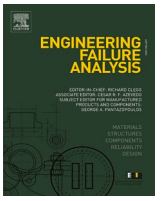
Engineering Failure Analysis 159 (2024) 108115

Contents lists available at ScienceDirect

Engineering Failure Analysis

journal homepage: www.elsevier.com/locate/engfailanal



Application of generative AI-based data augmentation technique in transformer winding deformation fault diagnosis

Yu Chen [a,b], Zhongyong Zhao [a,c,*], Jiangnan Liu [d], Shan Tan [a], Changqing Liu [a]

[a] College of Engineering and Technology, Southwest University, Chongqing 400716, China
[b] State Key Laboratory of Advanced Electromagnetic Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[c] Yibin Academy of Southwest University, Southwest University, Sichuan 644000, China
[d] Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd, Guangzhou 510620, China

ARTICLE INFO

Keywords:
Conditional-WGAN-GP
Transformer winding
Frequency response analysis
Fault diagnosis
Data augmentation

ABSTRACT

Accurately diagnosing various winding deformation faults is significant in power transformer maintenance. Among the different fault diagnosis methods, frequency response analysis (FRA) is widely used but still poses challenges. Artificial intelligence (AI)-based methods have recently been proposed to interpret FRA data. Nevertheless, these approaches are either complicated or exhibit limited generalization performance due to real-world FRA fault data scarcity. Inspired by AI-generated content (AIGC), this study proposes a data augmentation technique named conditional Wasserstein generative adversarial network with gradient penalty (Conditional-WGAN-GP) combined with fault diagnosis model. Numerous FRA-based data are automatically generated using the proposed data augmentation technique based on real FRA data obtained from a specially designed 10 kV transformer. The augmented dataset is then used to train fault diagnosis models to detect winding deformation faults. The trained fault diagnosis model is subsequently applied to assess two actual transformers. Experimental results demonstrate that when combined with the proposed method, even simpler fault diagnosis models can achieve high accuracy, exhibiting an improvement of approximately 5 % compared to the previous baseline model. The fault diagnosis models combined with the proposed data augmentation technique demonstrate improved generalization and robustness. (GitHub code: https://github.com/cy1034429432/Diagnosing-Transformer-Winding-Deformation-Fault-Types-from-FRA-Based-on-Conditional-WGAN-GP-/tree/main).

1. Introduction

The transformer is one of the crucial components of the power system, which is installed in all stages of power generation, transmission, substation, distribution, and consumption [1–5]. Since transformers are widely used in the power system, ensuring their stable operation plays a pivotal role in voltage transformation and energy transmission. With the rapidly increasing power system's capacity, the instantaneous short-circuit current of the power system is growing, and the power transformers are even more fault-prone [6]. Moreover, given the high cost of large power transformers, the timely detection of power transformer faults is critical. According to the International Council on Large Electric Systems (CIGRE), winding deformation faults account for 30 % of transformer faults.

* Corresponding author.
E-mail address: zhaozy1988@swu.edu.cn (Z. Zhao).

https://doi.org/10.1016/j.engfailanal.2024.108115
Received 29 August 2023; Received in revised form 21 January 2024; Accepted 9 February 2024
Available online 12 February 2024

Although minor winding deformation may not significantly affect the regular operation of the transformer, it can accumulate and lead to catastrophic failure if not addressed promptly. Therefore, when transformers are disconnected from the power system, an accurate diagnosing method during periodic inspections is necessary.

Several methods exist for diagnosing transformer winding deformation faults, which involve measuring different signals. For instance, the vibration method measures the vibration information of the transformer's core and winding [7,8]; the short circuit impedance method measures the variation of leakage inductance component of the transformer [9,10]; the low voltage impulse method measures the output impulse waveform of transformer winding [11]; frequency response analysis (FRA) method measures the frequency response characteristics of the equivalent two-port network of transformer winding [12,13], etc. The FRA is accurate, simple, economical, effective, and non-destructive [1,14]. Several FRA standards have been introduced, including IEC standard 60076–18 [15], CIGRE-342 [16], CIGRE-812 [17], IEEE Std C57.149 [18], Chinese power industry standard DL/T 911–2016 [19], etc. The traditional FRA uses mathematical-statistical indicators, such as the correlation coefficient (CC), to diagnose winding deformation faults. However, this threshold-based approach offers insufficient diagnostic accuracy and cannot directly distinguish between fault types.

Thus, other FRA interpretation methods have recently been proposed, which could be divided into two categories: (1) Establishing transformer equivalent circuit models for understanding the winding fault mechanisms. (2) Adopting data-driven techniques, including intelligent algorithms such as machine learning (ML) or deep learning (DL) for winding faults diagnosis.

In the equivalent circuit model-based diagnosis method, Ref. [20] establishes a circuit model of the transformer winding and calculates its parameters based on the finite element method (FEM), which provides a theoretical basis for interpreting winding radial deformation (RD) faults. In Ref. [21], the physical geometry of a single-phase transformer is simulated using three-dimensional FEM, which can be used to diagnose RD and disk space variation (DSV) faults. In addition to using FEM, the equivalent lumped high-frequency model can also be obtained through terminal FRA measurement and optimization algorithms. Moreover, some physical parameters of simple equivalent models can be extracted from frequency response measurements [17]. In summary, building the equivalent circuit model is suitable for understanding the fault mechanism of transformer winding deformation. Different transformers have diverse structural configurations and operating characteristics, affecting their equivalent circuit models. Therefore, an equivalent circuit model that accurately represents one type of transformer may not apply to a different kind. Consequently, researchers need to build various models according to diverse transformer structures. In summary, this approach requires prior knowledge of the transformer structure and much time to establish an equivalent circuit model, making it unsuitable for diagnosing winding faults.

In the data-driven-based method, data-driven models are used to classify transformer winding deformation fault types. Before using these models, their input should be carefully selected. Some FRA mathematical-statistical indicators are used as the evaluation criteria—for instance, Refs. [1,22–24] propose some indicators, such as Euclidean Distance (ED), Complex Distance (CD), etc., and corresponding thresholds. These indicators are proportional to the transformer winding deformation fault degree. Afterward, many researchers use various classifiers to diagnose fault types based on the calculated indicators, such as binary tree support vector machine (SVM) [25], SVM based on genetic algorithm (SVM-GA), SVM based on particle swarm optimization algorithm (SVM-PSO) [26], k-means based on grasshopper optimization algorithm [27]. There are some other Refs. [2,3,28–35] that do not use mathematical-statistical indicators. Instead, they directly use FRA data or images as the input [36,37]. In addition, many other methods exist depending on the input data types [23,38]. For data-driven models, collecting enough data is the first step. However, it is impractical to implement massive artificially made fault experiments on actual in-service transformers, which would be destructive, costly, and resource-wasteful. Moreover, the FRA fingerprints of the different transformers exhibit significant variations, making it nearly impossible to collect data with various faults at different degrees. Although numerous transformers are in the power system, none of the datasets is publicly accessible. In this context, obtaining a host of faulty transformers' FRA data has been challenging [32]. The insufficiency of faulty data will further hinder the intelligent process of transformer fault diagnosis.

Thus, for data-driven diagnostic methods, there exist the following limitations:

1) Due to the cumbersome experimental process, the collection of FRA data is difficult, resulting in a small dataset (Researchers in the Refs. [6,22,25,26,31,32,39–41] may refer to datasets of less than 100 samples for transformer winding diagnostics.), which leads to the trained model may not be reliable. Moreover, it prevents the use of some advanced data-driven techniques.
2) The FRA data insufficiency could result in an overfitting problem. The framework of diagnosis models and the modeling process is complicated. Still, the trained models have poor generalization performance, which cannot show high accuracy in detecting transformer windings except the experimental one.

Data augmentation techniques behind AI-generated content (AIGC) can be a viable solution. Data augmentation techniques have been successful in improving the performance of fault diagnosis models—for instance, Ref. [42] proposes a generative adversarial network (GAN)-based method that enhances the performance of fault diagnosis models by 2.07 %; Ref. [43] uses GAN to propose a power system stability assessment method by completing and recovering missing data; Ref. [44] presents a novel data augmentation method that enhances the model's performance in mechanical fault diagnosis by effectively handling noise labels. Data augmentation techniques can effectively improve the performance of fault diagnosis models by expanding the quantity and diversity of training data.

In contrast to conventional methods that heavily rely on advanced and complicated algorithms to improve accuracy, this study presents a novel data-centric method for diagnosing transformer winding deformation faults. The key contributions of this study are as follows:

1) Unlike traditional methods of generating image data, we present a suitable processing scheme for transformer FRA data incorporating data augmentation technology. This scheme expedites model training and allows for direct application across different transformers using the FRA-based mathematical-statistical indicators. Remarkably, we pioneer the exploration of data augmentation techniques explicitly tailored to transformer FRA data.
2) We propose a hybrid model named conditional Wasserstein generative adversarial network with gradient penalty (Conditional-WGAN-GP), which combines the existing conditional GAN (CGAN), WGAN-GP, and deep convolution GAN (DCGAN), designed explicitly for small FRA datasets. This model generates substantial fault data that conforms to the actual frequency response data distribution. By addressing the issue of limited fault data, this data augmentation technique enhances the performance of data-driven methods. Moreover, the proposed approach simplifies detection models in diagnosing winding deformation faults, demonstrating its robustness and generalizability compared to previous approaches.

The remainder of this study is organized as follows: Section II presents the methodology and model construction process. Section III provides experimental results. Section IV discusses this study's results. Section V presents the conclusions.

2. Methodology and model construction

2.1. Principle of FRA

FRA, initially proposed by Dick and Erven, involves the representation of a transformer winding as a cascaded ladder network when subjected to high-frequency excitation signals [1,20,21,41]. This network comprises fundamental components such as resistance, inductance, and capacitance. Typically, sinusoidal signals with varying frequencies (ranging from 1 kHz to 1000 kHz) are applied to the input terminal of the transformer winding. The corresponding response signal is then measured at the output terminal, deriving a frequency response curve. In the presence of a fault within the transformer winding, alterations in the structure and parameters of the equivalent circuit lead to changes in the FRA curve accordingly.

$$H_f(w) = 20\log\frac{|U_2(w)|}{|U_1(w)|} \tag{1}$$

where $H_f(w)$ is the frequency response curve while $U_1(w)$ and $U_2(w)$ are the winding's excitation and response voltage signals, and w represents different frequencies.

2.2. FRA-based transformer winding deformation fault experiment

A platform has been established for conducting FRA-based experiments using a specially designed 10 kV transformer with a nameplate value shown in Table 1, and more detailed information can be found in Ref. [26]. Three typical transformer winding deformation faults, including DSV, inter-disk short circuit (IDSC), and RD, are artificially made based on the experimental platform [6,40]. The experimental wiring diagram is depicted in Fig. 1, which shows the end-to-end open circuit connection meeting the wiring requirements of standards [15,16,18,19]. Fig. 1(a) shows the measurement wiring diagram, while Fig. 1(b) shows the actual wiring diagram. All experimental data are measured by a commercial FRA analyzer (Model: TDT6U, Manufacturer: Beijing Shengtai Real-Time Electric Technology Co., Output: 25 Vpp (maximum, adjustable), Output impedance: 50 Ω, Input impedance: 1 MΩ/50 Ω (optional), Frequency sweep range: 20 Hz to 2 MHz, Frequency accuracy: 0.005 %, Dynamic range: −120 dB ~ 20 dB, Gain accuracy: ±0.5 dB, Test speed: no more than 1 min (1–1000 kHz/1000 points)).

When measuring, the following guidelines [17,19] need to be followed:

Table 1
Nameplate value of specially designed transformer.

Parameters		Nominal value
Rated voltage (kV)		10/0.4
Capacity (kVA)		400
Rated current (A)		23/577
Frequency (Hz)		50
Number of phases		3
Connection type		Ynyn0
Tank (mm)		1705*740*1415
HV winding	Outer radius (mm)	250
	Inner radius (mm)	210.5
	Height (mm)	519
LV winding	Outer radius (mm)	174.5
	Inner radius (mm)	158
	Height (mm)	363
Iron core	Yoke radius (mm)	150
	Yoke height (mm)	1190
	Yoke length (mm)	1390

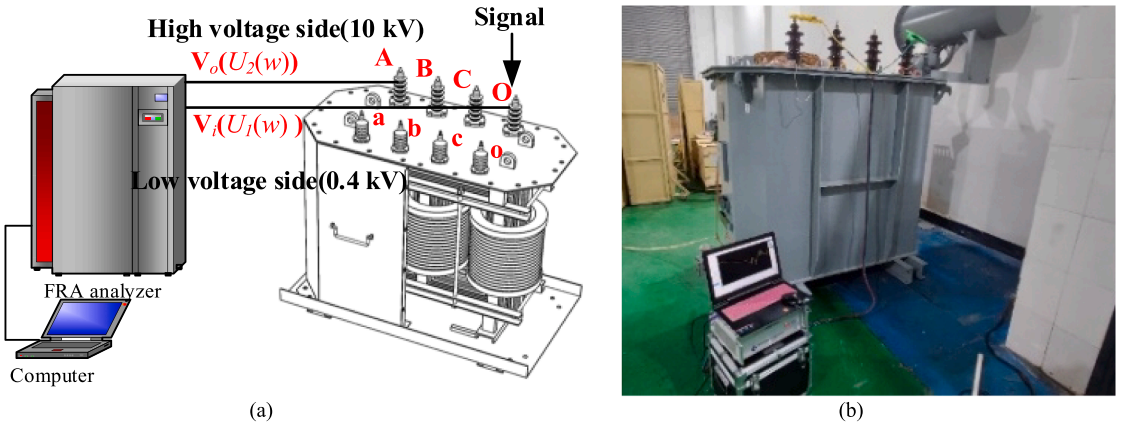


Fig. 1. Experimental measurement diagram. (a) measurement wiring diagram. (b) actual wiring diagram.

1) Each winding should be individually measured according to the wiring requirements, with the corresponding recording FRA curve. It is important to note that after any operational, repair, or maintenance history of the unit, such as general repairs affecting the inner geometry, bushing replacement, changes in the position of internal links (e.g., series versus parallel winding connections), or past through-faults, necessitate obtaining new benchmark FRA data.
2) To ensure measurement accuracy, it should maintain a short and straight connection between the screen of the measuring lead and the bushing flange. When using transformer earth as a signal reference, the earthing path should not introduce any tangible impedance into the circuit. Before conducting experiments, transformer winding necessitates the complete discharge and

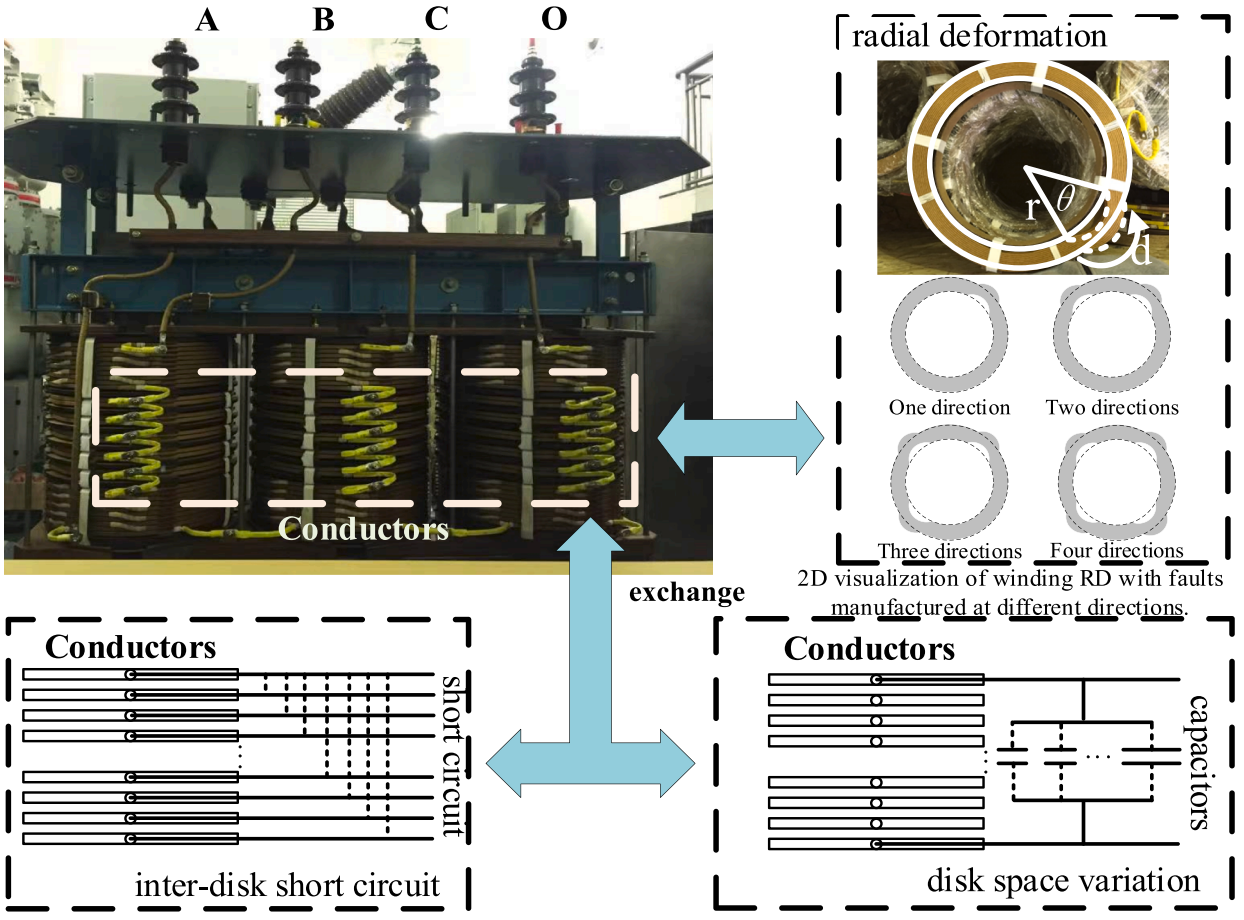


Fig. 2. Diagrammatic sketch of winding deformation faults.

demagnetization. Furthermore, maintenance personnel must inspect for any mismatches between the characteristic impedance of the measurement cables and the busbar, as well as ensure there are no poor connections. In addition, it is necessary to disconnect all leads connected to the end of the transformer bushing and keep them at a considerable distance from the experimental transformer.

3) The transformer winding FRA curves are easily affected by the position of the on-load tap changer. Measuring the transformer in the maximum tap position is advisable, or ensuring that the on-load tap-changer is in the same tap position for each experiment. The specially designed experimental transformer does not have the tap winding.

Fig. 2 and Table 2 illustrate the method of artificially made winding deformation, and please refer to Refs. [6,22,26,32,39,40] for more detail. Fig. 3(a)-(c) show typical FRA curves of DSV, IDSC, and RD faults. Compared to the normal situation, the low-frequency band exhibits minimal changes, while mid and high-frequency bands show significant changes for RD and DSV faults. In the case of IDSC fault, the FRA curves undergo a resonance point shift in the low-frequency band and show significant differences in the mid and high-frequency bands. These trends align with the results presented in Refs. [4,20,21,25]. In addition, Fig. 4 shows the FRA curves of one disk with 7 % and 10 % RD faults, as well as those obtained from repeated experiments. As repeated experiments are almost identical and do not benefit the generalization performance of data-driven models, only one set of FRA data is collected after replacing each faulty winding. It is important to note that replacing a faulty winding takes one day, and collecting massive faulty FRA data is challenging due to the complex, destructive, and time-consuming nature of experimental settings. Therefore, 53 FRA samples are obtained, including 21 DSV faults, 17 IDSC faults, and 15 RD faults.

2.3. Principle of Conditional-WGAN-GP

Proposed by Ian Goodfellow in 2014 [45]. GAN is a data augmentation technology that employs two neural networks (generator and discriminator). Through competition, GAN generates data resembling real data distribution. The generator in GAN converts Gaussian noise into a distribution similar to the real data distribution. The loss function L_G is as follows:

$$L_G = -\underset{z\sim p_z(z)}{E}(D(G(z))) \tag{2}$$

where $E(\cdot)$ represents the expected function; $G(\cdot)$ represents the generator function; $D(\cdot)$ represents the discriminator function; p_z(z) represents the distribution of Gaussian noise z.

The discriminator takes both real data and data generated by the generator as input, accurately distinguishing between the real and generated data. The loss function L_D is defined as:

$$L_D = -\left(\underset{x\sim p_{data}(x)}{E}(\log D(x)) + \underset{z\sim p_z(z)}{E}\log\left(1 - D\left(G\left(z\right)\right)\right)\right) \tag{3}$$

where $p_{data}(x)$ represents the distribution of the real data x.

However, GANs often suffer from training instability. To address this issue, researchers then proposed WGAN [46]. Although WGAN has shown improved stability during training, it may still generate poor samples and face convergence challenges [47]. This is because WGAN utilizes weight clipping to enforce the Lipschitz constraint on the discriminator, which can lead to undesirable training behaviors [47]. Instead, WGAN-GP uses another strategy called gradient penalty, penalizing the discriminator with two-norm of $\hat{x}$ gradients ($\hat{x}$ is a uniform sample on the line between the real and generated data). By incorporating the gradient penalty into the loss function, the training of WGAN becomes more stable and produces higher-quality generations. The proposed method involves generating data based on fault-type labels. Thus, this study combines CGAN [48] and WGAN-GP to propose Conditional-WGAN-GP. The loss functions $\widetilde{L}_G$ and $\widetilde{L}_D$ are as follows:

$$\widetilde{x} = G(z, \widetilde{y}) \tag{4}$$

$$\widetilde{L}_D = \underset{\widetilde{x}\sim p_g(\widetilde{x})}{E}(D(\widetilde{x})) - \underset{x\sim p_{data}(x)}{E}(D(x,y)) + \rho\underset{\hat{x}\sim p_{\hat{x}}}{E}\left(\left(\|\nabla_{\hat{x}}D(\hat{\mathrm{x}})\|_2 - 1\right)^2\right) \tag{5}$$

Table 2
The method of artificially made winding faults.

Fault type	Artificially made winding faults	Number
DSV	The influence of DSV fault on a transformer winding is simulated by connecting conductors of several disks with paralleled capacitors, where the ratio of parallel capacitance value to equivalent disk capacitance of winding (2nf [6]) represents the degree of DSV fault ranging from 5 % to 40 % in Fig. 3(a).	21
IDSC	Shorting conductors between different disks can simulate various IDSC faults. The notation "#1-#2" in Fig. 3(b) denotes that the first and second connectors of the middle sequential twist windings, viewed from top to bottom, are short-circuited.	17
RD	The transformer factory produces RD fault windings with varying degrees and directions of deformation to simulate RD faults, depicted in Fig. 2. In Fig. 2, r is the winding radius, d is the variable radial deformation, and θ represents the corresponding angle of the deformed part. The ratio of d to r indicates the degree of RD fault in Fig. 3(c).	15

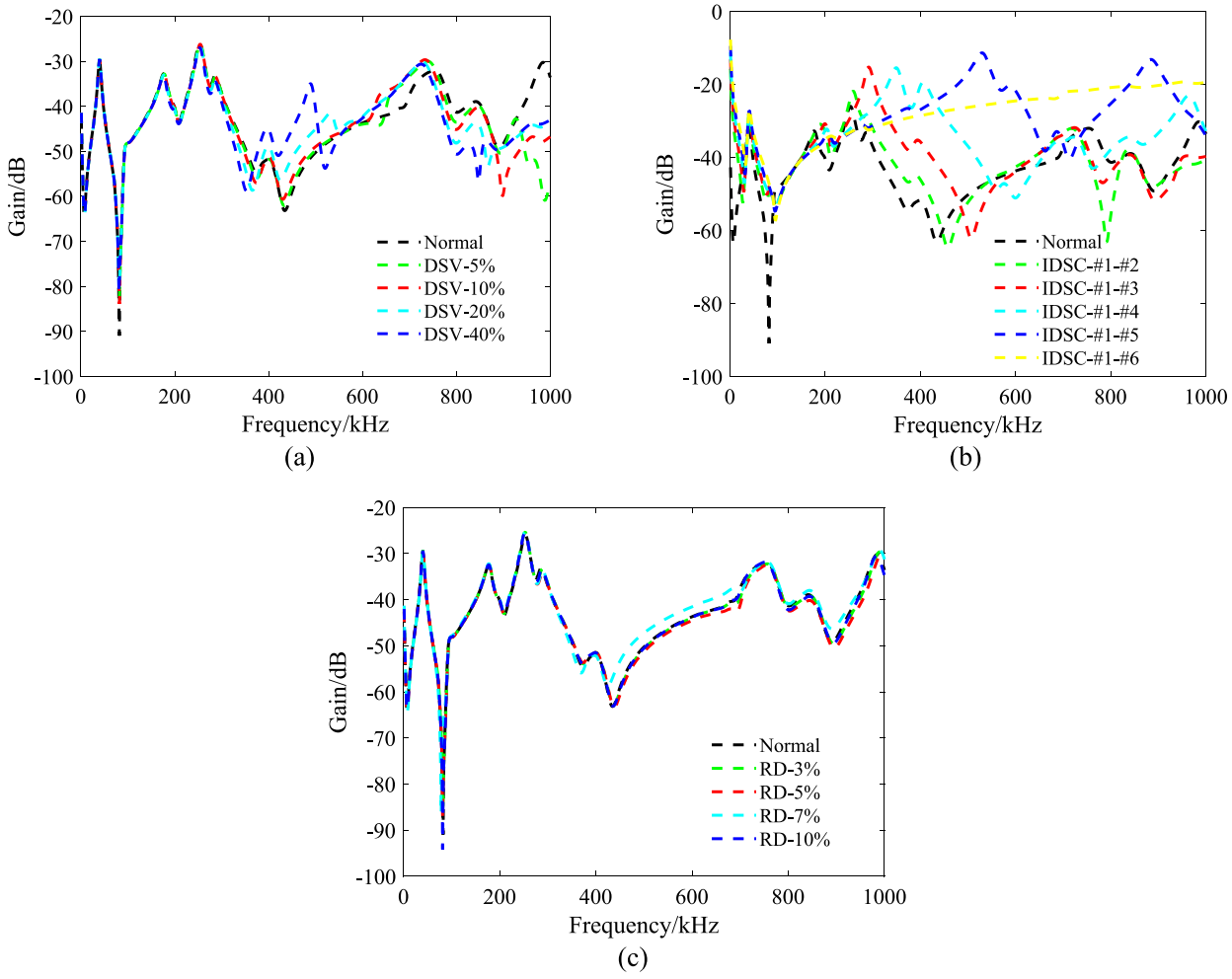


Fig. 3. Several typical FRA curves of winding faults. (a) DSV faults. (b) IDSC faults. (C) RD faults.

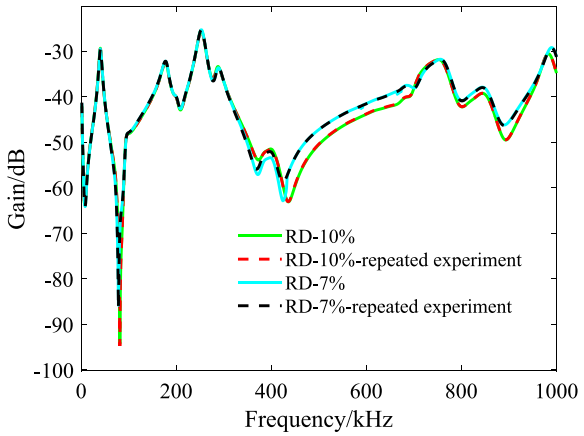


Fig. 4. FRA curves under repeated experiments.

$$\widetilde{L}_G = -E(D(G(z,\widetilde{y}))) \tag{6}$$

where z is the Gaussian noise, $\widetilde{y}$ is the label required to generate data, $\widetilde{x}$ is the generated data, x is the real data, y is the real label, $\widehat{x}$ is a uniform sample on the line between the real and generated data, $|\nabla D(\bullet)\|_2$ represents the two-norm of the gradient; $\widetilde{x} \sim p_g(\widetilde{x})$, $x \sim p_{data}(x)$ and $\widehat{x} \sim p_{\widehat{x}}$ represent the distribution of the corresponding data. In the formula (5), $\underset{\widehat{x}\sim p_{\widehat{x}}}{E}\left(\left(\|\nabla_{\widehat{x}} D(\widehat{x})\|_2 - 1\right)^2\right)$ is the gradient penalty term of WGAN-GP, and ρ is a hyperparameter, usually taken as 10. The specific backbone models of the discriminator and generator will be presented in the following subsection.

2.4. Backbone model construction combined with FRA

Typically, GAN is used for image generation, so this study initially attempted to generate the FRA curve image. Fig. 5(a) illustrates the input FRA image, and this study uses the CGAN and original DCGAN [48,49] with the loss functions (5)~(6). An example of the FRA image produced by the generator is displayed in Fig. 5(b). It is evident from Fig. 5(b) that the quality of the generated image is low and does not differ from the real FRA image, rendering it unsuitable as training data for the fault diagnosis model. The root causes are: (1) Resizing the input FRA images to 64*64 pixels during preprocessing leads to blurring of the generated image. (2) The limited real dataset makes it ineffective to update the model with larger parameters, and using a deeper model may result in overfitting, which generates data that closely resembles the real data. Furthermore, the limited reference significance hinders the generalizability of diagnosis models trained by these images for different transformers.

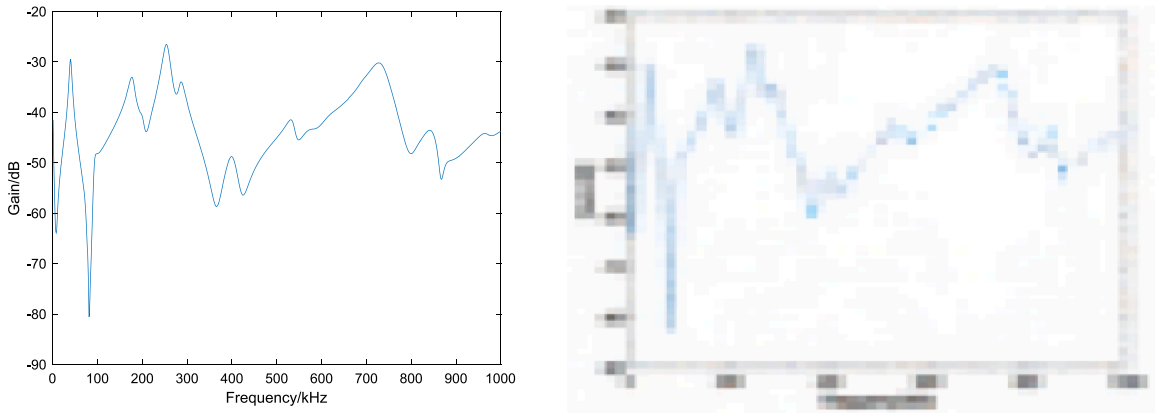
Hence, this study utilizes FRA-based mathematical-statistical indicators as input for the discriminator and generator, which offers the following advantages: (1) The mathematical-statistical indicators effectively measure the difference between the measured FRA curve and the reference one. The FRA standards, such as IEC 60076–18 and Chinese standard DL/T911-2016 [15,19], derive threshold values by analyzing numerous transformer cases with winding deformation faults. While FRA curves may vary across transformers, the changing trends of mathematical-statistical indicators for the same fault type remain consistent, thus overcoming poor generalization observed in FRA image-based fault diagnosis. (2) Regarding training, the indicators' dimension is much smaller than that of images, making it more suitable for few-shot learning using small networks [31]. This study summarizes some common mathematical-statistical indicators, as shown in Table 3, and they are shown in formulas (7)~(22).

$$CC = \frac{\sum_{w=1}^{N} H_{f1}(w) H_{f2}(w)}{\sqrt{\sum_{w=1}^{N} |H_{f1}(w)|^2 \sum_{f=1}^{N} |H_{f2}(w)|^2}} \tag{7}$$

$$ASLE = \frac{\sum_{w=1}^{N} |20\log_{10} H_{f1}(w) - 20\log_{10} H_{f2}(w)|}{N} \tag{8}$$

$$DABS = \frac{\sum_{w=1}^{N} |H_{f1}(w) - H_{f2}(w)|}{N} \tag{9}$$

$$ED = \sqrt{\sum_{w=1}^{N} \left(H_{f2}(w) - H_{f1}(w)\right)^2} \tag{10}$$



(a) (b)

Fig. 5. FRA images generated by GAN. (a) Input FRA image. (b) Generated FRA image.

Table 3
The mathematical-statistics indicators.

Abbreviation	Definition	Description	References
CC	Correlation coefficient	The most common evaluation indicator, on which most current standards are based, is to determine the severity of the winding faults.	[15,19,35]
ASLE	Absolute sum of logarithmic error	ASLE has the best relationship with the visual changes in the traces, which overcomes weak sensitivities to variations around the trough points.	[1,13]
DABS	Absolute difference	Other indicators do not show obvious changes in some frequency bands compared to DABS, demonstrating changes in all ranges.	[1,50]
ED	Euclidean distance	ED against noise, changes in the resistors due to temperature variations, differences in cable lengths, variances in measurement parameters, and differences in grounding practice.	[51]
E	Expectation	E allows a difference between two FRA curves due to noise.	[1]
Delta	Spectrum deviation	The same as E.	[1]
SSRE	Sum squared ratio error	SSRE follows the failure progression by demonstrating different sensitivity towards intermediary conditions.	[52]
MM	Minimum-maximum ratio	It's the same as SSRE.	[52]
Rou	Normalized correlation coefficient	Rou shows monotonic behaviors toward axial and radial shifts, while their behaviors are similar in the reverse direction.	[1]
SD	Standard deviation	SD shows differences when RD occurs in different directions.	[53]
CSD	Comparative standard deviation	CSD is more reliable than other indicators.	[24]
SDA	Standardized difference area	SDA measures the severity of the winding faults.	[1]
ID	Integral of difference	IA helps to detect the location of winding faults.	[1,24]
RMSE	Root mean square error	RMSE measures the degree of difference in regression algorithms.	[1,24]
IA	Integral of Absolute difference	IA helps to detect mechanical defects.	[1,24]
SPD	Stochastic spectrum deviation	SPD helps to detect aging phenomena.	[1,24]

$$Delta = \frac{1}{N}\sum_{w=1}^{N}\sqrt{\left(\frac{H_{f1}(w) - a/2}{a}\right)^2 + \left(\frac{H_{f2}(w) - a/2}{a}\right)^2} \tag{11}$$

$$a = H_{f1}(w) + H_{f2}(w)$$

$$MM = \frac{\sum_{w=1}^{N}\min\left(H_{f1}(w), H_{f2}(w)\right)}{\sum_{w=1}^{N}\max\left(H_{f1}(w), H_{f2}(w)\right)} \tag{12}$$

$$E[\Delta] = \frac{1}{N}\sum_{w=1}^{N}\Delta(w), \Delta(w) = \frac{H_{f2}(w) - H_{f1}(w)}{1/N\sum_{w=1}^{N}H_{f1}(w)} \tag{13}$$

$$Rou = \frac{\sum_{w=1}^{N}\widetilde{H}_{f1}(w)\widetilde{H}_{f2}(w)}{\sqrt{\sum_{w=1}^{N}|\widetilde{H}_{f1}(w)|^2\sum_{w=1}^{N}|\widetilde{H}_{f2}(w)|^2}} \tag{14}$$

$$SD = \sqrt{\frac{\sum_{w=1}^{N}\left(H_{f1}(w) - H_{f2}(w)\right)}{N-1}} \tag{15}$$

$$SSRE = \frac{\sum_{w=1}^{N}\left(H_{f2}(w)/H_{f1}(w) - 1\right)^2}{N} \tag{16}$$

$$CSD = \sqrt{\frac{\sum_{w=1}^{N}\left(\left(H_{f1}(w) - \overline{H}_{f1}(w)\right) - \left(H_{f2}(w) - \overline{H}_{f2}(w)\right)\right)}{N-1}} \tag{17}$$

$$SDA = \frac{\int\left(H_{f1}(w) - H_{f2}(w)\right)dw}{\int H_{f1}(w)dw} \tag{18}$$

$$ID = \int\left(H_{f1}(w) - H_{f2}(w)\right)dw \tag{19}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{w=1}^{N}\left(\frac{|H_{f1}(w)| - |H_{f2}(w)|}{\frac{1}{N}\sum_{w=1}^{N}|H_{f1}(w)|}\right)^2} \tag{20}$$

$$IA = \int |H_{f1}(w) - H_{f2}(w)|dw \tag{21}$$

$$SPD = \frac{100}{N}\sum_{w=1}^{N}\left|\frac{H_{f1}(w) - H_{f2}(w)}{H_{f1}(w)}\right| \tag{22}$$

where w is the frequency and N is the sampling point;$H_{f2}(w)$, $\overline{H}_{f2}(w)$ and $\widetilde{H}_{f2}(w)$ are the FRA curve, the FRA average value, and the normalized FRA curve of the measured winding; $H_{f1}(w)$, $\overline{H}_{f1}(w)$ and $\widetilde{H}_{f1}(w)$ are the FRA curve, the FRA average value, and the normalized FRA curve of the reference or healthy winding. It should be noted that there are numerous FRA-based mathematical-statistical indicators, and it is not possible to list them all in this study. The selected indicators should measure FRA curves' correlation, offset, and stability. This study focuses on winding faults detection methodology, and the start-of-the-art selection of indicators can be referred to [1,24].

According to Refs. [1,22,26,32] and standards [15,16,19], this study divides the FRA curve into three sub-frequency bands, including 1 ~ 100 kHz (low-frequency band), 101 ~ 600 kHz (middle-frequency band), and 601 ~ 1000 kHz (high-frequency band). The mathematical-statistical indicators in (7)~(22) are calculated within these sub-frequency bands and then normalized, resulting in 48 dimensions of the input matrix after frequency division, as shown below:

$$Input = [CC_{LF}, CC_{MF}, CC_{HF}, ..., SPD_{LF}, SPD_{MF}, SPD_{HF}]_{1\times 48} \tag{23}$$

where the subscripts LF, MF, and HF represent the low-, middle-, and high-frequency bands, respectively.

Referring to CGAN [48], the backbone models incorporate Word Embedding for inputting fault-type labels. The backbone models of the generator and discriminator are constructed based on the DCGAN framework, replacing conventional CNN with one-dimensional CNN and traditional transposed convolution with one-dimensional transposed convolution [49]. Utilizing a CNN-based structure instead of fully connected layers reduces trainable parameters through weight sharing, improving training speed. Since the input data lacks sequence features, the RNN-based model is not considered. To address overfitting and the dataset size, backbone models with fewer parameters are deemed more suitable in this study. The entire backbone models are shown in Fig. 6, and the specific steps are shown in Algorithm 1, while all hyperparameters are shown in the given GitHub code. In addition, the software and hardware configurations are shown in Table 4.

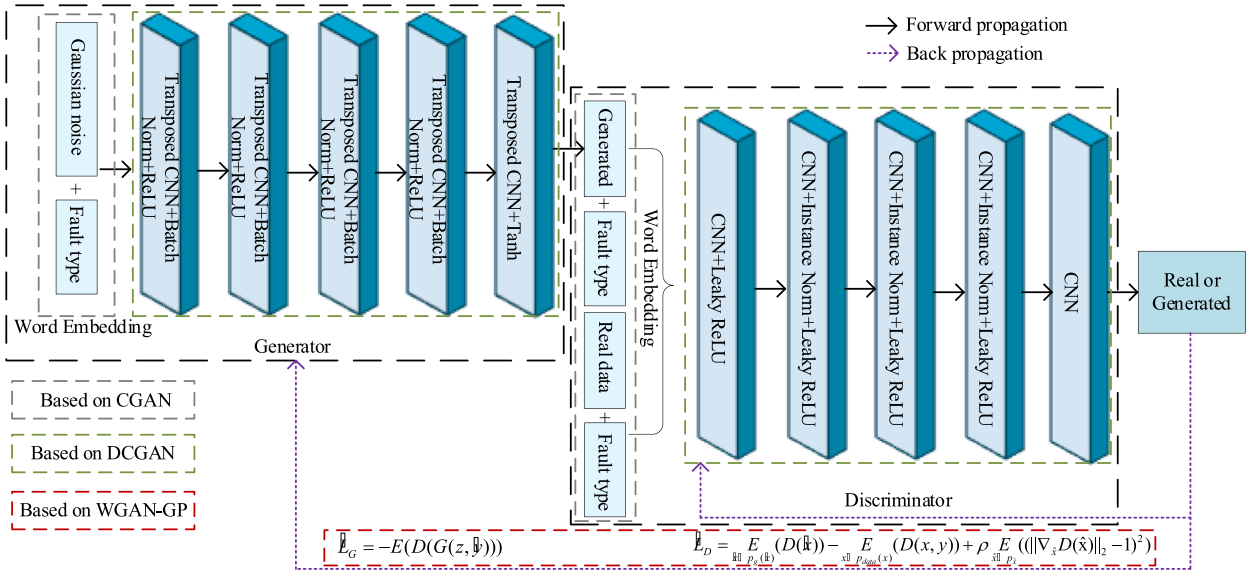


Fig. 6. The backbone models of discriminator and generator. Gaussian noise is 100 dimensions, fault type (Word Embedding) is 3 × 10 dimensions in the generator; Generated and real data are shown in (23), fault type (Word Embedding) is 3 × 30 dimensions in the discriminator, and the output of discriminator (Real or Generated) is 1 dimension. More information is presented on the GitHub code.

Algorithm 1. (*The main step of the proposed method for transformer winding deformation fault diagnosis. We use default values of* $\rho = 10$, $N_{discriminator} = 5$.)

Require: The initial discriminator D_0, the initial generator G_0, the learning algorithm L, the fault diagnosis model F, the gradient penalty ρ, the epoch T, the number of discriminator iterations per generator iteration $N_{\text{discriminator}}$, the real data x and corresponding fault label y, the uniform sample $\hat{x}$, the Gaussian noise z, the generated label $\tilde{y}$ and corresponding generated data $\tilde{x}$, the desired accuracy target Acc_{tar}, the accuracy of fault diagnosis model Acc_F, and the fault diagnosis model training set (x_{train}, y_{train}).

$x_{Norm} \leftarrow Norm(x)$ // The FRA-based mathematical-statistical indicators are calculated and then normalized.
$\theta \sim Gassian(0, 0.02),\ \theta \in G_0, D_0$ // Initialize the weights of the generator and discriminator.
$t = 1$
while $t \leqslant T$ or $Acc_F \leqslant Acc_{tar}$ **do**
  for $n = 1, \ldots, N_{\text{discriminator}}$ **do**
    $\tilde{x} \leftarrow G_{t-1}(z, \tilde{y})$ and get $\hat{x}$ // Use the generator to convert Gaussian noise to generated data and calculate the $\hat{x}$.
    $D_{5\times(t-1)+n} \overset{L}{\leftarrow} \tilde{L}_D(\tilde{x}, \hat{x}, x_{Norm}, y, \rho, D_{5\times(t-1)+n-1})$ // Use the formula (5) to update the current discriminator.
  end for
  $G_t \overset{L}{\leftarrow} \tilde{L}_G(z, \tilde{y}, G_{t-1})$ // Use the formula (6) to update the current generator.
  $x_{train} \leftarrow G_t(z, y_{train})$ // Use the trained generator to generate a training set for the fault diagnosis model.
  $F_t \leftarrow (x_{train}, y_{train})$ // Use generated training set to get fault diagnosis model.
  $Acc_F \leftarrow (F(x_{Norm}), y)$ // Use real data as the test set and obtain accuracy on real data.
  $t = t + 1$
end while
return the F

3. Experimental results

3.1. Fault diagnosis results

Machine learning is driven by three primary factors: computing power, algorithms, and data. Previous works [25,26,32] focused on fault diagnosis of transformer winding deformation without utilizing data augmentation techniques. Instead, researchers predominantly employed advanced algorithms, such as SVM-GA, SVM-PSO, multi-classification SVMs based on polar plot images, and binary tree SVM. This study uses six common SVMs (shown in Table 5) as fault diagnosis models, combined with the proposed data augmentation method, to detect winding deformation fault types. The generator produces 1000 samples for each fault, resulting in a total training set of 3000 samples. As mentioned in Section II, the test set comprises 53 samples of real experimental data.

Table 3 presents various FRA-based mathematical-statistical indicators. Fig. 7 shows the relationship between the number of indicators and the average performance of fault diagnosis models, indicating that using ten indicators is sufficient. This study recommends using CC, ASLE, DABS, ED, E, Delta, SSRE, MM, Rou, and SD as input. It is important to note that the results are obtained through 20 repeated experiments, and all subsequent experiments in this text utilize these ten indicators.

Table 5 presents the performance of six SVMs, while Fig. 8 shows the accuracy during the training process. Results in Table 5 and Fig. 8 show: (1) Using the data augmentation technique leads to higher performance, with an average accuracy improvement of 5 %, even when using simpler SVMs. (2) Simpler SVMs are more robust than complex fault diagnosis models like SVM-PSO and SVM-GA. (3) Improved SVMs with optimization algorithms typically require longer training time (>1min) compared to common SVMs (<5s).

To illustrate that data augmentation can generally improve the performance of data-driven fault diagnosis models, some typical classifiers, such as K-Nearest Neighbors (KNN), Tree, artificial neural network (ANN), and Bayesian classifier, are combined with the proposed data augmentation method. The results are shown in Table 6. From Table 6, it can be seen that: (1) The performance of data-driven models is improved significantly after using data augmentation, indicating their general applicability across various models. (2) The performance improvement of KNN is the most obvious. For a classifier like ANN, which requires a large amount of data as the training set, the data augmentation technique provides sufficient data for ANN, thus improving the model performance. In contrast, the performance improvements for classifiers like Tree and Bayesian classifiers, more suitable for small datasets, are less apparent due to inherent model characteristics. Since many researchers employ SVM-based models for transformer winding fault diagnosis [25,26,32], SVM remains the primary model in the subsequent sections.

The trained generator generates 100 samples for DSV, IDSC, and RD, respectively. These generated data are mixed with the real data and downscaled to 2 dimensions using the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm [54]. The visualization of all data is presented in Fig. 9. In Fig. 9, the real data are in green, and the generated data are in red. The generated data exhibit

Table 4
Hardware and software configuration.

Device	Configuration
CPU	Inter(R) Xeon(R) Gold 6268CL × 2
GPU	NVIDIA RTX A4000
RAM	128G
SOFTWARE	Python
PACKAGE	Numpy, Pytorch, Sklearn and Pandas

Table 5
Performance of common SVM-based diagnosis models.

Type	Kernal	Abbreviation	Best accuracy	Best recall
SVM	linear	SVM-1	0.9623	0.9608
NuSVM	rbf	SVM-2	1	1
LinearSVM	linear	SVM-3	1	1
SVM	poly	SVM-4	1	1
SVM	rbf	SVM-5	0.9623	0.9683
SVM	sigmoid	SVM-6	1	1

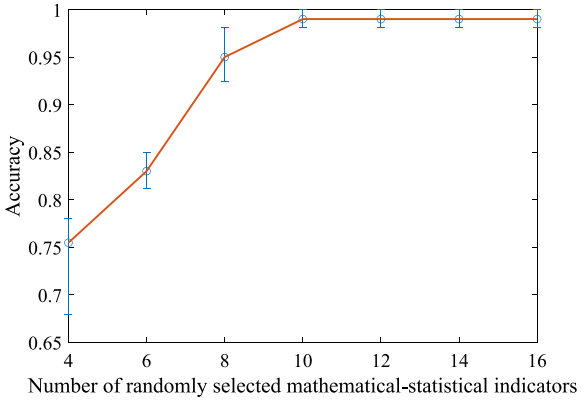


Fig. 7. The relationship between the number of selected indicators and the accuracy of the fault diagnosis model.

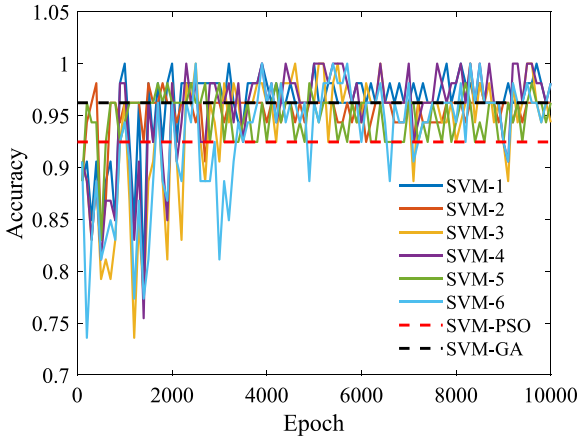


Fig. 8. Accuracy of the test set during training.

Table 6
Performance of typical machine learning classifiers.

With or without data augmentation technique	KNN	Tree	Bayesian classifier	ANN
Without (accuracy)	0.8867	0.9623	0.9433	0.9433
With (accuracy	1(↑)	0.9623(=)	0.9623(↑)	1(↑)

similarities to the real data but do not fully overlap, filling the gaps between the experimental representative data and generating related data across different degrees of fault. By incorporating diverse training data, this approach can effectively enhance data-driven fault diagnosis models' robustness and generalization ability. The t-SNE algorithm effectively extracts feature information for distinguishing between different fault types, aligning with the result of KNN in Table 6.

3.2. Ablation experiment

During the construction of the backbone models, two crucial network structures are found to be significant: the *Batch Norm* in the generator and the *Instance Norm* in the discriminator. Fig. 10 displays the recorded average accuracy changes of six SVMs on the test set during the training process. The results reveal that the *Instance Norm* enhances the stability of GAN while the *Batch Norm* accelerates the convergence of GAN. In addition, the parameter of Word Embedding is a tunable hyperparameter. This study suggests selecting 1/3 to 1 of the indicator's dimensions. A small size affects feature extraction of different fault types, while a large size dominates the weight update of the generator and discriminator. Table 7 displays the highest accuracy attained by multiplying the in_channels and out_-channels of both the generator and discriminator. The results indicate that larger networks do not produce better outcomes, likely due to the small dataset used in this study. Furthermore, larger networks increase trainable parameters, leading to a higher computational complexity. To demonstrate the superiority of using the CNN-based structure as the backbone model, this study replaces them with the linear layer, ensuring the same dimensional inputs and outputs as the CNN-based model. The statistical results are presented in Table 8, and the visualization of mixed data is displayed in Fig. 11. Using the linear layer as the main structure significantly reduces the performance, primarily due to the limited dataset size. Also, Fig. 11 shows the backbone models fail to extract features. In contrast, weight sharing is an inherent feature of the CNN-based structure, reducing trainable parameters and benefiting few-shot learning. In summary, the networks and Word Embedding parameters affect network performance, while the *Batch* and *Instance Norms* significantly impact the training process.

This study trains the generator and discriminator using GAN and WGAN loss functions to compare different loss functions under the same hyperparameter setting. Fig. 12(a)-(b) present the visualization results based on t-SNE and the test set accuracy during training with the GAN loss functions. From Fig. 12(a)-(b), it is evident that although the generator achieves better results on the test set, the generated data noticeably differs from the real data. This high accuracy is attributed to the smaller difference between the generated fault data and corresponding real data compared to non-corresponding data. Fig. 12(c)-(d) present the visualization results based on t-SNE and the test set accuracy during training with the WGAN loss functions. Fig. 12(c)-(d) illustrate that WGAN yields better results than GAN, with the generated data being more similar to real data. However, compared to WGAN-GP, WGAN exhibits instability during training, resulting in larger fluctuations in test set accuracy.

3.3. Case study

The first case study involves a 500 kV transformer manufactured in September 2016. Routine FRA experiments were conducted during the handover and recorded the FRA fingerprint. Following a fault occurrence in April 2021, the transformer was disconnected from the power system, and subsequent FRA experiments were carried out. Fig. 13(a) displays the FRA curves before and after the fault. Comparing the FRA curves before and after the fault, it is found that the FRA curves of the same phase in the low voltage (LV) winding are not similar, while the variation of the c-phase was most apparent. That is because the LV is triangularly connected, and three-phase windings would be affected. An IDSC fault was confirmed in the LV c-phase winding after disassembling the power transformer. After preprocessing the FRA data of the LV c-phase winding, the six trained SVMs all predict an IDSC fault, while the trained SVM-PSO model predicts an RD fault. The second case study is about a 110 kV transformer manufactured in December 2010, whose FRA fingerprints were recorded during the routine FRA experiments. In July 2017, the transformer was faulty for emergency repairs. The FRA experiments were carried out during the emergency repairs, and the corresponding FRA fingerprints were recorded. The FRA curves before and after the fault are shown in Fig. 13(b). The FRA data was pre-processed and inputted into six SVMs and SVM-PSO. The trained six SVMs all make the correct decision (DSV), while the SVM-PSO gives an IDSC fault. It should be noted that the related machines are actual large transformers that once operated in the power system. Based on the above results, fault diagnosis models combined with the proposed data augmentation method show better performance and greater generalizability, indicating potential for practical engineering applications.

4. Discussion

Currently, numerous works focus on detecting transformer winding deformation faults based on advanced classification models. In contrast to previous research, this study emphasizes enhancing performance through data-centric approaches. Based on existing FRA data, this study conducts related experiments based on well-known methods, and performance comparison is shown in Table 9. Without the use of classification models, using the FRA polar plot only can determine winding fault occurrences based on established thresholds. The dataset used in this study is currently too limited, rendering classic image or sequence classification models ineffective for realizing high-performance fault detection. These results unequivocally demonstrate that the proposed method has certain advantages in accuracy, generalization, speed, and intelligence.

While the proposed method works well in detecting transformer winding deformation faults, it has some shortcomings. In actual data collection, fewer faulty transformer winding data are available for diagnosis, typically only one set of samples per fault type rather than the more than ten samples used in this study. The method involves three tedious and complex steps: data processing, data augmentation, and fault diagnosis. A fault diagnosis model suitable for end-to-end few-shot learning should be proposed. Future efforts will focus on improving the experimental platform to obtain a wider range of fault data, including axial displacement (AD) and core deformation, and associating critical information such as the extent and location of faults with existing fault types.

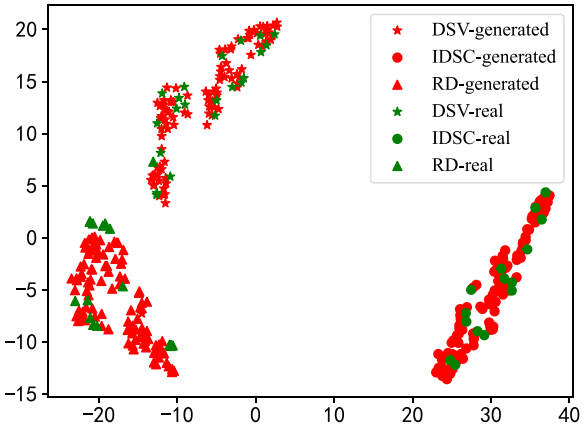


Fig. 9. Visualization result of generated fault and real data based on t-SNE.

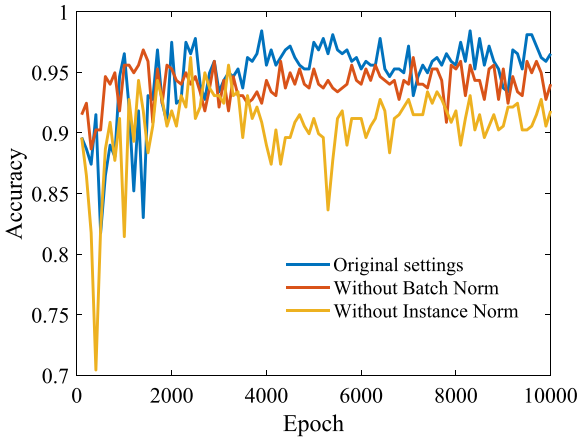


Fig. 10. Average accuracy without Batch Norm and Instance Norm.

Table 7
Performance statistics of fault diagnosis models under Conditional-WGAN-GP with different parametric multiples.

Type	Original	Channel*2	Channel*4	Channel*8	Channel*16
Total params	12,411	34,444	107,742	371,266	1,366,026
SVM-1 (acc)	1	0.9623	0.9623	0.9433	0.9245
SVM-2 (acc)	0.9623	1	1	0.9623	0.9623
SVM-3 (acc)	1	1	1	1	0.9623
SVM-4 (acc)	1	1	0.9623	0.9623	0.9623
SVM-5 (acc)	0.9623	0.9623	0.9623	0.9433	0.9433
SVM-6 (acc)	1	1	0.9623	0.9623	0.9433

Table 8
Performance statistics of fault diagnosis models under the linear layer.

Type	Same	Dimension*2	Dimension*4	Dimension*8	Dimension*16
Total params	108,869	395,749	1,504,037	5,858,725	23,120,549
SVM-1 (acc)	0.9623	0.9623	0.9623	0.9623	0.9245
SVM-2 (acc)	0.9623	0.9433	0.9433	0.9245	0.9433
SVM-3 (acc)	0.9623	0.9623	0.9623	0.9433	0.9245
SVM-4 (acc)	0.9634	0.9433	0.9433	0.9245	0.9245
SVM-5 (acc)	0.9634	0.9634	0.9623	0.9623	0.9433
SVM-6 (acc)	0.9434	0.9433	0.9433	0.9433	0.9433

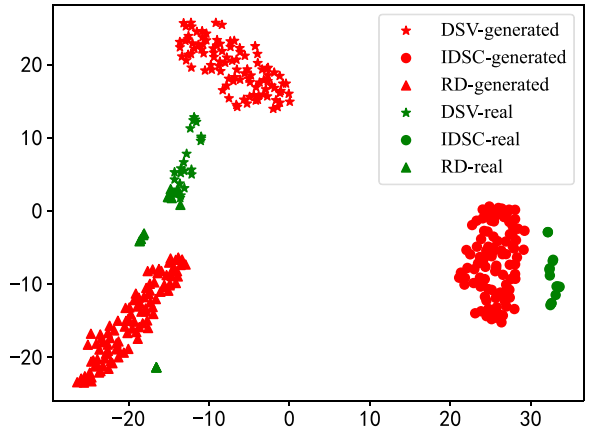


Fig. 11. Visualization result of the mixed data under the linear layer.

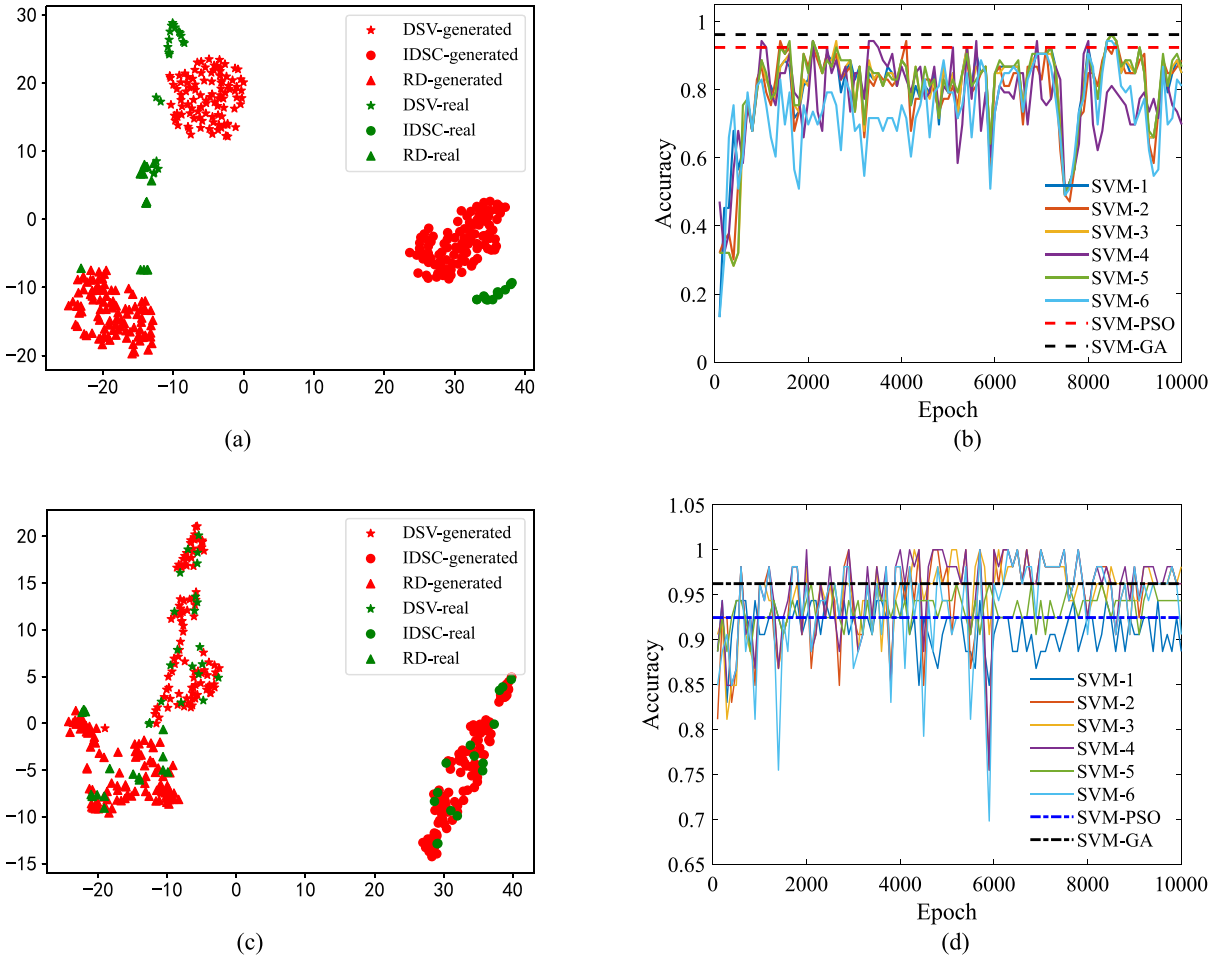


Fig. 12. Results under GAN and WGAN. (a) Visualization results under GAN based on t-SNE. (b) Accuracy of the test set under GAN. (c) Visualization results under WGAN based on t-SNE. (d) Accuracy of the test set under WGAN.

5. Conclusion

This study proposes a transformer winding deformation fault type diagnosis method based on data augmentation to solve the insufficiency of FRA fault samples and facilitate the generalization of data-driven models. The following conclusions are obtained

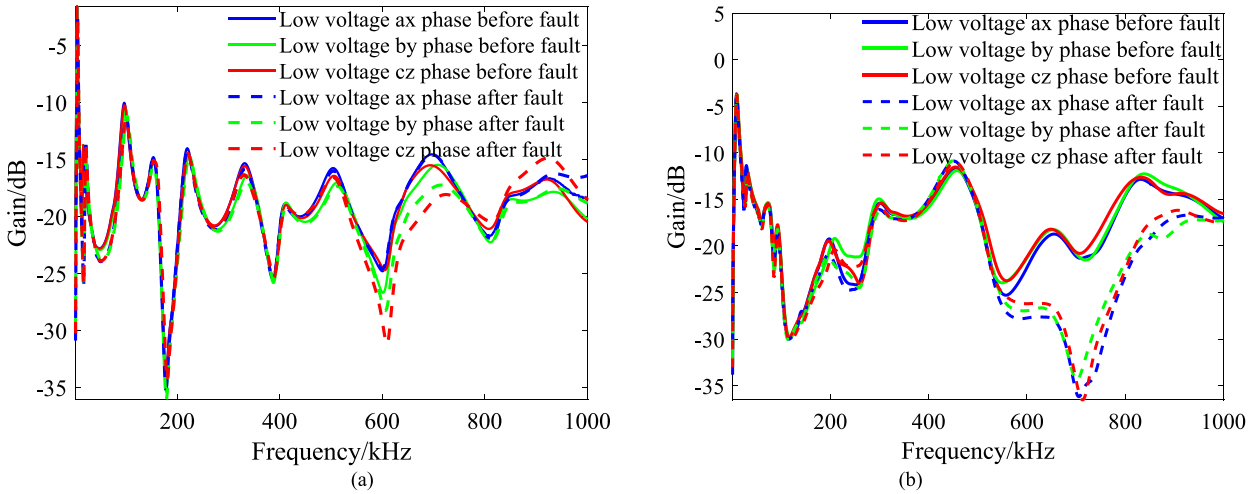


Fig. 13. FRA curves before and after the fault. (a) 500 kV transformer. (b) 110 kV transformer.

Table 9
Performance comparison with well-known methods.

Type	Method	Accuracy	Rapidity (rank)	Generalization (rank)	Intelligence	Whether can determine the fault types
Numerical indices combined with artificial intelligence	The proposed method	>95 %	2	1	√	√
Numerical indices	Polar plot	>75 %	1	2	×	×
Artificial intelligence	Image classification model	>50 %	3	3	√	√
Artificial intelligence	Sequence classification model	>50 %	3	3	√	√

based on the experimental results:

1) A hybrid data augmentation model named Conditional-WGAN-GP is proposed by fusing CGAN, DCGAN, and WGAN-GP. It can improve the performance of data-driven fault diagnosis models with FRA and enable simple models to exceed or reach the performance of complicated models, potentially improving the generalization and robustness of fault diagnosis models.
2) For small transformer winding fault datasets, mathematical-statistical indicators of FRA are more suitable than the FRA image when using data augmentation techniques. For the backbone models of representation learning, using a network structure with small parameters is beneficial for training small datasets. Compared with other structures, CNN-based models with weight sharing are more suitable for the limited data size. In addition, using the *Batch* and *Instance Norms* during the training of backbone models can make them converge faster and train more stably.
3) Compared to complicated fault diagnosis models, SVM-based models exhibit the expected accuracy with minimal training time. Although advanced models (such as ANN) can be used to replace SVM, the training time will be significantly improved, and the performance of advanced models has almost no apparent improvement compared to SVM.

CRediT authorship contribution statement

Yu Chen: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Zhongyong Zhao:** Writing – review & editing, Writing – original draft, Project administration, Funding acquisition. **Jiangnan Liu:** Writing – review & editing, Methodology, Investigation. **Shan Tan:** Writing – review & editing. **Changqing Liu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.
The raw/processed data required to reproduce these findings cannot be shared as the data also forms part of an ongoing study.

Acknowledgments

This work was supported by the Sichuan Science and Technology Program under Grant 2023NSFSC0829, by the Fundamental Research Funds for the Central Universities under Grant SWU-KT22027, and by the Venture and Innovation Support Program for Chongqing Overseas Returnees under Grant cx2019123.

References

[1] M.H. Samimi, S. Tenbohlen, FRA interpretation using numerical indices: State-of-the-art, Int. J. Electr. Power Energy Syst. 89 (2017) 115–125.
[2] Z. Zhao, C. Yao, C. Li, S. Islam, Detection of power transformer winding deformation using improved FRA based on binary morphology and extreme point variation, IEEE Trans. Ind. Electron. 65 (4) (2018) 3509–3519.
[3] A. Moradzadeh, H. Moayyed, B. Mohammadi-Ivatloo, G.B. Gharehpetian, A.P. Aguiar, Turn-to-turn short circuit fault localization in transformer winding via image processing and deep learning method, IEEE Trans. Ind. Inf. 18 (7) (2022) 4417–4426.
[4] M. Mahvi, V. Behjat, H. Mohseni, Analysis and interpretation of power auto-transformer winding axial displacement and radial deformation using frequency response analysis, Eng. Fail. Anal. 113 (2020) 104549.
[5] N.S. Beniwal, D.K. Dwivedi, H.O. Gupta, Life estimation of distribution transformers considering axial fatigue in loose winding conductors, Eng. Fail. Anal. 18 (1) (2011) 442–449.
[6] Z. Zhao, C. Tang, C. Li, Q. Zhou, L. Xia, C. Yao, Transformer winding deformation fault diagnosis method based on frequency response binary image, High Voltage Technology 45 (05) (2019) 1526–1534.
[7] J. Zheng, H. Huang, J. Pan, Detection of winding faults based on a characterization of the nonlinear dynamics of transformers, IEEE Trans. Instrum. Meas. 68 (1) (2019) 206–214.
[8] S. Gao, L. Sun, X. Wang, Y. Tian, J. Geng, H. Liu, Quantitative research on accumulative effect of transformer winding deformation and its influence degree based on time-frequency analysis of vibration signal, IEEE Access 10 (2022) 133451–133460.
[9] A. Abu-Siada, S. Islam, A novel online technique to detect power transformer winding faults, IEEE Trans. Power Delivery 27 (2) (2012) 849–857.
[10] F. Yang, S. Ji, Y. Liu, F. Zhang, "Research of sweep frequency impedance to determine transformer winding deformation after short-circuit impact,", IEEE International Power Modulator and High Voltage Conference (IPMHVC) 2016 (2016) 68–72.
[11] V.A. Lavrinovich, A.V. Mytnikov, Development of pulsed method for diagnostics of transformer windings based on short probe impulse, IEEE Trans. Dielectr. Electr. Insul. 22 (4) (2015) 2041–2045.
[12] S. Pramanik, A. Ganesh, V.S.B.C. Duvvury, Double-End Excitation of a Single Isolated Transformer Winding: An Improved Frequency Response Analysis for Fault Detection, IEEE Trans. Power Delivery 37 (1) (2022) 619–626.
[13] J.-W. Kim, B. Park, S. Jeong, S.W. Kim, P. Park, Fault Diagnosis of a Power Transformer Using an Improved Frequency-Response Analysis, IEEE Trans. Power Delivery 20 (2005) 169–178.
[14] Y. Akhmetov, V. Nurmanova, M. Bagheri, A. Zollanvari, G.B. Gharehpetian, A Bootstrapping solution for effective interpretation of transformer winding frequency response, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.
[15] "IEC/IEEE International Draft Standard - Power Transformers - Part 18: Measurement of frequency response," *IEEE P60076-18_D1,* 2012.
[16] D. Bormann and P. Picher, "Mechanical condition assessment of transformer windings using frequency response analysis (FRA)," *CIGRE Brochure 342,* 2008.
[17] P. Picher, "Advances in the interpretation of transformer Frequency Response Analysis (FRA)," *CIGRE Brochure 812,* 2020.
[18] "IEEE Guide for the Application and Interpretation of Frequency Response Analysis for Oil-Immersed Transformers," *IEEE Std C57.149-2012,* pp. 1-72, 2013.
[19] "High Voltage Test Technology Standardization Technical Committee of Electric Power Industry. Frequency response analysis on winding deformation of power transformers," *DL/T 911—2016,* 2016.
[20] N. Hashemnia, A. Abu-Siada, S. Islam, Improved power transformer winding fault detection using FRA diagnostics – part 2: radial deformation simulation, IEEE Trans. Dielectr. Electr. Insul. 22 (1) (2015) 564–570.
[21] N. Hashemnia, A. Abu-Siada, S. Islam, Improved power transformer winding fault detection using FRA diagnostics – part 1: axial displacement simulation, IEEE Trans. Dielectr. Electr. Insul. 22 (1) (2015) 556–563.
[22] J. Ni, Z. Zhao, S. Tan, Y. Chen, C. Yao, C. Tang, The actual measurement and analysis of transformer winding deformation fault degrees by FRA using mathematical indicators, Electr. Pow. Syst. Res. 184 (2020) 106324.
[23] Y. Chen, Z. Zhao, H. Wu, X. Chen, Q. Xiao, Y. Yu, Fault anomaly detection of synchronous machine winding based on isolation forest and impulse frequency response analysis, Measurement 188 (2022) 110531.
[24] M. Bigdeli, D. Azizian, G.B. Gharehpetian, Detection of probability of occurrence, type and severity of faults in transformer using frequency response analysis based numerical indices, Measurement 168 (2021) 108322.
[25] L. Zhou, T. Lin, X. Zhou, S. Gao, Z. Wu, C. Zhang, Detection of winding faults using image features and binary tree support vector machine for autotransformer, IEEE Trans. Transp. Electrif. 6 (2) (2020) 625–634.
[26] J. Liu, Z. Zhao, C. Tang, C. Yao, C. Li, S. Islam, classifying transformer winding deformation fault types and degrees using FRA Based on Support Vector Machine, IEEE Access 7 (2019) 112494–112504.
[27] M. Bigdeli, A. Abu-Siada, Clustering of transformer condition using frequency response analysis based on k-means and GOA, Electr. Pow. Syst. Res. 202 (2022) 107619.
[28] Q. Cheng, Z. Zhao, C. Tang, G. Qian, S. Islam, Diagnostic of transformer winding deformation fault types using continuous wavelet transform of pulse response, Measurement 140 (2019) 197–206.
[29] Z. Zhao, et al., Improved Method to Obtain the Online Impulse Frequency Response Signature of a Power Transformer by Multi Scale Complex CWT, IEEE Access 6 (2018) 48934–48945.
[30] Z. Wu, et al., A New Testing Method for the Diagnosis of Winding Faults in Transformer, IEEE Trans. Instrum. Meas. 69 (11) (2020) 9203–9214.
[31] J. Lin, J. Ma, J.G. Zhu, Y. Cui, A transfer ensemble learning method for evaluating power transformer health conditions with limited measurement data, IEEE Trans. Instrum. Meas. 71 (2022) 1–10.
[32] J. Liu, Z. Zhao, K. Pang, D. Wang, C. Tang, C. Yao, Improved winding mechanical fault type classification methods based on polar plots and multiple support vector machines, IEEE Access 8 (2020) 216271–216282.
[33] T. Hong, D. Deswal, F.d. León, An Online Data-Driven Technique for the Detection of Transformer Winding Deformations, IEEE Trans. Power Delivery 33 (2) (2018) 600–609.
[34] B. Gustavsen, P.Á.R. Ronchi, A. Mjelve, High-frequency resonant overvoltages in transformer regulating winding caused by ground fault initiation on feeding cable, IEEE Trans. Power Delivery 33 (2) (2018) 699–708.
[35] A.J. Ghanizadeh, G.B. Gharehpetian, ANN and cross-correlation based features for discrimination between electrical and mechanical defects and their localization in transformer winding, IEEE Trans. Dielectr. Electr. Insul. 21 (5) (2014) 2374–2382.

[36] Y. Chen, Z. Zhao, Y. Yu, W. Wang, C. Tang, Understanding IFRA for Detecting Synchronous Machine Winding Short Circuit Faults Based on Image Classification and Smooth Grad-CAM++, IEEE Sens. J. 23 (3) (2023) 2422–2432.
[37] Y. Chen, Z. Zhao, Y. Yu, Y. Guo, C. Tang, Improved Interpretation of Impulse Frequency Response Analysis for Synchronous Machine Using Life long Learning Based on iCaRL, IEEE Trans. Instrum. Meas. 72 (2023) 1–10.
[38] C. Zhang, H. Zhang, G. Sun, X. Ma, Transformer Anomaly Detection Method Based on MDS and LOF Algorithm, in: In 2022 7th Asia Conference on Power and Electrical Engineering (ACPEE), 2022, pp. 987–991.
[39] Z. Zhao, C. Yao, C. Tang, C. Li, F. Yan, S. Islam, Diagnosing transformer winding deformation faults based on the analysis of binary image obtained from FRA Signature, IEEE Access 7 (2019) 40463–40474.
[40] Z. Zhao, C. Tang, Y. Chen, Q. Zhou, C. Yao, S. Islam, Interpretation of transformer winding deformation fault by the spectral clustering of FRA signature, Int. J. Electr. Power Energy Syst. 130 (2021) 106933.
[41] J. Liu, C. Yao, L. Yu, S. Dong, Y. Liu, "An attempt of transformer winding fault location based on digital twin," in, IEEE International Conference on High Voltage Engineering and Applications (ICHVE) 2022 (2022) 1–4.
[42] D. Zhou, D. Huang, J. Hao, Y. Ren, P. Jiang, X. Jia, Vibration-based fault diagnosis of the natural gas compressor using adaptive stochastic resonance realized by Generative Adversarial Networks, Eng. Fail. Anal. 116 (2020) 104759.
[43] X. Deng, Y. Hu, Y. Jia, M. Peng, Power system stability assessment method based on GAN and GRU-Attention using incomplete voltage data, IET Gener. Transm. Distrib. 17 (16) (2023) 3692–3705.
[44] X. Zhang, B. Wu, X. Zhang, Q. Zhou, Y. Hu, J. Liu, A novel assessable data augmentation method for mechanical fault diagnosis under noisy labels, Measurement 198 (2022) 111114.
[45] I. Goodfellow, et al., Generative adversarial nets, Adv. Neural Inf. Proces. Syst. 27 (2014).
[46] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein Generative Adversarial Networks, in: International conference on machine learning, PMLR, 2017, pp. 214–223.
[47] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. J. A. i. n. i. p. s. Courville, "Improved training of wasserstein gans," vol. 30, 2017.
[48] M. Mirza, S. Osindero, Conditional Generative Adversarial Nets, ArXiv 1411 (2014) 1784.
[49] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ArXiv 1511 (2015) 06434.
[50] V. Behjat, M. Mahvi, Statistical approach for interpretation of power transformers frequency response analysis results, IET Sci. Meas. Technol. 9 (3) (2015) 367–375.
[51] M.H. Samimi, S. Tenbohlen, A.A. Shayegani Akmal, H. Mohseni, Improving the numerical indices proposed for the FRA interpretation by including the phase response, Int. J. Electr. Power Energy Syst. 83 (2016) 585–593.
[52] W.C. Sant'Ana, et al., A survey on statistical indexes applied on frequency response analysis of electric machinery and a trend based approach for more reliable results, Electr. Pow. Syst. Res. 137 (2016) 26–33.
[53] P. M. Nirgude, D. Ashokraju, A. D. Rajkumar, and B. P. Singh, "Application of numerical evaluation techniques for interpreting frequency response measurements in power transformers," *IET Science, Measurement & Technology*. vol. 2(5) pp. 275-285.
[54] D. Liu, T. Guo, M. Chen, "Fault Detection Based on Modified t-SNE", CAA Symposium on Fault Detection Supervision and Safety for Technical Processes (SAFEPROCESS) 2019 (2019) 269–273.