

Article

Vibration Trend Prediction of Pumped Storage Unit Based on Temporal-Enhanced GAN and Improved Bidirectional LSTM

Ziwei Zhong ^{1,*}, Lingkai Zhu ¹, Lei Deng ², Fei Zhang ² , Junshan Guo ¹, Kai Liang ¹ and Jun Xie ³ ¹ State Grid Shandong Electric Power Research Institute, Jinan 250003, China; sddky_zhulingkai@163.com (L.Z.); junshan_guo@sina.com (J.G.); 13260076400@163.com (K.L.)² State Grid Electric Power Engineering Research Institute Co., Ltd., Beijing 100069, China; denglei@ei.sgcc.com.cn (L.D.); spiritgiant@126.com (F.Z.)³ College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China; jxie@hhu.edu.cn

* Correspondence: dkyzhongziwei@126.com

Abstract

Accurate prediction of the state trend of pumped storage units (PSUs) is essential for timely anomaly detection and preventive maintenance to improve the overall economic performance of power plants. Nevertheless, the complex and time-varying characteristics of PSU vibration data increase the difficulty of accurately modeling their dynamic evolution. In response to this problem, an integrated vibration trend prediction (VTP) method for PSUs is developed by combining a temporal-enhanced generative adversarial network (TEGAN) with an improved bidirectional long short-term memory network (IBiLSTM). Firstly, TEGAN expands the original dataset by synthesizing artificial samples, thereby improving the structural diversity and representativeness of vibration data. Within TEGAN, a temporal characterization (TC) module is designed to collaboratively guide the generator and the discriminator, while a data processing module is adopted to incorporate structural priors into the learning process. Secondly, variational mode decomposition (VMD) is applied to decompose the original vibration data into intrinsic modes, followed by PSR to reconstruct the components of each modality into a higher-dimensional state space representation. Subsequently, by incorporating the proposed multi-order Kolmogorov–Arnold network (M-KAN) for high-order nonlinear fitting, IBiLSTM is employed to model each reconstructed sub-sequence. Finally, the outputs of all sub-sequences are aggregated to produce the final VTP results. The comparative evaluation verifies the advantages of the developed method in terms of prediction accuracy and robustness for PSU vibration trend forecasting.

Keywords: PSUs; vibration trend prediction; temporal generative adversarial network; VMD; IBiLSTM



Academic Editor: Chenliang Liu

Received: 16 July 2026

Revised: 5 August 2026

Accepted: 12 August 2026

Published: 21 August 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In recent decades, the increasing frequency of extreme weather events has led to greater variability and uncertainty in energy supply, thereby accelerating the transition toward low-carbon and sustainable energy systems [1]. Driven by global climate strategies, large-scale deployment of renewable energy has been actively promoted worldwide [2]. Among various renewable sources, wind and solar power have achieved rapid development and large-scale integration. Nevertheless, their inherent intermittency and variability introduce new challenges to grid stability and operational security. To address these issues, pumped storage plants, as a mature and economically viable energy storage solution, have been employed to enhance power system stability, including peak shaving, valley filling, and frequency regulation [3,4].

In the operation and maintenance of pumped storage power stations, as the operational life of pumped storage units (PSUs) increases, the structural integrity and operational condition of PSUs progressively degrade. This poses a potential threat to the operational safety of the power station and the overall stability of the power system [5]. Generally, this degradation is manifested in vibration data, which serve as a reliable indicator of unit health status. Therefore, accurate vibration trend prediction of PSUs is critical. However, the vibration data of PSUs is inherently nonlinear and non-stationary due to multiple sources of interference, structural coupling, and varying operating conditions, which poses significant challenges to accurate trend prediction. Thus, developing an efficient vibration trend prediction model with robust generalization capability has become essential for ensuring the long-term stable operation of PSUs.

Existing studies on vibration trend modeling of PSUs can be broadly categorized into machine learning (ML) methods and deep learning (DL) methods. ML methods typically rely on feature engineering applied to historical vibration data and regression models such as support vector machine (SVM), random forest (RF), and kernel extreme learning machine (KELM) for trend prediction [6–8]. These methods achieve high computational efficiency in cases with small sample sizes, well-defined features, and low noise levels. However, given the inherent nonlinearity of vibration data, ML methods often struggle to deliver precise forecasts while maintaining reliable adaptability across different operating conditions.

With the rapid development of AI, DL has emerged as a key approach for time series prediction and has been extensively studied in wind power, solar energy, and other energy-related domains [9–14]. In pumped storage applications, the integration of deep learning with signal processing and optimization techniques has attracted growing attention for the vibration trend forecasting of PSUs. For example, to overcome the sensitivity of conventional variational mode decomposition (VMD) to parameter configuration, Lu et al. [15] introduced an adaptive decomposition scheme that enhances the quality of signal decomposition. Yan et al. [16] combined VMD with a recursive method to develop a multi-step prediction model for forecasting the vibration trends in hydropower units. Hu et al. [17] developed a novel framework that integrated VMD optimized by bat algorithm with wavelet threshold denoising. This framework enhanced the quality of signal decomposition, effectively suppressed noise, and achieved high-precision vibration trend prediction.

Although the aforementioned methods can accurately predict vibration trends, they are constrained by the limited diversity of vibration data distribution, leading to poor generalization to previously unseen dynamic patterns under real-world operating conditions. Generative adversarial networks (GANs), known for their powerful sample generation capabilities, have been widely applied in data augmentation tasks [18]. Through the synergy between the discriminator and the generator, GANs effectively enhance the diversity of data distributions, thereby improving the model's ability to capture complex vibration patterns. However, traditional GANs rely solely on random noise vectors for sample generation and lack mechanisms to guide the generation toward specific target characteristics, resulting in highly stochastic and uncontrollable outputs. Conditional GANs (CGANs) address this limitation by introducing class labels or domain features as conditional variables, which are incorporated into both the generator and the discriminator during training [19]. Nevertheless, training instability and mode collapse remain significant challenges for CGANs, particularly under complex or sparsely sampled conditional spaces. To further alleviate these issues, Wasserstein GAN with gradient penalty (WGAN-GP) incorporates gradient penalty term into the discriminator loss function, thereby enhancing discriminator stability and effectively mitigating mode collapse [20]. Although these generative models can produce synthetic samples, their limited ability to capture latent data characteristics restricts the quality of the generated samples.

To further clarify the research gap between existing vibration trend prediction (VTP) methods and the proposed method, a comparative summary is provided in Table 1.

Table 1. Comparison of representative vibration trend prediction methods.

Reference	Method	Characteristics	Limitations
[17]	BA-VMD + LSTM	LSTM is employed to model temporal dependencies for vibration trend prediction.	LSTM has limited capability in capturing complex nonlinear degradation characteristics.
[21]	EEMD + IDBO-CNN-LSTM	CNN-LSTM is introduced to enhance local feature extraction and temporal dependency modeling.	CNN-LSTM mainly focuses on local feature extraction and lacks sufficient capability for complex nonlinear representation.
[22]	WGAN-GP + VMD + FBiLSTM	WGAN-GP is used for data augmentation, and KAN enhances the nonlinear mapping ability of the model	Conditional generation can provide richer temporal guidance, while the model still lacks effective local feature extraction capability.
This paper	TEGAN + VMD + IBiLSTM	TEGAN performs temporal-guided data augmentation, and M-KAN adaptively integrates convolutional features with multi-order Chebyshev representations.	

As summarized in Table 1, recent studies have demonstrated that integrating LSTM, CNN, or KAN with signal decomposition techniques can effectively improve VTP accuracy. Specifically, LSTM-based models exhibit superior capability in modeling temporal dependencies, CNN-based architectures enhance local feature extraction, and KAN further strengthens nonlinear function approximation. Nevertheless, an effective fusion strategy that combines CNN-based local feature extraction with KAN-based multi-order nonlinear mapping within a unified module remains insufficiently explored. To address this gap, the proposed M-KAN incorporates deep convolution, multi-order Chebyshev polynomial mapping, and a gated fusion mechanism into a unified architecture, enabling adaptive fusion and joint optimization of local patterns and high-order nonlinear representations.

Meanwhile, WGAN-GP-based data augmentation has been proven effective in enriching vibration data distribution and improving prediction robustness. Building upon this framework, incorporating temporal structural representations as conditional information can provide the generator with richer temporal information and improve the quality of synthesized samples for vibration trend prediction. Based on the preceding analysis, a VTP method for PSUs is developed by integrating a temporal-enhanced generative adversarial network (TEGAN) with improved bidirectional long short-term memory (IBiLSTM). Firstly, TEGAN is employed to generate synthetic vibration samples, thereby effectively increasing diversity and representativeness of the vibration dataset. Then, the original vibration data is decomposed into multiple intrinsic mode components via VMD, followed by phase space reconstruction (PSR) to embed each subsequence into a higher-dimensional state space representation. Subsequently, the reconstructed subsequences are fed into IBiLSTM for temporal modeling, enhanced by multi-order Kolmogorov–Arnold network (M-KAN) for high-order nonlinear fitting, yielding the final prediction results. Finally, the final PSU vibration trend is obtained by aggregating the prediction results of all subsequences. The main contributions are summarized as follows:

- (1) TEGAN is proposed to augment time series samples and enhance the diversity of data distributions. Firstly, K-means clustering is employed to categorize the original

vibration sequences into distinct patterns. Then, a temporal characterization (TC) module is designed to learn the feature distributions associated with each category. Guided by the information from TC, the generator performs conditional synthesis of samples corresponding to distinct structural patterns, thereby producing more targeted and diverse data. Meanwhile, the temporal representations extracted by the TC module are embedded into the discriminator as conditional conditions to jointly assess the structural consistency and authenticity of the generated data.

- (2) By embedding the multi-order Kolmogorov–Arnold network (M-KAN) module within IBiLSTM, the proposed method can more effectively capture the complex nonlinear and non-stationary characteristics of PSUs' vibration patterns.
- (3) A novel vibration trend forecasting method is developed, integrating TEGAN-based time series data augmentation with IBiLSTM network. Experiment results demonstrate that the proposed method significantly achieves the predictive accuracy of vibration trends, providing an effective approach for subsequent engineering applications.

2. Methodological Foundations

2.1. VMD

As an adaptive decomposition method, VMD separates complicated non-stationary data into a limited set of intrinsic mode functions, all of which possess narrow bandwidths and well-separated central frequencies [23]. Its formulation is as follows:

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum_{k=1}^K u_k(t) = f(t) \end{cases} \quad (1)$$

where $f(t)$ is raw data, $u_k(t)$ represents the k -th decomposed mode, ω_k is its corresponding center frequency, and t denotes time.

The constrained optimization problem is transformed into augmented Lagrange function L , which is calculated as follows:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \quad (2)$$

The mode functions $u_k(t)$ and center frequency ω_k are obtained via ADMM-based iterative updates as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (4)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \gamma \left[\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right] \quad (5)$$

where n denotes the number of iterations, and $\hat{f}(\omega)$, $\hat{u}_i(\omega)$, $\hat{\lambda}(\omega)$ and $\hat{u}_k^{n+1}(\omega)$ denote the Fourier-domain representations of $f(t)$, $u_i(t)$, $\lambda(t)$ and $u_k^{n+1}(t)$, respectively.

2.2. PSR

PSR transforms the original series into a reconstructed state space using delay embedding, thereby revealing the underlying system dynamics [24]. The expression is as follows:

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_L \end{bmatrix}^T = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_i & x_{i+\tau} & \cdots & x_{i+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-(d-1)\tau} & x_{N-(d-2)\tau} & \cdots & x_N \end{bmatrix} \quad (6)$$

The corresponding target matrix is constructed as

$$Y = \begin{bmatrix} X_{1+(d-1)\tau} & X_{2+(d-1)\tau} & \cdots & X_N \end{bmatrix}^T \quad (7)$$

3. TEGAN-IBiLSTM

3.1. TEGAN

3.1.1. Data Processing Module

To endow TEGAN with initial classification capability in the absence of labeled data, an unsupervised data processing module based on pseudo-labels is proposed. Specifically, the vibration data is segmented into N equal-length time series segments, each of length T , and each segment is then flattened into one-dimensional vector for subsequent analysis. K-means clustering is then applied to assign pseudo-labels, which are used as supervisory signals for subsequent model training [25]. Considering that the pseudo-labels are only intended to provide weak structural guidance for learning temporal characterization representations, they are exclusively used to pretrain the TC module. After pretraining, the TC module is frozen during adversarial training, ensuring that the pseudo-labels only affect the representation learning stage and do not directly constrain the generator optimization.

To reduce bias caused by imbalanced pseudo-label distributions, weighted sampling mechanism based on class occurrence count is introduced. The sampling weight for the i -th sample is defined as follows:

$$w_i = \frac{1}{f_{y_i} + \varepsilon} \quad (8)$$

where f_{y_i} represents the number of instances belonging to category y_i and ε is a smoothing term introduced to prevent division by zero. The sampling probabilities are obtained by normalizing the sampling weights w_i , enhancing the classifier's ability to identify underrepresented trend categories.

3.1.2. TC Module

The TC module is designed to extract representative temporal features that characterize the structural patterns of vibration sequences. First, an LSTM network is employed to capture sequential dependencies, followed by a multi-layer perceptron (MLP) to map these dependencies into high-level feature representations. To enhance feature discriminability, classification loss based on pseudo-labels derived from K-means is introduced to guide the TC module in learning discriminative temporal representations. Finally, learned temporal representations are used to guide the generator in producing structure-consistent samples, while being incorporated into the discriminator as conditional information to improve the quality of the generated data.

By introducing a gating mechanism, LSTM effectively improves the stability of RNNs [26]. As a result, the LSTM is well-suited for modeling vibration data with temporal dependencies, as it preserves relevant information while effectively filtering out irrelevant components. The architecture of the LSTM is shown in Figure 1.

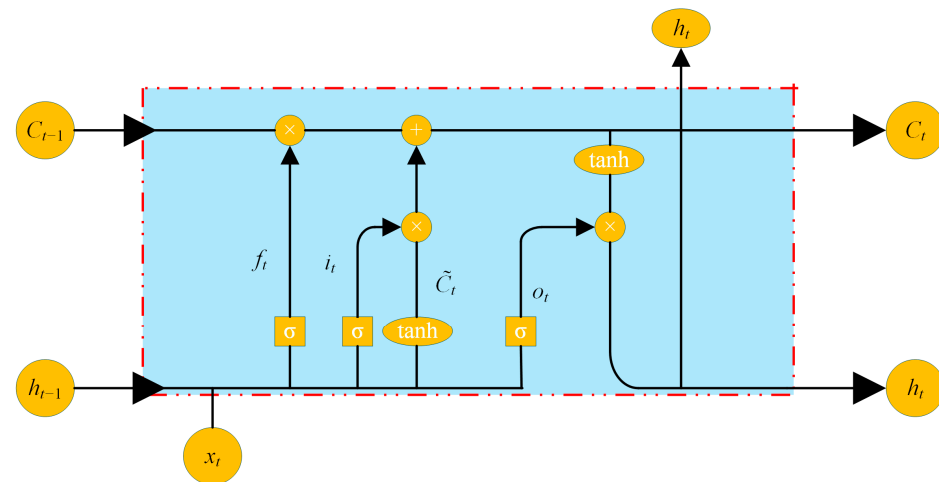


Figure 1. LSTM architecture diagram.

Forget gate controls what information from the previous cell state C_{t-1} should be discarded, and its output is given by

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

where x_t represents the current input vector, h_{t-1} contains the contextual information transferred, W_f and b_f are optimized during training to control the gate behavior, and σ enables adaptive filtering of historical information.

The input gate regulates the storage of newly introduced information in the cell state. It is given by

$$C_t = f_t \cdot C_{t-1} + i_t \tilde{C}_t \quad (10)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (11)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$

The output gate controls the contribution of the cell state to the current output representation h_t , which is expressed as

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (13)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (14)$$

where C_t preserves historical information. In contrast, h_t summarizes the current network state and serves as the immediate representation used for output generation.

MLP architecture includes fully connected layer with ReLU activation, Dropout, and final linear output layer that produces K-dimensional logits, which are transformed into a probability distribution via softmax function. This distribution is used as conditional information for both the generator and discriminator in TEGAN. The architecture of MLP is illustrated in Figure 2.



Figure 2. MLP architecture diagram.

3.1.3. CGAN

To enhance the generator's control over structural patterns in the data, a conditional mechanism is introduced within GAN, leading to the construction of CGAN [27]. Specifically, the probability vectors output by the TC module are used as conditional inputs to guide the generator in synthesizing samples that reflect specific temporal structural characteristics. Meanwhile, the discriminator incorporates these conditional probabilities as conditional information, improving its capacity to jointly evaluate the authenticity of generated samples and their consistency with the specified structural categories. The architecture of CGAN is shown in Figure 3.

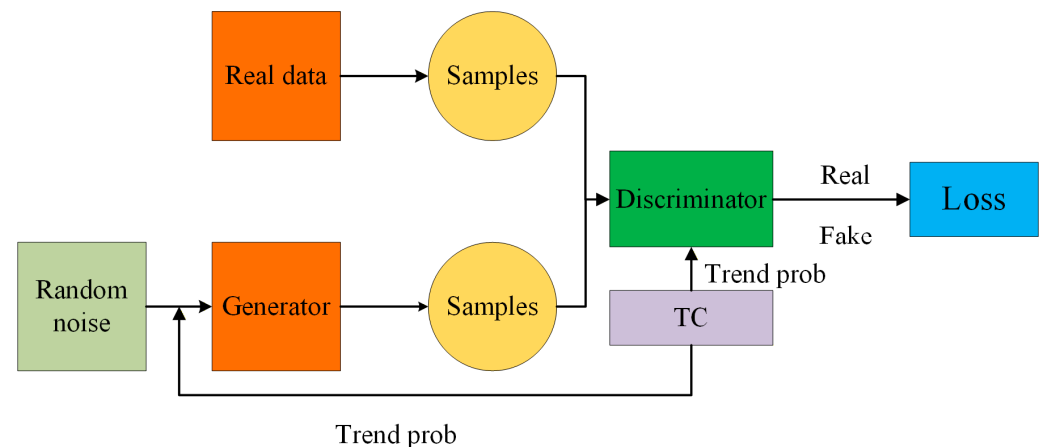


Figure 3. CGAN architecture diagram.

Traditional CGAN optimizes the adversarial objective using a probability-distribution discrepancy measure. When the generated and real distributions are significantly separated, the optimization process may suffer from weak gradients, resulting in unstable convergence and reduced sample diversity. To address the issue, TEGAN introduces Wasserstein distance as an optimization objective, enhancing the stability of training and improving the convergence between the generator and the discriminator [28,29]. The corresponding formulation is given as

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y) \sim \gamma} [c(x, y)] \quad (15)$$

where P_r and P_g correspond to the distributions of observed and synthesized samples, respectively. γ denotes the collection of admissible coupling distributions linking the two domains, and $c(x, y)$ measures the transportation effort required to transform a sample x into y .

To enforce the Lipschitz constraint, the following penalty term is introduced:

$$L_{GP} = E_{\hat{x} \sim P_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (16)$$

where $\hat{x}$ denotes a point sampled along the line segment connecting real and synthesized instances. $D(\hat{x})$ is the discriminator output evaluated at $\hat{x}$, and $\nabla_{\hat{x}} D(\hat{x})$ denotes its local derivative with respect to the input.

3.1.4. Self-Attention Module

As illustrated in Figure 4, self-attention is a fundamental mechanism in DL that dynamically adjusts the relevance of each element by computing its relationship with all other elements in the input sequence [30]. The core principle of self-attention is to capture dependencies between different positions in a sequence by computing a weighted sum of the value vectors, where the weights are derived from the similarity between query and key representations.

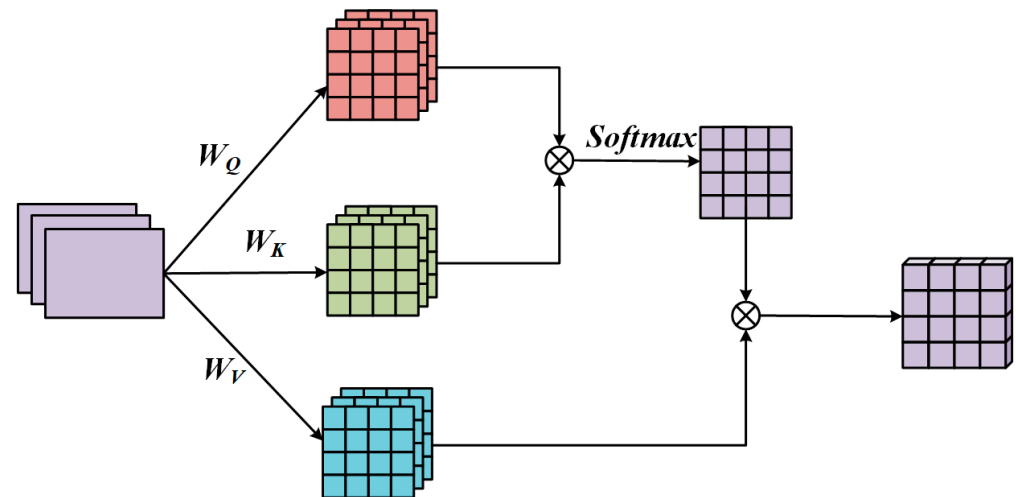


Figure 4. Self-attention architecture diagram.

Let $X = \{x_1, x_2, \dots, x_n\}$ denote an input vibration sequence with length n , where each element $x_i \in \mathbb{R}^d$ corresponds to the feature vector at the i -th position in the sequence. In a standard self-attention mechanism, the query, key, and value are first obtained through three separate linear transformations, which are defined as follows:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (17)$$

where W_Q , W_K and W_V are the learnable parameters.

Subsequently, the similarities between the queries and keys are computed to determine the relationships among elements. The dot product is typically used to measure the similarity between a query and a key. The corresponding formula is defined as follows:

$$\text{Attention}(Q, K) = \frac{QK^T}{\sqrt{d_k}} \quad (18)$$

where $\sqrt{d_k}$ is the scaling factor.

3.1.5. Overall Architecture of TEGAN

TEGAN incorporates an unsupervised data processing stage to provide supervisory signals, based on which the TC module models vibration sequences and captures temporal dependencies via LSTM. The extracted features are then passed through MLP to output a probability distribution over structural category. This distribution serves as a conditional input for both the generator and the discriminator. Conditioned on both random noise and the probability distribution provided by the TC module, the generator first employs a self-attention module to focus on critical time series segments. Subsequently, representative vibration data samples exhibiting temporal characteristics are generated. Furthermore, the discriminator, while evaluating the authenticity of input vibration data, incorporates the probability vectors from the TC module as conditional inputs, thereby enhancing the

consistency between the generated samples and the learned temporal structural distribution. The architecture of TEGAN is illustrated in Figure 5.

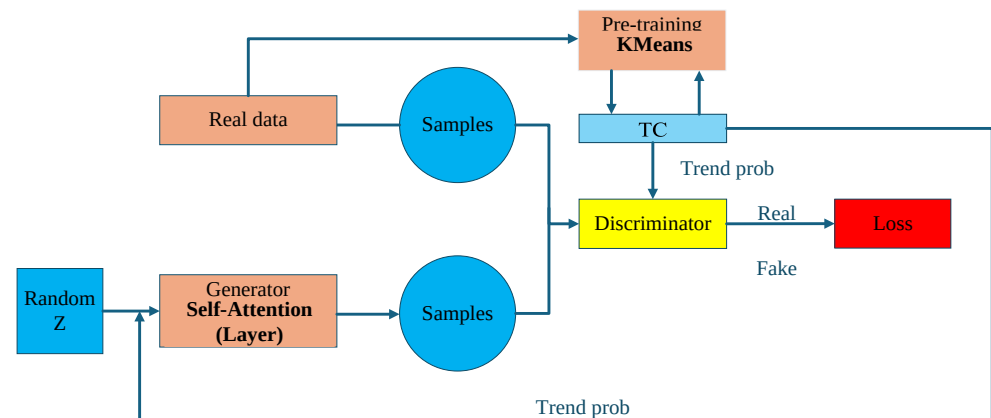


Figure 5. TEGAN architecture diagram.

3.2. IBiLSTM

3.2.1. BiLSTM

BiLSTM models temporal dependencies using forward and backward LSTM branches, enabling more comprehensive temporal representation [31]. The structure of BiLSTM is shown in Figure 6, comprising two parallel LSTMs, which process forward sequence $\vec{h}_t$ and reverse sequence $\overleftarrow{h}_t$, respectively. The calculation is as follows:

$$\vec{h}_t = f(W_f x_t + U_f \vec{h}_{t-1} + b_f) \quad (19)$$

$$\overleftarrow{h}_t = f(W_b x_t + U_b \overleftarrow{h}_{t-1} + b_b) \quad (20)$$

where W , U , and b denote the learnable parameters associated with the bidirectional recurrent structure. The arrows indicate the information flow in opposite temporal directions.

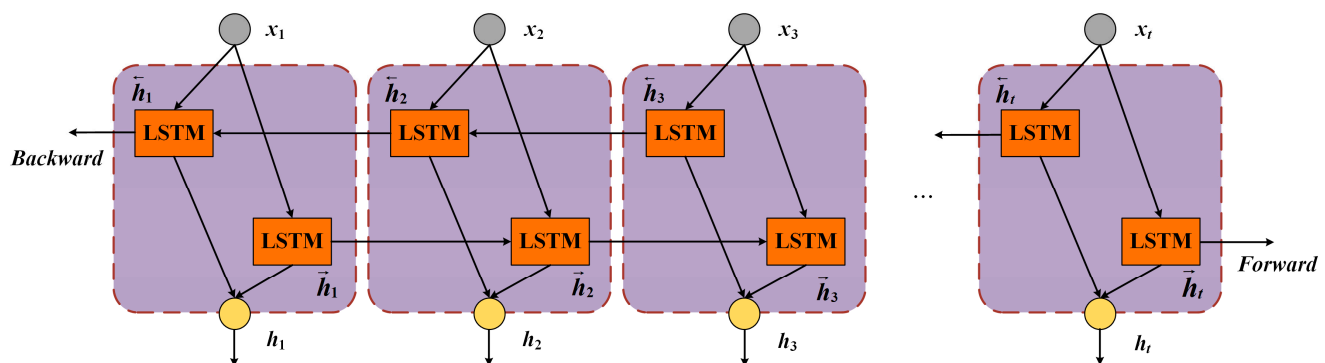


Figure 6. BiLSTM structure diagram.

The network integrates contextual features learned from past and future sequences to generate the hidden representation h_t . The resulting output can be expressed as

$$h_t = \left[\vec{h}_t; \overleftarrow{h}_t \right] \quad (21)$$

3.2.2. M-KAN Layer

By representing a multivariate function as a combination of several univariate functions, the Kolmogorov–Arnold network (KAN) effectively enhances the modeling of high-

dimensional data [32]. The overall structure is depicted in Figure 7. KAN can be expressed as follows:

$$f(x) = f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^n \varphi_{q,p}(x_p) \right) \quad (22)$$

where $\varphi_{q,p}$ and ϕ_q denote nonlinear transformation functions involved in the representation process.

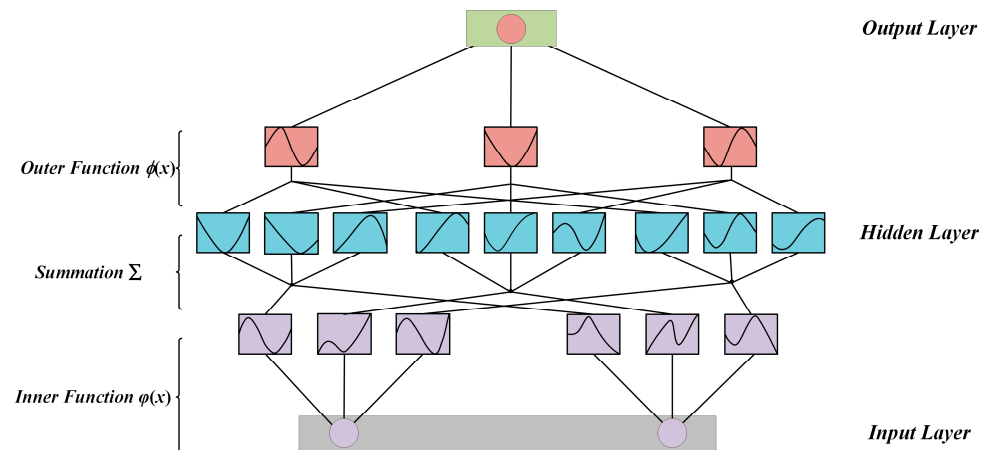


Figure 7. KAN structure diagram.

During training, spline coefficients are automatically refined to enhance the model's representation flexibility. The corresponding mapping function is given by

$$\phi(x) = w(b(x) + spline(x)) \quad (23)$$

$$b(x) = silu(x) = \frac{x}{(1 + e^{-x})} \quad (24)$$

$$spline(x) = \sum_i c_i B_i(x) \quad (25)$$

where $B_i(x)$ denotes the corresponding B-spline basis function.

To improve the nonlinear representation capability, Huang et al. [33] proposed TimeKAN, which includes an M-KAN module that integrates convolutional features with multi-order Chebyshev KAN representations. Considering the computational cost and the characteristics of PSU vibration data, the frequency decomposition and frequency mixing components in the original framework are removed to reduce computational complexity. Based on this simplified architecture, the proposed M-KAN further improves the feature fusion mechanism of the original framework by introducing an adaptive weighting strategy. Different from the original M-KAN, which directly combines convolutional features and KAN representations, the proposed method enables more flexible integration of nonlinear and local features according to the characteristics of input vibration data. Furthermore, a learnable residual weighting mechanism and layer normalization are incorporated to enhance feature fusion stability and improve model convergence. The structure of the MKAN layer is shown in Figure 8. As illustrated in Figure 8, the M-KAN layer comprises several key components, including M-KAN, deep one-dimensional convolution (D-Conv), residual weighting module, and layer normalization (Layer Norm). Specifically, M-KAN employs multiple feature branches with varying orders, where each branch models the nonlinear relationships between features using Chebyshev polynomials of the corresponding order,

enabling multi-level modeling from low to high orders. The Chebyshev polynomial used can be defined as follows:

$$T_n(x) = \cos(n \cdot \arccos(x)) \quad (26)$$

where n is the highest degree of the Chebyshev polynomial, with model complexity increasing as the polynomial degree rises.

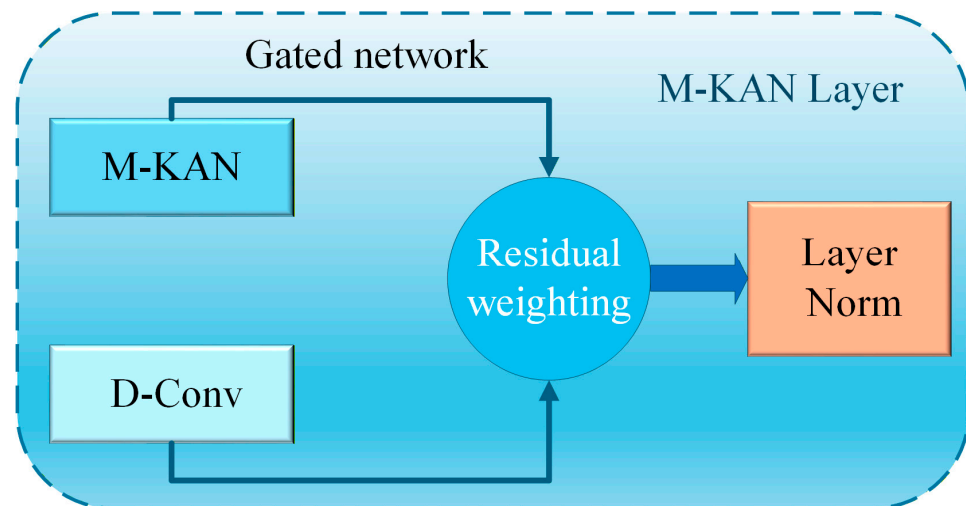


Figure 8. M-KAN Layer structure diagram.

The Chebyshev KAN of one layer can be formulated as follows:

$$\phi_o(x) = \sum_{j=1}^D \sum_{i=0}^n \theta_{o,j,i} T_i(\tanh(x_j)) \quad (27)$$

$$KAN(x) = \begin{Bmatrix} \phi_1(x) \\ \vdots \\ \phi_D(x) \end{Bmatrix} \quad (28)$$

where o denotes the output neuron, D represents the polynomial order, and θ is the learnable coefficient in the linear combination of Chebyshev polynomial.

A gated network is introduced to assign dynamic weights to the outputs of Chebyshev-KAN branches corresponding to different polynomial orders. Specifically, global statistical features of the input, such as the mean and maximum, are extracted and processed by a fully connected layer with Softmax normalization to obtain fusion weights. These weights are then used to adaptively fuse the outputs of each branch. To enhance the ability of the model to capture local patterns in time series, the M-KAN layer incorporates a D-Conv module. The process of D-Conv is as follows:

$$f_{i,1} = \text{Conv}_{D \rightarrow D}(f_i, \text{group} = D) \quad (29)$$

Simultaneously, the outputs of the multi-order Chebyshev KAN branches and the D-Conv branch are combined via residual fusion, where the fusion weights are learnable parameters optimized during training. A layer norm is further applied to enhance training stability and ensure feature consistency. The formula for residual weighting is as follows:

$$Y = \alpha X_{KAN} + (1 - \alpha) X_{Conv} \quad (30)$$

where α is a learnable parameter that adaptively controls the proportion of information contributed by the M-KAN and D-Conv branches during training.

3.2.3. Multi-Network Integration

To enhance the modeling capability for non-stationary vibration data of PSUs, this paper proposes IBiLSTM that integrates BiLSTM with M-KAN. The model first employs BiLSTM to capture temporal dependencies in the input sequence. The extracted features are then fed into the M-KAN layer, which adaptively models multi-order nonlinear relationships. Finally, the refined feature representations generated by the M-KAN layer are subsequently passed to a fully connected layer to produce the final prediction results. The structure of IBiLSTM is shown in Figure 9.

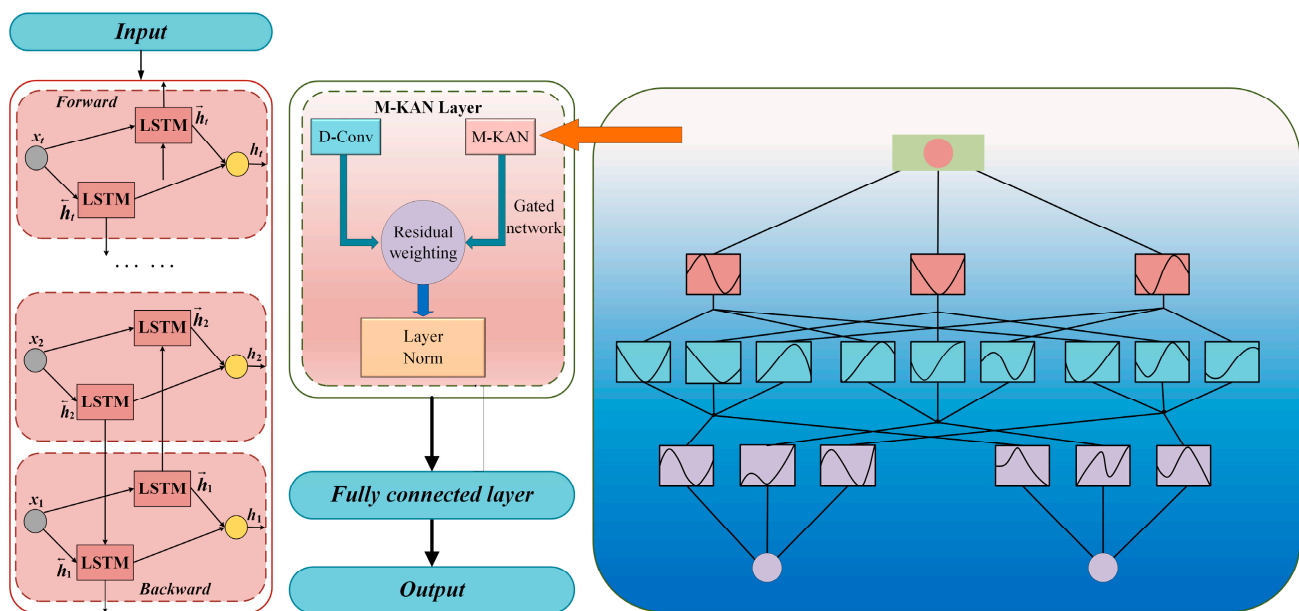


Figure 9. IBiLSTM structure diagram.

4. The Overall Framework of the Proposed Method

Non-stationary characteristics in PSUs vibration data pose significant challenges for accurate trend prediction. To address this issue, TEGAN and IBiLSTM are combined to construct an effective vibration forecasting method. The overall process is illustrated in Figure 10. The procedure can be summarized as follows:

- Step 1: The raw vibration data of PSUs are acquired, and TEGAN is then applied to generate additional samples, thereby enriching the data distribution.
- Step 2: Decomposition of the composite data by VMD.
- Step 3: Phase space reconstruction of each subsequence via PSR.
- Step 4: Feeding the processed subsequences into the proposed IBiLSTM to obtain individual predictions.
- Step 5: Aggregate the VTP results of all subsequences to obtain the final VTP of PSUs.

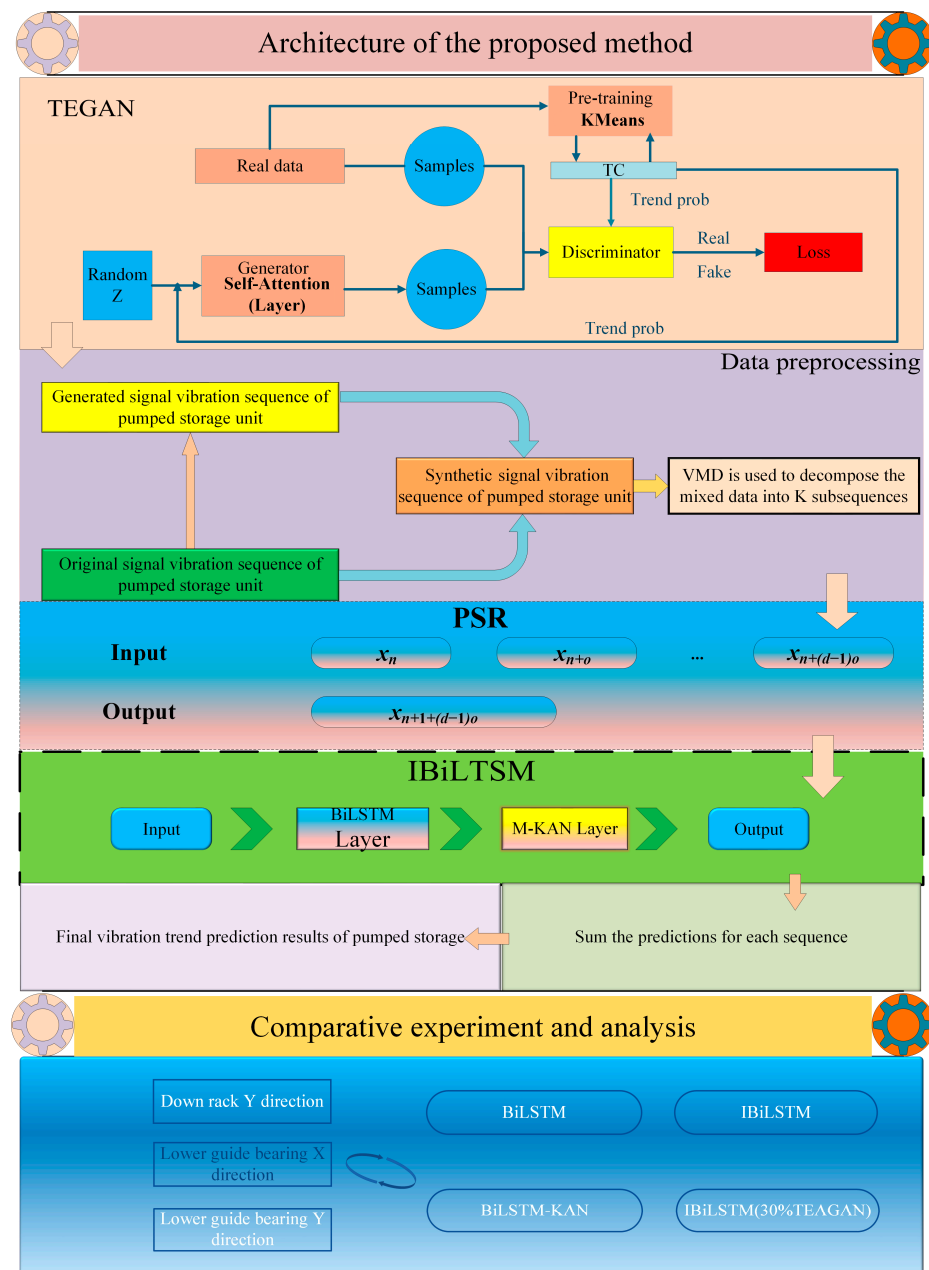


Figure 10. Overall prediction workflow.

5. Experimental Analysis

5.1. Data Description and Preprocessing

Vibration data collected from a pumped storage unit in China are used in this study. Specifically, Y-axis vibration data from the lower frame structure (LFS) of Unit 2#, along with the X- and Y-axis vibration data from the lower guide bearing (LGB) of Unit 3#, are selected for vibration trend prediction analysis. Prior to model training, the proposed TEGAN is employed to generate additional vibration samples with enhanced distribution diversity. These generated samples are then fused with the original measured data to construct an augmented dataset, which is subsequently used for model training and evaluation.

The vibration data were collected from the PSUs during 1 March and 30 June 2024 under normal operating conditions with sampling interval of 15 min. Considering the influence of unit start–stop processes and abnormal data, incomplete and abnormal operating periods were removed during data preprocessing, and only continuous and valid operating

data were retained. A total of 3045 valid samples were obtained. Considering the continuity of vibration trends in time series prediction, the first 1200 consecutive samples were selected as the original training dataset. In addition, vibration data collected from 1 December to 15 December 2024 were processed using the same data screening procedure, and 300 valid samples were reserved as an independent testing dataset for final performance evaluation. To prevent data leakage, the testing samples were completely isolated before the data augmentation process and were not involved in the training of either TEGAN or WGAN-GP. Based on the original training dataset, both TEGAN and WGAN-GP were employed for data augmentation, and 450 synthetic samples were incorporated into the training dataset. Before model training, the vibration data were preprocessed using LOESS smoothing for noise reduction and Z-score normalization for feature scaling.

5.2. The Value of K

To effectively evaluate the impact of the clustering number K on the quality of generated samples in TEGAN and determine an appropriate clustering configuration, the silhouette coefficient was employed to assess the clustering validity on three datasets. The results are presented in Figure 11. Considering both intra-cluster compactness and inter-cluster separability, the candidate range of K was set to [3,12]. As shown in Figure 11, by jointly considering the silhouette coefficient evaluation and practical engineering knowledge regarding vibration pattern categorization, the clustering number was finally determined as $K = 4$.

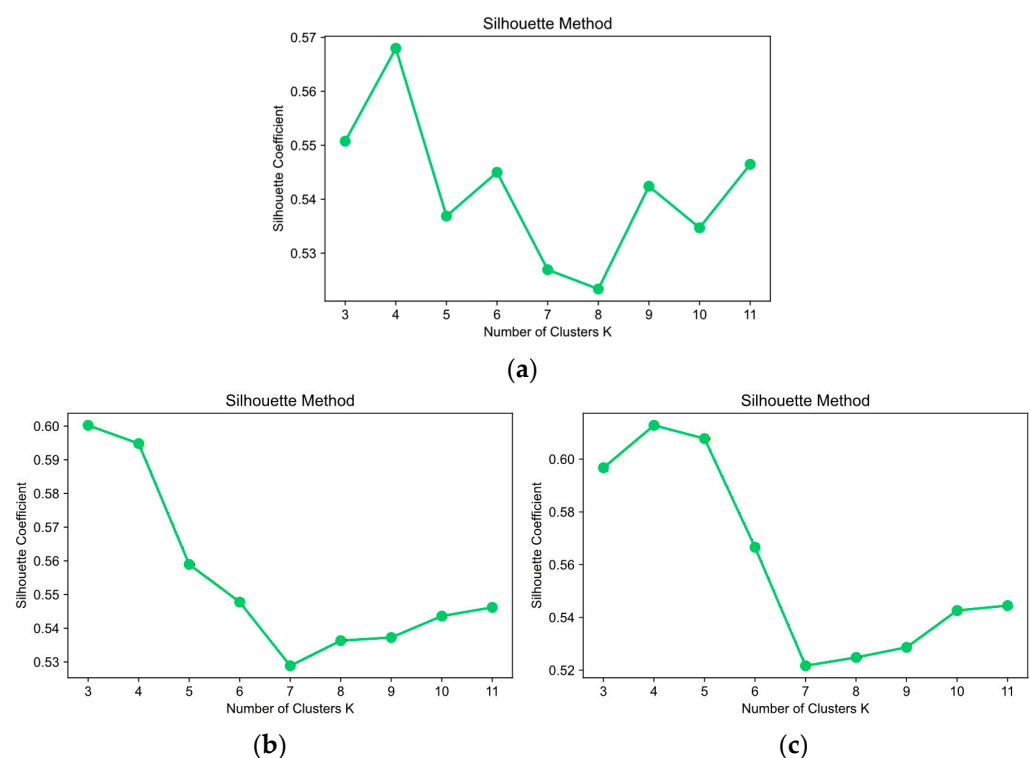


Figure 11. Silhouette coefficient: (a) Y-axis of LFS (b) X-axis of LGB; (c) Y-axis of LGB.

5.3. Vibration Data Decomposition

To address the nonlinearity and non-stationarity of vibration data, VMD is applied to decompose complex data. To improve parameter selection, particle swarm optimization is integrated with VMD, and permutation entropy is utilized as the objective function during the search process. The optimized parameters are determined as follows. For Y-axis vibration data of the lower frame structure, the optimal values are $K = 9$ and $\alpha = 2822$, as shown in Figure 12a. For the lower guide bearing, the X-axis and Y-axis data yield optimal

configurations of $K = 6$ and $\alpha = 2513$ and $K = 10$ and $\alpha = 2777$ respectively, as shown in Figure 12b,c.

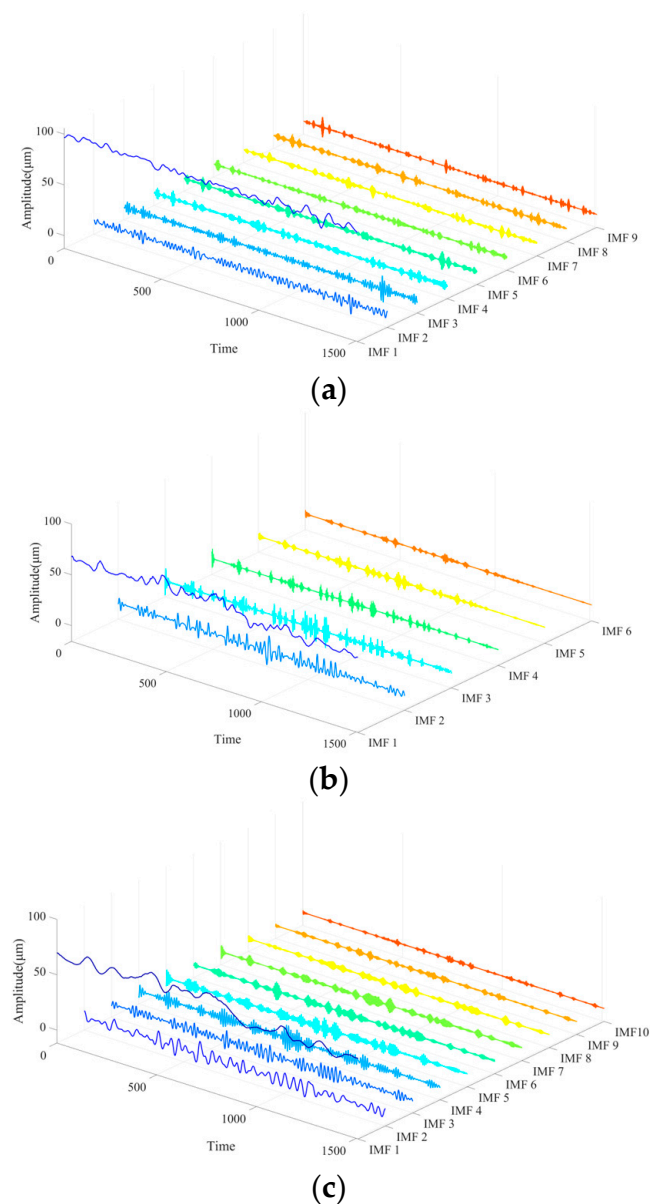


Figure 12. VMD decomposition results: (a) Y-axis of LFS; (b) X-axis of LGB; (c) Y-axis of LGB.

5.4. Parameter Settings

As model performance is highly dependent on parameter configurations, appropriate parameter settings are essential for reliable and accurate results. For IBiLSTM, key training parameters are updated using an Adam optimizer, while other conditional parameters are determined empirically through extensive experimentation and iterative testing to balance model accuracy and stability during training. All models including BiLSTM, BiLSTM-KAN, IBiLSTM, and the proposed method are all configured with aligned parameter settings. The learning rate across all models is uniformly set to 4×10^{-4} . Detailed parameter configurations are provided in Table 2.

Table 2. Method parameter settings.

	Parameters	Values
TEGAN	Cluster number K	4
	The maximum number of iterations for data processing	50
	drop rate	0.2
	Learning rate	1×10^{-3}
	Number of LSTM layers	256
IBiLSTM	Batch size	32
	Hidden layer dimension	128
	Network depth	2
	Maximum order	s5
	Time step number	10
PSR	τ	1
	d	10

5.5. Prediction Accuracy Metrics

The evaluation metrics used to assess prediction performance include MAE, RMSE and MAPE and R^2 . The corresponding formulations are presented below:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (31)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (32)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (33)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (34)$$

where n denotes the number of samples; y_i is the actual value of the i -th sample; $\hat{y}_i$ represents the predicted value of the i -th sample; and $\bar{y}$ is the average of the actual values, that is $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

5.6. Ablation Experiment

To comprehensively evaluate the effectiveness of each component in the proposed method, an ablation study was conducted by progressively introducing different modules. Specifically, BiLSTM, BiLSTM-KAN, IBiLSTM (BiLSTM-MKAN), and IBiLSTM-W (incorporating WGAN-GP-based data augmentation) were selected as the ablation variants. Since the ablation experiments mainly focus on investigating the contributions of the proposed prediction architecture and data augmentation strategy, VMD and PSR were not included as ablation factors. To ensure a fair comparison, identical VMD and PSR procedures were adopted for all prediction models. The prediction results and evaluation metrics of all ablation variants are presented in Table 3. As summarized in Table 3, the proposed method consistently achieves the lowest RMSE, MAE, and MAPE together with the highest R^2 across all three VTP tasks, demonstrating superior prediction accuracy and generalization capability compared with the other ablation variants.

Table 3. Evaluation metric results of all methods in the ablation experiments.

Case	Methods	RMSE	MAE	MAPE	R ²
Y-axis of the LFS	BiLSTM	4.2447	3.2304	3.3897	0.7562
	BiLSTM-KAN	3.0805	1.9445	2.0745	0.8716
	IBiLSTM	1.5495	0.9513	1.0222	0.9675
	IBiLSTM-W	1.3659	0.9532	0.9856	0.9747
	The proposed	1.0902	0.8256	0.8509	0.9839
X-axis of the LGB	BiLSTM	0.8992	0.6692	1.1247	0.9523
	BiLSTM-KAN	0.7408	0.5536	0.9246	0.9676
	IBiLSTM	0.7414	0.5306	0.8897	0.9676
	IBiLSTM-W	0.6748	0.4973	0.8392	0.9731
	The proposed	0.6377	0.4632	0.7873	0.9760
Y-axis of the LGB	BiLSTM	2.8484	2.3307	4.0003	0.7580
	BiLSTM-KAN	0.9033	0.7348	1.2279	0.9757
	IBiLSTM	0.5943	0.4683	0.7775	0.9895
	IBiLSTM-W	0.7209	0.5783	0.9735	0.9845
	The proposed	0.4907	0.3990	0.6706	0.9928

Furthermore, the comparison among BiLSTM, BiLSTM-KAN, and IBiLSTM verifies the effectiveness of the proposed M-KAN architecture. By incorporating MKAN as a multi-level nonlinear mapping module, the proposed method achieves more effective nonlinear feature representation and temporal dependency modeling compared with the conventional BiLSTM and BiLSTM-KAN. In addition, although WGAN-GP-based augmentation improves prediction performance, the proposed TEGAN consistently achieves further reductions in prediction errors and higher R² on all datasets. These results indicate that the trend-aware conditional generation mechanism can generate higher-quality vibration samples, thereby providing more effective augmented data for VTP of PSUs.

Overall, the experimental results demonstrate that the proposed method achieves superior prediction accuracy and robustness for VTP of PSUs, thereby providing an effective approach for intelligent health management of pumped storage systems.

The fitting performance of representative method variants across three VTP tasks is presented in Figure 13. As shown in Figure 13, in terms of the Y-axis of the lower frame structure, BiLSTM captures the overall trend reasonably well. However, it exhibits notable prediction deviations at sharp amplitude transitions, particularly within the range of 130 to 160. BiLSTM-KAN shows slight improvements over BiLSTM in the early and middle intervals, notably within the 50–130 range, but it still underperforms in accurately capturing certain local variations. Compared with the above two models, IBiLSTM demonstrates improved local fitting within the 100–150 interval but still falls short of optimal performance, especially in segments with abrupt changes where prediction deviations persist. In contrast, the proposed method exhibits the most consistent overall performance and demonstrates a superior ability to capture underlying temporal trends and handle non-stationary behaviors.

For X-axis of the LGB, several regions exhibit pronounced oscillatory behavior, particularly around sample point 50. While BiLSTM and BiLSTM-KAN capture the overall trend effectively within the initial segment (0–50), their performance degrades significantly in segments characterized by intense data fluctuations, such as between sample points 80 and 100. IBiLSTM demonstrates improved fitting in most regions but still struggles to capture localized spike patterns. In contrast, the proposed method achieves the most accurate fitting performance, particularly at critical sample points such as 80 and 140, where it effectively captures subtle amplitude variations.

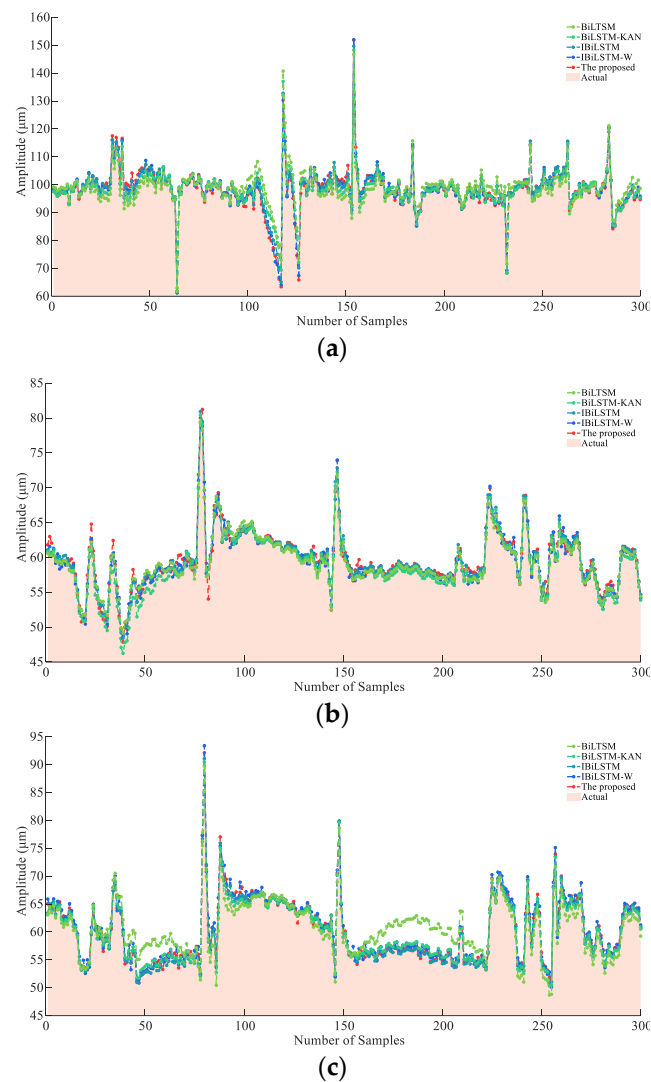


Figure 13. The forecasting outputs of all methods: (a) Y-axis of LFS; (b) X-axis of LGB; (c) Y-axis of LGB.

Regarding the Y-axis of the LGB, the signal exhibits a complex structural pattern with frequent intense fluctuations, especially sharp peak variations occurring between sample points 80 and 100. BiLSTM and BiLSTM-KAN perform reasonably well in relatively stable segments. However, they exhibit notable deviations in regions characterized by strong oscillations. Specifically, near sample points around 85 and 145, predicted peaks are significantly misaligned with the actual values, indicating limited capability in accurately capturing abrupt changes. IBiLSTM maintains relatively low prediction errors across most segments but lacks sufficient sensitivity to local transient features. In contrast, the proposed method demonstrates superior fitting accuracy across the entire time series and effectively aligns with the actual signal at multiple abrupt change points.

Figure 14 presents error comparisons for three different vibration trend prediction tasks. The Y-axis error trend of the lower frame structure reveals that BiLSTM exhibits substantial prediction errors between time steps 60 and 120, with the maximum error exceeding 12 μm . Although BiLSTM-KAN reduces error compared to BiLSTM, its peak error remains close to 10 μm . In contrast, IBiLSTM achieves a significantly lower overall error and exhibits improved stability, especially between time steps 180 and 300. The proposed method consistently maintains the lowest prediction error across all time steps, and its performance is especially notable in the interval between time steps 120 and 150, where it demonstrates superior predictive accuracy.

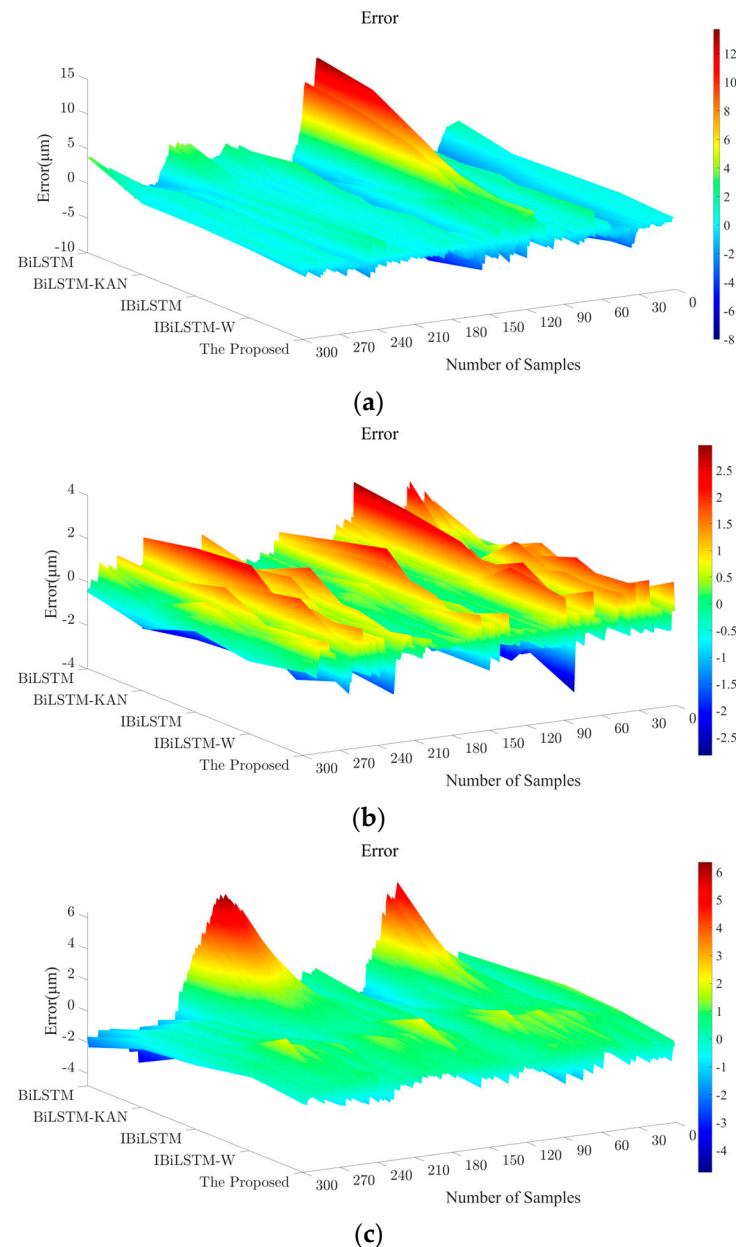


Figure 14. Error distribution charts for all methods: (a) Y-axis of LFS; (b) X-axis of LGB; (c) Y-axis of LGB.

The X-axis error trend of the LGB indicates that BiLSTM exhibits pronounced oscillations during the early sample indices (20–60), reflecting a certain degree of instability. In contrast, both IBiLSTM and the proposed method maintain consistently low error magnitudes throughout the entire time period, with absolute errors not exceeding $1.5 \mu\text{m}$. Among them, the proposed method achieves the smallest error overall and demonstrates minimal fluctuation, indicating superior prediction stability.

With respect to the Y-axis error trend of LGB, the results reveal that BiLSTM exhibits the highest error at sample indices 60 and 180, with the maximum error approaching $6 \mu\text{m}$. The error surface shows multiple oscillations, notably around sample indices 90 and 300. Although the error of BiLSTM-KAN is reduced compared to BiLSTM, distinct error peaks persist, and oscillations are not fully suppressed. IBiLSTM shows a further reduction in peak errors. However, localized fluctuations remain observable, such as around sample index 240. The proposed method demonstrates the best overall performance, maintaining errors consistently within $\pm 3 \mu\text{m}$ and exhibiting no prominent spikes throughout the entire time range. In summary, the proposed method achieves superior comprehensive performance

across all three experimental cases, consistently achieving the lowest prediction errors and the smoothest error surfaces, highlighting its strong accuracy and robustness.

5.7. Comparative Experiment

To comprehensively evaluate the prediction performance of the proposed method, several representative deep learning-based prediction methods were selected for comparison, including BiLSTM, BiGRU, TCN, Transformer, and BiLSTM-KAN. Specifically, the channel numbers of TCN were set to 32, 64, and 128. BiLSTM and BiGRU adopted the same hyperparameter settings as the proposed method, while Transformer was implemented using its standard architecture. The remaining experimental configurations were kept consistent with those in the ablation study. To ensure fair comparison, identical VMD and PSR were applied to all comparison methods. The prediction results and evaluation metrics of each method are presented in Table 4.

Table 4. Evaluation metric results of all methods in the comparative experiments.

Case	Methods	RMSE	MAE	MAPE	R ²
Y-axis of the LFS	BiLSTM	4.2447	3.2304	3.3897	0.7562
	TCN	1.2408	0.8338	0.8905	0.9792
	Transformer	1.8388	1.7180	1.7810	0.9542
	BiGRU	4.1673	3.0945	3.2517	0.7650
	BiLSTM-KAN	3.0805	1.9445	2.0745	0.8716
	The proposed	1.0902	0.8256	0.8509	0.9839
X-axis of the LGB	BiLSTM	0.8992	0.6692	1.1247	0.9523
	TCN	1.4954	1.3156	2.2318	0.8680
	Transformer	0.9422	0.7250	1.2414	0.9476
	BiGRU	0.7752	0.6031	1.0077	0.9645
	BiLSTM-KAN	0.7408	0.5536	0.9246	0.9676
	The proposed	0.6377	0.4632	0.7873	0.9760
Y-axis of the LGB	BiLSTM	2.8484	2.3307	4.0003	0.7580
	TCN	1.0177	0.8354	1.3691	0.9691
	Transformer	1.8513	1.7688	2.9832	0.8978
	BiGRU	2.0749	1.6694	2.8596	0.8716
	BiLSTM-KAN	0.9033	0.7348	1.2279	0.9757
	The proposed	0.4907	0.3990	0.6706	0.9928

As summarized in Table 4, the proposed method achieves the best prediction performance across all three VTP tasks with different datasets. It obtains the lowest RMSE, MAE, and MAPE values while achieving the highest R² values among all comparison methods. These results demonstrate that the proposed method effectively improves prediction accuracy and generalization capability, providing reliable technical support for the health management of PSUs.

5.8. Sensitivity Analysis of the K Value

To evaluate the robustness of the model prediction performance under different K values, sensitivity analysis was conducted with $K = 3, 4$, and 5 . The experiments were performed using the X-direction vibration data collected from the LGB, and the corresponding results are summarized in Table 5.

Table 5. Sensitivity analysis of K value on prediction performance.

K	RMSE	MAE	MAPE	R^2
3	0.6318	0.4789	0.8045	0.9764
4	0.6377	0.4632	0.7873	0.9760
5	0.6989	0.5073	0.8621	0.9712

As shown in Table 5, when $K = 5$, all evaluation metrics achieve the worst performance, indicating that excessive clustering may weaken the effectiveness of temporal characterization. The performance of $K = 3$ and $K = 4$ is highly similar, demonstrating the stability of the proposed framework within this range and supporting the selection of $K = 4$.

6. Conclusions

To improve prediction accuracy for the non-stationary and nonlinear vibration data of PSUs, a novel prediction method integrating TEGAN and IBiLSTM is proposed. TEGAN is first introduced to synthesize representative vibration data with enhanced diversity, improving the quality of training data. Subsequently, IBiLSTM, incorporating the M-KAN module, strengthens the representation of nonlinear dynamic characteristics. The experimental evaluation confirms the reliability and robustness of the developed method in PSU vibration trend prediction. The main conclusions are as follows:

- (1) The comparison among BiLSTM, BiLSTM-KAN, and IBiLSTM demonstrates the effectiveness of M-KAN. Compared with BiLSTM, IBiLSTM achieves lower RMSE, MAE, and MAPE values and higher R^2 values across all datasets, indicating that M-KAN effectively enhances the nonlinear fitting capability of the model and improves prediction accuracy. Moreover, IBiLSTM consistently outperforms BiLSTM-KAN, demonstrating that M-KAN is more effective than the conventional KAN in capturing complex higher-order nonlinear relationships.
- (2) Compared with IBiLSTM, the proposed method achieves lower prediction errors across all datasets. This improvement indicates that TEGAN can effectively enrich the diversity of training samples and provide a more comprehensive representation of the underlying data distribution, thereby facilitating more effective model training.
- (3) The proposed model achieves the best performance in all evaluation metrics across all datasets, with the highest R^2 values of 0.9839, 0.9760, and 0.9928, respectively, demonstrating its superior robustness and generalization capability.

Future work will focus on improving model interpretability and further enhancing predictive performance.

Author Contributions: Conceptualization, Z.Z. and L.Z.; methodology, Z.Z.; software, L.Z.; validation, L.D. and F.Z.; formal analysis, J.G.; investigation, L.D. and K.L.; resources, Z.Z.; writing—original draft preparation, Z.Z. and L.Z.; writing—review and editing, Z.Z.; visualization, F.Z. and J.G.; supervision, Z.Z. and J.X.; project administration, Z.Z. and L.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Science and Technology Project of State Grid Corporation of China “Technology Research on Enhancing the Regulation and Stability Support Capabilities and Coordinated Operation of Pumped Storage Units” (No. 4000-202499363A-3-1-DG).

Data Availability Statement: The data underlying this study are available in the article. Further details can be requested from the corresponding author.

Conflicts of Interest: Author Lei Deng, Fei Zhang was employed by the company State Grid Electric Power Engineering Research Institute Co., Ltd. The authors declare no conflicts of interest.

References

1. Zhang, Y.; Feng, Q.; Guan, X. Enablers of Carbon Neutrality in China's Energy Sector: A Review. *Sustainability* **2025**, *17*, 2657. [\[CrossRef\]](#)
2. Zou, Y.; Gao, D.; Yang, B. Array optimization for maximum power output of offshore wind-solar-wave hybrid systems via Aquila optimization algorithm. *Shandong Electr. Power* **2026**, *53*, 47–60.
3. Hou, J.; Qiu, L.; He, M.; Ji, K.; Fei, Y.; Fu, Y. Feasibility evaluation method of variable speed pumped storage transformation based on improved AHP-TOPSIS. *Shandong Electr. Power* **2026**, *53*, 110–120.
4. Pang, M.; Du, Y.; Pei, W.; Zhang, P.; Yang, J.; Zhang, L. Pumped hydro energy storage plants in China: Increasing demand and multidimensional impacts identification. *Energies* **2025**, *18*, 1801. [\[CrossRef\]](#)
5. Sun, W.; Zou, Y.; Wang, Y.; Xiao, B.; Zhang, H.; Xiao, Z. Fault diagnosis in hydroelectric units in small-sample state based on Wasserstein generative adversarial network. *Water* **2024**, *16*, 454. [\[CrossRef\]](#)
6. Zhang, Z.; Liu, X.; Wang, Y.; Li, E.; Zhang, Y. Stability Prediction Model of Transmission Tower Slope Based on ISCSO-SVM. *Electronics* **2024**, *14*, 126. [\[CrossRef\]](#)
7. Cui, L.; Hu, G.; Zhu, Y. Multi-strategy improved snow ablation optimizer: A case study of optimization of kernel extreme learning machine for flood prediction. *Artif. Intell. Rev.* **2025**, *58*, 181. [\[CrossRef\]](#)
8. Zhang, T.; Hou, Z.; Li, X.; Chen, Q.; Wang, Q.; Lüddeke, C.; Sun, W. A novel multivariable prognostic approach for PEMFC degradation and remaining useful life prediction using random forest and temporal convolutional network. *Appl. Energy* **2025**, *385*, 125569. [\[CrossRef\]](#)
9. Zhou, D.; Liu, Y.; Wang, X.; Wang, F.; Jia, Y. Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM-TCN-BiLSTM. *Energy* **2025**, *318*, 134847. [\[CrossRef\]](#)
10. Huang, Y.; Li, X.; Li, D.; Zhang, Z.; Yin, T.; Chen, H. Probabilistic prediction of wind farm power generation using non-crossing quantile regression. *Control Eng. Pract.* **2025**, *156*, 106226. [\[CrossRef\]](#)
11. Gong, J.; Qu, Z.; Zhu, Z.; Xu, H. Parallel TimesNet-BiLSTM model for ultra-short-term photovoltaic power forecasting using STL decomposition and auto-tuning. *Energy* **2025**, *320*, 135286. [\[CrossRef\]](#)
12. Yang, M.; Han, C.; Zhang, W.; Fang, G.; Jia, Y. A short-term power prediction method based on numerical weather prediction correction and the fusion of adaptive spatiotemporal graph feature information for wind farm cluster. *Expert Syst. Appl.* **2025**, *274*, 126979. [\[CrossRef\]](#)
13. Li, T.; Ai, L.; Yang, Q.; Zhang, X.; Li, H.; Lu, D.; Shen, H. Short-term wind power prediction based on multiscale numerical simulation coupled with deep learning. *Renew. Energy* **2025**, *246*, 122951. [\[CrossRef\]](#)
14. Gong, M.; Wang, Y.; Huang, J.; Cui, H.; Jing, S.; Zhang, F. A novel wind power prediction model based on PatchTST and temporal convolutional network. *Environ. Prog. Sustain. Energy* **2025**, *44*, e14584. [\[CrossRef\]](#)
15. Lu, Z.; Tao, R.; Xiao, R.; Li, P. Forecasting the hydropower unit vibration based on adaptive variational mode decomposition and neural network. *Appl. Soft Comput.* **2024**, *150*, 111040. [\[CrossRef\]](#)
16. Yan, S.; Chen, F.; Yang, J.; Zhao, Z. A multistep prediction model for the vibration trends of hydroelectric generator units based on variational mode decomposition and stochastic configuration networks. *Sensors* **2023**, *23*, 9762. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Hu, N.; Kong, L.; Zheng, H.; Zhou, X.; Wang, J.; Tao, J.; Lin, J. Trend prediction of vibration signals for pumped-storage units based on BA-VMD and LSTM. *Energies* **2024**, *17*, 5331. [\[CrossRef\]](#)
18. Fan, L.; Zhou, C.; Neri, F. Mt-gan: A multitask gan for severe convective weather nowcasting. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 1000605. [\[CrossRef\]](#)
19. Bu, X.; Wu, Q.; Zhou, B.; Li, C. Hybrid short-term load forecasting using CGAN with CNN and semi-supervised regression. *Appl. Energy* **2023**, *338*, 120920. [\[CrossRef\]](#)
20. Liu, Z.; Xuan, L.; Gong, D.; Xie, X.; Liang, Z.; Zhou, D. A WGAN-GP approach for data imputation in photovoltaic power prediction. *Energies* **2025**, *18*, 1042. [\[CrossRef\]](#)
21. Li, X.; Zeng, Y.; Qian, J.; Wang, Y.; Zhang, X.; Wang, F.; Li, D. A new multi-strategy optimization stability prediction framework based on DBO-Case study of pumped storage units. *J. Energy Storage* **2025**, *125*, 117005. [\[CrossRef\]](#)
22. Fu, W.; Chen, J.; Zheng, B.; Zhu, X.; Wen, B.; Chi, B. WGAN-GP enhanced ITOC-VMD-FBiLSTM for vibration trend prediction in pumped storage units. *Meas. Sci. Technol.* **2025**, *36*, 086206. [\[CrossRef\]](#)
23. Li, S.; Guo, L.; Zhu, J.; Chen, J.; Liu, M.; Cui, X.; Li, L. Medium-term offshore wind speed multi-step forecasting based on VMD and GRU-MATNet model. *Ocean Eng.* **2025**, *325*, 120737. [\[CrossRef\]](#)
24. Zou, H.; Yang, Q.; Chen, J.; Chai, Y. Short-term power load forecasting based on phase space reconstruction and EMD-ELM. *J. Electr. Eng. Technol.* **2023**, *18*, 3349–3359. [\[CrossRef\]](#)
25. Wang, S.; Guo, X.; Sun, T.; Xu, L.; Zhu, J.; Li, Z.; Zhang, J. Short-term photovoltaic power forecasting based on the VMD-IDBO-DHKELM model. *Energies* **2025**, *18*, 403. [\[CrossRef\]](#)
26. Wang, D.; Xu, M.; Guangming, Z.; Luo, F.; Gao, J.; Chen, Y. Enhancing wind power forecasting accuracy through LSTM with adaptive wind speed calibration (C-LSTM). *Sci. Rep.* **2025**, *15*, 5352. [\[CrossRef\]](#) [\[PubMed\]](#)

27. Zhang, J.; Zhao, Z.; Yan, J.; Cheng, P. Ultra-short-term wind power forecasting based on CGAN-CNN-LSTM model supported by lidar. *Sensors* **2023**, *23*, 4369. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, W.; Luo, Y.; Zhang, Y.; Srinivasan, D. SolarGAN: Multivariate solar data imputation using generative adversarial network. *IEEE Trans. Sustain. Energy* **2020**, *12*, 743–746. [[CrossRef](#)]
29. Lee, J.; Park, S.Y. WGAN-GP-Based Conditional GAN (cGAN) With Extreme Critic for Precipitation Downscaling in a Key Agricultural Region of the Northeastern US. *IEEE Access* **2025**, *13*, 46030–46041. [[CrossRef](#)]
30. He, L.; Zheng, K.; Ruan, H.; Yang, S.; Zhang, J.; Luo, C.; Cheng, J. A spatiotemporal mixed-enhanced generative adversarial network for radar-based precipitation nowcasting. *Comput. Geosci.* **2025**, *200*, 105919. [[CrossRef](#)]
31. Zhang, R.; Xu, Z.; Liu, S.; Fu, K.; Zhang, J. Prediction of Ultra-Short-Term Photovoltaic Power Using BiLSTM-Informer Based on Secondary Decomposition. *Energies* **2025**, *18*, 1485. [[CrossRef](#)]
32. Zhang, J.; Zhao, L. Efficient greenhouse gas prediction using IoT data streams and a CNN-BiLSTM-KAN model. *Alex. Eng. J.* **2025**, *123*, 261–270. [[CrossRef](#)]
33. Huang, S.; Zhao, Z.; Li, C.; Bai, L. TimeKAN: KAN-based frequency decomposition learning architecture for long-term time series forecasting. *arXiv* **2025**, arXiv:2502.06910.

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.