[1] Qi WANG
[2] Yue TONG
[3] Jin ZHANG
[4] Lingjie JIA,
[5] Tian XIA
[6] *Yifan XU*

Intelligent GIS Partial Discharge Diagnosis with Limited Samples via Conditional Data Augmentation and Modern CNNs

Journal of Electrical Systems

Abstract: - To address the challenges of defect sample scarcity and reliance on manual feature extraction in partial discharge (PD) diagnosis of Gas-Insulated Switchgear (GIS), this study proposes a small-sample intelligent diagnostic method that integrates a Conditional Generative Adversarial Network (CGAN) with the ConvNext architecture. Under conditional constraints, the CGAN is employed to perform data augmentation on PRPD spectrograms, effectively mitigating issues of sample imbalance and overfitting. Subsequently, the ConvNext network is introduced to extract multi-scale texture and morphological features from PRPD data, enabling efficient end-to-end classification. The model's performance is validated using a 220 kV GIS typical discharge dataset, and the influence of key architectural parameters on diagnostic performance is further investigated. Experimental results show that the proposed method achieves remarkable improvements in small-sample classification accuracy — with data augmentation enhancing accuracy by more than 5%, and the optimal configuration achieving diagnostic precision exceeding 97% — demonstrating strong potential for practical engineering applications.

Keywords: GIS,Partial Discharge, Spectrogram, Small Sample, GAN-ConvNext,Diagnosis

I. INTRODUCTION

The insulation condition of power equipment is directly related to grid security and supply reliability. Partial discharge (PD) is a key indicator of insulation degradation; effective diagnosis of PD defects enables early warning of potential insulation hazards, preventing fault escalation and enhancing operational safety [1–2]. In Gas-Insulated Switchgear (GIS), PD phenomena often signify internal defects. Accurate identification of PD types is therefore essential for assessing GIS insulation health and improving power supply reliability [3–4]. Once insulation failure occurs in GIS, it can easily lead to large-scale outages and significant economic losses. Consequently, developing high-precision and engineering-applicable PD pattern recognition techniques is of urgent practical importance.

Existing PD pattern recognition methods generally follow two technical pathways. The first is traditional machine learning approaches, including Bayesian Networks (BN [5]), Support Vector Machines (SVM [6–7]), Random Forests (RF [8–9]), Backpropagation Neural Networks (BPNN [10–11]), and XGBoost [12–13]. These methods possess certain recognition capabilities but heavily rely on manually crafted features. Their performance is thus limited by prior signal-processing expertise and feature selection, and their model expressiveness is relatively weak.

The second pathway is deep learning-based methods, which have emerged in recent years. Representative studies treat PD maps as inputs and employ Convolutional Neural Networks (CNN [14–16]) or Deep Residual Networks [17–18] for end-to-end feature learning and classification, achieving superior performance on multiple typical defect types. Although such models can automatically extract features, simplify manual processes, and exhibit stronger representation ability, they require large-scale, high-quality, and class-balanced labeled datasets. This dependence limits their applicability in practical engineering scenarios with small-sample data. Furthermore, high-quality and accurately labeled PD samples are difficult to obtain in practice [19–21]: field acquisition and labeling are costly, and different defect conditions are inherently imbalanced. The combination of small sample size and class imbalance easily leads to biased decision boundaries and model overfitting. In addition, PRPD data carry explicit physical meanings along their coordinate axes (phase–amplitude), and conventional data augmentations such as flipping, rotation, or scaling can distort this physical structure, making them unsuitable for direct use [18].

To address these challenges, this paper proposes a "CGAN-ConvNeXt" diagnostic framework. First, a Conditional Generative Adversarial Network (CGAN) is utilized to generate synthetic PRPD samples under class-conditioned constraints, ensuring consistency with the statistical characteristics of real data and achieving

[1] China Electric Power Research Institute Co., Ltd., Wuhan, Hubei 430074, China.

"physically semantic-friendly" data augmentation. Then, ConvNeXt is adopted as the backbone classifier. Its stage-wise convolutional structure and modernized architecture enhance multi-scale texture and morphological feature representation, enabling high diagnostic accuracy even under small-sample conditions. Furthermore, the effects of different ConvNeXt architectural parameters on diagnostic performance are investigated, leading to the establishment of an optimal CGAN-ConvNeXt configuration for small-sample GIS PD diagnosis based on PRPD information. The effectiveness of the proposed framework is validated and analyzed through experimental studies.

II. CGAN AND CONVNEXT PRINCIPLES

Generative Adversarial Networks (GANs) [22] and Convolutional Neural Networks (CNNs) [23] are two important models in deep learning. GANs are primarily used to generate realistic data samples and have been widely applied in image generation and data augmentation tasks. In contrast, CNNs have demonstrated outstanding performance in visual tasks such as image classification and object detection.

In this study, the advantages of Conditional GAN (CGAN) and ConvNeXt are combined to address the issues of insufficient samples and feature extraction in partial discharge defect diagnosis. CGAN generates diverse training data to alleviate the small-sample problem, while ConvNeXt, through its advanced convolutional design, enhances the capability of extracting discriminative features from complex image data.

A. Conditional Generative Adversarial Network (CGAN)

Generative Adversarial Networks (GANs) were first proposed by Goodfellow in 2014. In a GAN model, the generator and discriminator are jointly optimized through adversarial learning until reaching a Nash equilibrium. This model has achieved remarkable success in fields such as image generation and data augmentation, becoming one of the most common techniques for synthetic data generation.A particularly effective variant, the Conditional Generative Adversarial Network (CGAN) [24], introduces conditional information (y), enabling the generator to produce synthetic data that correspond to specific conditions. This mechanism enhances the controllability and relevance of the generated data. Compared with the conventional GAN, CGAN can generate more precise and targeted synthetic samples.The basic principle of CGAN is illustrated in Figure 1: the generator G combines random noise z and conditional information y to generate images, which—after preprocessing such as grayscale conversion—are fed into the discriminator D together with real samples. When the generated data meet certain criteria, the generator G outputs the corresponding synthetic PRPD patterns.

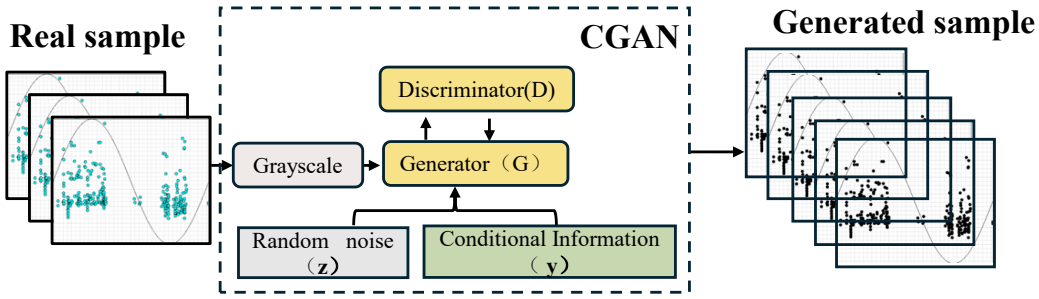


Figure 1. Schematic Diagram of CGAN

Given the conditional information y, the standard adversarial objective function between the generator G and discriminator D is formulated as follows:

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x \mid y)] + E_{z\sim p_z(z)}[\log(1 - D(G(z \mid y)))] \tag{1}$$

where z is the noise vector, and $p_{data}(x \mid y)$ indicates the real data distribution conditioned on class information y.

The Fréchet Inception Distance (FID) is adopted to evaluate how closely the generated samples approximate the real samples. Features are extracted from the real image set and the CGAN-generated image set using a pretrained Inception network. Let the mean vectors and covariance matrices of the two sets be μ_r, Σ_r and μ_g, Σ_g, respectively.

$$FID = \| \mu_r - \mu_g \|_2^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}) \tag{2}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, and $\|\cdot\|_2^2$ represents the squared Euclidean norm. A lower FID value indicates that the feature distributions of the generated and real sets are more similar, implying that the generated samples are more "realistic."The pretrained Inception network is not involved in the generation or discrimination

process of the CGAN; instead, it serves solely as an evaluation tool. By extracting features from both datasets, it facilitates the computation of the FID metric, thereby quantitatively assessing the quality of images generated by the CGAN.

B. ConvNeXt Convolutional Neural Network (ConvNeXt)

The ConvNeXt model, developed from the traditional convolutional neural network ResNet, was introduced by Zhuang Liu *et al.* in 2022 in their paper *"A ConvNet for the 2020s"* [26]. Over the past few years, ConvNeXt has evolved into a representative and effective classification technique in the field of image recognition [27–29].

ConvNeXt performs a systematic reconstruction based on ResNet. Its overall architecture consists of four sequential stages, each containing multiple ConvNeXt Blocks that progressively extract image features. Spatial downsampling is achieved using 2×2 convolutions. The ConvNeXt Block serves as the core module of the architecture and is composed of Layer Normalization (LayerNorm), a 7×7 depthwise convolution, two 1×1 pointwise convolutions (including GELU activation functions), and a LayerScale mechanism. Residual connections are employed to maintain gradient flow throughout the network.

The architecture adopts a channel-last data format and a pre-normalization design to enhance computational efficiency and training stability. Finally, global average pooling, layer normalization, and a linear classifier are applied to produce the final category prediction. A typical ConvNeXt system architecture is illustrated in **Figure 2**, where *K*, *S*, and *P* denote the convolution kernel size, stride, and padding, respectively.

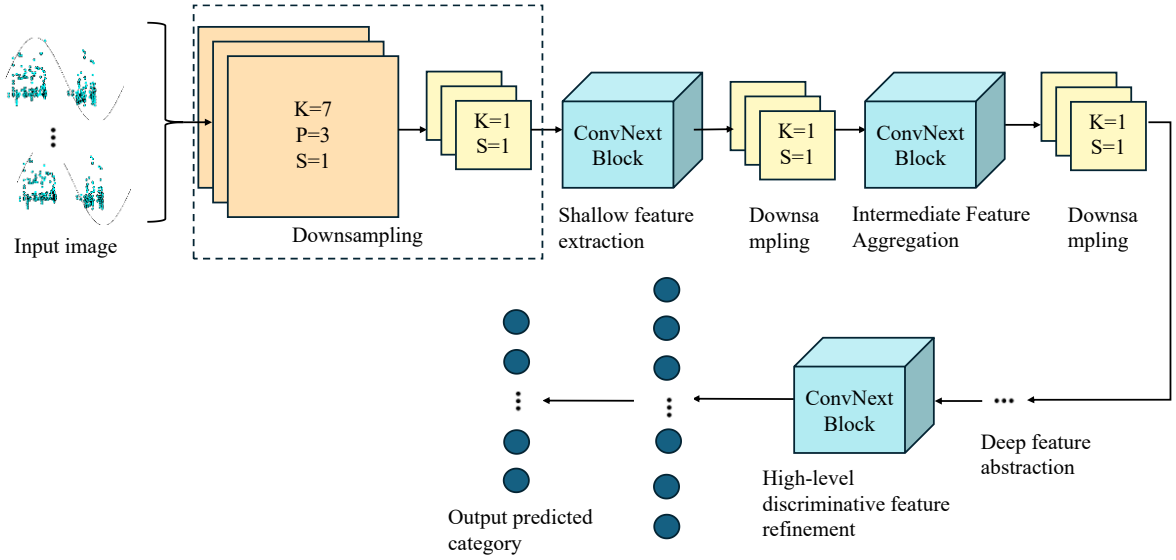


Figure 2.Typical ConvNext System Architecture

III. GIS DISCHARGE SMALL-SAMPLE CGAN-CONVNEXT DIAGNOSTIC MODEL

During the operation of on-site Gas-Insulated Switchgear (GIS), acquiring labeled partial discharge (PD) data remains a significant challenge. The collected samples are typically limited in number and exhibit severe class imbalance [30]–[33]. Such imbalance and data scarcity result in two major issues. First, the model tends to overfit the majority classes, leading to a noticeable reduction in recall for minority classes and, consequently, a decline in overall macro-averaged performance metrics. Second, when the number of samples is insufficient, commonly used data augmentation techniques such as arbitrary rotation or flipping can distort the physical semantics of phase-resolved partial discharge (PRPD) patterns, producing images inconsistent with the underlying discharge mechanisms.

To overcome these challenges, this study proposes a hybrid diagnostic framework that integrates a Conditional Generative Adversarial Network (CGAN) with a lightweight ConvNeXt backbone, as illustrated in Fig. 3. The proposed framework is designed to maintain high diagnostic accuracy while effectively alleviating the limitations posed by small-sample and imbalanced-class conditions, with reduced parameter count and computational complexity.

In the data preprocessing stage, PRPD maps are constructed from ultra-high-frequency (UHF) PD signals acquired via an experimental GIS platform. To reflect realistic operating conditions, signal segments with amplitudes less than 1.2 times the background noise level under non-discharge conditions are removed. The processed PRPD data are

subsequently normalized and resized to 224 × 224 grayscale images, minimizing domain shifts caused by variations in experimental environments.

In the data augmentation stage, the CGAN generator takes a noise vector zand a PD defect class label y(e.g., particle, tip, gap, or surface discharge) as conditional inputs to synthesize additional PRPD samples. The discriminator learns to evaluate the correspondence between each image and its associated class label, while the Fréchet Inception Distance (FID) metric is employed to quantify the similarity between real and generated samples. When the FID value is less than or equal to 15, the generated PRPD patterns become visually indistinguishable from real ones and remain consistent with the physical discharge mechanisms in terms of cluster morphology, phase distribution, and pulse density. Therefore, an FID threshold of 15 is adopted in this study as the convergence criterion. This strategy not only enriches the minority-class samples (e.g., particle and surface discharges) but also preserves the underlying physical semantics of the data.

In the pattern recognition stage, the augmented PRPD dataset is randomly divided into training and testing subsets with an 8:2 ratio. The lightweight ConvNeXt network is subsequently employed to perform feature extraction and classification on the enhanced PRPD images, achieving efficient and accurate PD pattern identification under small-sample conditions.

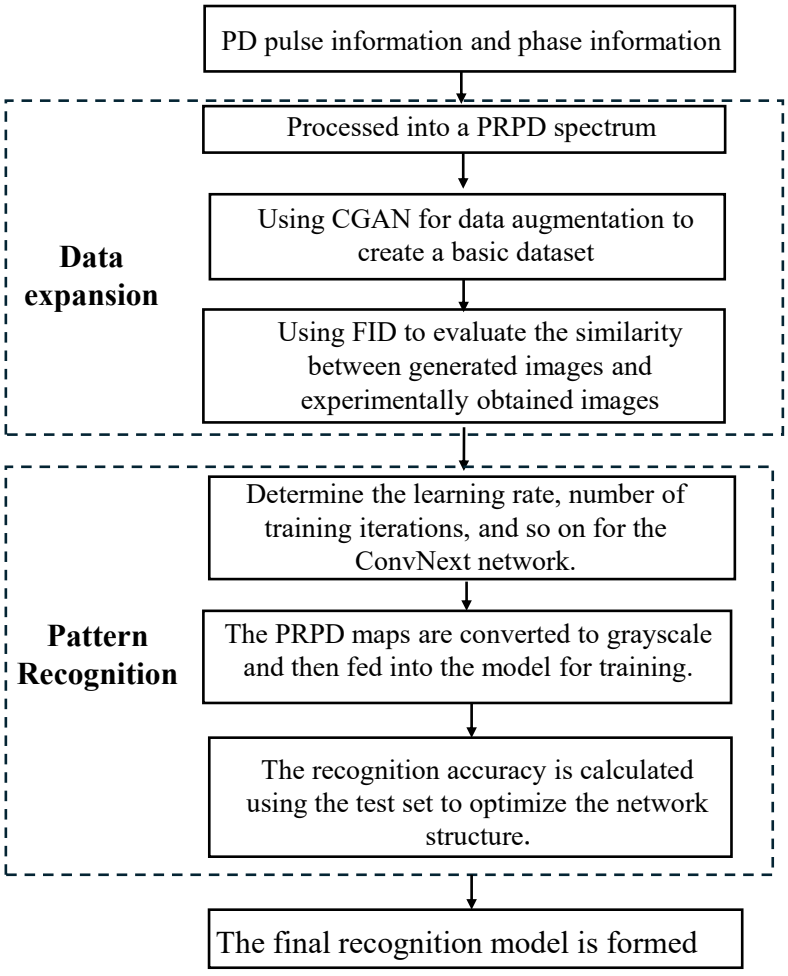


Figure 2.CGAN-ConvNext Diagnostic Model for Small-Sample GIS Discharge

The training process of the ConvNeXt model is described as follows. The input PRPD map is first converted into a tensor $\boldsymbol{X}$, where the map has a height $\boldsymbol{H}$, width $\boldsymbol{W}$, and three color channels $\boldsymbol{C}$, and all elements belong to the real number domain $\mathbb{R}$. The tensor is then processed by a $\mathbf{7 \times 7}$depthwise separable convolution to reduce the number of parameters required for feature extraction, thereby enhancing the computational efficiency of the model. After the depthwise convolution, the resulting feature map $\boldsymbol{U}$can be expressed as:

$$U = DWConv_{7\times7}\left(X\right) \tag{3}$$

where $DWConv_{7\times7}$ denotes a 7×7 convolution operation applied independently to each channel, and Xrepresents the input feature vector. Subsequently, $U$is converted into a channel-last format to accommodate the following normalization operation.

$$U' = Permute(U) \tag{4}$$

where U denotes the tensor converted into the $N \times H \times W \times C$ format.

In the normalization stage, the ConvNeXt Block replaces the conventional Batch Normalization (BatchNorm) used in ResNet with Layer Normalization (LN), which operates along the channel dimension and is more suitable for small-batch training.

$$V = LayerNorm(U') \tag{5}$$

where V represents the features after Layer Normalization.The normalized features are then fed into a two-layer Multi-Layer Perceptron (MLP) structure, equivalent to that used in the Vision Transformer (ViT) [25], which consists of two consecutive 1×1 pointwise convolutions. The first layer expands the channel dimension by a factor of four and employs the GELU activation function to enhance the model's nonlinear representation capability, while the second layer compresses it back to the original dimension. The detailed process can be expressed as follows:

$$Z_1 = Conv_{1\times1}^{c\to4c}(V) \tag{6}$$

$$Z_2 = GELU(Z_1) \tag{7}$$

$$Z_3 = Conv_{1\times1}^{4C\to C}(Z_2) \tag{8}$$

where V_1 denotes the feature map obtained after the first 1×1 convolution, with the channel dimension expanded from C to $4C$;V_2 represents the output after the GELU activation function;and V_3 denotes the feature map produced by the second 1×1 convolution, where the channel dimension is reduced back to C.To regulate the proportion of residual information, the ConvNeXt Block introduces a LayerScale module before the output stage. Specifically, a learnable parameter γ is applied to rescale each channel as follows:

$$Y' = \gamma \cdot Z_3 \tag{9}$$

where γ denotes the learnable channel-scaling parameter that enhances the control over the residual signal, and Y' represents the scaled output feature.

Finally, the module restores the original data layout from $N \times H \times W \times C$ to $N \times C \times H \times W$, and adds it to the initial input feature X to form the residual connection.

$$Y = X + Permute^{-1}(Y') \tag{10}$$

where Y denotes the final output obtained by adding the residual connection to the input X.

In summary, the ConvNeXt Block models global semantics through large convolution kernels, enhances training stability for small batches using Layer Normalization (LayerNorm), and controls residual information via the LayerScale mechanism. This design effectively improves the capability of modeling complex patterns while maintaining the computational efficiency of convolutional neural networks (CNNs). Consequently, the proposed structure achieves high diagnostic accuracy even under small-sample and class-imbalanced conditions.

IV. Construction of GIS Partial Discharge Experimental Platform and Data Acquisition

A. Establishment of the Experimental Platform

In accordance with IEC 60270 and GB/T 7354-2018 standards, a 220 kV GIS partial-discharge-free simulation experimental platform was constructed, as shown in Fig. 4. The platform employs a discharge-free power-frequency test transformer (model YD-50 kVA/100 kV), a discharge-free protection resistor R = 4 kΩ, and a voltage divider with capacitance ratio C_1/C_2 = 1000:1. The coupling capacitor is C_3 = KVOWF-100 kV, and the detection impedance is JF-3. The GIS test chamber is filled with SF_6 gas at a pressure of 0.5 MPa.

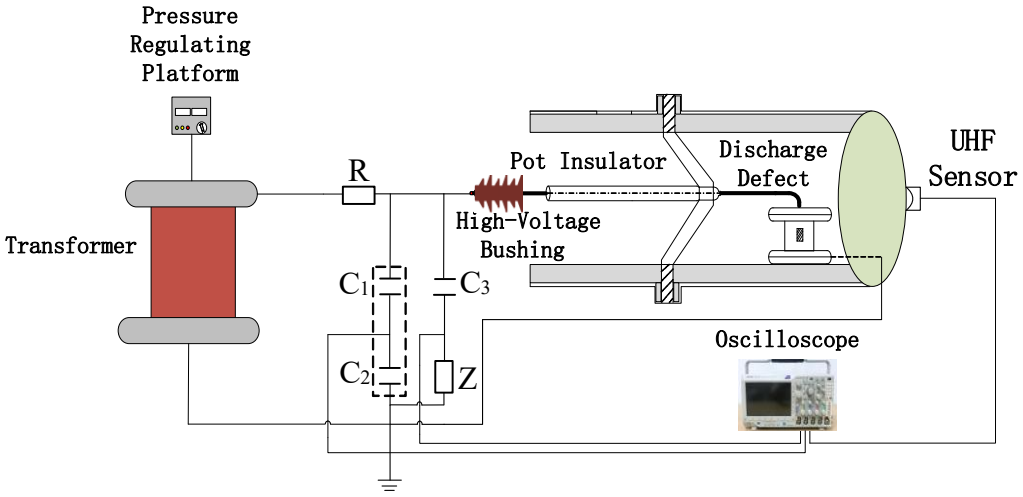
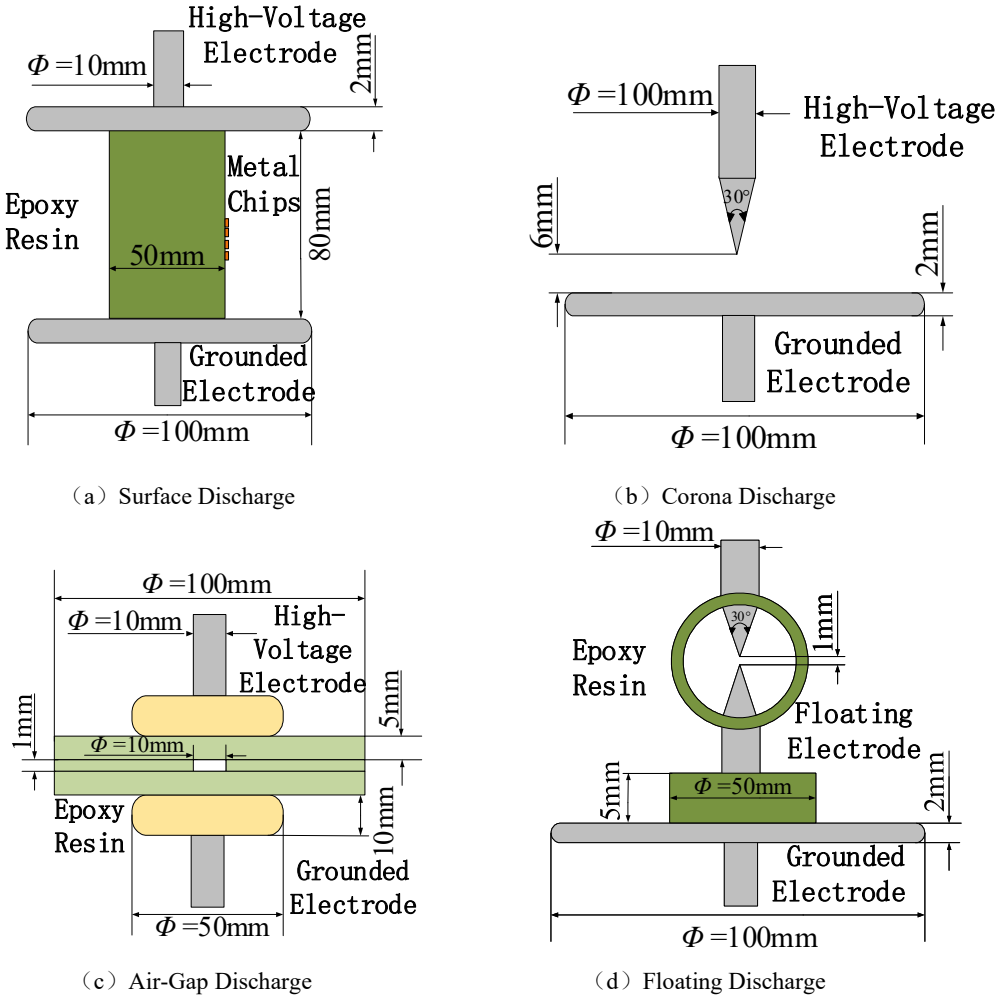


Figure 4 220kV GIS PD-Free Simulated Experimental Platform

Typical partial discharge (PD) defects within a GIS compartment include surface discharge, corona discharge, air-gap discharge, and floating discharge [34]. The structures and dimensions of the defect models are illustrated in **Fig. 5**.



（a）Surface Discharge　（b）Corona Discharge

（c）Air-Gap Discharge　（d）Floating Discharge

Figure 5 Typical Discharge Defect Model in GIS (with discharge environment of 0.5MPa SF_6

A stepwise voltage-increasing method was adopted to determine the initial partial discharge inception voltage (U_{PDIV}) for each defect type. The experimental voltage (U_{PDV}) was set to 1.5 times U_{PDIV} to ensure stable discharge. The discharge test voltages corresponding to different defect types are listed in **Table 1**

Table 1 Experimental Voltages of Typical Discharge Defects in GIS

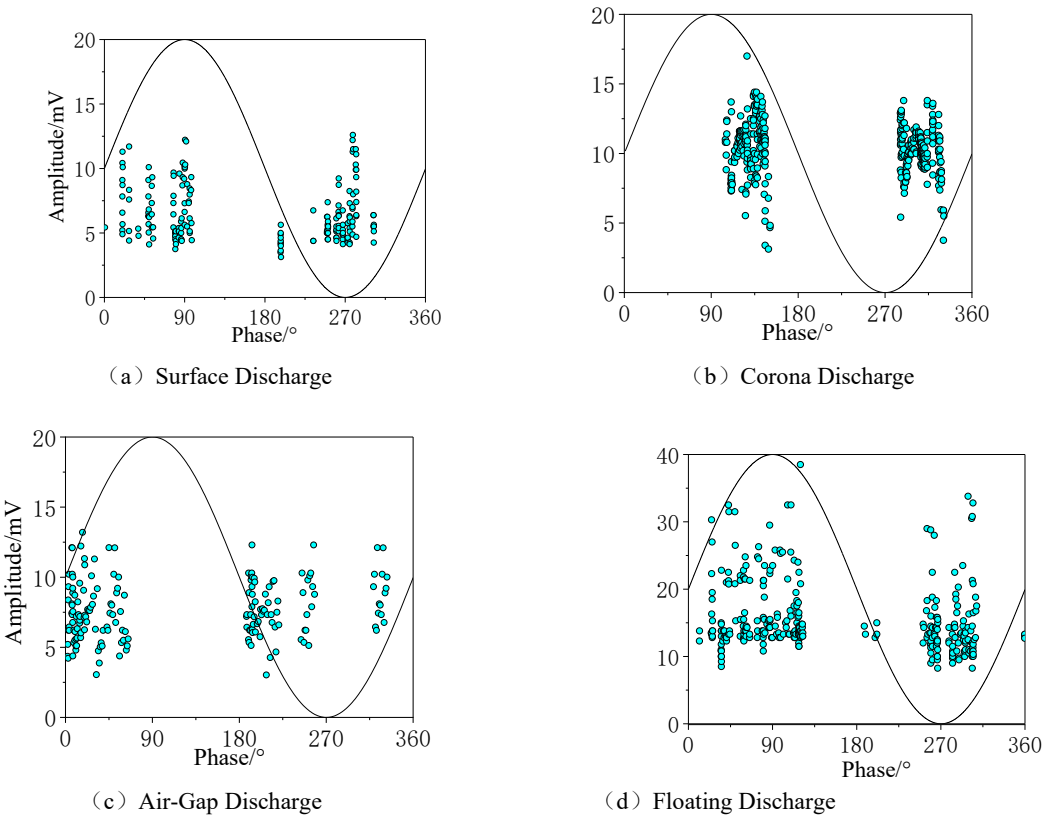
Defect type	U_{PDIV}/kV	U_{PDV}/kV
Floating Discharge	6.3	9.45
Air-Gap Discharge	4.6	6.9
Corona discharge	3.8	5.7
Surface Discharge	6.5	9.75

B. Experimental Data Acquisition

The UHF sensor used in the experiment is a resonant-cavity monopole UHF antenna sensor, with a detection frequency range covering 0.3–3 GHz. The standing wave ratio (SWR) is less than 5, and the maximum gain reaches 7.16 dB at the primary resonant frequency of 1.0 GHz. The average gain across the 0.3–3 GHz frequency band is 4.92 dB [35].

A Tektronix MSO44 high-performance digital oscilloscope (four channels, 1.5 GHz bandwidth) was employed to capture the partial discharge (PD) signals. The oscilloscope was configured with a sampling rate of 5 MS/s, a sampling duration of 1 s, and a sampling length of 5,000,000 points per signal acquisition.

Discharge data from the four types of defect models were collected, and the corresponding phase-resolved partial discharge (PRPD) patterns were plotted, as shown in Fig. 6.



（a）Surface Discharge　（b）Corona Discharge

（c）Air-Gap Discharge　（d）Floating Discharge

Figure 6 Discharge Spectrograms of Typical Discharge Defects in GIS Compartment

V. Analysis of Diagnostic Performance of the CGAN-ConvNeXt Model

To evaluate the effectiveness of the proposed CGAN-ConvNeXt model, both the CGAN and ConvNeXt networks were implemented in Visual Studio Code (VS Code) and trained and tested on an Ubuntu 20.04 operating system equipped with an NVIDIA RTX 4090 GPU. The experiments were conducted using Python 3.8 and CUDA 11.3. The total number of training epochs was set to 100, with learning rates of 0.003 for the generator and 0.001 for the discriminator. The Adam optimizer was used to accelerate convergence, with $\beta_1 = 0.5$, and the batch size was set to 16, meaning that each training batch contained 16 images.

The diagnostic performance of the CGAN-ConvNeXt model was quantitatively evaluated on the PRPD dataset of four typical GIS partial discharge defects—surface discharge, corona discharge, air-gap discharge, and floating discharge—using diagnostic accuracy as the primary evaluation metric. The proposed model was compared with conventional machine learning and deep learning approaches, including SVM, ANN, and CNN, under small-sample defect diagnosis scenarios.

Furthermore, to verify the effectiveness of data augmentation, a controlled experiment was conducted on the same dataset. The CGAN-ConvNeXt model was trained separately using the original small-sample dataset and the CGAN-augmented dataset, with diagnostic accuracy again serving as the evaluation metric.

Finally, to further confirm the contribution of the CGAN module, ablation experiments were performed on the small-sample defect subsets. The analysis examined how the diagnostic accuracy of the CGAN-ConvNeXt model varied with parameters such as data augmentation scale, convolution kernel size, LayerScale coefficient, activation function type, and normalization method, ultimately determining the model's optimal structural configuration.

A. Performance Comparison of Different Diagnostic Models

The proposed CGAN-ConvNeXt model was compared with a traditional machine learning method (SVM) and a classical deep learning method (ANN). Each model was trained and tested on both the original small-sample dataset and the CGAN-augmented dataset. The overall diagnostic accuracy and classification accuracy for the four types of defects were evaluated, and the results are presented in Table 2. The average accuracy represents the mean value of diagnostic accuracies across the four partial discharge (PD) defect types.

As shown in Table 2, the CGAN-ConvNeXt model consistently achieved higher average diagnostic accuracy than both the SVM and ANN models, regardless of whether CGAN-based data augmentation was applied. It also demonstrated an advantage over the CNN model, particularly under small-sample diagnostic conditions. On the original small-sample dataset, the CGAN-ConvNeXt model achieved an average accuracy of 94.9%, which is 3.6% higher than that of SVM (91.3%), 2.9% higher than ANN (92.0%), and 0.5% higher than CNN (94.4%). On the CGAN-augmented dataset, the average diagnostic accuracy further increased to 97.3%, significantly outperforming SVM (92.75%) and ANN (93.3%), and maintaining a clear advantage over CNN (95.5%). Moreover, the CGAN-based data augmentation notably enhanced the model's diagnostic performance for air-gap discharge, floating discharge, and surface discharge, as illustrated in Fig. 7.

Table 2 Comparison of Diagnostic Accuracy of Different Models on Two Types of Datasets

Type of Dataset	Diagnosis Model	Average Accuracy	Corona Discharge	Air-Gap Discharge	Floating Discharge	Surface Discharge
Original Small-Sized Dataset	SVM	91.3%	100%	88.3%	90%	87%
	ANN	92%	95%	90%	93%	86%
	CNN	94.4%	100%	96.2%	94.5%	87%
	CGAN-ConvNext	**94.9%**	**100%**	**96.4%**	**95%**	**88.2%**
CGAN-Augmented Dataset	SVM	92.75%	100%	90%	92%	89%
	ANN	93.3%	98%	91.6%	95%	88.6%
	CNN	95.5%	100%	97%	95%	90%
	CGAN-ConvNext	**97.3%**	**100%**	**97.5%**	**97%**	**95%**

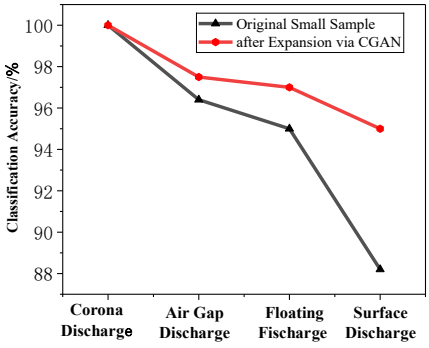


Figure 7 Recognition Accuracy of Four Types of Defects Before and After Expansion

B. Optimal Structural Parameter Analysis of the CGAN-ConvNeXt Diagnostic Model for GIS Partial Discharge

To systematically optimize the CGAN-ConvNeXt model, this study further analyzes the variation of average diagnostic accuracy with respect to fluctuations in the key architectural parameters from multiple dimensions, including data augmentation scale, convolution kernel size, LayerScale coefficient, activation function type, and normalization method.The relationships between model diagnostic accuracy and these parameters are summarized in Table III. The results indicate that the optimal performance is achieved when using 300 augmented samples, a 7×7 convolution kernel, $\gamma = 1\times10^{-5}$, the GELU activation function, and LayerNorm normalization.

Table 3 Comparison of Key Parameter Adjustments and Diagnostic Accuracy of the CGAN-ConvNeXt Model

	Value Range / Configuration	Variation Trend of Average Accuracy	Optimal Value	Description
Number of Augmented Samples	100、200、300、400、500	94.9→96.2%→97.3%	300	When the augmentation scale reaches 300 samples, the data coverage is optimal; too few samples lead to underfitting, while too many introduce redundancy or spurious features.
Convolution Kernel Size	3×3、5×5、7×7、9×9、11×11	90.5%→97.3%（7×7）→97.3%（11×11）	7×7	The receptive field matches the distribution of defect features most effectively; a larger convolution kernel increases computational cost without further performance improvement.
LayerScale Coefficient (γ)	0、1e-6、1e-5、1e-4、1e-3	When $\gamma = 0$, the residual connection is absent, resulting in low accuracy; the model achieves optimal performance when $\gamma$is around 1×10^{-5}; an excessively large γleads to unstable training.	1e-5	Balancing the residual information flow ensures stable convergence and high diagnostic accuracy.
Type of Nonlinear Activation Function	ReLU、LeakyReLU、SiLU、GELU	ReLU shows a relatively low baseline performance, while LeakyReLU performs slightly better; both SiLU and GELU achieve superior results, with GELU providing the best overall performance.	GELU	Smooth nonlinear activation functions are more suitable for complex defect distributions, providing stable convergence and the highest diagnostic accuracy.
Type of Normalization Layer	BatchNorm、LayerNorm、GroupNorm	BatchNorm is unstable under small-sample conditions; GroupNorm provides moderate performance; LayerNorm achieves the best results.	LayerNorm	It is friendly to small-batch training and alleviates training fluctuations under sample-imbalance conditions.

Specifically, the influence of the number of augmented PRPD maps on the diagnostic performance of the model is shown in Table IV. When each defect type is augmented with only 100 images, insufficient sample coverage leads to underfitting, resulting in a diagnostic accuracy of only 94.5%. As the number of augmented samples increases to 200 and 300, the accuracy improves to 96.2% and 97.2%, respectively. The peak performance at 300 samples indicates that the generated maps sufficiently capture the representative distribution of typical defect features. However, when the augmentation size further increases to 400 or 500, the accuracy slightly decreases, as

excessive augmentation introduces feature redundancy and even spurious features, which interfere with the model's discrimination capability.

The effect of convolution kernel size on diagnostic accuracy is presented in Table V. The baseline experiment with the conventional 3×3 kernel achieved an accuracy of 90.5%. Expanding the kernel to 5×5 improved the accuracy to 94.3%, and further enlargement to 7×7 achieved 97.0%, yielding the best performance. This demonstrates that the 7×7 convolution kernel provides the most suitable receptive field for matching the spatial distribution of GIS discharge defect features—capturing global information effectively while maintaining computational efficiency. When the kernel size increases to 9×9 or 11×11, no further accuracy improvement is observed; instead, performance slightly declines due to higher computational complexity and feature redundancy.

In the comparison of LayerScale coefficients (γ), when $\gamma = 0$, the residual path of the model is suppressed, leading to insufficient information flow and a significant decrease in accuracy.When $\gamma \approx 1\times10^{-5}$, the model achieves optimal performance, as this value effectively balances the proportion of residual information and maintains training stability.However, an excessively large γ (e.g., 1×10^{-3}) causes training oscillations, resulting in degraded accuracy.Therefore, $\gamma = 1\times10^{-5}$ is identified as the optimal choice, as shown in Table 4.

Table 4 Impact of the Size of Expanded Graph Dataset on Recognition Accuracy

Number of Augmented Samples	Average Accuracy
100	94.5%
200	96.2%
300	97.2%
400	96.7%
500	95.8%

Table 5 Impact of Different Convolutional Kernel Sizes on Recognition Accuracy

Kernel Size	Accuracy
3×3（基准）	90.5%
5×5	94.3%
7×7	97%
9×9	96.9%
11×11	96%

Table 6 Impact of Different LayerScale Coefficients on Recognition Accuracy

γ Value	Accuracy
0	92.1%
1e-6	95.3%
1e-5	96.8%
1e-4	95.9%
1e-3	94.7%

As shown in Table 7, the choice of activation function significantly affects the diagnostic performance of the model. Using ReLU as the baseline yields relatively low performance, while LeakyReLU provides a slight improvement. Both SiLU and GELU considerably enhance the nonlinear fitting capability of the model, with GELU achieving the best overall results. This is because GELU offers a smoother nonlinear mapping in the low-value region, making it more suitable for complex defect pattern distributions and ensuring more stable convergence.

Table 7 Impact of Different Activation Functions on Recognition Accuracy

Activation Function	Accuracy
ReLU	93.5%
LeakyReLU	94.6%
SiLU	96.2%

The effect of normalization methods on diagnostic performance is presented in Table 8. BatchNorm exhibits instability under small-sample conditions, leading to significant fluctuations in accuracy. GroupNorm provides

moderate performance, while LayerNorm maintains stability in small-batch training and demonstrates stronger robustness to class imbalance, thereby achieving the best performance.

Therefore, the CGAN-ConvNeXt model achieves the optimal balance between diagnostic accuracy and training stability when configured with 300 augmented samples, a 7×7 convolution kernel, $\gamma = 1\times10^{-5}$, the GELU activation function, and LayerNorm normalization.

The influence of different normalization methods on diagnostic performance is shown in Table 8. BatchNorm exhibits instability under small-sample conditions, resulting in large fluctuations in accuracy. GroupNorm provides moderate performance, while LayerNorm maintains stability during small-batch training and demonstrates stronger robustness to class imbalance. Consequently, LayerNorm achieves the best overall performance, as shown in Table 8.

Therefore, the CGAN-ConvNeXt model achieves optimal performance when configured with 300 augmented samples, a 7×7 convolution kernel, $\gamma = 1\times10^{-5}$, the GELU activation function, and LayerNorm normalization. This configuration provides the best balance between diagnostic accuracy and training stability.

Table 8 Impact of Different Activation Functions on Recognition Accuracy

Normalization Method	Accuracy
BatchNorm	93.8%
GroupNorm	95.4%
LayerNorm	97.1%

VI. CONCLUSION

In this study, a CGAN-ConvNeXt diagnostic framework is proposed to address the limitations in GIS partial discharge (PD) pattern recognition, where traditional methods rely heavily on handcrafted features, deep learning approaches are constrained by small-sample and class-imbalance conditions, and conventional PRPD map augmentations often distort physical semantics. The main conclusions are summarized as follows:

- By employing a **Conditional Generative Adversarial Network (CGAN)** under class-conditioned constraints, synthetic PRPD samples consistent with the statistical characteristics of real data are generated, enabling *"physically semantic-friendly"* data augmentation. This effectively mitigates the issues of small sample size and class imbalance. The **ConvNeXt** network, serving as the classification backbone, leverages stage-wise convolution and a modernized architecture to efficiently extract multi-scale texture and morphological features from PRPD maps, significantly improving diagnostic accuracy under small-sample conditions.
- To address the lack of engineering guidelines for ConvNeXt architecture parameter selection and to meet the dual requirements of **high precision** and **engineering applicability** in GIS PD recognition, the **CGAN-ConvNeXt framework** achieves enhanced diagnostic accuracy and stability through the synergy of data augmentation and architectural optimization. This approach provides a **reliable technical solution for early insulation fault warning** in power grid equipment.

REFERENCES

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first . . ."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] J. Li, X. Han, H. Wang, et al., "Review of partial discharge detection technology for electrical equipment: 2015–2025," High Volt. Eng., vol. 51, no. 7, pp. 3132–3158, Jul. 2025.

[2] H. Li, C. Fu, H. Hu, X. Wu, and W. Si, “Review of pulse current method for partial discharge detection in electrical equipment,” High Volt. Apparatus, vol. 59, no. 11, pp. 1–14, Nov. 2023.
[3] Z. Hao, C. Jiang, H. Li, et al., “High-frequency partial discharge monitoring technology based on current transformer at transformer bushing riser,” High Volt. Eng., vol. 51, no. 3, pp. 1126–1134, Mar. 2025.
[4] X. Zhang, J. Tang, C. Pan, et al., “Research on partial discharge recognition based on deep belief nets,” Power Syst. Technol., vol. 40, no. 10, pp. 3272–3278, Oct. 2016.
[5] P. H. D. S. Palhares, C. D. J. Ribeiro, L. D. C. Brito, et al., “Bayesian network and compact genetic algorithm approach for classifying partial discharges in power transformers,” J. Control, Autom. Electr. Syst., vol. 29, no. 5, pp. 605–613, Oct. 2018.
[6] X. Wang, S. Guan, L. Hua, et al., “Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method,” Ultrasonics, vol. 91, pp. 161–169, Jan. 2019.
[7] Y. Xu, F. Yang, Y. Qian, et al., “Partial discharge pattern recognition of DC cable terminations based on improved ECOC classifier,” Proc. CSEE, vol. 37, no. 4, pp. 1260–1268, Apr. 2017.
[8] S. Govindarajan, J. A. Ardila-Rey, K. Krithivasan, et al., “Development of hypergraph-based improved random forest algorithm for partial discharge pattern classification,” IEEE Trans. Qual. Control, vol. 9, no. 1, pp. 96–109, Jan. 2021.
[9] J. Hu, X. Zhang, Z. Song, et al., “Feature recognition method for partial discharge patterns of generator stator bars based on random forest,” High Volt. Eng., vol. 50, no. 3, pp. 1272–1280, Mar. 2024.
[10] J. Li, C. Sun, L. Du, et al., “Combined feature extraction method for partial discharge images,” High Volt. Eng., vol. 30, no. 6, pp. 11–13, Jun. 2004.
[11] S. Liu, Z. Xu, K. Yang, et al., “Partial discharge characteristics and recognition of power cables under typical defects,” High Volt. Eng., vol. 49, no. S1, pp. 36–39, Jan. 2023.
[12] Y. Liu, J. Su, S. Chen, et al., “Industry load forecasting method based on GRDCC-BiGRU prediction model under the synergy effect of industrial loads,” Proc. CSEE [Online]. pp. 1–21, Sep. 2025.
[13] Y. Zhang, B. Feng, Y. Chen, et al., “Fault diagnosis method for oil-immersed transformers based on XGBoost optimized by genetic algorithm,” Electr. Power Autom. Equip., vol. 41, no. 2, pp. 200–206, Feb. 2021.
[14] Y. Chen, X. Peng, H. Wang, et al., “Partial discharge pattern recognition based on morphological features and Faster R-CNN–AlexNet for generator stator,” IEEE Trans. Dielectr. Electr. Insul., vol. 31, no. 6, pp. 2956–2965, Jun. 2024.
[15] R. Kitani, et al., “Verification of interpretability of phase-resolved partial discharge using a CNN with SHAP,” IEEE Access, vol. 11, pp. 4752–4762, Jan. 2023.
[16] Y. Wang, H. Chen, J. Liu, et al., “Partial discharge recognition of oil-paper insulation in converter transformers based on Markov transition field and CBAM-CNN,” J. Wuhan Univ. (Eng. Ed.), vol. 56, no. 10, pp. 1245–1254, Oct. 2023.
[17] H. Jin, L. Gao, J. Pan, et al., “Enhancing partial discharge pattern recognition via WGAN-GP and Inception-ResNet-V2,” Electr. Eng., vol. 107, no. 6, pp. 6805–6814, Jun. 2025.
[18] C. Xu, J. Chen, W. Liu, et al., “Pattern recognition of PRPD spectra for GIS partial discharge based on deep residual network,” High Volt. Eng., vol. 48, no. 3, pp. 1113–1123, Mar. 2022.
[19] Y. Wang, J. Yan, Z. Yang, et al., “GAN and CNN for imbalanced partial discharge pattern recognition in GIS,” High Volt., vol. 7, no. 3, pp. 845–854, Jun. 2022.
[20] J. Xie, “Partial discharge data enhancement and pattern recognition method based on a CAE-ACGAN and ResNet,” Symmetry, vol. 17, no. 1, p. 55, Jan. 2025.
[21] Y. Wu, W. Jiang, Y. Luo, D. Shen, and J. Yang, “Partial discharge pattern recognition of GIS based on improved SSD,” High Volt. Eng., vol. 49, no. 2, pp. 812–821, Feb. 2023.
[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., “Generative adversarial nets,” in Advances in Neural Information Processing Systems (NeurIPS), Montreal, Canada, Dec. 2014, pp. 2672–2680.
[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[24] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, Nov. 2014.
[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, Oct. 2020.
[26] Z. Liu, H. Mao, C. Y. Wu, et al., “A ConvNet for the 2020s,” arXiv preprint arXiv:2201.03545, Jan. 2022.
[27] Y. Han, C. Zhang, X. Zhan, et al., “Fine-grained recognition of crop pests based on improved ConvNeXt model,” Trans. Chin. Soc. Agric. Eng., vol. 41, no. 4, pp. 185–192, Apr. 2025.
[28] H. Liu, F. Wen, H. Du, et al., “Adversarial example detection with dual thresholds based on image transformation,” J. Softw. [Online]. pp. 1–16, Sep. 2025.
[29] H. Liu, Y. Li, L. Zhang, et al., “Apple leaf disease recognition based on MA-ConvNeXt network and stepwise relational knowledge distillation,” J. Zhejiang Univ. (Eng. Ed.), vol. 58, no. 9, pp. 1757–1767, 1780, Sep. 2024.
[30] F. Chen, M. Li, X. Wang, et al., “Few-shot partial discharge pattern recognition method for GIL based on prototype network and meta-learning,” Insulating Mater. [Online]. pp. 1–9, Sep. 2025.
[31] Y. Wang, J. Yan, Z. Yang, et al., “GAN and CNN for imbalanced partial discharge pattern recognition in GIS,” High Volt., vol. 7, no. 3, pp. 452–460, Jun. 2022.

[32] Q. Jing, et al., “A novel partial discharge pattern recognition for GIS with unbalanced sample based on conditional variational autoencoder,” in Proc. 18th Int. Conf. AC and DC Power Transmission (ACDC 2022), Xi’an, China, Jul. 2022, pp. 1314–1319.

[33] Y. Fu, K. Zhou, G. Zhu, et al., “A method for improving recognition accuracy of partial discharge in cable terminations based on improved WGAN model,” Power Syst. Technol., vol. 46, no. 5, pp. 2000–2008, May 2022.

[34] S. Ji, D. Zhao, Y. Jia, et al., “Research status and prospect of discharge development characteristics and detection methods for transformer insulation defects,” High Volt. Eng., vol. 50, no. 10, pp. 4297–4314, Oct. 2024.

[35] C. Liao, L. Zhang, G. Zhang, et al., “Partial discharge wideband full-band high-gain resonant cavity UHF sensor research,” Sensors, vol. 23, no. 15, Art. no. 6422, Aug. 2023.