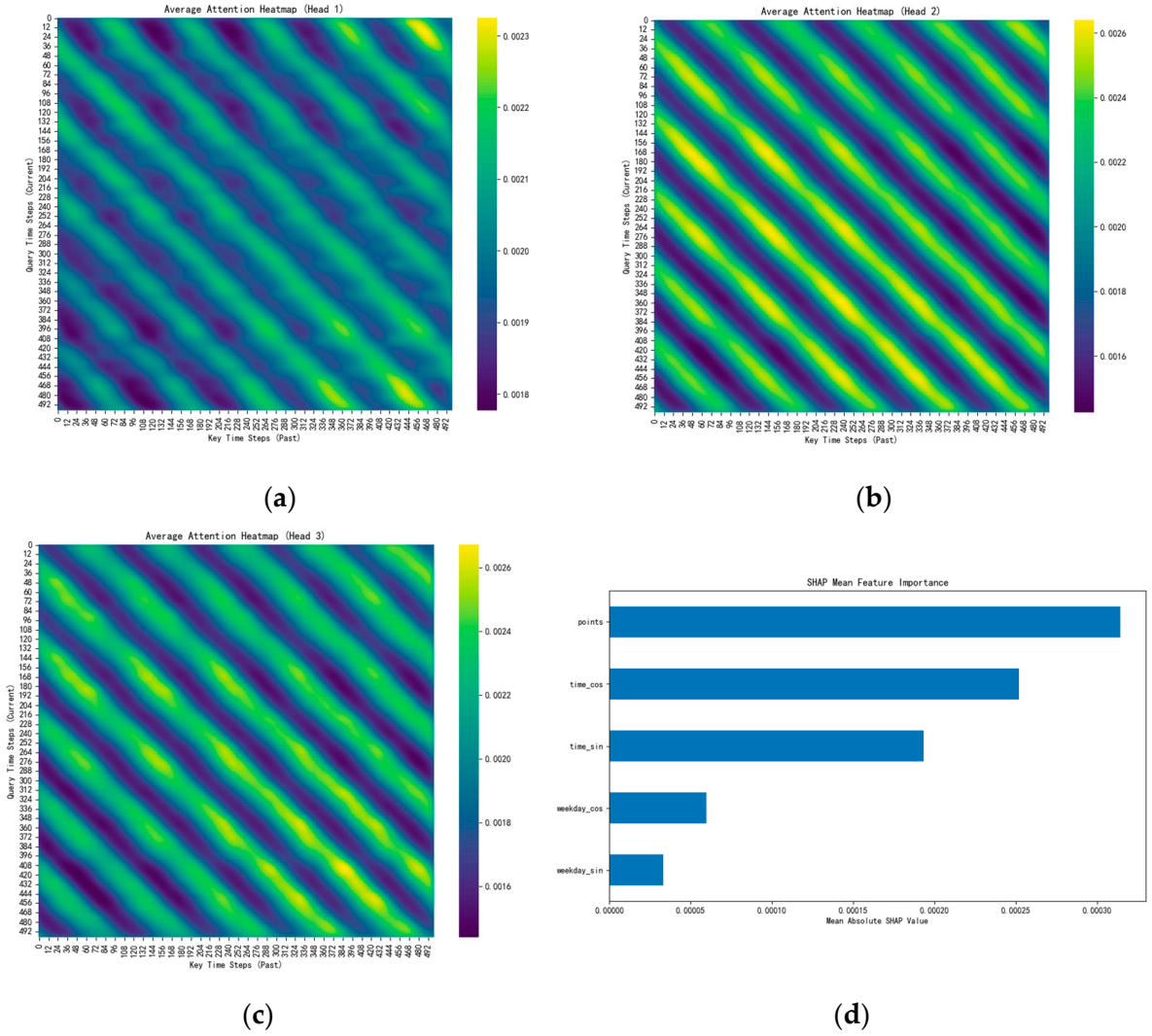


Figure 13. Case 1 for (**a**) Head 1, (**b**) Head 2, (**c**) Head 3, (**d**) Feature Importance.

4.6. Feature Comparison Experiments

After conducting the aforementioned forecasting model comparisons and ablation experiments, the integrated model exhibited the best predictive performance in electrical load forecasting. To further examine the influence of feature inputs on the proposed model, this paper compares the performance metrics of models incorporating COM encoding with those excluding COM encoding as feature inputs. The corresponding results are presented in Table 3 and Figure 15. For instance, under the experimental conditions of Case 1, the experimental results show that the integrated BiLSTM-Att-KAN model with COM encoding achieves an RMSE of 141.403, a reduction of 13.606 compared to the model without COM encoding, demonstrating an enhanced ability to capture load peaks and troughs. The MAE is 106.687, 10.389 lower than that of the model without COM encoding, indicating improved stability in short-term fluctuation forecasting. In terms of relative error, the model's MAPE is reduced from 7.535% to 6.958%. This improvement provides compelling, scale-independent evidence that COM encoding makes a substantial contribution to predictive accuracy. The R^2 reaches 0.962, an increase of 0.007, reflecting stronger fitting performance. These results confirm that COM encoding enables the model to effectively extract and utilize multi-scale periodic features, thereby improving forecasting accuracy.

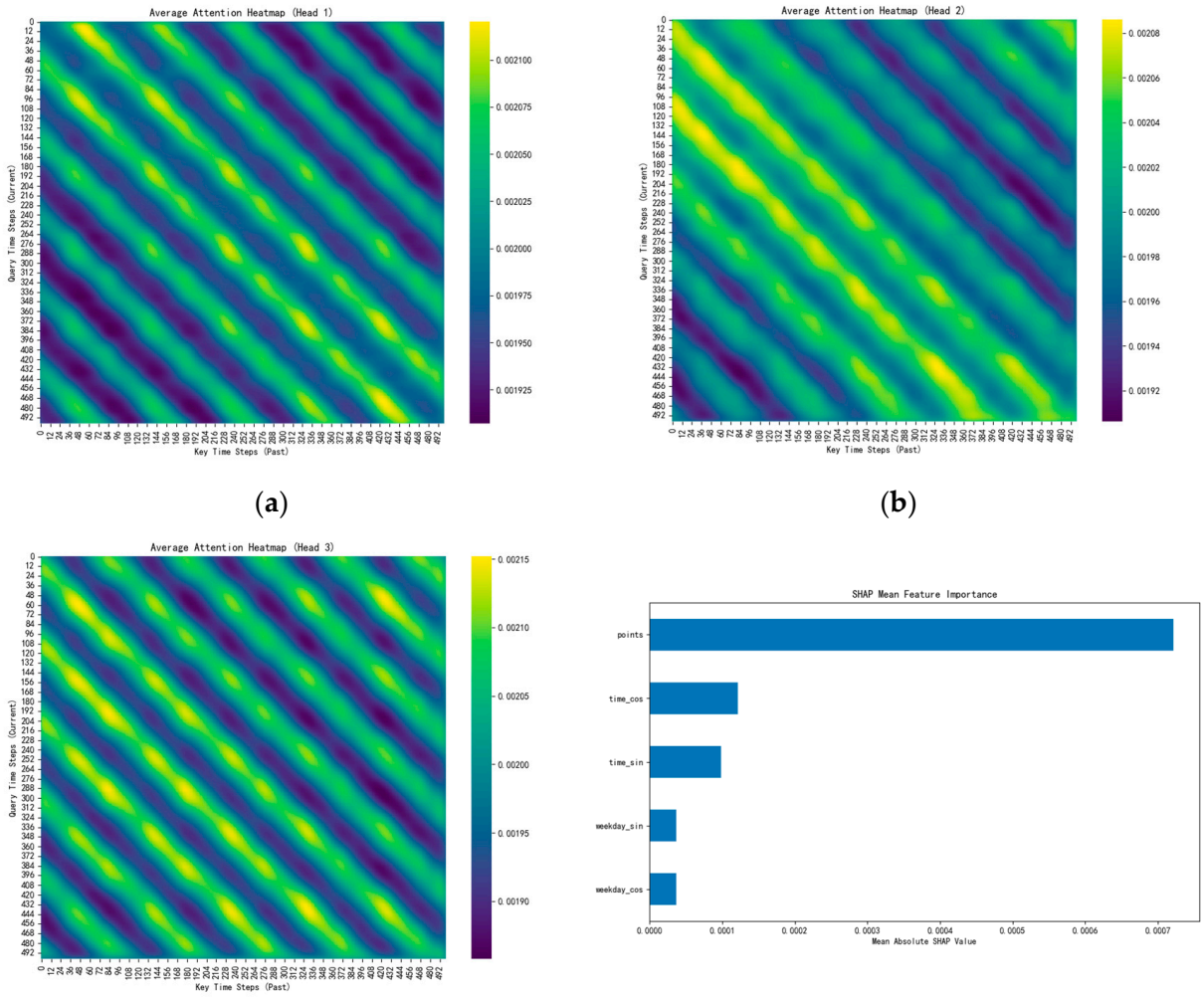


Figure 14. Case 2 for (**a**) Head 1, (**b**) Head 2, (**c**) Head 3, (**d**) Feature Importance.

Table 3. Forecasting results for different feature inputs.

Case	Models	RMSE	MAE	MAPE (%)	R^2
Case 1	With COM encoding	141.403	106.687	6.958	0.962
	Without COM encoding	155.009	117.076	7.535	0.955
Case 2	With COM encoding	53.489	39.145	10.329	0.950
	Without COM encoding	58.706	43.807	12.270	0.939

4.7. Comprehensive Performance Analysis

The relative error distributions for each model are illustrated Figure 16. In Case 1, the proposed BiLSTM-Att-KAN model exhibits the narrowest box plot, with an interquartile range or IQR of 165.42588, which is significantly lower than that of all baseline models. This provides clear visual evidence of a more concentrated error distribution and, consequently, higher prediction stability. To statistically substantiate this visual observation, a paired *t*-test was conducted to compare the prediction error distributions between the proposed model and key baseline models. The results, presented in Table 4, reveal a highly statistically significant difference with a *p*-value less than 0.001. This rigorous statistical evidence confirms that the model's accuracy improvement is substantive and not a result of random chance. Furthermore, the superior IQR compared to both LSTM and Transformer models validates the selection of BiLSTM as the foundational architecture. However, despite its

outstanding overall performance, the model exhibits a limitation in tracking extreme load peaks, where its predictions tend to underestimate the actual spikes. This phenomenon is primarily attributed to the sparse and atypical nature of such peaks within the dataset. To achieve optimal generalization, the model prioritizes learning prevalent patterns, which results in a statistical smoothing effect on these extreme values. This behavior represents a fundamental trade-off between a model's global generalization capability and its precision in capturing rare, high-magnitude events.

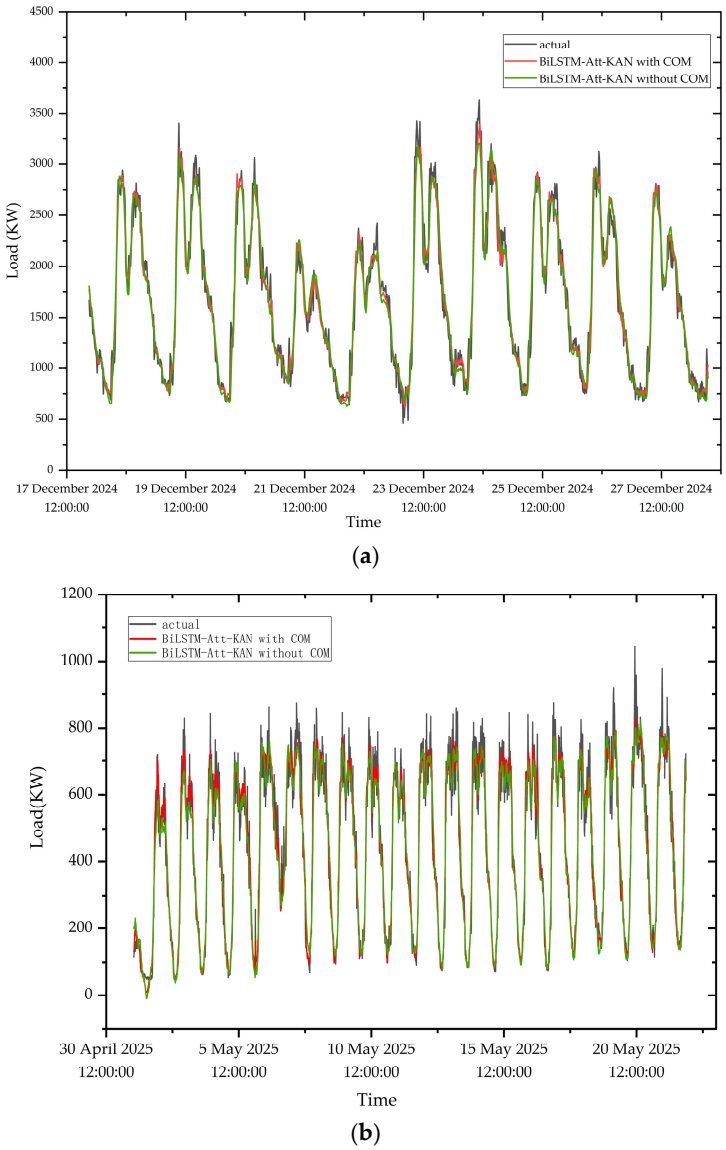
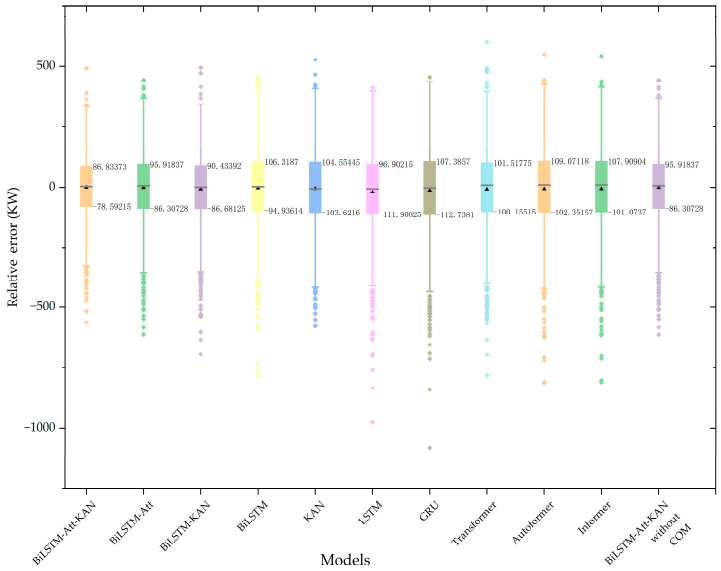


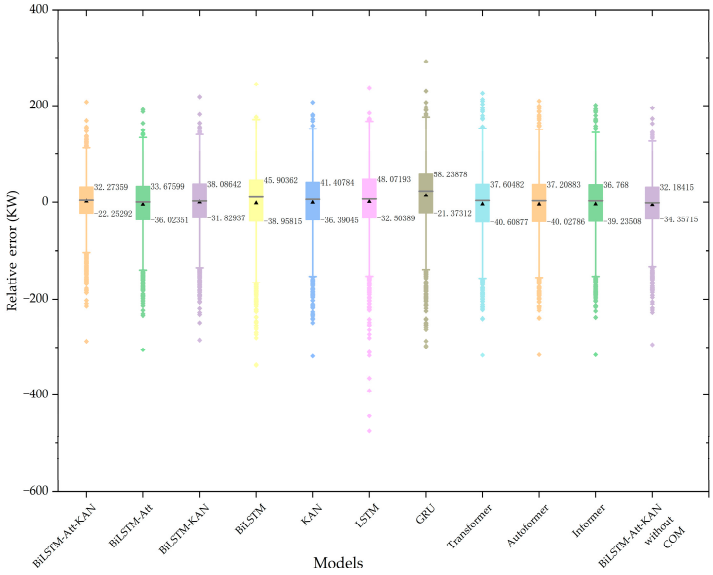
Figure 15. COM encoding analysis results for (**a**) Case 1, (**b**) Case 2.

4.8. Rolling-Origin Cross-Validation

To investigate the impact of training data size on model performance, a validation experiment was conducted with varying proportions of the training set. The results for Case 1, presented in Table 5, show a clear trend: as the training set proportion increases from 10% to 80%, the model's performance exhibits a significant and monotonic improvement across all metrics. Notably, even with a training set as small as 30%, it achieves a highly robust predictive performance. This finding suggests that the model's synergistic architecture, which combines advanced feature extraction with powerful nonlinear mapping, enables it to effectively capture critical temporal dependencies even under limited sample conditions.



(**a**)



(**b**)

Figure 16. Box plot of relative errors for (**a**) Case 1, (**b**) Case 2.

Table 4. Baseline model statistical significance.

Case	Baseline Models	*p*-Value
Case 1	BiLSTM-Att	<0.001
	BiLSTM-KAN	<0.001
	BiLSTM	<0.001
	KAN	<0.001
	LSTM	<0.001
	GRU	<0.001
	Transformer	<0.001
	Informer	<0.001
	Autoformer	<0.001
	BiLSTM-Att-KAN without COM	<0.001

Table 4. *Cont.*

Case	Baseline Models	*p*-Value
Case 2	BiLSTM-Att	<0.001
	BiLSTM-KAN	<0.001
	BiLSTM	<0.001
	KAN	<0.001
	LSTM	<0.001
	GRU	<0.001
	Transformer	<0.001
	Informer	<0.001
	Autoformer	<0.001
	BiLSTM-Att-KAN without COM	<0.001

Table 5. Forecasting results for different training set Ratios.

Case	Training Set Ratio	RMSE	MAE	MAPE (%)	R^2
Case 1	10%	381.674	297.382	21.397	0.698
	30%	241.521	188.411	13.144	0.880
	50%	216.928	165.490	10.233	0.905
	80%	141.403	106.687	6.958	0.962
	90%	156.116	120.019	7.824	0.960
Case 2	10%	123.191	96.673	76.216	0.698
	30%	71.820	53.734	29.088	0.909
	50%	65.616	48.736	36.035	0.925
	80%	53.489	39.145	10.329	0.950
	90%	56.266	41.210	9.464	0.946

4.9. Model Deployability

All experiments were performed in a Python 3.9 environment on a system equipped with an Intel® Core™ Ultra 9 CPU and an NVIDIA GeForce RTX 4060 GPU. Key model hyperparameters include a primary BiLSTM layer with 64 units and a secondary layer with 32 units. The self-attention mechanism is configured with three heads of size 36. The KAN layer is specified by 5 units, a grid size of 5, a spline order of 3, a grid range of $[-50, 50]$, and a tanh basis activation. To assess the model's practical deployability, its computational performance was evaluated. Under this hardware configuration, the model achieved a single-step inference time of approximately 156 ms.

Furthermore, a simulated deployment workflow was designed to validate the model's adaptability and efficiency. After an initial training phase on 80% of the data, the complete model was saved. Subsequently, a validation set was used to simulate the arrival of new data, upon which the model's parameters were rapidly fine-tuned. The fine-tuned model demonstrated robust performance on an independent test set, achieving RMSE, MAE, MAPE, and R^2 values of 151.900, 118.564, 7.706%, and 0.963, respectively. Crucially, the entire fine-tuning and prediction workflow was completed in approximately 157 s, a duration significantly shorter than the 15 min data sampling interval. This result not only validates the model's generalization capability but also provides compelling evidence of its robust potential for rapid iteration, efficient adaptation, and online deployment in real-world operational settings.

5. Conclusions

This paper proposed a novel feature extraction method, COM encoding, and an integrated predictive model, the BiLSTM-Att-KAN, to address critical challenges in time-series forecasting. The COM encoding technique resolves the issue of feature discontinuity

by mapping discrete time points to a continuous two-dimensional space, providing the model with an unambiguous representation of temporal proximity. The BiLSTM-Att-KAN is designed as a synergistic, multi-stage architecture: a BiLSTM layer captures local features, an efficient self-attention mechanism establishes global long-term dependencies, and a KAN module leverages its learnable activation functions to perform a powerful nonlinear mapping. The efficacy of both the encoding method and the model was rigorously validated through a comprehensive experimental framework across two distinct case studies. The technical contributions of this study translate directly into significant economic value for the efficient operation of VPP. High-precision forecasting enables VPPs to submit declaration curves to the grid that are highly consistent with actual demand, thereby minimizing financial penalties arising from prediction errors. Moreover, improved accuracy reduces the need for expensive reserve capacity, which is typically maintained to hedge against uncertainty. Ultimately, by enhancing both market transaction strategies and internal resource scheduling, the proposed model significantly improves the overall operational economy of VPPs while ensuring system balance.

Future research will proceed in two primary directions. First, future study will systematically integrate multi-source external variables, such as meteorological data and electricity price signals, to construct a multimodal input framework. This will allow for a more comprehensive characterization of the factors driving load fluctuations and is expected to further enhance the model's predictive capability for extreme events. Second, the model will be deployed and tested in heterogeneous VPP environments characterized by diverse geographical features and energy structures. This will allow for a thorough evaluation of its generalization and robustness across different regions and operational scenarios.

Author Contributions: Conceptualization, Y.Z., L.P. and D.Y.; Methodology, T.K. and M.P.; Software, C.L. and M.P.; Validation, Y.Z., L.P. and C.L.; Formal analysis, T.K.; Investigation, Y.Z., L.P., D.Y. and C.L.; Resources, M.P. and C.Z.; Data curation, D.Y. and T.K.; Writing—original draft, Y.Z.; Writing—review and editing, L.P.; Visualization, D.Y. and C.Z.; Supervision, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the "CUG Scholar" Scientific Research Funds at China University of Geosciences (Wuhan) (Project No. 2020138).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Yong Zhu, Liangyi Pu, Di Yang and Chao Liang were employed by the company Chongqing Huizhi Energy Co., Ltd. Author Tun Kang was employed by the company SPIC Chongqing Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COM	Cyclic Order Mapping
DERs	Distributed Energy Resources
Att	Efficient self-attention mechanism
VPP	Virtual Power Plant
MLP	Multi-Layer Perceptron
BiGRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short-Term Memory
KAN	Kolmogorov-Arnold Network
RMSE	Root Mean Square Error
MAE	Mean Absolute Error

MAPE	Mean Absolute Percentage Error
R^2	Coefficient of Determination

List of Symbols

d_w	The week ordinal
t_d	The intraday time ordinal
f_t	The output of the forget gate
i_t	The output of the input gate
C_t	Cell state
o_t	The output of the output gate
h_t	Hidden state
σ	The sigmoid activation function
tanh	The activation function
Q	The query vector
K	The key vector
V	The value vector

References

1. Venegas-Zarama, J.F.; Muñoz-Hernandez, J.I.; Baringo, L.; Diaz-Cachinero, P.; De Domingo-Mondejar, I. A Review of the Evolution and Main Roles of Virtual Power Plants as Key Stakeholders in Power Systems. *IEEE Access* **2022**, *10*, 47937–47964. [CrossRef]
2. Alahyari, A.; Ehsan, M.; Mousavizadeh, M.S. A Hybrid Storage-Wind Virtual Power Plant (VPP) Participation in the Electricity Markets: A Self-Scheduling Optimization Considering Price, Renewable Generation, and Electric Vehicles Uncertainties. *J. Energy Storage* **2019**, *25*, 100812. [CrossRef]
3. Liu, X.; Gao, C. Review and Prospects of Artificial Intelligence Technology in Virtual Power Plants. *Energies* **2025**, *18*, 3325. [CrossRef]
4. Jin, W.; Wang, P.; Yuan, J. Key Role and Optimization Dispatch Research of Technical Virtual Power Plants in the New Energy Era. *Energies* **2024**, *17*, 5796. [CrossRef]
5. Nti, I.K.; Teimeh, M.; Nyarko-Boateng, O.; Adekoya, A.F. Electricity Load Forecasting: A Systematic Review. *J. Electr. Syst. Inf. Technol.* **2020**, *7*, 13. [CrossRef]
6. Lindberg, K.B.; Seljom, P.; Madsen, H.; Fischer, D.; Korpås, M. Long-Term Electricity Load Forecasting: Current and Future Trends. *Util. Policy* **2019**, *58*, 102–119. [CrossRef]
7. Düzgün, B.; Bayındır, R.; Köksal, M.A. Estimation of Large Household Appliances Stock in the Residential Sector and Forecasting of Stock Electricity Consumption: Ex-Post and Ex-Ante Analyses. *Gazi Univ. J. Sci. Part C Des. Technol.* **2021**, *9*, 182–199. [CrossRef]
8. Jahan, I.S.; Snasel, V.; Misak, S. Intelligent Systems for Power Load Forecasting: A Study Review. *Energies* **2020**, *13*, 6105. [CrossRef]
9. Yazici, I.; Beyca, O.F.; Delen, D. Deep-Learning-Based Short-Term Electricity Load Forecasting: A Real Case Application. *Eng. Appl. Artif. Intell.* **2022**, *109*, 104645. [CrossRef]
10. Aguilar Madrid, E.; Antonio, N. Short-Term Electricity Load Forecasting with Machine Learning. *Information* **2021**, *12*, 50. [CrossRef]
11. Hafeez, G.; Alimgeer, K.S.; Khan, I. Electric Load Forecasting Based on Deep Learning and Optimized by Heuristic Algorithm in Smart Grid. *Appl. Energy* **2020**, *269*, 114915. [CrossRef]
12. Huang, S.; Zhang, J.; He, Y.; Fu, X.; Fan, L.; Yao, G.; Wen, Y. Short-Term Load Forecasting Based on the CEEMDAN-Sample Entropy-BPNN-Transformer. *Energies* **2022**, *15*, 3659. [CrossRef]
13. Yang, Y.; Che, J.; Deng, C.; Li, L. Sequential Grid Approach Based Support Vector Regression for Short-Term Electric Load Forecasting. *Appl. Energy* **2019**, *238*, 1010–1021. [CrossRef]
14. Baesmat, K.H.; Shokoohi, F.; Farrokhi, Z. SP-RF-ARIMA: A sparse random forest and ARIMA hybrid model for electric load forecasting. *Glob. Energy Interconnect.* **2025**, *8*, 486–496. [CrossRef]
15. Baur, L.; Ditschuneit, K.; Schambach, M.; Kaymakci, C.; Wollmann, T.; Sauer, A. Explainability and Interpretability in Electric Load Forecasting Using Machine Learning Techniques—A Review. *Energy AI* **2024**, *16*, 100358. [CrossRef]
16. Islam, B.; Ahmed, S.F. Short-term electrical load demand forecasting based on LSTM and RNN deep neural networks. *Math. Probl. Eng.* **2022**, *2022*, 2316474. [CrossRef]
17. Kwon, B.S.; Park, R.J.; Song, K.B. Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer. *J. Electr. Eng. Technol.* **2020**, *15*, 1501–1509. [CrossRef]
18. 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]

19. Ke, K.; Hongbin, S.; Chengkang, Z.; Brown, C. Short-term electrical load forecasting method based on stacked auto-encoding and GRU neural network. *Evol. Intell.* **2019**, *12*, 385–394. [CrossRef]
20. Wu, K.; Peng, X.; Chen, Z.; Su, H.; Quan, H.; Liu, H. A Novel Short-Term Household Load Forecasting Method Combined BiLSTM with Trend Feature Extraction. *Energy Rep.* **2023**, *9*, 1013–1022. [CrossRef]
21. Xu, Y.; Jiang, X. Short-term power load forecasting based on BiGRU-Attention-SENet model. *Energy Sources Part A Recovery Util. Environ. Eff.* **2022**, *44*, 973–985. [CrossRef]
22. Zhang, G.; Wei, C.; Jing, C.; Wang, Y. Short-term electrical load forecasting based on time augmented transformer. *Int. J. Comput. Intell. Syst.* **2022**, *15*, 67. [CrossRef]
23. Chan, J.W.; Yeo, C.K. A Transformer-Based Approach to Electricity Load Forecasting. *Electr. J.* **2024**, *37*, 107370. [CrossRef]
24. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
25. Zhang, J.; Wei, Y.M.; Li, D.; Tan, Z.; Zhou, J. Short-Term Electricity Load Forecasting Using a Hybrid Model. *Energy* **2018**, *158*, 774–781. [CrossRef]
26. Bashir, T.; Haoyong, C.; Tahir, M.F.; Liqiang, Z. Short-Term Electricity Load Forecasting Using Hybrid Prophet-LSTM Model Optimized by BPNN. *Energy Rep.* **2022**, *8*, 1678–1686. [CrossRef]
27. Jiang, B.; Wang, Y.; Wang, Q.; Geng, H. A Novel Interpretable Short-Term Load Forecasting Method Based on Kolmogorov–Arnold Networks. *IEEE Trans. Power Syst.* **2024**, *40*, 1180–1183. [CrossRef]
28. Yi, Z.; Xu, Y.; Wang, H.; Sang, L. Coordinated Operation Strategy for a Virtual Power Plant with Multiple DER Aggregators. *IEEE Trans. Sustain. Energy* **2021**, *12*, 2445–2458. [CrossRef]
29. Li, K.; Huang, W.; Hu, G.; Li, J. Ultra-Short-Term Power Load Forecasting Based on CEEMDAN-SE and LSTM Neural Network. *Energy Build.* **2023**, *279*, 112666. [CrossRef]
30. Mounir, N.; Ouadi, H.; Jrhilifa, I. Short-Term Electric Load Forecasting Using an EMD-BiLSTM Approach for Smart Grid Energy Management System. *Energy Build.* **2023**, *288*, 113022. [CrossRef]
31. Liu, F.; Liang, C. Short-Term Power Load Forecasting Based on AC-BiLSTM Model. *Energy Rep.* **2024**, *11*, 1570–1579. [CrossRef]
32. Liu, C.L.; Chang, T.Y.; Yang, J.S.; Huang, K.-B. A Deep Learning Sequence Model Based on Self-Attention and Convolution for Wind Power Prediction. Renew. *Energy* **2023**, *219*, 119399.
33. Somvanshi, S.; Javed, S.A.; Islam, M.M.; Pandit, D.; Das, S. A Survey on Kolmogorov–Arnold Network. *ACM Comput. Surv.* **2024**, *58*, 55. [CrossRef]
34. Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T.Y.; Tegmark, M. KAN: Kolmogorov–Arnold Networks. *arXiv* **2024**, arXiv:2404.19756.

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.