



ARTICLE

The Missing Data Recovery Method Based on Improved GAN

Su Zhang¹, Song Deng^{1,*} and Qingsheng Liu²

¹College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China

²State Grid Jiashan Power Supply Company, Jiaxing, 314100, China

*Corresponding Author: Song Deng. Email: dengsong@njupt.edu.cn

Received: 03 September 2025; Accepted: 24 November 2025; Published: 10 February 2026

ABSTRACT: Accurate and reliable power system data are fundamental for critical operations such as grid monitoring, fault diagnosis, and load forecasting, underpinned by increasing intelligentization and digitalization. However, data loss and anomalies frequently compromise data integrity in practical settings, significantly impacting system operational efficiency and security. Most existing data recovery methods require complete datasets for training, leading to substantial data and computational demands and limited generalization. To address these limitations, this study proposes a missing data imputation model based on an improved Generative Adversarial Network (BAC-GAN). Within the BAC-GAN framework, the generator utilizes Bidirectional Long Short-Term Memory (BiLSTM) networks and Multi-Head Attention mechanisms to capture temporal dependencies and complex relationships within power system data. The discriminator employs a Convolutional Neural Network (CNN) architecture to integrate local features with global structures, effectively mitigating the generation of implausible imputations. Experimental results on two public datasets demonstrate that the BAC-GAN model achieves superior data recovery accuracy compared to five state-of-the-art and classical benchmark methods, with an average improvement of 17.7% in reconstruction accuracy. The proposed method significantly enhances the accuracy of grid fault diagnosis and provides reliable data support for the stable operation of smart grids, showing great potential for practical applications in power systems.

KEYWORDS: Power system; data recovery; generative adversarial network; bidirectional long short-term memory network; multi-head attention mechanism; convolutional neural network

1 Introduction

Driven by increasing intelligence and digitization in power systems, modern grid data have grown significantly in complexity. The widespread deployment of sensors, smart meters, and IoT devices enables real-time collection and transmission of operational, power delivery, weather, and business data [1–3], supporting enhanced grid monitoring and management [4]. However, data integrity is frequently compromised by missing or incomplete values resulting from equipment failures and communication network issues [5]. This data loss severely impacts critical functions including real-time monitoring, fault diagnosis, and load forecasting [6]. A notable example is the 2018 California grid outage, where missing monitoring data from substations significantly hampered fault diagnosis and restoration efforts, ultimately affecting approximately 200,000 households and businesses [7]. Consequently, effective recovery of missing data is essential for ensuring power system reliability and security.

Methods for recovering missing data in power systems are broadly classified as statistical, machine learning (ML), and deep learning (DL) approaches. Statistical methods encompass techniques such as linear interpolation [8], multiple imputation [9], and regression imputation [10]. Time-series specific approaches



include the ARMA model (combining autoregressive (AR) and moving average (MA) components) for prediction [11], ARIMA for non-stationary time series recovery [12], and spatial interpolation using Delaunay triangulation [13]. While these methods primarily infer missing values based on underlying statistical characteristics, they often struggle to effectively capture the complex nonlinear dynamics and temporal dependencies inherent in power system data due to its intricate nature.

As recognition of power system data complexity grows, machine learning (ML) has been increasingly adopted for missing data imputation. Compared to statistical methods, ML excels at automatically learning intricate data patterns. For instance, K-Nearest Neighbors (KNN) has been applied to impute missing wind power data [14], hybrid approaches combine statistical and ML techniques like kernel canonical correlation analysis to enhance accuracy [15], and random forest-based multiple imputation algorithms have been developed [16]. However, ML methods typically depend on explicit feature engineering. Given the high dimensionality, nonlinearity, and complex temporal dependencies inherent in power system data, these approaches often struggle to capture deep underlying relationships, limiting their effectiveness in complex scenarios.

Deep learning (DL) advances beyond conventional ML by employing sophisticated neural architectures to autonomously extract latent patterns and correlations from raw data. Representative approaches include: bidirectional RNNs (M-RNN) for data stream interpolation [17]; GRU-D models incorporating masking and time intervals [18]; graph networks enhanced with sparse spatiotemporal attention for improved robustness to sparsity [19]; conditional score diffusion models for time series imputation [20]; Gaussian mixture models parameterized via maximum likelihood estimation [21]; and denoising autoencoders for missing data handling [22]. Despite these innovations, DL-based imputation methods remain challenged by complex missing patterns and limitations in model robustness.

To address these challenges, generative adversarial networks (GANs) have emerged as a promising approach for missing data imputation, leveraging their strong generative capabilities to produce realistic and plausible imputation results. Notably, GAIN (Generative Adversarial Imputation Network) [23] introduced a GAN-based framework that generates high-quality imputations through adversarial training. However, the GAIN model exhibits two key limitations for power system applications: (1) Its reliance on fully connected neural networks hinders effective capture of long-term temporal dependencies essential for complex power system time series characterized by strong periodicity and temporal correlations, and (2) It inadequately models local data correlations, limiting its ability to exploit local patterns and variation trends inherent in power system data.

To overcome these limitations, this paper proposes BAC-GAN (Bidirectional LSTM with Multi-Head Attention and CNN-based Discriminator Generative Adversarial Network), an enhanced GAN architecture for power system data imputation. The generator integrates bidirectional long short-term memory (BiLSTM) networks with multi-head attention (MA) mechanisms to capture complex temporal dependencies. Simultaneously, the discriminator employs a convolutional neural network (CNN) architecture to effectively model local features and global structures, mitigating the generation of implausible imputations.

The principal contributions are:

- We propose BAC-GAN, a novel generative adversarial network framework incorporating Bidirectional LSTM with Multi-Head Attention and a CNN-based Discriminator, specifically designed for high-accuracy imputation of missing data in power systems.
- The generator integrates Bidirectional LSTM (BiLSTM) networks with a Multi-Head Attention (MA) mechanism to comprehensively capture complex long-term temporal dependencies, while the discriminator employs a Convolutional Neural Network (CNN) architecture to effectively model local

features and integrate them with global structural context, thereby ensuring the plausibility of generated imputations.

- Comprehensive experimental evaluation on two public power system datasets demonstrates that BAC-GAN achieves significantly superior imputation accuracy compared to five state-of-the-art and classical benchmark methods.

The remainder of this paper is organized as follows. [Section 2](#) provides prerequisite knowledge. [Section 3](#) highlights the BAC-GAN model. [Section 4](#) presents experimental analysis. The conclusions are shown in [Section 5](#).

2 Prerequisite Knowledge

Consider a d -dimensional space. Assume that $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_d$ is a random variable with a distribution denoted as $P(X)$. X is referred to as the power data vector, representing observed data in the power system. Meanwhile, a random variable $M = (M_1, \dots, M_d)$ takes values in $\{0, 1\}^d$ and M is called the mask vector, which identifies which components in X are observed values ($M_i = 1$ represents observed values, and $M_i = 0$ represents missing values). For each $i \in \{1, \dots, d\}$, the original space $\mathcal{X}_i$ is extended to define a new space $\tilde{\mathcal{X}} = \mathcal{X}_i \cup \{*\}$, where $*$ is a special symbol used to represent unobserved values.

Based on this, a new random variable $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_d) \in \tilde{\mathcal{X}}$ is defined, with its computation rule as follows:

$$\tilde{X}_i = \begin{cases} X_i, & M_i = 1 \\ *, & M_i = 0 \end{cases} \quad (1)$$

The power system data matrix is $X \in R^{N \times C}$, and M is its corresponding mask matrix. The input missing data matrix is $\tilde{X} = X \odot M$, where $\tilde{X} = X \odot M$ denotes element-wise multiplication.

3 The BAC-GAN Model

Given the temporal and high-dimensional nature of power system data, this paper proposes a missing data recovery model named BAC-GAN (as shown in [Fig. 1](#)), which is based on an improved Generative Adversarial Network.

The generator adopts a structure that combines BiLSTM with a multi-head attention mechanism. BiLSTM integrates historical information and future trends through its bidirectional information flow mechanism, aiding in a comprehensive understanding of the dynamic characteristics of time series data. The multi-head attention mechanism captures global dependencies and key time points in the sequence through weighted allocation, further enhancing the model's ability to model temporal features and compensating for BiLSTM's limitations in capturing global contextual information.

The discriminator employs a CNN architecture. Power system data often exhibit high dimensionality, containing multivariate information such as voltage, current, frequency, and load. CNNs can capture local correlations and underlying structures within the data. When processing high-dimensional data, the CNN models interdependencies among variables through multi-channel input mechanisms, reducing the impact of redundant information. This provides more reliable feedback signals to the generator, prompting it to continuously optimize the imputation performance. Through adversarial training between the generator and the discriminator, high-precision recovery of missing data is ultimately achieved.

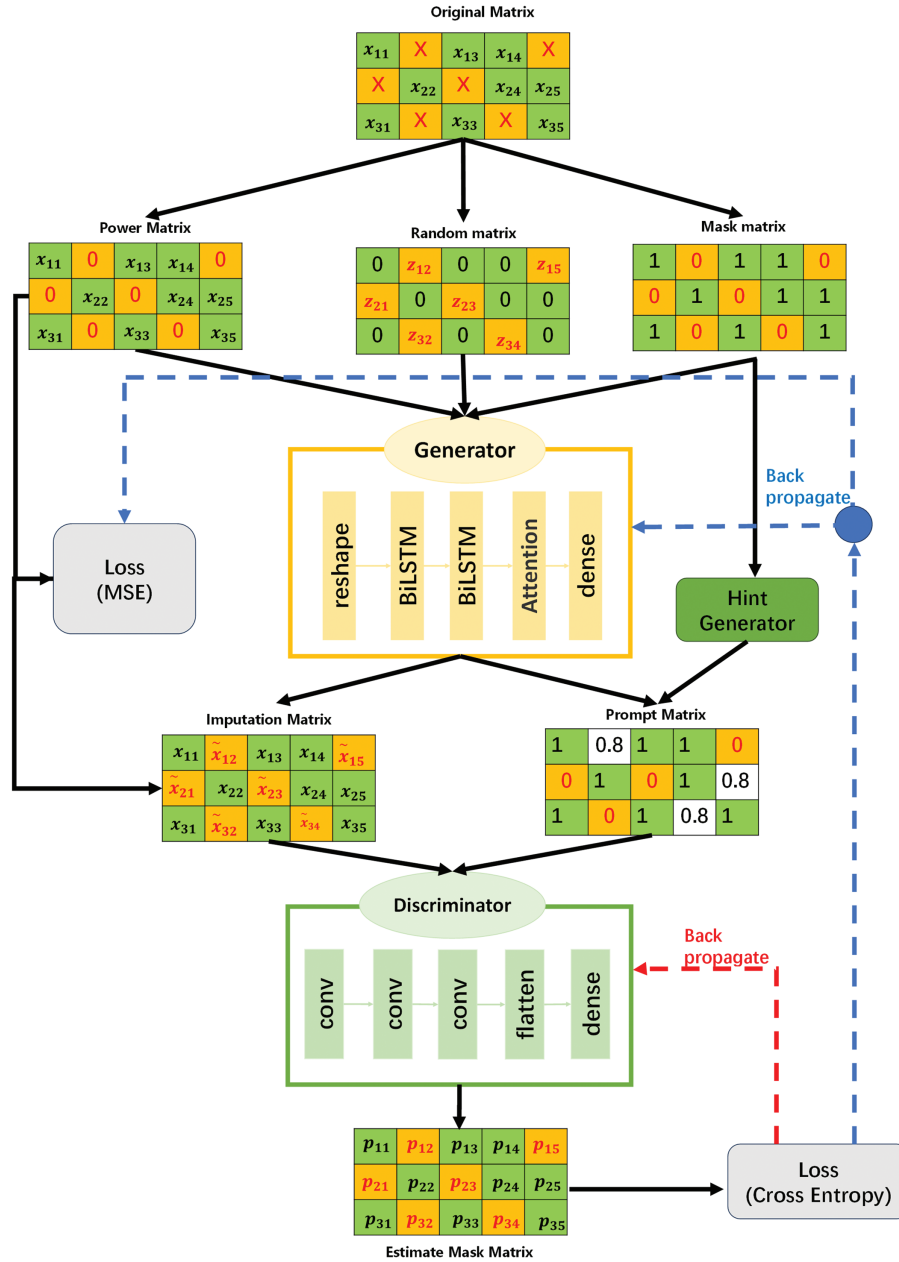


Figure 1: Recovery model

3.1 Generator

Conventional generators typically employ fully connected networks, which fail to effectively capture bidirectional temporal dependencies and global features, resulting in low-quality imputed data. To address this issue, this paper proposes a generator architecture based on a BiLSTM and multi-head attention mechanism. The core of our temporal modeling is a two-layer BiLSTM. This stacked design aims to learn hierarchical temporal representations: the first BiLSTM layer processes the input sequence and generates preliminary hidden states containing fundamental bidirectional patterns; the second BiLSTM layer builds upon this foundation to model more complex and abstract long-term dependencies at a higher level of feature abstraction. This hierarchical processing enables the network to capture intricate temporal dynamics

that might be missed by a single-layer BiLSTM. Subsequently, the output from the second BiLSTM layer is fed into a single-layer multi-head attention mechanism. We employ a single attention layer not for deep feature transformation, but to serve as a powerful global weighting and aggregation module. Its function is to reweight the temporal features refined by the BiLSTM, identifying and emphasizing the most critical time steps across the entire sequence from multiple representation subspaces. This single-layer structure effectively captures global contextual relationships while avoiding the computational overhead and overfitting risks associated with multi-layer attention stacking. The detailed structure is illustrated in Fig. 2.

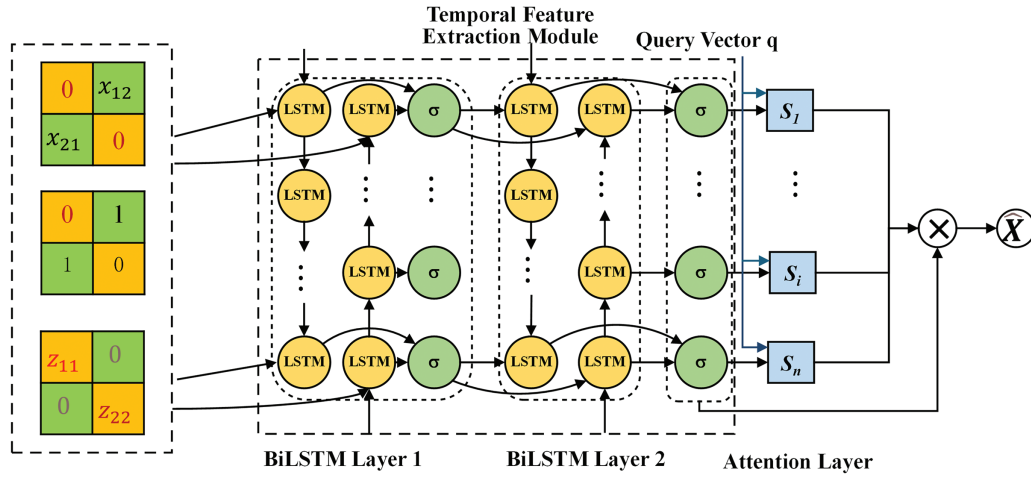


Figure 2: Improvement of generator structure

Generator G receives the power data with missing values X , the mask matrix M , and the noise-processed Z after masking as inputs. Through adjustments by the mask matrix, randomness is introduced only at the missing value positions, thereby aligning with the target distribution dimensions, i.e., the input data $\tilde{X} = G(X, M, (1 - M) \odot Z)$.

BiLSTM is an extension of LSTM, consisting of two LSTMs: one processes information in the forward direction, while the other handles information in the backward direction. BiLSTM can capture bidirectional temporal information in sequences, making it suitable for tasks that require consideration of both past and future information.

Based on the temporal features extracted by BiLSTM, the multi-head attention mechanism further enhances the feature representation capability. The multi-head attention mechanism employs multiple attention heads to capture dependencies between different time steps of the input sequence from distinct subspaces. For the query (Query), key (Key), and value (Value) of the input sequence, denoted as Q , K and V , respectively, the computation for a single head of attention is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where $\frac{1}{\sqrt{d_k}}$ is the scaling factor, $\sqrt{d_k}$ is the dimension of the key, and the softmax function ensures the normalization of attention weights. The multi-head attention mechanism computes attention in parallel through multiple independent attention heads and concatenates the results of all heads. By leveraging this mechanism, the model can capture global dependencies in temporal data from diverse perspectives, further refining feature representation.

3.2 Discriminator

In the model proposed in this paper, a hint matrix (Hint Mechanism) and a CNN-based network architecture are introduced to enhance the discriminator's capability, as illustrated in Fig. 3.

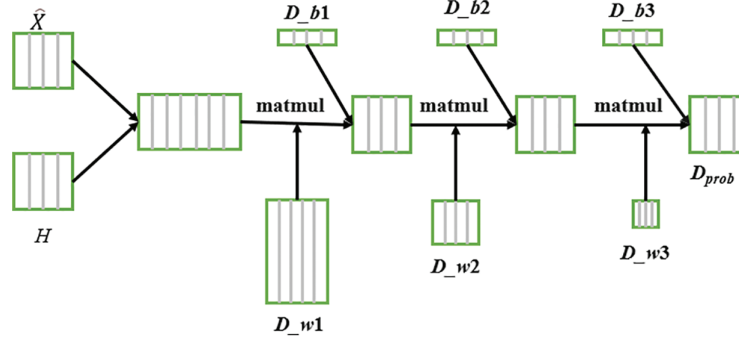


Figure 3: Improvement of discriminator structure

The hint matrix provides information about the validity of the original data, helping the discriminator more accurately distinguish between observed and imputed values. Particularly in cases of severe data missingness, the generator may produce multiple plausible imputation results, making it challenging for the discriminator to make effective judgments. The calculation formula for the hint matrix is:

$$H = B \odot M + 0.8 \odot (1 - B) \quad (3)$$

where B is a defined random variable and M is the mask matrix.

The input to the discriminator consists of the generated sample $\hat{X}$ and the hint matrix H . To better capture the features in the input data $\hat{X}$ and enhance the discriminator's ability to distinguish between observed and imputed values, a CNN architecture is adopted for the discriminator. In this model, the discriminator is structured with three convolutional layers, which are systematically designed to extract hierarchical features from the input data through sequential convolutional operations. The initial convolutional layer functions as a low-level feature detector, capturing fundamental patterns and local structures within the input. As the network depth increases, the subsequent layers progressively learn more abstract and complex representations: the second layer integrates local features to identify intermediate patterns, while the third layer synthesizes global contextual information, enabling the discriminator to effectively distinguish subtle inconsistencies between observed and imputed values. This hierarchical feature extraction mechanism significantly enhances the discriminator's capacity to evaluate data authenticity across multiple scales.

3.3 Loss Function

3.3.1 Discriminator Loss Function

The discriminator D aims to distinguish between observed values and generated values. The loss function of the discriminator is defined as cross-entropy:

$$L_D = -\frac{1}{n} \sum_{i=1}^n [M_i \log(D(x_i)) + (1 - M_i) \log(1 - D(x_i))] \quad (4)$$

where $D(x_i)$ represents the probability predicted by the discriminator that the i -th component is real data, M_i is the i -th value in the mask matrix (with missing positions marked as 0), and n is the total number of data samples.

3.3.2 Generator Loss Function

The objective of the generator is to impute missing data by generating values that are as realistic as possible, such that the discriminator cannot easily distinguish between observed and imputed values. Furthermore, to ensure the accuracy of the imputed data at the positions of observed values, a reconstruction error term is incorporated.

(1) Adversarial Loss. The generator aims to mislead the discriminator into classifying the imputed data as observed data. The adversarial loss is applied exclusively to the positions of missing values:

$$L_G^{adv} = -\frac{1}{n} \sum_n^{i=1} [(1 - M_i) \log(D(x_i))] \quad (5)$$

where $1 - M_i$ ensures that the loss is computed only for missing values, and $D(x_i)$ represents the probability predicted by the discriminator that the i -th component is real data.

(2) Reconstruction Loss. To ensure that the imputed values produced by the generator align closely with the ground truth data at observed positions, the reconstruction loss is formulated as follows:

$$L_M = \frac{\frac{1}{n} \sum_n^{i=1} [(1 - M_i)(x_i - G(x_i))^2]}{\frac{1}{n} \sum_n^{i=1} (1 - M_i)} \quad (6)$$

where x_i represents the ground truth data values, and $G(x_i)$ denotes the imputed values generated by the generator G . The total loss function of the generator is defined as a weighted sum of the adversarial loss L_G^{adv} and the reconstruction loss L_M :

$$L_G = L_G^{adv} + \alpha L_M \quad (7)$$

where α is a hyperparameter that balances the two loss functions.

4 Experimental Analysis

This project runs on the Ubuntu 20.04 operating system, equipped with an Intel® Xeon® Platinum 8474C processor (16-core vCPU) and an NVIDIA GeForce RTX 4090D GPU (24 GB VRAM). The deep learning framework used is PyTorch 2.0.0, with Python version 3.8 and CUDA version 11.8. Our experimental datasets and source code will be available at <https://github.com/zhangsul234/BAC-GAN> (accessed on 23 November 2025). In the subsequent experiments, this study aims to address the following research questions:

RQ1: How does the performance of the proposed model compare to other methods as the proportion of missing data increases?

RQ2: How does the proposed model perform across different datasets?

4.1 Datasets

Datasets are shown in Table 1.

(1) Residential Load Dataset (RLD): The dataset comprises residential household electricity consumption data, with daily load profiles sampled at 15-min intervals, resulting in a feature dimension of 96. This dataset provides detailed documentation of typical residential electricity usage patterns and holds significant value for analyzing and predicting household electricity consumption behaviors. The time series plot is shown in Fig. 4.

(2) Smart Meter Dataset (SMD): The SMD aggregates hourly energy consumption and multi-dimensional electrical parameters—including voltage, current, and power factor—from smart meters

deployed across urban and suburban regions in a regional power distribution network. Covering a full calendar year of 2020 (8760 hourly records), the dataset integrates metadata such as timestamps and device status, enabling comprehensive analysis of load profiles, anomaly detection, and missing data imputation under realistic grid operation scenarios. The time series plot is shown in Fig. 5.

Table 1: Experimental dataset

Acronym	Dataset name	Instance	Feature
RLD	Residential load dataset	1580	96
SMD	Smart meter data	8760	134

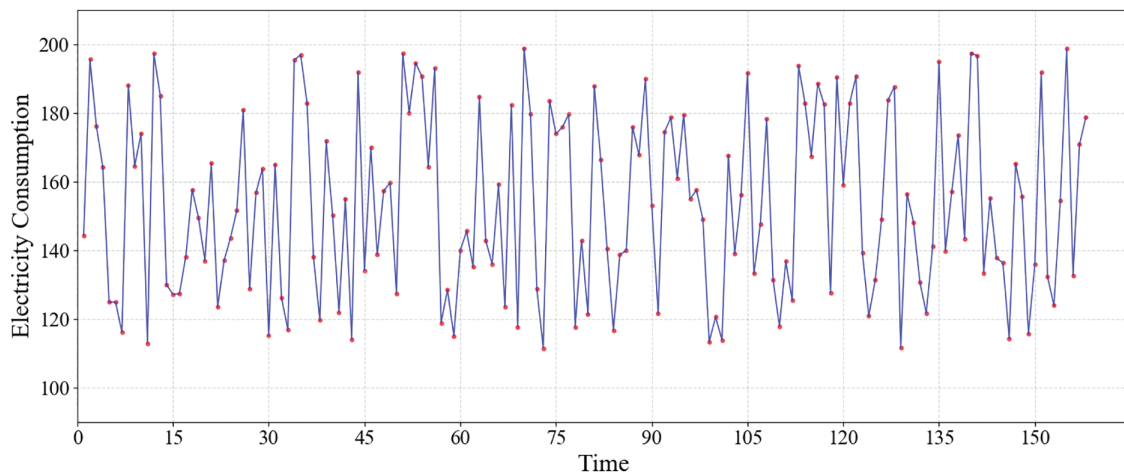


Figure 4: Time series plot of RLD (15-min intervals)

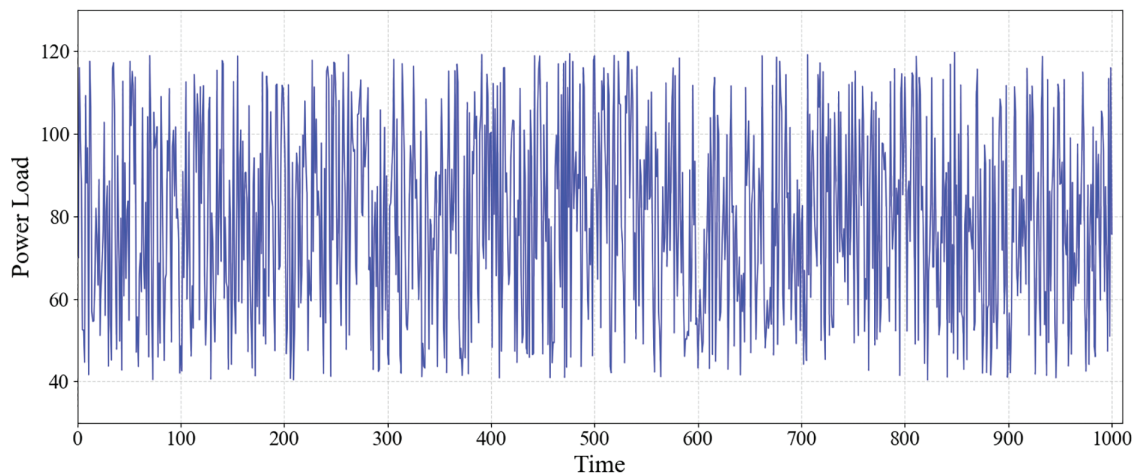


Figure 5: Time series plot of SMD (1-h intervals)

4.2 Compared Algorithms

1. KNN [14]: This method identifies the K nearest observations to a sample with missing data using Euclidean distance and performs a distance-weighted average to impute the missing values.

2. VAE [24]: By maximizing the likelihood in the latent space and minimizing the discrepancy between the generated data and the original data, this approach recovers missing values.

3. GAIN [23]: Built upon the Generative Adversarial Network (GAN) framework, this model recovers missing data by performing conditional modeling of the missing locations and adversarial training.

4. M-RNN [19]: This method constructs a generative module based on Recurrent Neural Networks (RNNs) to model temporal dependencies by integrating historical and contextual information. It achieves data recovery through recursive prediction.

5. MIVAE [25]: By jointly optimizing the posterior likelihood of latent variables and the reconstruction error, this approach generates multiple probabilistic imputations within a variational framework, effectively modeling the uncertainty of missing values and enabling high-accuracy data recovery.

4.3 Evaluation Metric

Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2) are commonly used evaluation metrics for assessing the accuracy of predictive models or algorithms. The formulas for each metric are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (8)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (9)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (11)$$

where y_i is the actual value at the missing position, $\bar{y}$ is the mean value, $\hat{y}_i$ is the imputed value at the missing position obtained by the imputation algorithm, and N is the total number of missing values.

4.4 Experimental Analysis

Sixty percent of the dataset was used as the training set, twenty percent was allocated to the validation set, and the remaining twenty percent was used for the test set. The experimental data were subjected to random missingness at missing rates ranging from 10% to 90%.

The selection of the optimal hyperparameters, specifically Hint = 0.9 and Alpha = 400, was driven by a systematic grid search on the validation set with a 50% missing rate, with the results detailed in Table 2.

Table 2: Hyperparameter selection experiment

(Alpha, Hint)	MSE	(Alpha, Hint)	MSE	(Alpha, Hint)	MSE	(Alpha, Hint)	MSE
(100, 0.6)	0.0985	(200, 0.5)	0.0875	(400, 0.5)	0.0458	(800, 0.5)	0.0652
(100, 0.6)	0.0974	(200, 0.6)	0.0812	(400, 0.6)	0.0412	(800, 0.6)	0.0612
(100, 0.7)	0.0875	(200, 0.7)	0.0789	(400, 0.7)	0.0398	(800, 0.7)	0.0545

(Continued)

Table 2 (continued)

(Alpha, Hint)	MSE	(Alpha, Hint)	MSE	(Alpha, Hint)	MSE	(Alpha, Hint)	MSE
(100, 0.8)	0.0814	(200, 0.8)	0.0745	(400, 0.8)	0.0354	(800, 0.8)	0.0446
(100, 0.9)	0.0755	(200, 0.9)	0.0678	(400, 0.9)	0.0215	(800, 0.9)	0.0401

Our analysis of the experimental data reveals a clear trend of consistent performance improvement as the Hint value increased from 0.5 to 0.9 across all Alpha groups. For instance, within the critical Alpha = 400 group, the MSE decreased monotonically from 0.0458 at Hint = 0.5 to 0.0215 at Hint = 0.9, indicating that providing more precise location information to the discriminator is crucial for guiding the generator to produce more accurate imputations in high-missing-rate scenarios. The optimality of Alpha = 400 emerges from its synergistic interaction with Hint = 0.9, as evidenced by the performance comparison with other combinations. While the combination of Alpha = 200 with Hint = 0.9 yields a relatively high MSE of 0.0678, and Alpha = 800 with Hint = 0.9 achieves a better but still suboptimal MSE of 0.0401, the pair of Alpha = 400 with Hint = 0.9 achieves the lowest MSE of 0.0215 across the entire parameter space tested. This demonstrates that Alpha = 400 provides the ideal weight for the reconstruction loss, ensuring the generator not only fools the discriminator but also accurately reconstructs the known values—a capability fully leveraged when the discriminator is well-informed through Hint = 0.9. Therefore, this specific combination was selected as it represents the configuration where the generator receives the most effective guidance from the discriminator while being simultaneously constrained to achieve the highest factual accuracy. The parameters of the network model are configured as shown in the [Table 3](#).

Table 3: Model parameters

Parameters	Value
Hint	0.9
Missing_rate	0.1–0.9
Alpha	400
num_epochs	1000
Batch_size	64

The imputation results of the RLD dataset across a range of missing rates are shown in [Tables 4–8](#). The time series after imputation for a missing rate of 0.5 is shown in [Fig. 6](#).

Table 4: RMSE of RLD under different missing rate level

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.1163	0.1321	0.1327	0.1414	0.1558	0.1611	0.1648	0.1874	0.2185
GAIN	0.0825	0.1080	0.1215	0.1750	0.1898	0.2453	0.3198	0.3370	0.3726
KNN	0.1586	0.1645	0.1653	0.1696	0.1712	0.1734	0.2105	0.2235	0.2339
VAE	0.2041	0.2103	0.2409	0.2740	0.2753	0.2781	0.2790	0.2804	0.2822

(Continued)

Table 4 (continued)

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
MRNN	0.1208	0.1800	0.2275	0.2305	0.2430	0.2906	0.2974	0.2998	0.3345
MIVAE	0.1341	0.1715	0.2074	0.2171	0.2201	0.2597	0.2607	0.2654	0.2697

Table 5: MSE of RLD under different missing rate level

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.0135	0.0174	0.0176	0.0200	0.0243	0.0260	0.0272	0.0351	0.0477
GAIN	0.0068	0.0117	0.0148	0.0306	0.0360	0.0602	0.1023	0.1136	0.1388
VAE	0.0252	0.0270	0.0273	0.0288	0.0293	0.0301	0.0443	0.0499	0.0547
KNN	0.0416	0.0442	0.0580	0.0751	0.0758	0.0773	0.0779	0.0787	0.0796
MRNN	0.0146	0.0324	0.0517	0.0531	0.0590	0.0844	0.0885	0.0899	0.1119
MIVAE	0.0180	0.0294	0.0430	0.0471	0.0484	0.0674	0.0680	0.0704	0.0727

Table 6: MAE of RLD under different missing rate level

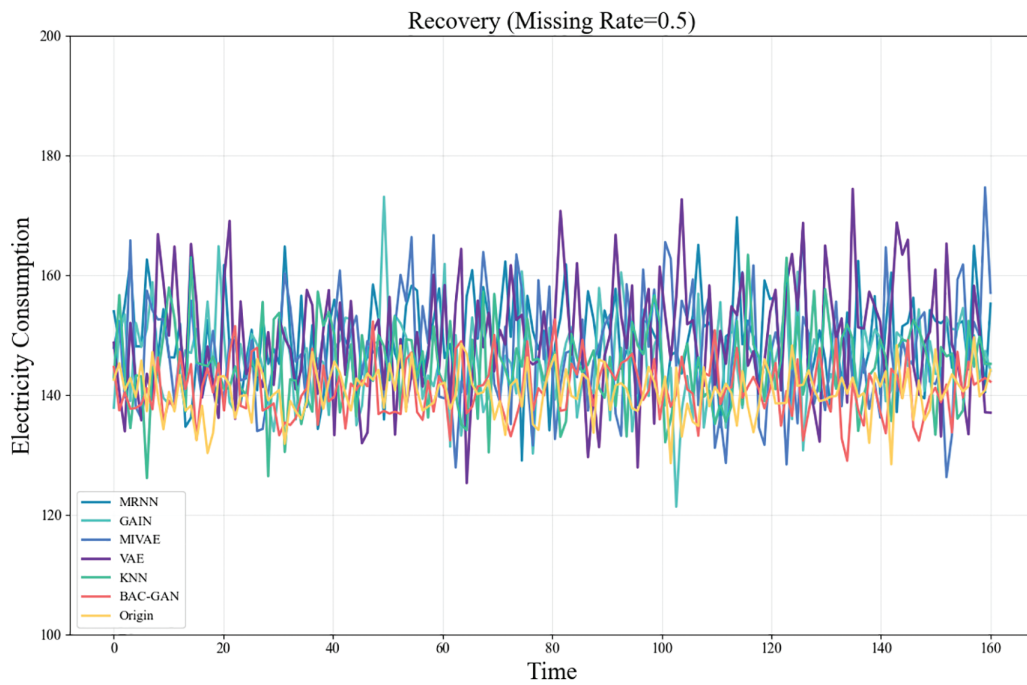
Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.0998	0.1144	0.1182	0.1195	0.1254	0.1382	0.1511	0.1654	0.1796
GAIN	0.0615	0.0825	0.0886	0.1328	0.1491	0.1847	0.2882	0.2936	0.3305
VAE	0.1258	0.1267	0.1279	0.1314	0.1406	0.1485	0.1762	0.1780	0.1852
KNN	0.1760	0.1799	0.2118	0.2473	0.2489	0.2512	0.2598	0.2642	0.2565
MRNN	0.1123	0.1566	0.1715	0.1804	0.1896	0.2685	0.2703	0.2775	0.2955
MIVAE	0.1194	0.1367	0.1304	0.1417	0.1569	0.1614	0.1789	0.1814	0.1914

Table 7: R^2 of RLD under different missing rate level

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.9865	0.9825	0.9824	0.9800	0.9757	0.9740	0.9728	0.9649	0.9523
GAIN	0.9732	0.9683	0.9552	0.9494	0.9440	0.9398	0.9377	0.9201	0.8907
MIVAE	0.9487	0.9354	0.9217	0.9201	0.9199	0.9155	0.9109	0.9001	0.8999
MRNN	0.9360	0.8527	0.8065	0.7770	0.7896	0.7684	0.7578	0.8356	0.7478
VAE	0.7858	0.7829	0.7858	0.7189	0.7210	0.6943	0.6899	0.6418	0.6039
KNN	0.6583	0.6558	0.5742	0.6049	0.5842	0.5727	0.5522	0.5314	0.5204

Table 8: Recovery time of RLD under different missing rate level (Unit: ms)

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	69	80	95	143	152	178	188	201	260
GAIN	78	85	99	103	110	138	165	198	241
VAE	84	98	107	118	136	187	196	219	264
KNN	120	134	174	189	201	213	239	287	334
MRNN	94	102	103	178	198	215	245	258	298
MIVAE	97	102	110	137	146	199	235	244	294

**Figure 6:** Time series plot of RLD after imputation (Missing rate = 0.5)

At low missing rates (0.1–0.3), the models were evaluated using metrics such as MSE, RMSE, and MAE. The results indicate that the proposed model performed second only to GAIN. At medium to high missing rates (0.4–0.9), the proposed model achieved the best performance across these metrics, significantly outperforming other methods, demonstrating its strong robustness in handling high missing scenarios. Furthermore, the model exhibited favorable R^2 values, indicating its ability to effectively capture the relationship between the imputed data and the ground truth. On the other hand, by comparing recovery times, the proposed model also showed favorable recovery times, suggesting that it not only excels in terms of accuracy but also efficiently handles the computational workload.

The imputation results of the RLD dataset across a range of missing rates are shown in Tables 9–13. The time series after imputation for a missing rate of 0.5 is shown in Fig. 7

Table 9: RMSE of SMD under different missing rate level

Method									
Missing rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.1363	0.1371	0.1378	0.1390	0.1436	0.1438	0.1542	0.1597	0.1627
GAIN	0.1064	0.1109	0.1292	0.1783	0.1820	0.1850	0.1968	0.2081	0.2087
VAE	0.2210	0.2211	0.2297	0.2326	0.2349	0.2370	0.2393	0.2421	0.2471
KNN	0.1733	0.2182	0.2681	0.2689	0.2736	0.2769	0.2820	0.2949	0.2968
MRNN	0.1295	0.1945	0.2195	0.2752	0.2870	0.2901	0.2993	0.3017	0.3130
MIVAE	0.1414	0.1517	0.1774	0.1801	0.1921	0.2017	0.2114	0.2152	0.2314

Table 10: MSE of SMD under different missing rate level

Method									
Missing rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.0186	0.0188	0.0190	0.0193	0.0206	0.0207	0.0238	0.0255	0.0265
GAIN	0.0113	0.0123	0.0167	0.0318	0.0332	0.0342	0.0387	0.0433	0.0436
VAE	0.0480	0.0489	0.0528	0.0541	0.0552	0.0561	0.0573	0.0586	0.0611
KNN	0.0300	0.0476	0.0719	0.0723	0.0749	0.0767	0.0795	0.0869	0.0881
MRNN	0.0168	0.0378	0.0482	0.0752	0.0823	0.0846	0.0896	0.0910	0.0979
MIVAE	0.0200	0.0230	0.0315	0.0324	0.0369	0.0404	0.0447	0.0463	0.0535

Table 11: MAE of SMD under different missing rate level

Method									
Missing rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.1014	0.1071	0.1160	0.1173	0.1207	0.1278	0.1348	0.1379	0.1390
GAIN	0.0780	0.0810	0.0989	0.1352	0.1393	0.1418	0.1567	0.1556	0.1563
VAE	0.1860	0.1863	0.2032	0.2064	0.2089	0.2112	0.2138	0.2168	0.2221
KNN	0.1374	0.1731	0.2133	0.2138	0.2175	0.2200	0.2238	0.2336	0.2353
MRNN	0.1001	0.1602	0.1835	0.2339	0.2471	0.2666	0.2881	0.2922	0.3001
MIVAE	0.1101	0.1317	0.1379	0.1401	0.1497	0.1501	0.1697	0.1897	0.2014

Table 12: R^2 of SMD under different missing rate level

Method									
Missing rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	0.9503	0.9400	0.9401	0.9317	0.9256	0.9220	0.9201	0.9129	0.9089
GAIN	0.9274	0.9117	0.9008	0.9017	0.8994	0.8954	0.8845	0.8514	0.8102
VAE	0.7895	0.7785	0.7565	0.7486	0.7341	0.7214	0.7100	0.6458	0.6130

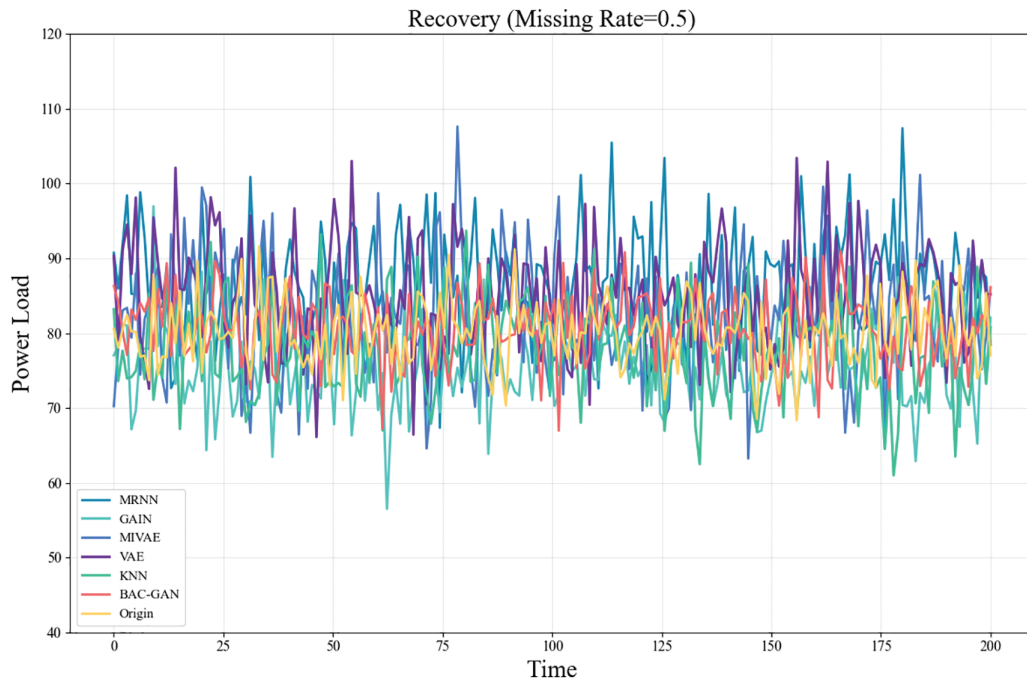
(Continued)

Table 12 (continued)

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
KNN	0.6624	0.5618	0.5012	0.4512	0.4529	0.4408	0.3917	0.3879	0.3540
MRNN	0.8971	0.8814	0.8543	0.8241	0.8214	0.8107	0.8047	0.7778	0.7521
MIVAE	0.9101	0.9087	0.9065	0.8958	0.8914	0.8814	0.8797	0.8701	0.8696

Table 13: Recovery time of SMD under different missing rate level (Unit: ms)

Missing rate	Method								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BAC-GAN	402	554	689	898	1201	1388	1590	1958	2164
GAIN	356	417	598	752	1024	1411	1741	2378	2575
VAE	312	401	498	665	997	1347	1607	2187	2413
KNN	845	1054	1254	1423	1665	1998	2385	2798	3041
MRNN	498	654	758	994	1474	1554	1745	2063	2369
MIVAE	641	875	956	1001	1417	1697	1956	2407	2598

**Figure 7:** Time series plot of SMD after imputation (Missing rate = 0.5)

The proposed model demonstrates outstanding performance across all missing-rate scenarios, with its superiority being particularly pronounced at medium to high missing rates (0.4–0.9). Under these conditions, the model significantly outperforms comparative methods in terms of imputation accuracy,

robustness, and ability to capture complex temporal dependencies. Meanwhile, BAC-GAN also exhibits excellent recovery time efficiency, maintaining fast response capabilities even under high missing-rate conditions, highlighting its strong applicability in practical situations where high proportions of data loss commonly occur.

Based on the four evaluation metrics mentioned above, the proposed model demonstrates excellent recovery performance under most missing-rate scenarios. The key to its success lies in the fact that BAC-GAN integrates BiLSTM and a multi-head attention mechanism, which effectively captures complex dependencies in time-series data and enhances the learning of global relationships. Meanwhile, the incorporation of CNN improves the model's ability to handle high-dimensional data and strengthens local feature extraction, thereby increasing the accuracy and robustness of data recovery.

4.5 Ablation Study

To validate the effectiveness of individual components in our proposed BAC-GAN framework, we conducted comprehensive ablation studies under various missing rate scenarios (0.2, 0.4, 0.6, 0.8). As demonstrated in [Tables 14](#) and [15](#), the complete BAC-GAN model consistently achieves superior or competitive performance across all missing rates, particularly excelling at higher missing scenarios (0.6–0.8).

Table 14: Ablation study results on RLD dataset under different missing rates

	0.2		0.4		0.6		0.8	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
BAC-GAN	0.0174	0.1144	0.0200	0.1195	0.0260	0.1382	0.0351	0.1654
B-GAN	0.0180	0.1230	0.0264	0.1264	0.0310	0.1614	0.0794	0.2497
BC-GAN	0.0178	0.1149	0.0213	0.1201	0.0278	0.1397	0.0456	0.1756
BA-GAN	0.0183	0.1201	0.0245	0.1214	0.0281	0.1412	0.0642	0.1879
A-GAN	0.0190	0.1268	0.0287	0.1307	0.0541	0.1798	0.1014	0.2789
AC-GAN	0.0109	0.0798	0.0254	0.1256	0.0297	0.1498	0.0754	0.2014
C-GAN	0.0186	0.1245	0.0275	0.1301	0.0419	0.1697	0.0851	0.2596
GAN	0.0117	0.0825	0.0306	0.1328	0.0602	0.1847	0.1136	0.2936

Table 15: Ablation study results on SMD dataset under different missing rates

Method	0.2		0.4		0.6		0.8	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
BAC-GAN	0.0188	0.1071	0.0193	0.1173	0.0207	0.1278	0.0255	0.1379
B-GAN	0.0193	0.1079	0.0259	0.1248	0.0265	0.1345	0.0312	0.1511
BC-GAN	0.0134	0.0910	0.0199	0.1189	0.0219	0.1298	0.0261	0.1412
BA-GAN	0.0156	0.0970	0.0208	0.1197	0.0232	0.1307	0.0278	0.1489
A-GAN	0.0201	0.1175	0.0304	0.1304	0.0327	0.1401	0.0412	0.1543
AC-GAN	0.0178	0.1041	0.0216	0.1211	0.0242	0.1316	0.0287	0.1501
C-GAN	0.0197	0.1085	0.0274	0.1297	0.0298	0.1365	0.0387	0.1525
GAN	0.0123	0.0810	0.0318	0.1352	0.0342	0.1418	0.0433	0.1556

The performance degradation observed in partial component models reveals the complementary nature of the designed modules in this work. The BA-GAN variant demonstrates consistently intermediate performance, indicating that the attention mechanism contributes significantly to feature extraction but requires the constraint module to achieve optimal results. The A-GAN and C-GAN models exhibit relatively weaker performance, confirming that individual components alone are insufficient to handle complex missing data patterns.

Furthermore, the progressive performance decline from BC-GAN to BAC-GAN highlights the challenging nature of high missing-rate scenarios and validates the robustness of the proposed model. The complete BAC-GAN model maintains the most stable performance trajectory, exhibiting the smallest performance degradation as the missing rate increases from 0.2 to 0.8. These findings collectively demonstrate that the synergistic integration of all components in BAC-GAN is crucial for achieving robust performance across diverse missing data scenarios, with each element playing a distinct yet interdependent role in the overall architecture.

5 Conclusion

With the continuous development of power systems, the complexity and diversity of power data have been steadily increasing. Due to factors such as sensor failures and communication interruptions, missing or anomalous data frequently occur in power systems, posing significant challenges to system monitoring and management. To address this issue, this paper proposes a data recovery method based on an improved Generative Adversarial Network (BAC-GAN). The generator of this method incorporates bidirectional LSTM and multi-head attention mechanisms to capture complex dependencies in time-series data, while the discriminator employs CNN to integrate local features with global structures, ensuring the rationality of data recovery. Without relying on complete datasets, this method achieves high-precision data recovery. Experiments conducted on three publicly available power system datasets demonstrate that the BAC-GAN model exhibits significant advantages in recovery accuracy compared to five state-of-the-art and classical data recovery methods, providing an effective solution for missing data recovery in power systems.

The study of nonlinear generation mechanisms for multidimensional data represents a critical direction for the future development of smart grids. Current generation models based on linear assumptions struggle to accurately capture the complex dynamic coupling characteristics among source-grid-load-storage components. This limitation is particularly evident in high-penetration renewable energy integration scenarios, where traditional methods fail to adequately characterize the spatiotemporal correlations of wind and solar power outputs and load response characteristics. Future research should focus on developing interpretable generation architectures that integrate domain knowledge, ensuring that the data generation process aligns with power system operational principles.

As the scale and complexity of power system data continue to grow, missing data recovery methods based on generative adversarial networks still hold broad prospects for development. On one hand, more efficient network architectures and training strategies can be explored to further enhance the recovery accuracy and generalization capability of the model. On the other hand, the integration of technologies such as edge computing and federated learning may enable efficient data recovery in distributed environments.

Furthermore, although this study does not delve deeply into the application of Generative Adversarial Networks in multidimensional data fusion within power systems, the potential of GANs in handling multidimensional power data deserves further exploration. GANs can integrate different types of data and generate high-quality recovery results, which would significantly improve the comprehensiveness and accuracy of data recovery in power systems. While this paper primarily focuses on time-series data recovery, future work

could extend the application of GANs to multidimensional data fusion. By integrating various types of data in power systems, the performance and robustness of missing data recovery could be further enhanced.

Acknowledgement: The authors acknowledge the support provided by the Advanced Ocean Institute of Southeast University, Nantong.

Funding Statement: This work was supported by the National Natural Science Foundation of China (No. 51977113) and the Science and Technology Project of State Grid Zhejiang Electric Power Co., Ltd. (No. 5211JX240001).

Author Contributions: Study conception and design: Su Zhang, Song Deng and Qingsheng Liu; data collection: Su Zhang, Song Deng and Qingsheng Liu; analysis and interpretation of results: Su Zhang, Song Deng and Qingsheng Liu; draft manuscript preparation: Su Zhang, Song Deng and Qingsheng Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Cheng X, Sun Y, Jara A, Song H, Tian Y. Big data and knowledge extraction for cyber-physical systems. *Int J Distrib Sens Netw*. 2015;11(9):231527. doi:10.32604/cmc.2022.026751.
2. Baccarelli E, Cordeschi N, Mei A, Panella M, Shojafar M, Stefa J. Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study. *IEEE Netw*. 2016;30(2):54–61. doi:10.1109/MNET.2016.7437025.
3. Wang K, Li H, Feng Y, Tian G. Big data analytics for system stability evaluation strategy in the energy Internet. *IEEE Trans Ind Inform*. 2017;13(4):1969–78. doi:10.1109/TII.2017.2692775.
4. Fang X, Misra S, Xue G, Yang D. Smart grid—the new and improved power grid: a survey. *IEEE Commun Surv Tutor*. 2011;14(4):944–80. doi:10.1109/SURV.2011.101911.00087.
5. Wang MC, Tsai CF, Lin WC. Towards missing electric power data imputation for energy management systems. *Expert Syst Appl*. 2021;174(1):114743. doi:10.1016/j.eswa.2021.114743.
6. Mohammadi M, Kavousi-Fard A, Dabbaghjamanesh M, Farughian A, Khosravi A. Effective management of energy internet in renewable hybrid microgrids: a secured data driven resilient architecture. *IEEE Trans Ind Inform*. 2021;18(3):1896–904. doi:10.1109/TII.2021.3081683.
7. Do V, McBrien H, Flores NM, Northrop AJ, Schlegelmilch J, Kiang MV, et al. Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA. *Nat Commun*. 2023;14(1):2470. doi:10.1038/s41467-023-38084-6.
8. Yuan Y, Zhang C, Wang Y, Liu C, Ji J, Feng C. Linear interpolation process and its influence on the secondary equipment in substations. In: 2017 China International Electrical and Energy Conference (CIEEC); 2017 Oct 25–27; Beijing, China. p. 205–9. doi:10.1109/CIEEC.2017.8388447.
9. Liu T, Wei H, Zhang K. Wind power prediction with missing data using Gaussian process regression and multiple imputation. *Appl Soft Comput*. 2018;71:905–16. doi:10.1109/SURV.2011.101911.00087.
10. Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1–67. doi:10.18637/jss.v045.i03.
11. Broersen PMT, Bos R. Time-series analysis if data are randomly missing. *IEEE Trans Instrum Meas*. 2006;55(1):79–84. doi:10.1109/TIM.2005.861247.
12. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. Hoboken, NJ, USA: John Wiley and Sons; 2015.

13. Li B, Lin YP, Zhou SW, Luo Q, Yin B. An interpolation algorithm based on sliding neighborhood in wireless sensor networks. *J Comput Res Dev*. 2012;49(6):1196–203.
14. Poloczek J, Treiber NA, Kramer O. KNN regression as geo-imputation method for spatio-temporal wind data. In: *Proceedings of International Joint Conference SOCO'14-CISIS'14-ICEUTE'14*; 2014 Jun 25–27; Bilbao, Spain. Cham, Switzerland: Springer International Publishing; 2014. p. 185–93.
15. Su T, Shi Y, Yu J, Yue C, Zhou F. Nonlinear compensation algorithm for multidimensional temporal data: a missing value imputation for the power grid applications. *Knowl Based Syst*. 2021;215:106743. doi:10.1016/j.knosys.2021.106743.
16. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–8. doi:10.1093/bioinformatics/btr597.
17. Yoon J, Zame WR, Van Der Schaar M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Trans Biomed Eng*. 2018;66(5):1477–90. doi:10.1109/TBME.2018.2874712.
18. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Sci Rep*. 2018;8(1):6085. doi:10.1038/s41598-018-24271-9.
19. Marisca I, Cini A, Alippi C. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Adv Neural Inf Process Syst*. 2022;35:32069–82. doi:10.48550/arXiv.2205.13479.
20. Tashiro Y, Song J, Song Y, Ermon S. Csd: conditional score-based diffusion models for probabilistic time series imputation. *Adv Neural Inform Process Syst*. 2021;34:24804–16. doi:10.48550/arXiv.2107.03502.
21. Mesquita DPP, Gomes JPP, Rodrigues LR. Artificial neural networks with random weights for incomplete datasets. *Neural Process Lett*. 2019;50(3):2345–72. doi:10.1007/s11063-019-10012-02.
22. Jiang B, Siddiqi MD, Asadi R, Regan A. Imputation of missing traffic flow data using denoising autoencoders. *Procedia Comput Sci*. 2021;184:84–91. doi:10.1016/j.procs.2021.03.122.
23. Yoon J, Jordon J, Schaar M. Gain: missing data imputation using generative adversarial nets. In: *Proceedings of the 35th International Conference on Machine Learning*; 2018 Jul 10–15; Stockholm Sweden. p. 5689–98. doi:10.48550/arXiv.1806.02920.
24. McCoy JT, Kroon S, Auret L. Variational autoencoders for missing data imputation with application to a simulated milling circuit. *IFAC-PapersOnLine*. 2018;51(21):141–6. doi:10.1016/j.ifacol.2018.09.406.
25. Ma Q, Li X, Bai M, Wang X, Ning B, Li G. MIVAE: multiple imputation based on variational auto-encoder. *Eng Appl Artif Intell*. 2023;123:106270. doi:10.1016/j.engappai.2023.106270.