

Transfer-learning-based BiLSTM-WGAN Approach for Synthetic Data Generation of Sub-synchronous Oscillations in Wind Farms

Shuang Feng, *Member, IEEE*, Zhirui Zhang, Yuhang Zheng, Jiaxing Lei, *Member, IEEE*, and Yi Tang, *Senior Member, IEEE*

Abstract—The phenomenon of sub-synchronous oscillation (SSO) poses significant threats to the stability of power systems. The advent of artificial intelligence (AI) has revolutionized SSO research through data-driven methodologies, which necessitates a substantial collection of data for effective training, a requirement frequently unfulfilled in practical power systems due to limited data availability. To address the critical issue of data scarcity in training AI models, this paper proposes a novel transfer-learning-based (TL-based) Wasserstein generative adversarial network (WGAN) approach for synthetic data generation of SSO in wind farms. To improve the capability of WGAN to capture the bidirectional temporal features inherent in oscillation data, a bidirectional long short-term memory (BiLSTM) layer is introduced. Additionally, to address the training instability caused by few-shot learning scenarios, the discriminator is augmented with mini-batch discrimination (MBD) layers and gradient penalty (GP) terms. Finally, TL is leveraged to fine-tune the model, effectively bridging the gap between the training data and real-world system data. To evaluate the quality of the synthetic data, two indexes are proposed based on dynamic time warping (DTW) and frequency domain analysis, followed by a classification task. Case studies demonstrate the effectiveness of the proposed approach in swiftly generating a large volume of synthetic SSO data, thereby significantly mitigating the issue of data scarcity prevalent in SSO research.

Index Terms—Sub-synchronous oscillation, wind farm, data scarcity, synthetic data generation, artificial intelligence, Wasserstein generative adversarial network, transfer learning.

I. INTRODUCTION

IN recent years, the integration of power electronic devices into power systems, coupled with the increasing penetration of renewable power generation (RPG), has become in-

creasingly prevalent. However, the complex interactions between these power electronic devices, RPG installations, and the power system have given rise to sub-synchronous oscillations (SSO) issues [1]–[3]. SSO incidents occurred frequently in power systems worldwide. For instance, in 2019, the Hornsea offshore wind farm in the UK experienced a series of SSOs at a frequency of 8.4 Hz, triggered by an asymmetrical voltage drop fault [4]. In 2015, SSOs were observed in the weak grid of Hami, China [5]. Compared with low-frequency oscillations, SSO events have a broader frequency range typically between 2.5 Hz and 50 Hz, and are characterized by the strong nonlinearity of multimodal and time-varying dynamics. SSO can disconnect wind turbine or damage electrical equipment, potentially leading to system collapse and jeopardizing the safe and stable operation of the power system [6]–[9].

Existing analysis approaches for SSO can be divided into two main categories: model-based and data-driven. The model-based approaches include frequency scanning [10], time-domain simulation [11], eigenvalue analysis [12], open-loop modal [13], etc. These approaches inevitably depend on an accurate physical model, which requires detailed and precise system parameters. However, accurate modeling is challenging, considering the proprietary nature of equipment parameters and the complexity of high-order system modeling. Therefore, it is particularly difficult to accurately analyze SSO problems in practical power systems using model-based approaches.

With the advancement of artificial intelligence (AI), numerous scholars have proposed data-driven approaches for oscillation analysis and suppression. Reference [14] suggests a dynamic time warping (DTW)-based approach to classify multivariate oscillation time-series, which reduces the location time of the oscillation source and enhances location accuracy. Reference [15] achieves accurate location of forced oscillation sources using augmented Lagrange multiplier approach. Reference [16] transforms the oscillation source location issue into an image recognition one and proposes a two-stage deep transfer learning (TL) approach for high-precision location. Reference [17] employs convolutional neural network (CNN) architecture and long short-term memory (LSTM) network to predict transient and small-signal stability, and uses LSTM to learn the oscillatory response over

Manuscript received: May 26, 2024; revised: October 8, 2024; accepted: December 24, 2024. Date of CrossCheck: December 24, 2024. Date of online publication: January 20, 2025.

This work was supported by the National Natural Science Foundation of China (No. 52377084) and the Zhishan Young Scholar Program of Southeast University, China (No. 2242024RCB0019).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

S. Feng (corresponding author), Z. Zhang, J. Lei, and Y. Tang are with the Department of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: sfeng@seu.edu.cn; 220232952@seu.edu.cn; jxlei@seu.edu.cn; tangyi@seu.edu.cn).

Y. Zheng is with the School of Software, Southeast University, Suzhou 215123, China (e-mail: zyh19990219@163.com).

DOI: 10.35833/MPCE.2024.000550



time. Reference [18] proposes an adaptive dynamic programming approach, i.e., goal representation heuristic dynamic programming (GrHDP), for designing a supplementary damping controller (SDC) of voltage source converter based high-voltage direct current (VSC-HVDC) to suppress inter-area oscillations in power systems. Nevertheless, due to their temporal features, high dimensionality, and volatility, deep learning approaches often require intricate structures and a large number of parameters to extract features from SSO data. As the complexity of the network increases, the demand for data grows exponentially [19], exacerbating the challenge of data scarcity. Additionally, most of the phasor measurement unit (PMU) measurement data are confidential and highly sensitive, which further compounds the issue of data scarcity. In model training, data scarcity can lead to poor generalization performance, overfitting, and other challenges [20].

Generative adversarial network (GAN) [21], an unsupervised generative deep learning model, can learn intrinsic features of various data and rapidly generate synthetic samples distinct from the real ones, offering a promising solution to the issue of data scarcity in data-driven approaches. Currently, in the field of power systems, many studies have applied GAN in smart grids to address the issue of insufficient load data, which hinders the implementation of data-driven approaches [22]. Reference [23] generates a large number of power signatures using GAN for training non-intrusive load monitoring (NILM) models. Reference [24] proposes a data augmentation approach based on GAN models, which enhances low-resolution load profiles (LRLPs) collected from smart meters into high-resolution load profiles (HRLP). Reference [25] encodes load curves into red-green-blue (RGB) images and utilizes GAN models to generate load curves with realistic spatiotemporal correlations. In addition to the generation of load data, GANs have also been applied to fault diagnosis and the synthesis of PMU data. References [26] and [27] propose a GAN-based approach for augmenting partial discharge data in high-voltage cables, which effectively addresses the problem of limited training samples in partial discharge pattern recognition. References [28] and [29] utilize GAN to generate synthetic PMU data, thereby improving the accuracy of fault diagnosis in scenarios where training data are limited. However, traditional GANs are prone to gradient vanishing and explosion, leading to unstable training. Moreover, GAN itself is also a deep learning model, which means it is susceptible to data scarcity and still requires a large amount of sample data to ensure high-quality data generation. It is critical to address the few-shot learning problem for optimizing the performance of GAN models.

To tackle the problem of data scarcity in SSO research, this paper proposes a TL-based Wasserstein generative adversarial network (WGAN) approach for synthetic data generation of SSOs in wind farms. The proposed approach enables the efficient generation of synthetic data for SSO research. The main contributions of this paper are as follows.

1) Considering the temporal features of SSO data, the bidirectional long short-term memory (BiLSTM) network is introduced into the generator and discriminator of the WGAN, respectively, which is capable of capturing the information

of the entire oscillation sequence and improving the ability of the network to extract the temporal features of oscillation.

2) Due to the difficulty in obtaining large amounts of samples of SSO in practice, the mini-batch discrimination (MBD) layer and gradient penalty (GP) term are introduced into the BiLSTM-WGAN to improve the stability of the model in few-shot learning.

3) The TL technique is combined with GAN to reduce the time and difficulty of training, minimizing the training samples of SSO required for the task of data generation.

4) Two indicators that characterize the effectiveness of the synthetic data are proposed, one based on the DTW algorithm and the other based on the frequency domain analysis (FDA).

5) The effectiveness of the proposed approach is demonstrated through classification tasks in different power systems, the results of which show that the classification accuracy can be improved with the synthetic SSO data.

The paper is structured as follows. Section II introduces the BiLSTM-WGAN for improving the temporal feature extraction of SSO data. Section III discusses the improvement of BiLSTM-WGAN and TL for the few-shot learning problem. Section IV presents the procedure and quality check of the the proposed approach. Section V demonstrates the effectiveness of the proposed approach through case study. Finally, Section VI concludes this paper.

II. BiLSTM-WGAN FOR IMPROVING TEMPORAL FEATURE EXTRACTION OF SSO DATA

A. Problem Formulation

In the calculation of the linearized state equation of the power system containing n permanent magnet synchronous generators (PMSGs) [30], the state variable matrix ΔX of the system can be expressed as the sum of the zero-input response $\Delta x_1(t)$ and the zero-state response $\Delta x_2(t)$. Let the initial time be t_0 , the diagonal array formed by all the eigenvalues of state matrix A be Λ , and the corresponding left and right eigenvector matrices be $\Phi = [\phi_1, \phi_2, \dots, \phi_s, \phi_1^*, \phi_2^*, \dots, \phi_s^*]$, $\Psi = \Phi^{-1} = [\psi_1, \psi_2, \dots, \psi_s, \psi_1^*, \psi_2^*, \dots, \psi_s^*]$, then the i^{th} component of $\Delta x_2(t)$ is [31]:

$$\Delta x_{2i}(t) = \sum_{r=1}^s \left(\phi_{ir} e^{\lambda_r t} \int_{t_0}^t e^{-\lambda_r \tau} \sum_{l=1}^m \psi_{rl} b_l \Delta u_l(\tau) d\tau + \phi_{ir}^* e^{\lambda_r^* t} \int_{t_0}^t e^{-\lambda_r^* \tau} \sum_{l=1}^m \psi_{rl}^* b_l \Delta u_l(\tau) d\tau \right) \quad (1)$$

where ϕ_{ir} , ϕ_{ir}^* , and ψ_{rl} , ψ_{rl}^* are the i^{th} and r^{th} entries of the right and left eigenvectors, respectively; (λ_r, λ_r^*) is the conjugate complex eigen root of matrix A ; s is the number of pairs of conjugate complex eigen roots; m is the number of perturbations in n generators; and $b_l \Delta u_l(t)$ is the l^{th} perturbation, $l = 1, 2, \dots, m$.

Considering that it is difficult to obtain the exact parameters of the system, the elements of matrix A cannot be fully determined. As a result, it becomes challenging to perform calculations using (1). Besides, the above temporal features of SSOs are obtained based on the linearized analysis ap-

proach, which are more complicated in practice due to the nonlinearity components of power electronic devices. The SSO data obtained from the linearized model may not reflect the true oscillation pattern of the actual system. Aiming to tackle the problems above, this paper proposes a TL-based BiLSTM-WGAN approach for synthetic data generation of SSOs in wind farms.

B. From GAN to WGAN-GP

The GAN model is composed of two neural networks, i.e., discriminator (D) and generator (G). The generator is used to generate new, namely fake data as real as possible, in order to disable the discriminator to accurately distinguish the effectiveness of the data. Conversely, the discriminator is used to distinguish the synthetic data from the real data. They are trained together in a competitive process until a Nash equilibrium is reached.

The overall objective function V of the GAN model is shown in (2):

$$\min_G \max_D V(G, D) = E_{x \sim P_r} [D(x)] - E_{z \sim P_g} [D(G(z))] \quad (2)$$

where $E[\cdot]$ is the mathematical expectation; $G(z)$ is the data output by the generator; $D[G(z)]$ is the probability value output by the discriminator; x is the synthetic data; P_r is the distribution of real data; and P_g is the distribution of synthetic data.

For the original GAN, its loss functions are prone to the phenomenon of gradient vanishing, resulting in high similarity with the real data. To address this problem, a WGAN is adopted in this paper [32]. The Wasserstein distance measures the difference in probability distribution between the synthetic data and the real data, and the convergence of the training at the same time. The closer the distance, the better the training effect. The Wasserstein distance is defined as:

$$W(P_r, P_g) = \inf_{\delta \in \Pi(P_r, P_g)} E_{(x,y) \sim \delta} [\|x - y\|] \quad (3)$$

where $\Pi(P_r, P_g)$ is the set of joint probability distributions δ with P_r and P_g as the marginal distributions; $\|x - y\|$ is the distance needed to shift synthetic data x to the real data y in order to fit P_g to P_r ; and the value of $W(P_r, P_g)$ is the distributional similarity between the real SSO data and the synthetic data of WGAN.

Since it is difficult to directly solve the Wasserstein distance, WGAN utilizes the Kantorovich-Rubinstein duality transformation, as shown in (4).

$$W(P_r, P_g) = \frac{1}{K} \sup_{\|\gamma\| \leq K} \{E_{x \sim P_r} [\gamma(x)] - E_{y \sim P_g} [\gamma(y)]\} \quad (4)$$

where $\sup\{\cdot\}$ is the upper definite bound of the function value; K is the Lipschitz constant; and $\|\gamma\| \leq K$ means that the function γ satisfies K -Lipschitz continuity, i.e., for any inputs x and y , there is $|\gamma(x) - \gamma(y)| \leq K\|x - y\|$. Function γ is the discriminator function obtained utilizing the fitting ability of the neural network.

The use of duality transformation makes the solution of the Wasserstein distance easier, and the magnitude of the Wasserstein distance can be directly equated to the difference between the real and synthetic data. To satisfy the Lipschitz condition, each batch of data performs the weight cropping, and the weight parameters of the discriminator network are restrict-

ed to a certain interval range $([-C, C])$ to achieve Lipschitz continuity. This makes the network parameters tend to extremes, and the performance of the neural network cannot be effectively utilized. In addition, the network parameters are usually set manually so that most of the weights have critical values after weight cropping, which may lead to the problems of gradient vanishing and gradient explosion again.

To solve the above problems, a new loss term GP is added to the original loss function of WGAN, so that WGAN satisfies the Lipschitz continuity condition [33].

$$GP = \lambda E_{x \sim \chi} [\|\nabla D(x)\|_p - 1]^2 \quad (5)$$

where λ is the penalty coefficient; χ is the probability distribution of the entire sample space, i.e., the set of probability distributions of the synthetic samples and the real samples; $\nabla D(x)$ is the gradient of the discriminator; and $\|\cdot\|_p$ is the norm.

The objective function of WGAN-GP model can be expressed by:

$$L_{GP} = E_{x \sim P_r} [D(x)] - E_{x \sim P_g} [D(x)] + \lambda E_{x \sim \chi} [\|\nabla D(x)\|_p - 1]^2 \quad (6)$$

The WGAN-GP model is able to provide more effective gradient information, which improves the network convergence performance, enhances the stability of generator and discriminator training, and enables the model to accomplish small-sample data generation tasks.

C. BiLSTM-WGAN for Extracting Time-series Feature of SSO Data

While WGAN is commonly employed for image generation tasks, it presents challenges in effectively capturing the inherent temporal features in SSO data. In this paper, the BiLSTM network is introduced into the WGAN model, so as to fully capture the forward and backward sequence of the oscillation data and to extract the temporal features.

SSO events have a wider frequency range possibly between 2.5 Hz and 50 Hz, which means that the value of a data point at a particular time is time-dependent and these dependencies can be short-term, long-term, or a combination of both. The LSTM network [34] can effectively extract the features of long interconnectedness in time-series during training [35]. Furthermore, SSOs in systems integrated with renewable energy are characterized by strong nonlinearity of multimodal and time-varying nature. The nonlinearity exhibits complex relationships between past and future values. BiLSTM network [36], by considering the information from both directions, can better capture these complex relationships and model the nonlinear dynamics of the signal. Besides, most SSO data are time-varying, meaning that their statistical properties change over time. BiLSTM network can adapt to these changes more effectively than LSTM network, as they can consider information from both the past and future to identify trends and patterns. BiLSTM network introduces two independent hidden layers based on LSTM network, splices them together to the same output layer, and then superimposes the forward and backward information. It not only considers the historical information of the time-series, but also utilizes the future information of the data [37].

The computational process of the BiLSTM network is shown in Fig. 1, where W_f is the weight of forward computation; W_b is the weight of backward computation; x_i is the in-

put layer; and O_i is the output layer.

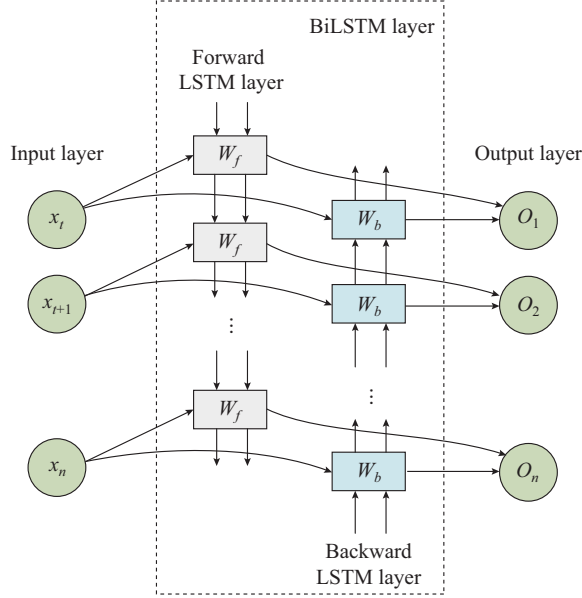


Fig. 1. Computational process of BiLSTM network.

To adequately capture the temporal features of the SSO data, in this subsection, a BiLSTM neural network is constructed and embedded into the generator and discriminator. The network structure of the BiLSTM-WGAN is shown in Fig. 2.

In the process, a high-dimensional random noise sequence following a Gaussian distribution is input into the generator. A single-layer BiLSTM neural network is used to process the input sequence from the generator. The time step of the

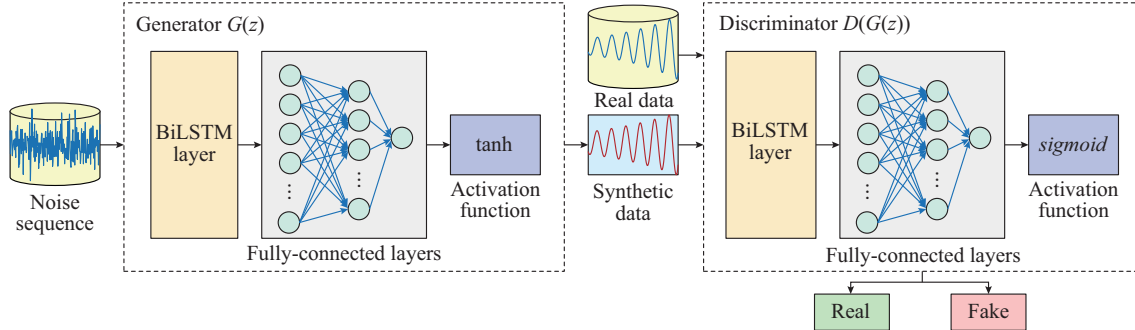


Fig. 2. Network structure of BiLSTM-WGAN.

III. IMPROVEMENT OF BiLSTM-WGAN AND TL FOR FEW-SHOT LEARNING PROBLEM

A. Improvement of BiLSTM-WGAN

The BiLSTM-WGAN used in Section II is able to extract temporal features of SSO and effectively improve the problem of unstable training of the model in few-shot learning, but there is still a risk of falling into local instead of global optimal solutions during the training process. Consequently, the BiLSTM-WGAN loses data diversity and leads to model collapse. To prevent the above problem, the MBD [38] layer is used in this subsection to increase the diversity and robustness of the BiLSTM-WGAN. The MBD layer calculates the

BiLSTM layer is consistent with the length of the generator's input sequence, and the number of neurons in its hidden cell units is adjusted. Then, the tensor output from the BiLSTM layer is fed into a fully-connected layer for feature extraction. After performing calculations through three fully-connected layers, the tensor is reshaped by using a Reshape layer. Layer normalization (LN) is applied between the layers, which is suitable for RNN as it stabilizes the training process by normalizing the mean and variance of the input data at each layer, thus accelerating convergence. To avoid potential negative values resulting from LN, the activation function used in the input layer and the middle layer is set as leaky rectified linear unit (LeakyReLU), and the dropout layer is added to randomly deactivate neurons with a specified probability, which effectively prevents overfitting. To ensure that the data output from the generator is diverse and avoids convergence to a specific distribution, the LN is omitted in the output layer. Finally, the output tensor is processed by using the tanh activation function, so that the values of the output data are all non-negative, thereby obtaining the final synthetic data that match the dimensionality of real SSO data.

The structure of the discriminator is similar to that of the generator, consisting of one BiLSTM layer and three fully-connected layers. The LeakyReLU activation function and dropout layer are also added between the layers, but the difference with the generator is that the LN is renounced to make the probability distribution of the output data more decentralized, which is conducive to the discriminator to distinguish the effectiveness of the data. In addition, the *sigmoid* activation function is used in the output layer to make the output be mapped between (0, 1).

similarity between the samples in the minibatches, and adds this information into the feature representation of the discriminator, which helps the discriminator to more accurately distinguish between real and synthetic data and facilitates the convergence of the WGAN model. The principle of the MBD layer is shown in Fig. 3.

The feature vector of input x_i is defined as $f(x_i) \in \mathbb{R}^A$, generated by one of the intermediate layers of the discriminator, and multiply $f(x_i)$ by a tensor $T \in \mathbb{R}^{A \times B \times C}$ to obtain the corresponding result matrix $M_i \in \mathbb{R}^{B \times C}$. Then the difference of the b^{th} feature of each sample x_i from the other samples in the same batch is calculated as:

$$c_b(x_i, x_j) = \exp(-\|M_{i,b} - M_{j,b}\|_{L1}) \in \mathbb{R} \quad (7)$$

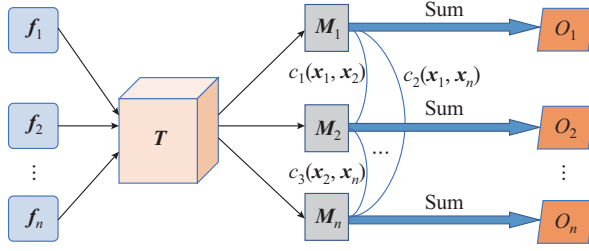


Fig. 3. Principle of MBD layer.

where $M_{i,b}$ is the b^{th} row of M_i ; and the $L1$ -norm is used to denote the difference between the two vectors.

The difference between the individual sample features is summed as the output of the MBD layer:

$$\begin{cases} o(x_i)_b = \sum_{j=1}^n c_b(x_i, x_j) \in \mathbb{R} \\ o(x_i) = [o(x_i)_1, o(x_i)_2, \dots, o(x_i)_B] \in \mathbb{R}^B \\ o(X) \in \mathbb{R}^n \times B \end{cases} \quad (8)$$

where n is the dimension of the eigenspace; o is the sum of c of each sample; B and b are the batch size and the batch index, respectively; and $o(x_i)_b$ is the combination of features of the real samples, which can diversify the samples generated by the WGAN.

In this subsection, based on BiLSTM-WGAN, an MBD layer is added to the discriminator network. The structure of the BiLSTM-WGAN with MBD layer is shown in Fig. 4.

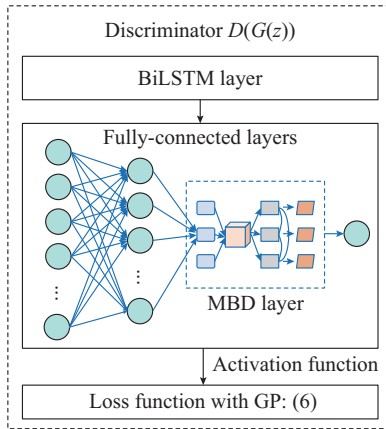


Fig. 4. Structure of BiLSTM-WGAN with MBD layer.

B. TL

In the previous introduction, a BiLSTM-WGAN with MBD layer for the generation of SSO data is proposed. Nevertheless, the scarcity of sub-synchronous samples from practical power systems poses a significant challenge in adequately training deep learning models for this specific task. Obtaining training data by constructing an electromagnetic transient model of the actual system faces the problem of inconsistency in topology and parameters between the actual system and the simulation system. To solve such problems, this paper utilizes the simulation data to train a base model first, and then adopts a small amount of data from the actual

system to fine-tune the system based on the TL approach, which solves the few-shot learning problem.

TL [39] aims to utilize knowledge learned in a related domain (called the source domain) to help accomplish a task in another domain (called the target domain). In deep learning, TL minimizes the number of labeled samples required for the target domain task, improves the generalization of the model, avoids training the neural network from scratch, and saves training time.

In this paper, the freeze discriminator (FreezeD) approach is utilized, which freezes the BiLSTM layer and fully-connected layers in the lower layers of the discriminator, and only fine-tunes the parameters of the fully-connected layer and that of the MBD layer in the higher layers for adversarial training. This is because the low-level network of the discriminator is responsible for learning the temporal features and other general features of the SSO data, while the high-level learning discriminates the data as real or fake based on the extracted features. The FreezeD approach accelerates the model training and avoids the overfitting problem arising from direct fine-tuning.

IV. PROCEDURE AND QUALITY CHECK OF PROPOSED APPROACH

A. Procedure

The specific procedure of the model and training approach proposed in Section III is described as follows.

- 1) Training data acquisition: build the electromagnetic transient simulation model of the power system. Obtain the training data from the simulation and preprocess them.
- 2) Model construction: build the BiLSTM-WGAN with MBD layer. The structure is shown in Fig. 4.
- 3) Model pretraining: input the SSO data obtained from the simulation system into the BiLSTM-WGAN with MBD layer. Train the basic discriminator model D_s and the basic generator model G_s to learn the temporal features with sufficient SSO data. The pre-trained models D_s and G_s are obtained.
- 4) Model transferring: transfer D_s and G_s . Fine-tune the transferred models that are then trained under small samples of oscillation data from different power systems or different oscillation modals of the same system. Obtain the discriminator model D_t and generator model G_t , which are applicable under the target domain.

5) Synthetic data generation: use G_t to synthesize a substantial volume of SSO data corresponding to a specific system configuration or oscillation modal. Subsequently, expand the real dataset in accordance with the synthetic data to effectuate data augmentation.

6) Quality check: the synthetic data are analyzed using the indexes based on DTW algorithm and FDA and the random forest classifiers to assess whether the temporal characteristics of the synthetic data are in line with the characteristics of real SSO data.

The procedure of the proposed TL-based BiLSTM-WGAN approach is shown in Fig. 5.

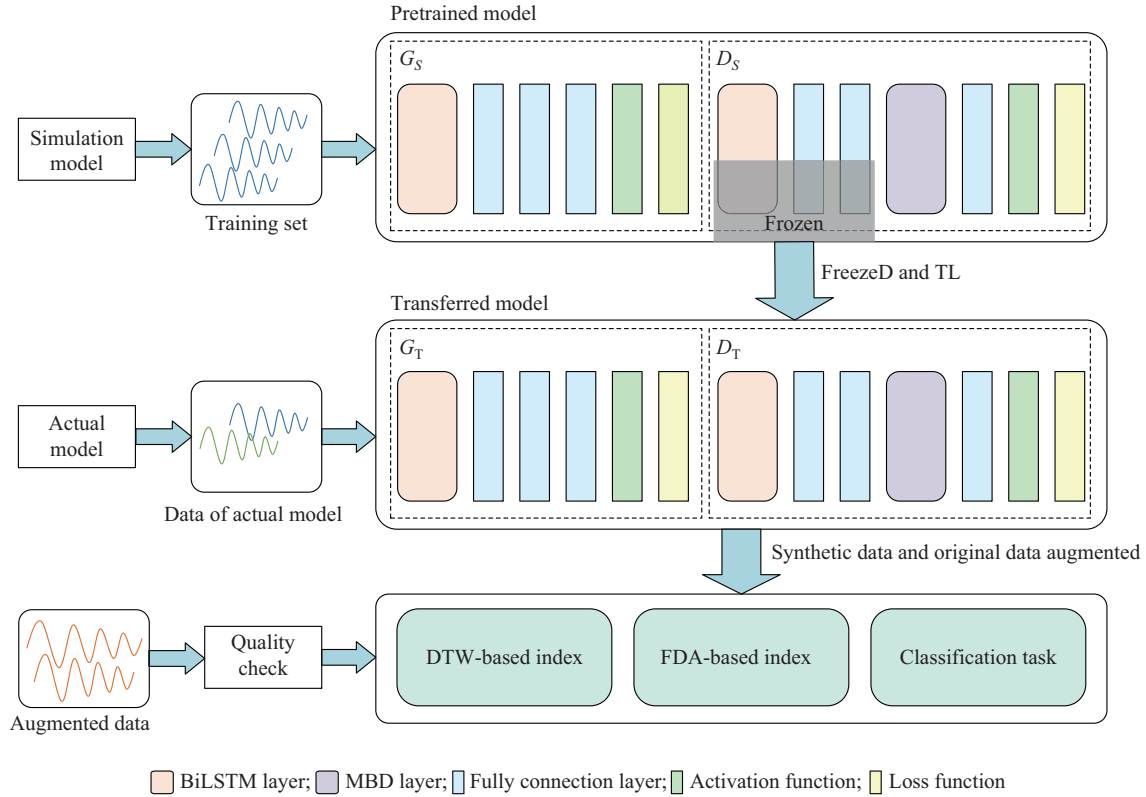


Fig. 5. Procedure of proposed BiLSTM-WGAN approach.

B. Synthetic Data Quality Check

The synthetic SSO data generated by the BiLSTM-WGAN need to satisfy the features of real SSO data as much as possible. In this subsection, one index based on the DTW and the other index based on the FDA are deployed to evaluate the quality of the synthetic data, respectively. Finally, the ability of the synthetic data to improve the classification accuracy of the BiLSTM-WGAN is verified based on actual classification examples.

1) DTW-based Index

The DTW [40] is classical and effective in measuring the difference of data, and its main principle is to perform repeated sampling by distorting one of the sequences when comparing two time-series, mapping and misaligning the sequences. Then, it is transformed into a classical dynamic programming problem by traversing all the possible transformed paths to find the ideal path that minimizes the total distance between the mapped points.

Suppose that $\mathbf{Q}=\{q_1, q_2, \dots, q_m\}$ is a sequence of primitive SSO of dimension m , and $\mathbf{C}=\{c_1, c_2, \dots, c_n\}$ is a model-generated sequence of dimension n . DTW tries to find an optimal alignment between $\mathbf{Q}$ and $\mathbf{C}$ that minimizes the cumulative difference between the mapped points [41]. In the mapping, one point in a sequence can be mapped to multiple points in another series. DTW finds such a mapping by creating a distance matrix $\mathbf{D}$ between each pair of points in $\mathbf{Q}$ and $\mathbf{C}$, where each element $d(i, j)$ of the matrix is the distance required to align q_i with c_j . A dynamic programming algorithm is then used to find a shortest path in $\mathbf{D}$ which is called $\mathbf{P}$, and to compute the dynamically regularized dis-

tance $DTW(i, j)$ to minimize the sum of the elements of the distance matrix over the path from the lower left corner to the upper right corner, as shown in the following recursive equation (9):

$$DTW(i, j) = d(i, j) + \min \{DTW(i-1, j), DTW(i, j-1), DTW(i-1, j-1)\} \quad (9)$$

where $d(i, j)$ is calculated using the Euclidean distance.

According to the definition above, the DTW distance is used to represent the similarity between the real and synthetic data. It provides a basis for testing the effectiveness of the BiLSTM-WGAN with MBD layer in data generation.

2) FDA-based Index

The time-domain expression for the occurrence of SSO in a power system can be expressed as:

$$x(t) = A_0 \cos(2\pi f_0 t + \delta_0) + \sum_{i=1}^s A_i e^{\alpha_i t} \cos(2\pi f_i t + \delta_i) \quad (10)$$

where A_0, f_0, δ_0 and A_i, f_i, δ_i are the initial amplitude, frequency, and phase of the fundamental frequency component and the i^{th} SSO component of the system, respectively; α_i is the damping ratio coefficient of the SSO component; and s is the number of oscillation modals.

The synthetic data and the real data are computed using the total least squares-estimating signal parameter via rotational invariance techniques (TLS-ESPRIT) [42] to obtain the SSO frequency f_s and damping ratio α . Polynomial fitting is performed on the two sets of data using the least squares approach, respectively. The function is fitted by minimizing S , which is the sum of the squares of e as shown below:

$$S = ne^2 = \sum_{i=1}^n (\alpha_i - \hat{\alpha}_i)^2 \quad (11)$$

$$\alpha(f) = \sum_{i=0}^n w_i f^i \quad (12)$$

where α_i is the sample value; $\hat{\alpha}_i$ is the actual value; e is the error between the sample value and the actual value; w_i is the polynomial coefficient; and f is the frequency. Subtract the two polynomials and then perform definite integration, as shown in (13).

$$Int = \left| \int_a^b (\alpha_s(f) - \alpha_r(f)) df \right| \quad (13)$$

where a and b are the upper and lower limits of the oscillation frequency, respectively; and α_s and α_r are the fitting functions of the synthetic and real data, respectively.

The magnitude of the integral difference reflects the proximity of the frequency-damping ratio [43] distributions of the two sets of data, which, in turn, reflects the authenticity of the oscillation modal of the synthetic data.

C. Classification Task of SSO Type for Synthetic Data

In addition, to verify the effectiveness of the synthetic data for model training, this paper adopts a classification algorithm for verification. Depending on their different triggering perturbations, the oscillations can be categorized into forced oscillations and natural oscillations. The two types of oscillations are generated and then used to augment the data and train the random forest (RF) model [44] for classification.

RF is a supervised machine-learning algorithm based upon the bootstrap aggregating (Bagging) idea with put-back in ensemble learning (EL), which randomly selects features to generate decision trees. Each decision tree gives the voting result of a category, and selects the category with the most votes from all the trees as the final prediction result. Compared with the single tree approach, the Bagging idea in RF reduces the variance of the prediction function. The principle of the RF model is shown in Fig. 6.

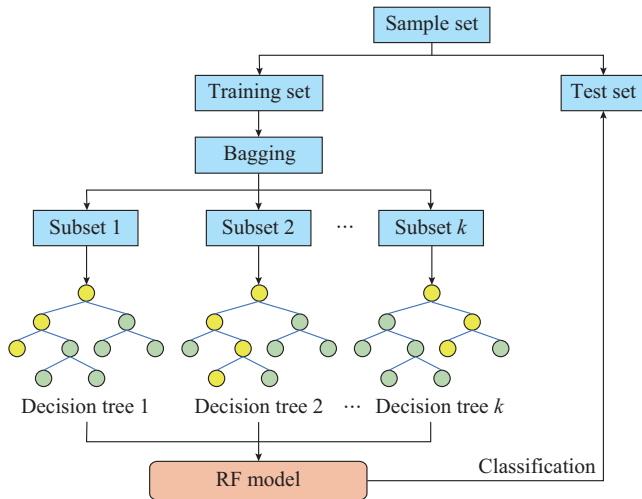


Fig. 6. Principle of RF model.

The evaluation process based on the classification of SSO type with RF model is as follows.

1) First, calculate the feature indexes in four aspects,

namely statistical domain, time domain, frequency domain, and dominant modal based on fast Fourier transform (FFT) spectral analysis. Twenty indexes in four aspects are used as the input of the RF classifier, as shown in Table I.

TABLE I
TWENTY FEATURE INDEXES IN FOUR ASPECTS

Aspect	Feature indexes
Statistical domain	Mean, standard deviation, median absolute deviation, Kurtosis, skewness, and root-mean-square value
Time domain	Peak, autocorrelation coefficient, Kurtosis index, margin index, waveform index, and pulse index
Frequency domain	Center frequency, variance, root mean square frequency, Skewness, waveform stability factor, and frequency center
Dominant modal	Frequency and damping ratio

2) Subsequently, train the RF algorithm with a few data from the actual system.

3) Finally, augment the real data with different amounts of the synthetic data, train the RF algorithm, and observe whether the augmented data are able to improve the classification accuracy, so as to verify the effectiveness of the synthetic data [45].

V. CASE STUDY

A. Test Systems

1) Training Sample Acquisition

To choose suitable test systems, two aspects must be taken into consideration: whether the system is able to simulate SSOs with different mechanisms, and whether it considers the influential factors of SSO features. Considering the different mechanisms, the samples generated in this paper need to encompass two types of oscillations: forced oscillations and negative damping oscillations. To investigate various factors influencing SSO, this paper verifies the effects by adjusting wind speed and turbine control parameters in both single-machine grid-connected systems and large-scale wind farm grid-connected systems.

The SSO data used for training and testing of the pre-trained model are obtained by modeling and simulating a wind farm system based on a single direct-drive PMSG connected to a weak power grid, the system structure of which is shown in Fig. 7. In Fig. 7, C_{dc} is the DC capacitor; R_f is the filter resistance; L_f is the filter inductance; R_g is the grid resistance; L_g is the grid inductance; I_g is the output current; MSC represents the machine-side converter; GSC represents the grid-side converter; and PCC represents the point of common coupling. Initially, the steady-state simulation parameters of the system are configured with twenty wind turbines, a wind speed of 6 m/s, a simulation time of 1 s, and the steady-state values are saved. Subsequently, the oscillation simulation parameters of the system are adjusted, with the wind speed incrementally increased from the steady-state value by 0.001 m/s as a disturbance, until reaching 16 m/s. The simulation duration is 1 s with a sampling interval of 0.005 s. Ten thousand sets of SSO active power data are collected from the PCC in Fig. 7, which are then used to train the pre-

trained model shown in Fig. 5.

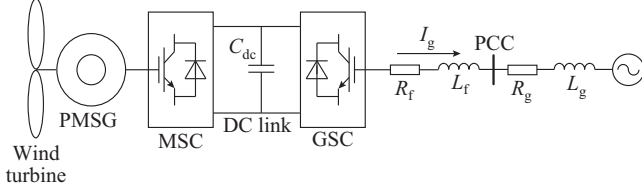


Fig. 7. Direct-drive PMSG connected to weak power grid.

To verify the feasibility of the proposed approach, three different cases are studied in this paper.

1) Case 1: parameter-modified PMSG grid-connected system. Adjust the control parameters of PMSG. Change the oscillation modal of the system. The wind speed perturbation is applied to stimulate the SSO phenomenon. By adjusting the voltage and current loop parameters, the damping and frequency of the system are changed to simulate different oscillation modals. One hundred sets of active power data at the PCC are obtained as the training data for the training of transferred model in Case 1.

2) Case 2: forced SSO of a four-machine two-area system connected with a large-scale wind farm. The structure of the four-machine two-area system with large-scale wind farm is shown in Fig. 8, which contains four 900 MW synchronous machines and a wind farm containing 200 PMSGs. By applying a periodic sinusoidal perturbation at the grid-side controller of the direct-drive turbine that is consistent with the frequency of the system oscillation modal, the system is excited to generate forced SSO. Eighty forced oscillation data between bus 11 and bus 10 in Fig. 8 are sampled by adjusting the control parameters, which are used as the training data for TL in Case 2.

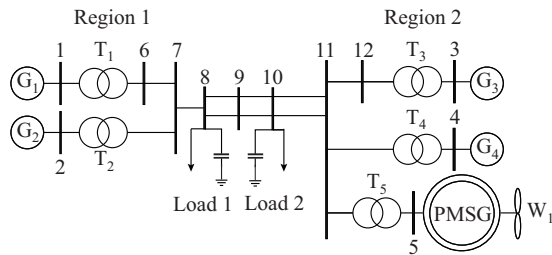


Fig. 8. Structure of four-machine two-area system with large-scale wind farm.

3) Case 3: on the basis of Case 2, the system is excited by a step perturbation to negatively generate damped oscillations with damping ratio close to 0. By adjusting the wind speed and the voltage and current loop parameters, 80 sets of active power outputs at bus 11 are obtained and used as the training data for Case 3.

In this paper, the BiLSTM-WGAN with MBD layer will be firstly trained using the PMSG grid-connected system data and then transferred to three cases respectively to validate the feasibility and advantages of the proposed approach through quality check. On this basis, the synthetic data generated in Cases 2 and 3 are used to augment the real data and train the RF classifier to discriminate the two types of oscillations and verify the effectiveness of the synthetic data.

2) Data Preprocessing

Considering the outliers in the measurement data that affect the training, this section uses the box plot approach in statistics to detect outliers and fills in the missing points using the median M . The box plot approach sets boundaries for the data and determines data points that exceed the upper boundary or are smaller than the lower boundary as outliers, as shown in (14) and (15).

$$X_i > U + K_b \cdot IQR \quad (14)$$

$$X_i < L - K_b \cdot IQR \quad (15)$$

where X_i is the value of the variable; U is the upper quartile, i.e., the median of the interval $[M, X_n]$; L is the lower quartile, i.e., the median of the interval $[X_1, M]$; K_b is the coefficient, which generally takes the value of 1.5; and IQR is the interquartile range, i.e., distance between the upper and lower quartiles.

The data after removing outliers are divided into training, validation, and test sets, with a total of 5000 sets of data in the training and validation sets and 1000 data in the test set.

The divided training data X are normalized by Z-score to transform the data into a uniform magnitude:

$$\hat{X} = \frac{X - \bar{X}}{\sigma} \quad (16)$$

where $\bar{X}$ is the mean; and σ is the standard deviation.

B. Training Process

By adjusting the hyperparameters, the convergence of the pretrained model is optimized. The final hyperparameter settings for the pretrained model are shown in Table II.

TABLE II
HYPERPARAMETER SETTINGS FOR PRETRAINED MODEL

Hyperparameter	Value
Learning rate of generator	0.0003
Learning rate of discriminator	0.0002
Batch size	128
Epoch	1000
Number of discriminator updates per generator update	2
Number of MBD layers	3

The change in Wasserstein distance over the course of training is shown in Fig. 9 (every five rounds).

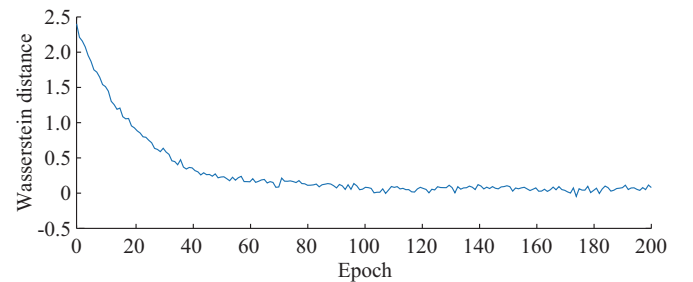


Fig. 9. Change in Wasserstein distance.

As shown in Fig. 9, the model converges quickly in training, and the Wasserstein distance between the synthetic data and the real data decreases gradually at the beginning of

training, and fluctuates slightly around 0 in the middle and late stages of training. This proves that the generator has sufficiently grasped the distribution of the data.

The synthetic data and the real data are subjected to modal identification using the TLS-ESPRIT to compute the frequency and damping ratio for each data set. Polynomial fitting using the least squares approach is then applied to the frequencies and damping ratios of both the synthetic and real data. The data distributions and fitting results for the two models are shown in Fig. 10.

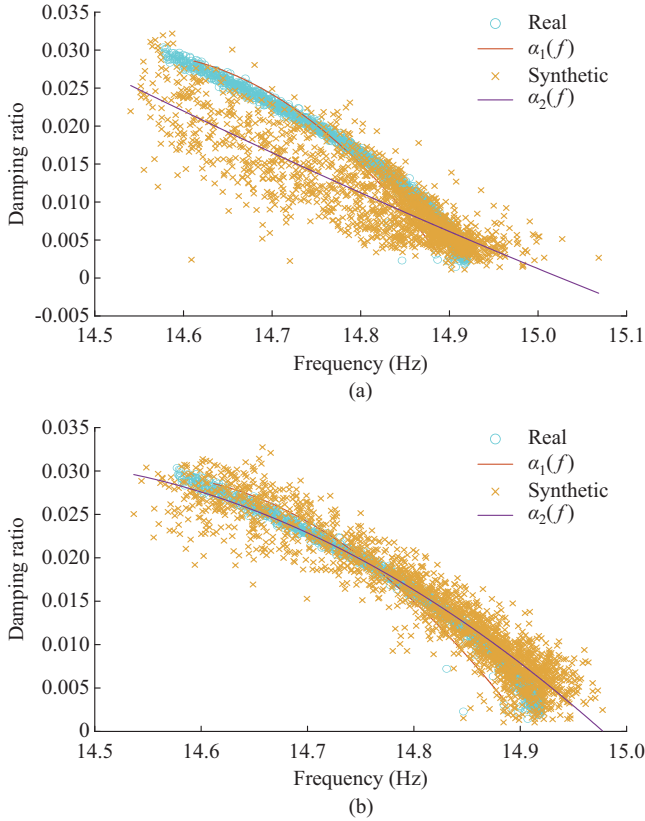


Fig. 10. Data distributions and fitting results for two models. (a) Without BiLSTM. (b) With BiLSTM.

As illustrated in Fig. 10, by introducing the BiLSTM layer, the synthetic data are more consistent with the temporal features of the real data, while the frequency-damping ratio distributions without BiLSTM are more deviated. Also, the oscillation modal of the synthetic data generated by the BiLSTM-WGAN exhibits greater diversity based on the real data. Consequently, this underscores the critical role of BiLSTM in capturing the temporal characteristics of SSO and validates the efficacy of the proposed approach.

Subsequently, the pre-trained discriminator model D_s and generator model G_s are transferred to Case 1, Case 2, and Case 3, respectively. The lower layers of the discriminator are frozen, and the hyper-parameters are fine-tuned for 10 iterations. For instance, in Case 1, Fig. 11 illustrates a batch of active power samples generated after one round of fine-tuning. Comparison of Fig. 11(a) and (b) reveals significant variability within the batch of samples produced by the model incorporating the MBD layer. Conversely, the synthetic samples generated by the model without the MBD layer ex-

hibit minimal diversity. This indicates that the model without the MBD layer may be overfitting, leading to the generation of numerous repetitive samples that lack diversity.

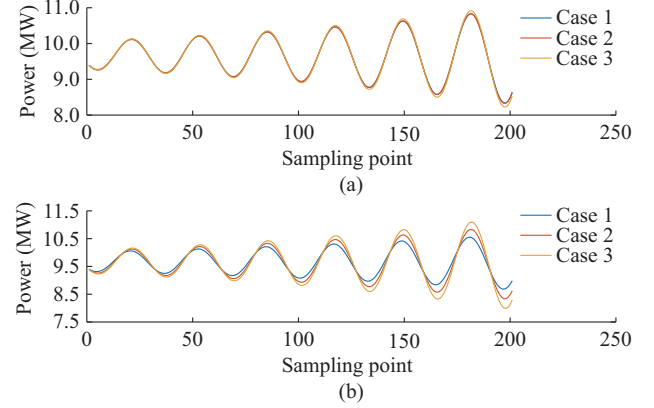


Fig. 11. Active power samples generated. (a) Without MBD layer. (b) With MBD layer.

C. Quality Check Results of Synthetic Data

To demonstrate the effectiveness of the TL in few-shot learning, the model without TL is used as a comparison, with the same small-sample dataset of the target domain used for both approaches.

The small-sample data of Cases 1-3 are each trained K times using 10-fold cross-validation, and each time the same amount of data as the test set are generated using the three trained target generation models G_t for the validation analysis. Each time, the DTW-based indices of the synthetic data and the real data are calculated and the mean value is taken as the final test result. Figure 12 shows the comparison results of the DTW-based indexes in Cases 1-3.

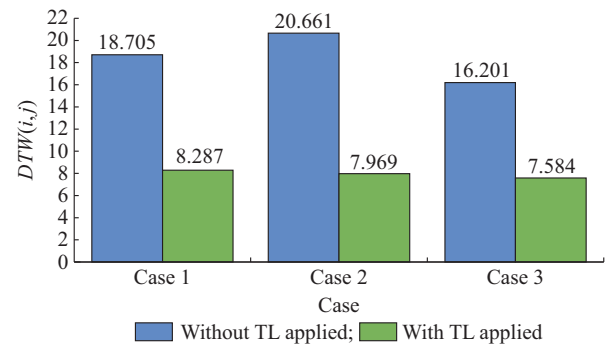


Fig. 12. Comparison results of DTW-based indexes in Cases 1-3.

As can be observed in Fig. 12, the DTW-based indexes of the synthetic data after TL are significantly reduced, indicating that the model learns the temporal features of the real data, while the model without TL applied is unable to generate realistic SSO data due to the small number of samples in the target domain, resulting in the failure of the training to converge.

The FDA-based indexes of Cases 1-3 are calculated according to (13), and the comparison results are shown in Fig. 13. As can be observed from Fig. 13, the FDA-based indexes of the oscillation modal of the TL-based data generated by BiLSTM-WGAN with MBD layer are smaller compared with those of the model without TL applied. This indi-

cates that the frequency-damping ratio distribution of the synthetic data is closer to that of the real data, and the oscillation modals of the synthetic data are more realistic. It is demonstrated that the transferred model shows better performance on few-shot learning, achieving the expansion of a small bunch of SSO data samples.

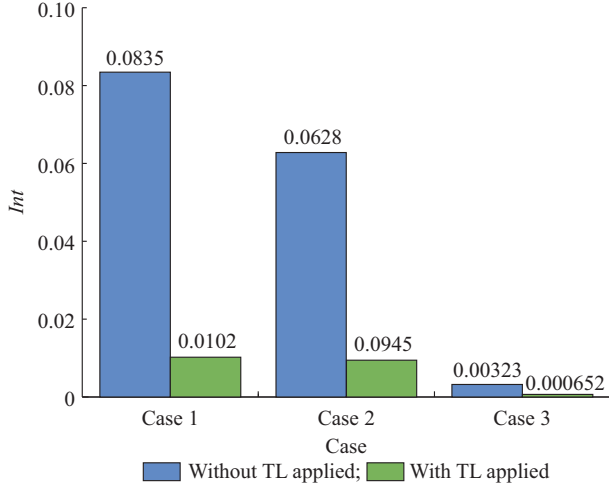


Fig. 13. Comparison results of FDA-based indexes in Cases 1-3.

The frequency and damping ratio of each piece of data are calculated after each fine-tuned training using the parameter identification approach in Section IV. Subsequently, the

polynomial fitting of the total frequency-damping ratio of the synthetic data and the real data is performed using the least squares approach. The distribution of the data and fitting results of Cases 1-3 are shown in Fig. 14. It can be seen that in the three cases, the damping ratio distribution of the synthetic data of the transferred model more closely matches real data than the distribution of the model without TL applied.

D. Classification Task Results for Synthetic Data

Since the amplitudes of the forced oscillations in Case 2 and the negatively damped oscillations in Case 3 are nearly equal, it is difficult to determine the type of SSOs by observing whether the waveforms are dispersed or attenuated.

In this subsection, the oscillation type discrimination approach based on the RF classifier is used to evaluate the quality of the synthetic data generated based on small samples for Cases 2 and 3. The data obtained from the simulation for the two cases are first processed to remove outliers, and then the data are labeled according to the two categories of negatively damped and forced oscillations. The real dataset, which contains 20 features, is divided into a training set and a test set using the 10-fold cross-validation approach. The training set is feature-normalized and input into the RF model, and its hyperparameters are adjusted for training. Calculate the classification accuracy on the test set and average over the K tests. The classification accuracy formula is shown as:

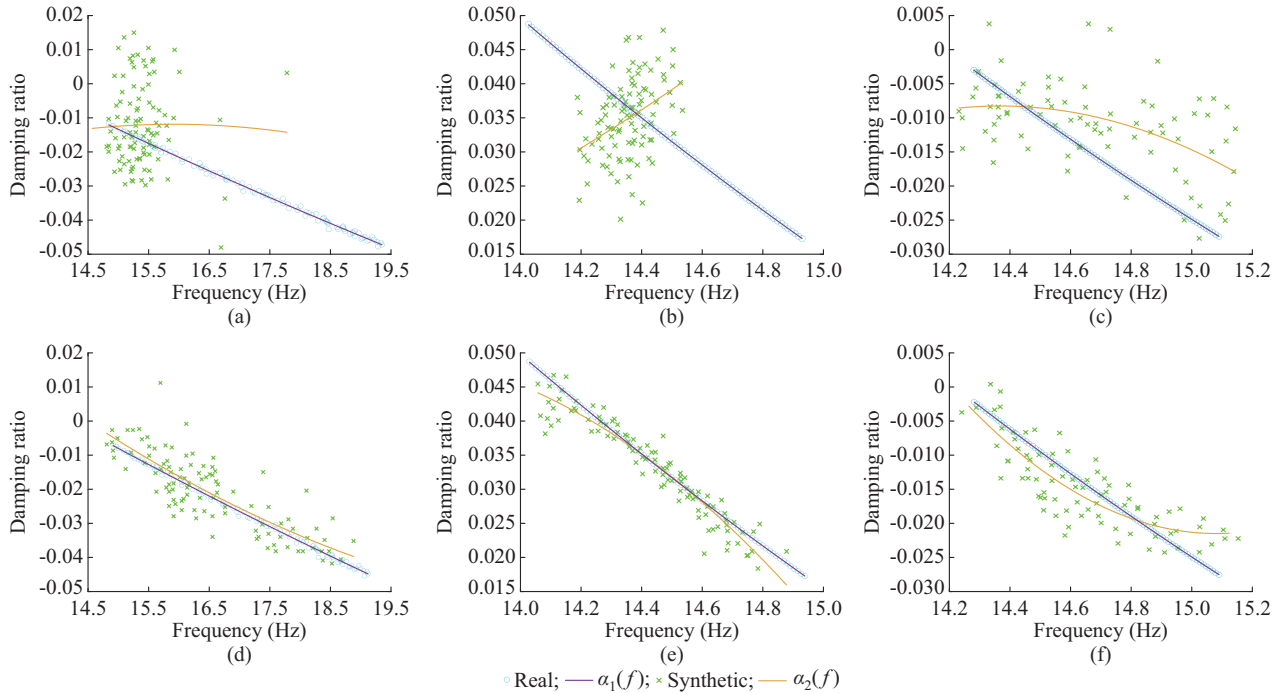


Fig. 14. Distribution of data and fitting results of Cases 1-3. (a) Case 1 without TL applied. (b) Case 2 without TL applied. (c) Case 3 without TL applied. (d) Case 1 with TL applied. (e) Case 2 with TL applied. (f) Case 3 with TL applied.

$$Accuracy = \frac{TP + TN}{P + N} \quad (17)$$

where P and N are the numbers of negatively damped oscillation and forced oscillation samples, respectively; TP is the number of correctly predicted negatively damped oscillation

samples; and TN is the number of correctly predicted forced oscillation samples.

The synthetic data generated by the BiLSTM-WGAN with MBD layer are used to expand the real dataset with different proportions, and the classification accuracies of the RF mod-

el before and after augmenting the dataset with varying proportions of synthetic data are compared as an evaluation index of the quality of the synthetic data. Table III shows the comparison of the classification accuracy with different sample expansion proportions.

TABLE III
COMPARISON OF CLASSIFICATION ACCURACY WITH DIFFERENT SAMPLE EXPANSION PROPORTIONS

Sample expansion proportion (%)	Classification accuracy (%)
0	87.5
30	89.0
60	91.5
100	92.1

As can be observed from Table III, after augmenting the small-sample data of Cases 2 and 3 by the proposed approach in this paper, the classification accuracies of the RF model on the test set are higher than those generated by the proposed approach in this paper. When the expansion ratio reaches 100%, the classification accuracy reaches the maximum value of 92.1%, which is 4.6% higher than the real dataset.

VI. CONCLUSION

In this paper, a TL-based BiLSTM-WGAN approach for synthetic data generation of SSO in wind farms is proposed. The proposed approach enriches the temporal feature extraction capacity of GANs through the incorporation of a BiLSTM layer, coupled with the integration of a MBD layer and GP term to enhance model performance in few-shot learning scenarios. To address the challenge of discrepancies between actual and simulation models, a limited dataset sourced from the actual model is harnessed for TL purposes. Case studies illustrate that the proposed approach can efficiently generate brand-new samples, resolving issues related to few-shot learning and data scarcity in SSO research at the data level. In addition to solving the problem of data scarcity for machine learning models, synthetic data generation is also expected to address privacy protection concerns, such as potential leaks of user's behavior or grid topology and parameters.

The proposed approach is highly generalizable and can be adapted to other domains within the power system for time-series data generation, effectively addressing the issue of data scarcity. Further research will discuss the use of conditional generative adversarial networks (CGANs) to generate modal-specific SSO data with the addition of constraints, or replacing LSTM with temporal convolutional network (TCN) to mitigate potential gradient vanishing issues in LSTM networks when processing high-frequency oscillation signals.

REFERENCES

- [1] R. N. Damas, Y. Son, M. Yoon *et al.*, "Subsynchronous oscillation and advanced analysis: a review," *IEEE Access*, vol. 8, pp. 224020-224032, Dec. 2020.
- [2] K. Gu, F. Wu, and X. Zhang, "Sub-synchronous interactions in power systems with wind turbines: a review," *IET Renewable Power Generation*, vol. 13, no. 1, pp. 4-15, Sept. 2018.
- [3] J. Zheng, B. Li, Q. Chen *et al.*, "HPF-LADRC for DFIG-based wind farm to mitigate subsynchronous control interaction," *Electric Power Systems Research*, vol. 214, p. 108925, Jan. 2023.
- [4] J. Bialek, "What does the GB power outage on 9 August 2019 tell us about the current state of decarbonised power systems?" *Energy Policy*, vol. 146, p. 111821, Aug. 2020.
- [5] Y. Cheng, L. Fan, J. Rose *et al.*, "Real-world subsynchronous oscillation events in power grids with high penetrations of inverter-based resources," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 316-330, Jan. 2023.
- [6] H. Liu, X. Xie, J. He *et al.*, "Subsynchronous interaction between direct-drive PMSG based wind farms and weak AC networks," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4708-4720, Nov. 2017.
- [7] N. P. W. Strachan and D. Jovcic, "Stability of a variable-speed permanent magnet wind generator with weak AC grids," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2779-2788, Oct. 2010.
- [8] Y. Li, L. Fan, and Z. Miao, "Wind in weak grids: low-frequency oscillations, subsynchronous oscillations, and torsional interactions," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 109-118, Jan. 2020.
- [9] X. Xie, X. Zhang, H. Liu *et al.*, "Characteristic analysis of subsynchronous resonance in practical wind farms connected to series-compensated transmissions," *IEEE Transactions on Energy Conversion*, vol. 32, no. 3, pp. 1117-1126, Sept. 2017.
- [10] M. S. Annakkage, C. Karawita, and U. D. Annakkage, "Frequency scan-based screening method for device dependent sub-synchronous oscillations," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1872-1878, May 2016.
- [11] S. K. Jain and S. N. Singh, "Exact model order esprit technique for harmonics and interharmonics estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 7, pp. 1915-1923, Jul. 2012.
- [12] L. Yang, Z. Xu, J. Østergaard *et al.*, "Oscillatory stability and eigenvalue sensitivity analysis of a DFIG wind turbine system," *IEEE Transactions on Energy Conversion*, vol. 26, no. 1, pp. 328-339, Mar. 2011.
- [13] W. Du, X. Chen, and H. Wang, "A method of open-loop modal analysis to examine the SSOs in a multi-machine power system with multiple variable-speed wind generators," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4297-4307, Jul. 2018.
- [14] Y. Meng, Z. Yu, N. Lu *et al.*, "Time series classification for locating forced oscillation sources," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1712-1721, Mar. 2021.
- [15] T. Huang, N. M. Freris, P. R. Kumar *et al.*, "A synchrophasor data-driven method for forced oscillation localization under resonance conditions," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3927-3939, Sept. 2020.
- [16] S. Feng, J. Chen, Y. Ye *et al.*, "A two-stage deep transfer learning for localisation of forced oscillations disturbance source," *International Journal of Electrical Power & Energy Systems*, vol. 135, p. 107577, Feb. 2022.
- [17] S. K. Azman, Y. J. Isbeih, M. S. E. Moursi *et al.*, "A unified online deep learning prediction model for small signal and transient stability," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4585-4598, Nov. 2020.
- [18] Y. Shen, W. Yao, J. Wen *et al.*, "Adaptive supplementary damping control of VSC-HVDC for interarea oscillation using GrHDP," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1777-1789, Mar. 2018.
- [19] R. A. de Oliveira and M. H. J. Bollen, "Deep learning for power quality," *Electric Power Systems Research*, vol. 214, p. 108887, Jan. 2023.
- [20] J. Duan, Y. He, and X. Wu, "A space hybridization theory for dealing with data insufficiency in intelligent power equipment diagnosis," *Electric Power Systems Research*, vol. 199, p. 107363, Oct. 2021.
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139-144, Oct. 2020.
- [22] S. E. Kababji and P. Srikantha, "A data-driven approach for generating synthetic load patterns and usage habits," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4984-4995, Nov. 2020.
- [23] A. Harell, R. Jones, S. Makonin *et al.*, "TraceGAN: synthesizing appliance power signatures using generative adversarial networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4553-4563, Sept. 2021.
- [24] L. Song, Y. Li, and N. Lu, "ProfileSR-GAN: a GAN based super-resolution method for generating high-resolution load profiles," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3278-3289, Jul. 2022.

- [25] Y. Hu, Y. Li, L. Song *et al.*, "MultiLoad-GAN: a GAN-based synthetic load group generation method considering spatial-temporal correlations," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 2309-2320, Mar. 2024.
- [26] Y. Wu, C. Lu, G. Wang *et al.*, "Partial discharge data augmentation of high voltage cables based on the variable noise superposition and generative adversarial network," in *Proceedings of 2018 International Conference on Power System Technology*, Guangzhou, China, Nov. 2018, pp. 3855-3859.
- [27] X. Zhu, K. Zhou, L. Lu *et al.*, "Partial discharge data augmentation based on improved Wasserstein generative adversarial network with gradient penalty," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6565-6575, May 2023.
- [28] X. Zheng, B. Wang, D. Kalathil *et al.*, "Generative adversarial networks-based synthetic PMU data creation for improved event classification," *IEEE Open Access Journal of Power and Energy*, vol. 8, pp. 68-76, Feb. 2021.
- [29] X. Zheng, A. Pinceti, L. Sankar *et al.*, "Synthetic PMU data creation based on generative adversarial network under time-varying load conditions," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 234-242, Jan. 2023.
- [30] S. Feng, K. Wang, J. Lei *et al.*, "Influences of DC bus voltage dynamics in modulation algorithm on power oscillations in PMSG-based wind farms," *International Journal of Electrical Power & Energy Systems*, vol. 124, p. 106387, Jan. 2021.
- [31] H. Ye, Y. Liu, P. Zhang *et al.*, "Analysis and detection of forced oscillation in power system," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1149-1160, Mar. 2017.
- [32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, Jul. 2017, pp. 214-223.
- [33] I. Gulrajani, F. Ahmed, M. Arjovsky *et al.*, "Improved training of Wasserstein GANs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, USA, Dec. 2017, pp. 5769-5779.
- [34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
- [35] Y. Yu, X. Si, C. Hu *et al.*, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235-1270, Jul. 2019.
- [36] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602-610, Jul. 2005.
- [37] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325-338, Apr. 2019.
- [38] T. Salimans, I. Goodfellow, W. Zaremba *et al.*, "Improved techniques for training GANs," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Red Hook, USA, Dec. 2016, pp. 2234-2242.
- [39] F. Zhuang, Z. Qi, K. Duan *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, Jan. 2021.
- [40] T. Rakthanmanon, B. Campana, A. Mueen *et al.*, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining*, New York, USA, Aug. 2012, pp. 262-270.
- [41] H. Li, J. Wan, S. Liu *et al.*, "Wetland vegetation classification through multi-dimensional feature time series remote sensing images using Mahalanobis distance-based dynamic time warping," *Remote Sensing*, vol. 14, no. 3, p. 501, Jan. 2022.
- [42] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984-995, Jul. 1989.
- [43] A. H. A. El-Kareem, M. A. Elhameed, and M. M. Elkholy, "Effective damping of local low frequency oscillations in power systems integrated with bulk PV generation," *Protection and Control of Modern Power Systems*, vol. 6, no. 41, Dec. 2021.
- [44] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197-227, Apr. 2016.
- [45] S. Feng, J. Chen, and Y. Tang, "Identification of low frequency oscillations based on multidimensional features and ReliefF-mRMR," *Energies*, vol. 12, no. 14, p. 2762, Jan. 2019.

Shuang Feng received the B.Sc. degree in electrical engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2012, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2017. She is currently an Associate Professor at the Department of Electrical Engineering, Southeast University. Her research interests include power system stability and control, and application of artificial intelligence in power systems.

Zhirui Zhang received the B.Sc. degree in electrical engineering from Nanjing University of Science and Technology, Nanjing, China, in 2023. He is currently pursuing the M.S. degree at the Department of Electrical Engineering, Southeast University, Nanjing, China. His research interest includes application of artificial intelligence in power systems.

Yuhang Zheng received the M.Sc. degree in software engineering from Southeast University, Suzhou, China, in 2023. His research interest includes application of artificial intelligence in power systems.

Jiaying Lei received the B.Sc. and Ph.D. degrees in electrical engineering from College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2012 and 2017, respectively. He is currently an Associate Professor at the Department of Electrical Engineering, Southeast University, Nanjing, China. His research interests include precision switching-mode power amplifier, high-performance topology, and control of power converters.

Yi Tang received the B.S., M.Eng., and Ph.D. degrees in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2000, 2002, and 2006, respectively. He is currently a Professor with the School of Electrical Engineering, Southeast University, Nanjing, China. His research interests include power system analysis, renewable energy, and application of artificial intelligence in power systems.